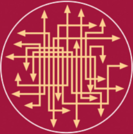
Demography

Analysis and Synthesis

A Treatise in Population



Graziella Caselli • Jacques Vallin
Guillaume Wunsch

Demography: Analysis and Synthesis

Demography: Analysis and Synthesis

Graziella Caselli, Jacques Vallin, and Guillaume Wunsch

with contributions by

Daniel Courgeau, Nico Keilman, Eva Lelièvre, James Vaupel,
Anatoli Yashin, and John Wilmoth

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ∞

Library of Congress Cataloging-in-Publication Data
Caselli, Graziella.
[Dimographie. English]
Demography : analysis and synthesis : a treatise in population studies / Graziella Caselli, Jacques Vallin, and Guillaume Wunsch ; with contributions by Daniel Courgeau . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-12-765660-X (set : alk. paper)—ISBN 0-12-765661-8 (volume 1 : alk. paper)—ISBN 0-12-765662-6 (volume 2 : alk. paper)—ISBN 0-12-765663-4 (volume 3 : alk. paper)—ISBN 0-12-765664-2 (volume 4 : alk. paper) 1. Demography. I. Vallin, Jacques. II. Wunsch, Guillaume J. III. Title.
HB871.C37513 2005
304.6—dc22 2005019413

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 13: 978-0-12-765660-1
ISBN 10: 0-12-765660-X

For all information on all Elsevier Academic Press publications
visit our Web site at www.books.elsevier.com.

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Demography: Analysis and Synthesis

Graziella Caselli, Jacques Vallin, and Guillaume Wunsch

with contributions by

Daniel Courgeau, Nico Keilman, Eva Lelièvre, James Vaupel,
Anatoli Yashin, and John Wilmoth

VOLUME I

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Contents
Volume I

CHAPTER

12

Cohort Analysis of Fertility 131

GUILLAUME WUNSCH

CHAPTER

13

Cohort Approach to External Migration Flows 149

GRAZIELLA CASELLI AND JACQUES VALLIN

PART

III

PERIOD ANALYSIS REVISITED

The Hypothetical Cohort and its Relations with Actual Birth Cohorts 161

GRAZIELLA CASELLI AND JACQUES VALLIN

CHAPTER

14

The Hypothetical Cohort as a Tool for Demographic Analysis 163

JACQUES VALLIN AND GRAZIELLA CASELLI

CHAPTER

15

Confounding Variables, Standardization, and the Problem of Summary Indices 197

GUILLAUME WUNSCH

CHAPTER

16

Relationships between Age-Specific Rates and Synthetic Indicators: Decomposition of a Difference 209

JACQUES VALLIN AND GRAZIELLA CASELLI

Contents of Other Volumes

Volume II Contents

Volume III Contents

Volume IV Contents

Contributors

Amoakon Anoh, Ècole nationale supérirure de statistique et d'économie appliqué (ENSEA), Abidjan, Côte d'Ivoire.

Philippe Antoine, Institut pur la recherché et le developpement (IRD), Dakar, Senegal.

Isabelle Attané, Institut national d'études démographiques (INED), Paris, France.

Alexandre Avdeev, Université de Moscou, Moscow, Russia, and Université Marc Bloch, Strasbourg, France.

Maria Avdeeva, Departement de bibliographie et de l'information, Centre d'etudes demographiques, Faculte d'economie, Université de Moscou, Moscow, Russia.

Brigitte Baccaïni, Institut national d'études démographiques (INED), Paris, France.

Stefano Baldi, Permanent Mission of Italy to the United Nations, United Nations Plaza, New York, United States.

Magali Barbieri, Institut national d'études démographiques (INED), Paris, France.

Odo Barsotti, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Françoise Bartiaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Gijs Beets, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Giovanni Berlinguer, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Elza Berquo, Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil.

Francesco C. Billari, Istituto di Metodi Quantitativi, Università Bocconi, and Innocenzo Gasparini Institute for Economic Research, Milan, Italy.

Jean-Noël Biraben, Institut national d'études démographiques (INED), Paris, France.

Anna-Maria Birindelli, Dipartimento di Statstica, Università degli Studi Milano-Bicocca, Milan, Italy.

Alberto Bonaguidi, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Corrado Bonifazi, Instituto di Ricerche sulla Popolazione e le Politiche Sociali, Rome, Italy.

Marco Bottai, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Michel Bozon, Institut national d'études démographiques (INED), Paris, France.

Vittoria Buratta, Istituto Nazionale di Statistica (ISTAT), Rome, Italy.

Raimondo Cagiano de Azevedo, Facoltà di Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Gérard Calot, Institut national d'études démographiques (INED), Paris, France.†

Emmanuelle Cambois, Institut national d'études démographiques (INED), Paris, France.

Graziella Caselli, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Jean-Claude Chasteland, Institut national d'études démographiques (INED), Paris, France.

Jean-Claude Chesnais, Institut national d'études démographiques (INED), Paris, France.

Luciano Ciucci, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

†Deceased

Robert Cliquet, Population and Social Policy Consultants, Brussels, Belgium.

David Coleman, Department of Social Policy and Social Work, University of Oxford, United Kingdom.

Maria-Eugenia Cosio-Zavala, Universite de Paris X-Nanterre, Nanterre Cedex, France.

Daniel Courgeau, Institut national d'études démographiques (INED), Paris, France.

Jean Coussy, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Gianpiero Dalla Zuanna, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy.

Patricia David, Harvard School of Public Health, Boston, United States.

Lorenzo Del Panta, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna, Italy.

Bart de Bruijn, Netherland Interdisciplinary Demographic Institute (NIDI), Den Haag (La Haye), Pays Bas, Netherlands.

Jean-Michel Decroly, Laboratoire de Géographie Humaine, Université Libre de Brussels, Brussels, Belgium.

Ainhoa de Federico de la Rúa, Institut Federatif de Recherche sur les Economies et les Societes Industrielles (IRESI), Universite des Sciences et Technologies de Lille 1, Lille, France.

Arna Dellis, University of Hawaii, Manoa, Hawaii, United States.

Paul Demeny, Population Council, New York, United States.

Alessandra De Rose, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Paolo De Sandre, Departimento de Scienze Statistiche, Università di Padova, Padova, Italy.

Gustavo De Santis, Facoltà di Scienze Politiche, Università di Messina, Messina, Italy.

Martine Deville, Institut national d'études démographiques (INED), Paris, France.

Manon Domingues Dos Santos, Centre de recherché en economie et statistique (CREST), Malakoff, France.

Josianne Duchêne, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Viviana Egidi, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Vincent Fauveau, United Nations Population Fund, New York, United States.

Alexis Ferrand, Institut de Sociologies et d'anthropologie, Universite des Sciences et Technologies de Lille, Villeneuve d'Ascq, France.

Patrick Festy, Institut national d'études démographiques (INED), Paris, France.

Judith A. Fortney, Family Health International, Research Triangle Park, North Carolina, United States.

Luisa Frova, Instituto Nationale di Statistica (ISTAT), Rome, Italy.

Alexia Fürnkranz-Prskawetz, Vienna Institute of Demography, Vienna, Austria.

Hubert Gérard, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Giuseppe Gesano, Instituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy.

Piero Giorgi, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Valérie Golaz, Institut national d'études démographiques (INED), Paris, France.

Antonio Golini, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Rosa Gomez-Redondo, Universidad Nacional de Educacion a Distancia (UNED), Facultad de CCPP y Sociologia, Madrid, Spain.

Catherine Gourbin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Carl Haub, Population Reference Bureau, Washington, D.C., United States.

François Héran, Institut national d'études démographiques (INED), Paris, France.

Kenneth Hill, Department of Population and Family Health Sciences, Johns Hopkins University, Maryland, United States.

Serguey Ivanov, Population Division, United Nations, New York, United States.

Carol Jagger, Department of Epidemiology and Public Health, University of Leicester, United Kingdom.

Heather Joshi, Centre for Longitudinal Studies, Institute of Education, University of London, London, United Kingdom.

Nico Keilman, Department of Economics, University of Oslo, Oslo, Norway.

Shigemi Kono, Faculty of International Economics, Reitaku University, Chiba-ken, Japan.

Marlène Lamy, Institut de demographie de l'Universite de Paris (IDUP), Paris, France.

Jacques Légaré, Department of Demography, University of Montreal, Montreal, Canada.

Luc Legoux, Institut de démographie de l'université de Paris (IDUP), Paris, France.

Eva Lelièvre, Institut national d'études démographiques (INED), Paris, France.

Henri Leridon, Institut national d'études démographiques (INED), Paris, France.

Ron Lesthaeghe, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Thérèse Locoh, Institut national d'études démographiques (INED), Paris, France.

Michel Loriaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Enzo Lucchetti, Dipartimento di Biologia Evolutiva, Università degli Studi di Parma, Parma, Italy.

Dionisia Maffioli, Università degli Studi di Bari, Bari, Italy.

Godelieve Masuy-Stroobant, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

France Meslé, Institut national d'études démographiques (INED), Paris, France.

Alain Monnier, Institut national d'études démographiques (INED), Paris, France.

Karel Neels, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Annunziata Nobile, Dipartimento di Istituzioni politiche e Scienze sociali, Università degli Studi Roma Tre, Rome, Italy.

Alberto Palloni, Center for Demography and Ecology, University of Wisconsin, United States.

Sophie Pennec, Institut national d'études démographiques (INED), Paris, France.

Pierre Pestieau, Université de Liège, Liège, Belgium and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Victor Piché, Inter-university Centre for Demographic Studies, University of Montreal, Montreal, Canada.

Marc Pilon, Institut de recherché pour le developpement (IRD), Piagadpigpi, Burkina Faso.

Gilles Pison, Institut national d'études démographiques (INED), Paris, France.

Antonella Pinnelli, Department of Demographic Science, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Denise Pumain, Institut national d'études démographiques (INED), Paris, France.

S. Irudaya Rajan, Centre for Development Studies (CDS), Kerala, India.

Jean-Louis Rallu, Institut national d'études démographiques (INED), Paris, France.

Rosella Rettaroli, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna Italy.

Jean-Marie Robine, Démographie et Santé, Institut national de la sante et de la recherché medicale (INSERM), Montpellier, France.

Paul-André Rosental, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Silvana Salvini, Dipartimento di Statistiche, Università degli Studi di Firenze, Florence, Italy.

Antonio Santini, Dipartimento di Statistiche, Univrsità degli Studi di Firenze, Florence, Italy.

Francis Sartor, Departement d'epidemiologie-toxicologie, Institut scientifique de la Sante Publique, Brussels, Belgium.†

Bruno Schoumaker, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Catherine Sermet, Institut national d'études démographiques (INED), Paris, France.

Jolande Siebenga, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Patrick Simon, Institut national d'études démographiques (INED), Paris, France.

Lamberto Soliani, Dipartimento de Scienze Ambientali, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Salvatore Strozza, Dipartimento di Scienze Statische, Università di Napoli Federico II, Napoli, Italy.

Pierre Surault, Groupe d'études démographiques, economiques et socials (GEDES), Universite de Poitiers, Poitiers, France.

Dominique Tabutin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michael Teitelbaum, Alfred P. Sloan Foundation, New York, United States.

†Deceased

Marc Termote, National Institute of Scientific Research, University of Quebec, and Department of Demography, University of Montreal, Canada.

Laurent Toulemon, Institut national d'études démographiques (INED), Paris, France.

Tapani Valkonen, Département de Sociologie, Université de Helsinki, Helsinki, Finland.

Jacques Vallin, Institut national d'études démographiques (INED), Paris, France.

Celine Vandermeersch, Institut national d'études démographiques (INED), Paris, France.

Jean-Pascal van Ypersele, Institut d'astronomie et de geophysique G. Lemaitre, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

James Vaupel, Max Planck Institute for Demographic Research, Rostock, Germany.

Jacques Véron, Institut national d'études démographiques (INED), Paris, France.

Éric Vilquin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Paolo Vineis, Dipartimento di Scienze Biomediche e Oncologia Umana, Università di Torino, Turin, Italy.

Anatoly Vishnevsky, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Tania Vishnievskaia, Institut national d'études démographiques (INED), Paris, France.

Carolyne Wanja Njue, Population Council, Nairobi I, Kenya.

Christine Wattelar, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michel Willems, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

John Wilmoth, Department of Demography, University of California, Berkeley, California.

Guillaume Wunsch, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Anatoli Yashin, Max Planck Institute for Demographic Research, Rostock, Germany.

Sergei Zakharov, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Hania Zlotnik, Population Division, United Nations, New York, United States.

General Introduction

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Six hundred million Chinese in 1950, and double that today. Eight hundred million Africans on the eve of the third millennium, and probably three times that in 40 years. Humankind under threat from a third world population *explosion*. Industrial societies enfeebled by *depopulation* and the opposing risk of a *population implosion.* Pension systems hard-pressed by aging populations. Northern societies living in entrenched fear of invasions from the east or south, be it history's *yellow peril* or the more recent *open floodgates* from across the Mediterranean or Rio Grande, turning a convenient blind eye to their own ancestral past as invaders, with their most overwhelming wave (toward North America) being a matter of recent history. For the past 50 years, population issues have dogged our contemporaries and they remain as live an issue as ever.

There is nothing new in this. Populations have been counted since the start of recorded time. But rational scientific methods to do so are of fairly recent date. Back in the 17th century, John Graunt (1662) tried estimating the population of London from registrations of deaths, while the late 18th century found Thomas Malthus publishing the first edition of his *Essay on the Principle of Population* (1798). But it was not until the 20th century that Alfred Lotka (1934, 1939) gave us a general theory of population dynamics, and Adolphe Landry (1934) a general explanation of the *demographic revolution,* which we now call "demographic transition" (Kirk, 1944; Notestein, 1945).

I. SCIENCE AND ITS PURPOSE

The problem with how demographers relate to the public may be that they now have so many different sophisticated tools for analysis that they risk shutting themselves up in an ivory tower of overtechnicality, even to forgetting what their science is all about: population. Too many handbooks give a sterile purchase on demography and, after lengthy expatiation on the relationships of *rates* to *probabilities*, the differences between *occurrence—exposure rates and frequencies,* or even the pros and cons of the n^{th} *parameter* of a descriptive or explanatory *model* of fertility or mortality, stop just where things start to get interesting: explaining the phenomena measured and assessing their impact on other aspects of economic and social life. The aim here is to combine a comprehensive and comprehensible presentation of all these methods of demographic analysis with a discussion of how they can be used to understand general population dynamics and the relationships (causes and effects) between them and their natural, economic, social, political, and cultural environments. In other words, we need to look not only at the science (demography), but also its subject (population) and not just what we can know about the latter, but what we can do with it.

The aim of scientific inquiry is not just to describe the universe and the laws that govern it; but also to try and tell us how it works. What it tells us enables us to act directly on the real world in some areas and change it for our benefit. So, modern physics has given us control of atomic energy and laser beams and medicine has produced antibiotics. In other areas (meteorology and astronomy, for example), science has only—at least as yet—enabled us to predict certain developments so that we can better accommodate them. For some scientific topics, we must be much less ambitious and content ourselves with describing past or current events and trying to understand their root causes. This, surely, is where we must class economics and politics, with the caveat that these disciplines are firmly geared toward action and control of the material world and the future, whereas the ultimate goal of science, regardless of performances, is to increase humankind's ability to take its own destiny in hand. Where does demography stand in this? Is it predictive? Can it be used for action?

Knowing, understanding, predicting, and controlling population development . . . Can this be done in isolation from other disciplines like history, geography, economics, sociology, psychology, medicine, biology, and genetics? Absolutely not. But, attempting to include every discipline would be as doomed to failure as would limiting our study to the techniques of demographic analysis alone: Co-opting everything potentially related to demography would lead us off course and risk muddying the paths we are trying to clarify.

The term "demography" first appeared in a work by Achille Guillard (1855), "*Eléments de statistique humaine ou démographie comparée*." Although more of a crusading pamphlet than a scientific discourse, and offering too narrow a definition, he must at least be credited with having coined the word used to describe the discipline today. But, it was more a late baptism than a birth. The foundations of the scientific edifice of demography were laid in the 17th and 18th centuries by a series of scientists working along the same lines: England's John Graunt (1662) and Edmond Halley (1693), Germany's Peter Süssmilch (1775), Holland's Willem Kersseboom (1742; INED, 1970), France's Antoine Deparcieux (1746), Sweden's Per Wargentin (1766). This nascent science as yet had no clear name or object. It was often referred to as *political arithmetic*. In fact, while its founding fathers came from quite different backgrounds (a cloth merchant, an astronomer, a theologian, an actuary, and two mathematicians), they agreed on at least one point: the need to address their *political* problem (the study of human populations) with *numerical* sciences—mathematics, statistics, and probability theory. And, that is precisely what Achille Guillard meant by linking his neologism with the expression *elements of human statistics*. We, too, must start by reducing reality to calculable and measurable elements in order to describe, analyze, and understand the mechanisms that govern the composition and development of a population. This is the foundation on which Alfred Lotka (1939) developed his general theory of *population dynamics* and this process is central to demography, even if it also seeks to study the dynamics of population from all angles connected with the various qualitative and quantitative aspects of the population at both individual and societal levels.

II. HUMAN POPULATIONS

What populations are we referring to? This is not a trivial question because in strictly statistical terms, a *population* is generally defined as a set of unspecified individual elements meeting a single definition. Once that group becomes subject to an ongoing process of entries and exits capable of changing its size and structure, it can be studied with the tools of demographic analysis. In this way, one can just as easily calculate the life expectancy of a *population* of electric light bulbs as measure the fertility of a *population* of butterflies, and so study the overall dynamics of these specific populations. *Animal population studies* in particular are a special field in which the principles of demography are widely used. Nevertheless, we exclude it from our area of study by limiting our discussion to human populations only. This is less because *human demography* is already a broad enough field in itself to be the subject of encyclopedic treatment than to highlight its singularity: Humans are not just statistical units, simple living beings, or merely social creatures like bees and ants; in Aristotle's[1] phrase, man is a *political animal* thinking and influencing his or her own individual and collective destiny, which situates the study of human population dynamics clearly in the field of social sciences rather than biology.

The most easily defined human population is all the human beings inhabiting the earth: All humanity, taking the term "humanity" in its full demographic sense of not just all human beings alive today, but also the entire course of the history of settlement, which is the core issue of population dynamics. Here, population dynamics relates only to one kind of entry and exit: births and deaths. But demography can just as easily apply to a small part of that group, set apart by reference to one or more criteria by which the individuals selected to form part of it can be identified: territory, religion, language, nationality, education, economic activity, disease, sex, blood group, and so forth. Any group thus defined can constitute a *population* in the demographic sense of the term. In many cases, the dynamics of this population will then be

[1] In his "Politics" Aristotle (384–322 BC) was seeking to emphasize the sophistication of humankind's social organization, embodied in the organized state. As Marcel Prélot (1961, p. 6) puts it: "Ancient man, as Aristotle defined him, was only a being or 'creature of the city.' The import of the definition is severely curtailed by translating 'zoon politikon' as 'social animal.' Animals are also social, but only man is political. His nature is not to live in flocks, herds or hordes but as an integrated part of that social organization represented by the Polis, the City; which is, for him, both a natural need and an ideal."

governed by more complex rules, according to many different types of entry and exit. In addition to the *natural changes* of births and deaths, we must take into account other events through which the individual characteristics defining the group studied are acquired or lost: crossing a border, acquiring or losing nationality, taking up or leaving an occupation, illness or cure, and so forth, unless they are (biologic or other) characteristics acquired through birth and lost only with death (gender, blood group, birthplace, etc.).

Of all the criteria used to identify a subpopulation whose dynamics are a matter for demographic study, the most important is residence in a socially significant territory. First, it must correspond to the simplest definition covering all humankind, which is self-defined by residence on a territory (the planet) whose social significance is beyond question in these times of globalization, the only difference being that this territory represents the entire area so far inhabitable by humankind. In addition, the concept of territory is inseparable from the social organization of human life, taken here as a building block of the subject of demography. So, territory is not just about geographically bounded areas—the area concerned must still be connected with the organized community life of human beings. Admittedly, the boundaries of that area may be determined by strict physical geography (the population of the Andean plateaux or the middle Senegal valley, for example) but the significant fact is that the ecologic conditions governing these territories influence the collective way of life. Consider too that these ecosystems are not the only source of socially organized territories: remember how many states of the United States are delimited by imaginary lines representing latitudes and longitudes!

III. POPULATION DYNAMICS

Whatever the population studied, demography focuses on its dynamics. Although it may attempt to identify the determinants and assess the impacts in the economic, social, cultural, political, and other fields, these factors and impacts fall within its purview only in so far as they offer explanations of these dynamics in view of informing government and society of the implications for the future. However, developing the specific analytic tools to do that has enabled demographers to acquire experience and skills in equally useful fields of study outside the specific focus of the discipline, not directly connected to population dynamics. Mortality, for example, one of the key factors of population dynamics, is obviously related to health and thus is a touchstone issue of demographic research in which demographers were interested even before the word demography was coined. But, the demographic study of health is supremely relevant to the social system and policy makers, be they general government, the medical profession, welfare funds, or charitable agencies. As a result, demographers are encouraged to engage with studies or research on subjects completely independent of the relevance they may have to explaining population dynamics or the effects of growth. For example, having developed abilities to measure morbidity, which has a clear impact on mortality (and hence population dynamics through tuberculosis, cancer, or cardiovascular diseases), the demographer may just as easily be prompted (by a study sponsor or simply to further his or her own scientific interest) to evaluate diseases or disabilities such as the common cold, vision problems, or mental illnesses whose connection with mortality are much more tenuous and in any event not the major concern.

Likewise, marriage entered the ambit of demography prima facie because it was or still is in many societies a factor for formally organizing the expression of fertility. However, it has subsequently remained a subject of demographic inquiry even where the marriage rate and fertility are increasingly uncoupled by the rise in nonmarital fertility and contraception within marriage. Demographers are increasingly working with sociologists, psychologists, political scientists, and lawyers in studying the diversifying forms of marriage, including those that clearly no longer have a link with fertility, such as same-gender unions.

The fact is that anything connected with fertility or mortality, however remotely, can have a direct bearing on population dynamics and it is only well upstream (factors) or downstream (consequences) that it can be considered completely outside the specific field of demographic research.

Migratory moves are another consideration. It is clear that what has just been said about mortality and fertility can readily be applied to external migrations—those across the boundaries of the area in which the study population lives. Internal migrations, which are a zero sum for the population concerned and so do not affect its dynamics directly, are a different matter altogether. Should they simply be excluded from the scope of demography? There are at least three good reasons why they should not.

First, the concepts and methods of analysis of migrations are self-contained, for one very good reason: The same migration flow can be considered as external or as internal according to how the study population is defined. A migration from Ohio to Oklahoma is an internal migration for the United States, but external for each of the two states concerned. The same argument can be applied at all levels and applies even more so where the study population is defined not by territory but by any other distinguishing criterion (school population, occupation, legal categories, etc.). Occupational, social, and legal migratory moves may be internal or external, depending on the study population profile. Such rapid change in the distinction between internal and external alone would justify including the study of all migrations of whatever kind in the scope of demography.

But, also, even restricting discussion to the strict confines of a given population—in the broadest case of the population of a geographic area—internal migrations are usually not just indirect factors. Rather, they are are also a direct consequence of population dynamics. The spatial reorganization effected through internal migrations, especially between town and countryside or relatively different cultural areas, is rarely without consequence on demographic behavior, health, and fertility. To this extent, internal migratory moves contribute at least indirectly, and often significantly, to the overall movement. But, even more clearly, internal (and external) migrations are often the consequences of population dynamics effects. Population pressure exerted on certain strata or regions of the territory (the countryside, for example, or areas with a higher natural increase) are often the source of internal migration flows, which, as a result, fall within the field of general population dynamics.

For all these reasons, far from excluding internal migrations from the scope of this book, external and internal migrations will be distinguished only according to methodologic or conceptual needs. It is quite clear that although only external migrations are central to the dynamic, internal migratory moves are equally important in studying determinants and consequences and that in the final analysis, they will in most cases be methodologically inseparable.

The division between the components of population dynamics and their relations with other aspects of social dynamics may be unclear, but so are those of the wide range of populations that can be the focus of demographic study. It was said earlier that the focus of this text will be on resident populations living in a socially constructed area, but that is not to exclude other populations susceptible to demographic analysis. However, other examples will be cited only briefly and selectively to illustrate the varied range of possible applications, which is so wide as to rule out a comprehensive approach.

IV. INDIVIDUALS AND VITAL EVENTS

As stated above, demography is about the study of the *human population*, although its techniques can be applied to other subjects. However, any population consists of a group of individuals. While demographic research does not focus on the individual *per se*, unlike psychology for example, demographers must bear in mind the obvious facts that a population's mortality is due to the death of some of its members during a given period, *natality*[2] is the result of births engendered by men and women, and migration stems from the mobility of human beings within or outside borders. The correspondence between population changes and individual vital events, however, is more complex than it appears at first sight. Take the case of mortality, for instance. In any year, an individual may die or survive; so his or her survival probability will be 0 or 1. During the same year, general mortality may be characterized by the ratio of the number of deaths to the population exposed to the risk at the start of the year. This ratio gives a frequency or probability of dying that will never be equal to 0 or 1 other than in the exceptional case of everyone surviving or dying in the year. When we say that a person has a 0.083 probability of dying in the year, we are actually applying a group characteristic to the individual that will *never* occur on an individual level. The fact is that if the person survives during the period of time, its probability of dying will be 0, and if the person does die, it will be 1; it will never be equal to 0.083. An individual probability cannot be estimated unless there are at least two individuals; we then apply to the individual a probability calculated for the group by assuming that the population is homogeneous in relation to the risk of dying. This is the basic principle of life insurance, enabling a measure taken on the population to be transposed to a specific individual.[3]

Another example of the differences between a population and an individual relates to the fact that a person may age comparatively quickly or slowly, but can never get younger. The *arrow of time*, as philoso-

[2] We generally talk of natality in relation to a population, and fertility for individuals of childbearing age.

[3] This principle depends on the concept of probability itself. Do we assume that a probability is the result of a series of identical experiences, where events are repeated in the same conditions (a barely tenable position for human events), or do we instead consider that probability expresses the degree of certainty about the truth of a proposal in the light of the available information? This second "Bayesian" approach seems to correspond more closely to the needs of demography.

phers call it, is a one-way street.[4] But a population can get younger if the birth rate rises. In this case, the proportion of young people in the population will rise (to the detriment of the other age groups) and the mean age of the population will fall. So a population may rejuvenate following a rise in the birth rate even if individual life spans lengthen. Conversely, populations have historically aged as a result of declining fertility and not because the mean life span of individuals increased.

A population is a point-in-time cross-section of a multitude of individual life paths. In a given period of time, people are added to this population through birth and immigration, while others die or emigrate. The total population will increase or decrease as a result of these events. To properly understand the period change in the population, therefore, we need to know how many people were born in the period under review, how many died, and how many emigrated or immigrated (as will be seen in more detail in Chapter 3). The cross-sectional *demographic structure*, therefore, will depend on the characteristics of the individuals comprising the population at that time (e.g., their age, marital status, gender, etc.). *Population change* will result from the number of births, deaths, emigrations, and immigrations occurring in that population during a period of time.

There is an infinite variety of possible demographic structures according to the number of characteristics selected. Individuals do not differ only by age or marital status; they also differ in dietary patterns, musical tastes, hair and eye color, and so forth. Not all of these characteristics and associated structures are useful to demography; only those characteristics capable of affecting population change must be taken into account. So, fertility, mortality, and migration vary with the individual's age, sex, and marital status; but educational level and social or ethnic group often affect an individual's fertility or life span too. On the other hand, a love of tulips or Sibelius symphonies probably has no impact on the components of population change, although these characteristics may be more specifically *associated* with certain social or ethnic groups. Any putative correlation between a passion for tulips and individual fertility would probably be spurious, therefore, and would almost certainly be eliminated by controlling for the social or ethnic group.

Generally, for an individual characteristic to be meaningful for demographic purposes, it must be individually discrete and affect at least one of the components of population change, after controlling for potential common causes. For example, individuals differ in age, and age influences fertility, mortality, and migration. The same applies to gender and many other variables. By contrast, musical tastes may vary from individual to individual, but do not influence population change; any correlation will be eliminated by controlling for *confounding variables* such as social status. Musical tastes may interest sociologists and musicologists, but are utterly irrelevant to demography.

Demographic behavior (i.e., giving birth to a child and dying or migrating within or outside of the area concerned) is clearly an individual characteristic. The aggregation of these events is what produces population and structural changes. The causes of these types of behavior may be physiologic. Sexual desire is important for the reproduction of the species; death is probably imprinted in our genes. Most behavior, however, is learned during life or childhood (i.e., it stems from factors related to the individual's past, family and social environment, and culture). Even gender may depend on assignment and upbringing, such as whether parents and peers treat the child as a boy or girl. This individual gender identification may even lead to pseudo-hermaphroditism (Fantino and Reynolds, 1975). The decisions to get married, have children, and/or migrate are very much dictated by learned behavior, parental and peer influence, social membership, and group culture. Dietary patterns, which have a significant influence on health and survival, are also largely culturally determined, as lovers of snails or frogs' legs will testify! Smoking and drinking, two major causes of death, are also socially conditioned.

Thus, inquiring into the possible causes of demographic behavior means going beyond individual factors and looking at the human being as a member of a household, family, social category, ethnic group, and so forth. In recent years, targeted surveys have enabled demography to sharpen its focus not just on individual characteristics (age, sex, education, etc.) that influence demographic behavior, but also on the multilayered aspects of demographic change stemming from the individual's membership of different networks of groups and institutions.[5] The subject of demographic research, therefore, is the individual *and* the society in which he or she lives. Explanations focused just on individual behavior, like the rational behavior of economic agents beloved of economists, cannot encompass the broad spectrum of influences

[4] Although some philosophers consider the forward march of time to be only a subjective characteristic of life as lived, and not an objective reality; they would argue that time stands still (see, for example, Michael Tooley, 1997).

[5] See D. Courgeau (1996).

and incentives that rule human needs and motives; the same would apply to a purely macrodemographic explanation (see, e.g., de Bruijn, 1999).

As will be seen throughout this book, demographic explanations and theories are increasingly incorporating the social value systems and factors that predicate demographic behavior. This is not to say, however, that all individuals act in an identical manner; even in very homogeneous societies, some individuals adopt nonconformist attitudes and styles of behavior. So, some couples in a low-fertility country may have large families, just as some couples may practice contraception in countries where fertility is highly valued. Nevertheless, behavior in most societies is sufficiently homogeneous for aggregated measurements like life expectancy and average parity to be meaningful.

V. DEMOGRAPHY AND SOCIETY: THE GREAT CHALLENGES

Demographers, therefore, cannot rely only on demographic variables to describe and explain population dynamics. Their conceptual framework must necessarily accommodate all individual and collective factors likely to influence the demographic processes of the individuals—and hence the population—studied.

Recent developments in the field are intertwined with the broader changes wrought to knowledge and practice in all the social sciences by turn-of-the-century political, economic, social, and demographic developments. Society today is in the throes of rapid change, which is affecting some geographic areas and population groups more than others and creating new (social, economic, and demographic) divides or reviving old inequalities. As a result of the rapid pace of social and cultural changes, major technologic advances, the growing power of the market economy, the decline of the welfare state, globalization, adaptability to new situations differs from one population to the other, even within countries, according to their cultures, access to new technologies, and, of course, economic performances. In almost every case, the strongest win out over the weakest. As a result, the gap—including the demographic gap—between the leaders and the laggards is widening. These processes have aggravated the world north/south split, just as economic globalization has widened the welfare divide between an increasingly developed north and the different components of an increasingly riven South. The most dynamic southern countries, which are also those with the earliest onset demographic transitions, have managed to make the best of the situation, and even compete with northern countries, whereas others, resource-starved and still tackling the consequences of overrapid population growth, are sinking deeper into crisis.

Population trends in the wealthy countries could become a new source of concern. The received wisdom that the sweeping historical change represented by demographic transition—the shift from high levels of fertility and mortality to the low levels attained between the two world wars—would be followed by a situation of equilibrium—near-zero growth and unchanging structures—seems not to have to been borne out in reality. Far from stabilizing, the distribution of the population between its three large components that typify the individual life cycle—childhood, adulthood, and old age—seems more fluid now than ever. In the classic posttransition scenario, the population aging that is inherent to the change from pre- to posttransitional stages could have stayed within reasonable enough limits to require only tinkering at the edges with the social systems established over the century; but the collapse of fertility levels and the unexpected decline in mortality among the oldest old is distancing the situation from the accepted model. Furthermore, short-term cyclical reasons have created an unprecedented accumulated potential among northern populations for the demographic aging process to accelerate. Basically, the working age populations have so far remained virtually stable thanks to the size of the post–World War II *baby boom;* before long, however, these large numbers of cohorts will reach retirement age, while the working population dwindles as the less populous cohorts reach working age.

It is easy to predict when the two waves of baby boomers—immediately postwar and the 1960s—will leave the labor force. With retirement ages of 60 or 65, the first wave will start exiting around 2005 to 2010, the second around 2025 to 2035. Logically, this will put immense pressure on health and social protection systems. Here, the veritable revolution in health care brought about by major advances in treatment and prevention, which give most people with degenerative diseases or cancer a longer life expectancy than before, must be taken into account. The possibilities of survival amongst the oldest old are expected to increase considerably in the coming decades, because today's older people—and the next generation even more so—are (or will be) of the generations that have drawn far more of the positive benefits of welfare than their predecessors: proper diet, improved quality of life, occupational health and safety, cultural development, and, especially, the goods and services necessary to stay in good physical and mental shape until entering the oldest age group. The increased number, and especially proportion, of the old elderly, however, will begin to push up not only social protection but also health costs: the prevention and treatment of chronic and degenerative diseases, assistance for disabled people, and, more generally, care for very old people with physical and mental disabilities. So, demographic analysis must accommodate not only declining mortality and lengthening life expectancy, but also the development of different disease patterns and morbid conditions affecting the population.

Entrenched very low fertility also has a major bearing on the increase in the proportion of old people. Basically, where the top of the population pyramid gets wider under the effect of increased longevity, the bottom part narrows each year with the declining birth rate. Furthermore, that low fertility is itself the result of shrinking generations, because women born after the baby boom and falling birthrate are now the child-bearing-age generation: Even supposing an average parity equal to that of their mothers, these women would produce fewer children in total than preceding generations. Fertility decline thus creates a dual cause of the falling birthrate. Were this knock-on effect to continue, it would produce an inescapable decline in the total population. Although immigration might provide a stopgap solution, only a significant rise in fertility can halt the process in the long term. The steadier the decrease in the youngest cohorts entering at the bottom of the pyramid, the harder it will be to turn the situation around and the longer any possible resurgence in fertility will take to restore age structure balance in the population. But it is far from certain that any such turnaround in the situation is imminent, so greatly does it depend on the position of women in society and, more generally, children's place in their scale of priorities.

Historically, women's lives were taken up with bearing and raising children; now, these activities occupy only a small part of their lives: More women work, leaving the labor force only for short periods (confinement and delivery), and have many fewer births, being more intent on pursuing their careers in historically male jobs. The world of work has not yet fully adjusted to the growing number of women, which is a competitive handicap to those trying to balance work and motherhood because job characteristics (working time and procedures) and promotion prospects are still patterned on male lifestyles. In many areas, therefore, career opportunities are dictated by the early years of working life and generally fixed around 30 years, just the age at which personal and family choices—partners and family size—are played out. Basically, women today want financial independence and a professional career and spend most of their youthful years in education and establishing a foothold in working life, which means deferring or abandoning plans to start a family. Obviously, until such time as work and fertility can be fully reconciled, women will almost always choose the former over the latter.

Overlaying this are the consequences of changing family patterns. Demographic transition evidenced a shift from the patriarchal family to the nuclear family; but far from being a foundation of permanent stability, the latter is now being called into question in turn by marriage break ups and new living arrangements. Rising divorce and separation rates are almost automatically increasing the proportion of lone, often marginalized, mothers. This severing of parental bonds has affected the postwar generations, especially the baby boomers (i.e., those who will soon be adding to the oldest age group). Rising insecurity related to the loosening of traditional ties of solidarity, places new demands on social protection, which could further undermine existing systems. Failure to address this new situation may hinder a return to demographic balance, as the fear of marriage breakdown may depress fertility. All in all, the only-child option, perhaps combined with deferred childbirth or even voluntary infertility, could well become an established societal norm with a series of repercussions on the future population dynamic.

This concern about the demographic future of the developed world may seem out of place, given the magnitude of the problems pulling in opposite directions today within many developing countries. Yet it appears likely that the globalization of demographic processes will turn it into the key issue for all humankind within the next few decades. However, this must not eclipse the fact that a significant proportion of the world's population will still have to contend with the problems of rapid population growth for some decades to come.

The completion of the demographic transition, foreshadowed and confirmed by each new United Nations (UN) population projection for the past 20 years, is naturally a great relief compared with the acute fears aroused less by the UN's late-1950s forecast of 6 billion people by 2000 and the resultant rash of far-fetched extrapolations made beyond that date. Right as it may be to condemn the overblown rhetoric of the past that sought to stigmatize rising births in the developing world as irresponsible or support the introduction of fertility regulation policies, it is equally important now to consider that although the impending end of the demographic transition is proof positive of the groundlessness of some past fears, many crucial problems remain to be solved.

With 6 billion men and women in 2000, the hardest part may seem to be over. We have gone from 2.5 to 6 billion in 50 years and have still to go from 6 to 9.5 in the next 50 years. We have therefore experienced an increase of 3.5 billion in a half-century, from a starting point of 2.5—a 140% rise—while the 3.5 billion still to come, also in a half-century, is a mere 60% increase over the current 6 billion! Granted, but two problems still remain. One is that the results to date are hardly impressive: 20% of the world's population has 80% of its income, while the overwhelming majority share the crumbs that remain. This is not a matter for pride, especially given the very unequal division of that remainder. Poor countries increasingly have to be divided between genuinely developing countries and very poor countries whose extremely fragile economies are on the brink of collapse.

Above all, however, a large share of the 3.5 billion people added in the last 50 years accreted directly to developed countries with growing populations (especially North America and the former Union of Soviet Socialist Republics [USSR]) while the bulk of that in the south was added by the most dynamic countries or regions (China, Southeast Asia, Latin America). By contrast, the next 3.5 billion will be almost entirely added to the world's poorest countries or areas, especially in sub-Saharan Africa and some South Asian countries. Setting this further increase against the current total population of the regions most affected (under 2 billion), we get a real growth figure of nearly 200%. However, not only are these countries the world's poorest, but unlike most of Asia or Latin America whose highest population growth came during the "30 boom years," today's poorest countries will incur theirs in adverse international economic conditions (Vallin, 2000).

Arguably, there is every reason for consigning to history what in fact we should never have feared, but there is now all the more reason to look seriously at what, all rhetoric aside, we have consistently put off to later: the economic and social development of the poorest regions. This remains a live and very urgent issue, like finding environmentally sound development models that respect the broad balance of nature on the planet.

VI. FUTURE POPULATION DYNAMICS

It is a racing certainty that once this exceptional growth period has peaked, all the world's populations will more or less attain a mortality—fertility equilibrium in the near or relatively near future. However, there are no guarantees that it will remain stable. What is most likely is that the end of the great historical process dubbed "demographic transition" will also see the end of the like-named paradigm.

Barely had the UN ventured to produce its fine simulation of a general stabilization of the world population based on the transition theory in the early 1980s than the fault lines began to show everywhere. Less because of its failure to predict the acquired immunodeficiency syndrome (AIDS) epidemic in Africa or the health crisis in Eastern Europe (which, however appalling, are finally just adverse developments that cast no doubt on the fundamentals of the model) than because of the growing body of evidence that "the end of the transition" will be anything but that expected. The postwar baby boom experience should have sounded the alarm bells, but the fact is that World War II upset so many apple carts that it might be reasonably thought to be only an interlude in the unfurling of an inevitable historical process. And that is how it was seen.

But, how could it be thought that declining mortality and fertility, the period of exceptional growth due to the time lag between the two, and the resulting upheaval in age structures, would end up in a period of general stability? How could it be thought that after the storm would come the eternal, untroubled calm on nothing more than the belief that life expectancy at birth would never extend beyond 85 and that fertility would inevitably level off at 2.1 children per woman? There was nothing, absolutely nothing, to suggest that it would be this way. To the contrary, all the signs today are that things could go very differently. Sadly, there is no new paradigm, model, or theory with which to accommodate not the end of the demographic transition, but its expected consequences. All that can be done when imagining the long-term future, therefore, is to devise possible alternative scenarios for fertility and mortality and their expected consequences for the total population and age structures. Admittedly, it can be done on the basis of currently observable facts, but the future import of those we cannot tell.

As to fertility, first, two facts give pause for thought. First, in a number of northern countries, the total fertility rate has been significantly below the replacement level of 2.1 children per woman for the past 20 years. In some, it has even been far below: 1.3 in Germany, Italy, and Spain, for example, and even 0.8 in Northern Italy! Second, mean age at childbirth is rising significantly. These two facts may not be unrelated. It could readily be assumed that a late first birth would reduce period fertility to very low levels simply as the combined result of very low fertility among women who completed their fertility when younger and the continuing very low fertility among younger women who are putting off motherhood. Overall lifetime fertility aspirations may be unchanged, but, in these conditions, period fertility could actually fall to very low levels at least for a certain time. Obviously, it may equally well be that what has really changed is desired family size.

On the basis of these different possible interpretations of the changes in progress, one chapter of this treatise will seek to explore various fertility trend scenarios that combine assumptions about level (from a return to replacement level to the only-child model) and timing (early or late, with one or two key family formation periods).

Still on the basis of current facts, questions arise about the lasting quality of a parameter that demographers have in the past always taken as stable—the sex ratio at birth. It is a universal fact that until very recently the number of male births was always slightly above female births (105 males to 100 females), with excess male mortality redressing the balance at childbearing age. Now, in some societies where male preference is particularly strong, this successful balance is being seriously undermined by antenatal gender diagnosis and the possibility of selective abortion. In some parts of India and China, the sex ratio at birth has now risen to 120, 150, and above. What would be the result of a long-term spread of such behavior?

But fertility is not the only question mark for the future. Contrary to expectations, the future of average duration of life is little more certain than that of fertility. Disregard the possibility of humankind being simply wiped out by a nuclear war, or an epidemic even worse than AIDS is proving for Africa. Leave to the gloom merchants the idea that life expectancy could fall on a long-term basis worldwide as it recently has in the Eastern European countries. The game still remains wide open. There are, in fact, two theories today setting biologists and demographers at odds with one another.

The first says that we are fast approaching the absolute limit of human longevity. Having raised life expectancy from 25 in the 18th century to 80 years now, all we have done is to bring the average (life expectancy) closer to the maximum possible life span (longevity). Only the former is susceptible to change, while the latter is irreversibly fixed by the initial life potential written in our genes. It is clear that the closer one gets to the maximum, the harder it is to raise the average. Furthermore, because of the variations in individual gene pools, the upper limit itself varies widely between individuals, and the absolute record recently set by Jeanne Calment, who died in 1997 slightly older than 122 years, is just that—a record (i.e., a score that only a tiny number of individuals will ever be able to match). This line of reasoning has led biologists like James Fries (1982) and demographers like Jay Olshanski (1990) to conclude that it is virtually impossible for life expectancy to exceed 85 years. This, indeed, is the underlying assumption of the most recent UN projections.

There is, however, another theory that human longevity is not immutable. First, we have no evidence that it has not increased in the past. Methuselah aside, anecdotal evidence of extreme longevity is not uncommon in historical records, but singularly lacking in hard proof (Vaupel and Jeune, 1994). No proof has ever been produced of any man or woman before Jeanne Calment attaining the age of 120 (Allard *et al*., 1994). However, there is a well-established tendency in almost all societies to exaggerate the age of very old people. By contrast, there is circumstantial evidence that human longevity may have increased in recent decades (Robine *et al*., 1997). The highest age at death has risen each year in the most advanced countries over the past 40 years (Wilmoth and Lundström, 1996). Also, since the 1970s, mortality among the oldest age groups (over 75 years), which had been relatively stable suddenly, began to recede (Kannisto, 1994, 1996; Meslé and Vallin, 1998). Finally, some biologists, like Roy Walford (1984), for example, believe that we are on the threshold of effectively delaying the human aging process. If so, anything becomes possible.

Without going as far as André Klarsfeld and Frédéric Revah (1999) with their suggestion that immortality does not exist, not because of biologic imperatives, as had always been thought in the past, but simply because serving no biologic purpose, it did not evolve, should we not, for example, admit the plausibility of Roy Walford's proposition that life expectancies of up to 150 years should not be beyond imagining in the 21st century? What effect would that have on populations?

All these scenarios would obviously have very different consequences in terms of total populations and demographic structures. The world population, for example, could just as easily continue to grow rapidly after 2050 or collapse to pre-Neolithic levels with the risk of being wiped out altogether. Alternatively, "reasonable" aging leading, for example, to 30% of the population aged over 60 is just as likely as a radical upheaval in age—sex structures going far beyond even the most radical predictions so far: Life expectancy of 150 years and a single-child family trend could, for example, produce a population with only 2% of "young people" (younger than 20 years) and 7% of "adults" (aged 20 to 60 years), but 91% of people older than 60 years and 74% older than 100 years! Obviously, the descriptions young person, adult, and old person could not then be applied to the same age groups as today. Regardless, given the difficulties we now have contemplating the prospect of living tomorrow (in 2050) in a society where at least 30% and perhaps even 40% of the population will be older than 60 years, how can we envisage a world where three fourths of the people will be centenarians? It is easy for demographers to make population forecasts, but it is much harder to envision the kind of society that they would produce. Sociologists, psychologists, economists, and political scientists would also have to think hard! Much will depend on their respective abilities to communicate and create synergies of knowledge. In any event, the size of the challenges, the complexity of development factors, and the still largely uncharted possible consequences require a new demography with a broad cultural focus capable of addressing issues holistically. The next stage will probably see new developments increasingly tracked through in-depth surveys that pave the way for research into the linkages between socioeconomic and environmental changes and the demographic—and even biologic—components of the population and the development of a framework of theories and assumptions capable of connecting all the components of demographic thought and experience with those of life as it is lived in an overall approach. For that to happen, demographers will obviously have to be more skilled in combining macroanalysis with microanalysis, moving from the forensic analysis of types of behavior and their determinants to interpreting them at the holistic level of the complex systems in which we live.

In relation to the state of the art in methods and knowledge, this view of demography lies more in the realms of science fiction than today's realities. At the very outside, it may perhaps be incipient. This book, therefore, does not seek to offer its readers what does not yet exist. Its aim is less lofty: to report on what does exist, pointing up the consistencies but also the emerging inconsistencies that may foreshadow the future shape of our discipline.

GENERAL LAYOUT

This treatise on demography comprises four volumes. This volume, the *first* in the series, falls into two sections. Section I discusses population dynamics and focuses on the essence of demography: the analysis of the components of population change related to changes in total population and age–sex structure. It falls into five parts, themselves directed by the overall approach taken in this treatise progressing from analysis to synthesis. Part one covers population growth, with an initial overall view of the principles of population dynamics. Part two is given over to the analysis of the three specific components of population change (fertility, mortality, migration) in the context of the life experiences of cohorts. In part three, the measurements made are reconstituted as period indicators to track demographic trends. This paves the way for an initial synthesis followed, in part four, by the construction of population models. Part five looks at ways in which to take account of the dependency between demographic phenomena and of population heterogeneity.

The following chapters analyse the factors of each of the three building blocks of population change: fertility, mortality, and migration. There are separate parts of each component, where the different types of factors (biological, socioeconomic, cultural) and levels of explanation (baseline variables and intermediate variables, theories, frameworks, and explanations) are addressed. The determinants of fertility are covered in Section II of Volume I, while those of mortality and migration are successively discussed in the two sections of Volume II.

The first section of Volume III addresses the history and consequences of demographic change, i.e., the history and geography of human settlement but also population forecasts. Section II considers the consequences of demographic changes in the biological and health fields, in the socioeconomic and cultural ones, on development, and on the environment.

The fourth and final volume covers, in Section I, population theories and doctrines and population policies. Section II falls into three parts and considers observation systems, auxiliary methods in demography, and the history and teaching of demography.

References

Allard Michel, Lèbre Victor and Robine Jean-Marie, 1994. *Les 120 ans de Jeanne Calment doyenne de l'humanité.* Paris, Le cherche midi, 162 p.

Courgeau Daniel, 1996. Vers une analyse multi-niveaux en sciences sociales. *in*: Jean-Pierre Bocquet-Appel, Daniel Courgeau and Denise Pumain (eds.), *Analyse spatiale de données biodémographiques*, p. 9–22. Paris, INED, John Libbey, xxi + 367 p. (Coll. Congrès et colloques, 16).

De Bruijn Bart. J., 1999. *Foundations of demographic theory*. Amsterdam, Thela Thesis, 298 p.

Deparcieux Antoine, 1746. *Essai sur les probabilités de la durée de la vie humaine*. Paris, H.L. Guérin and L.F. Delatour, 208 p.

Fantino Edmund and Reynolds George S., 1975. *Introduction to contemporary psychology*, San Francisco, W.H. Freeman, 610 p.

Fries James, 1982. Aging, natural death, and the compression of morbidity, *New England Journal of Medicine*, vol. 303, p. 130–135.

Graunt John, 1662. *Natural and political observations made upon the bills of mortality*. London, printed by T. Roycroft for J. Martin, J.

Allestry and T. Dicas, 150 p. (+ American edition by W.F. Wilcox, Baltimore, Johns Hopkins Press, 1939, + French edition by Eric Vilquin, Paris, INED, 1977).

Guillard Achille, 1855. *Éléments de statistique humaine, ou démographie comparée*. Paris, Guillaumin, 376 p.

Halley Edmond, 1693. An estimate of the degrees of the mortality of mankind, *Philosphical Transactions*, vol. XVII(196), p. 579.

Henry Louis, 1972. *Demographie: analyse et modèles.* Paris, Larousse, 341 p.

INED 1970. *Essais d'arithmétique politique contenant trois traités sur la population des provinces de Hollande et de Frise occidentale [...] par Willem Kersseboom et Kersseboom et son œuvre, la table de mortalité par M. Van Haaften*. Paris, Imprimerie Nationale, 172 p.

Kannisto Väinö, 1994. *Development of oldest-old mortality, 1950–1990: evidence from 28 develped countries.* Odense (Denmark), Odense University Press, 108 p. (Odense Monographs on Population Aging no. 1).

Kersseboom Willem, 1742. Eerste *Verhandeling tot een proeve om te weeten de probable menigte des wolks in de provintie van Holland en Westvrieslandt, en, specialyk tot aanleidinge van verder onderzoek, in de steden Haarem, Amsterdam en Gouda, als mede in's-Gravenhage.[...]. Tweede Verhandeling, bevestigende de proeve om te weeten de probable menigte des wolks in de provintie van Holland en Westvrieslandt. Beneevens een grondslag tot het fundeeren van een proeve om te ontdekken de probable leeftyt der weduwen . . . Derde Verhandeling over de probable menigte des wolks in de provintie van Holland en Westvrieslandt, bevattende, eerstelyk een vertoog over de proportie der meenigte des volks tegens het getal der geboorene*. The Hague, Jan van den Bergh, Librairie Wagestraat.

Kirk Dudley, 1944. Population change and the postwar world, *American Sociological Review*, February, p. 28–35.

Klarsfeld André and Revah Frédéric, 1999. *Biologie de la mort.* Paris, Odile Jacob, 290 p.

Landry Adolphe, 1934. *La révolution démographique. Étude et essai sur les problèmes de la population*. Paris, Sirey, 231 p.

Lotka Alfred J., 1934. *théorie analytique des associations biologiques. Première partie: principes*. Paris, Herman & Cie, 150 p. (Actualités Scientifiques et Industrielles, no. 187)

Lotka Alfred J., 1939. *Théorie analytique des associations biologiques. Deuxième partie : analyse démographique avec application particulière à l'espèce humaine*. Paris, Herman & Cie, 139 p. (Actualités Scientifiques et Industrielles, no. 780)

Malthus Thomas R., 1798. *An essay on the principle of population*. London, McMillan, 396 p.

Meslé France and Vallin Jacques, 1998. *Évolution de la mortalité aux âges élevés en France depuis 1950.* Paris, INED, 42 p. (Dossiers et recherches no. 68).

Notestein Frank, 1945. Population, the long view, *in*: T. Schultz (ed), *Food for the world*, p. 36–57. Chicago, University of Chicago Press, 367 p.

Olshansky S. Jay, Carnes Bruce A. and Cassel Christine, 1990. In search of Mathuselah: estimating the upper limits to human longevity, *Science*, vol. 250, p. 634–640.

Pressat Roland, 1961. *L'analyse démographique*, 1st ed. Paris, PUF, xii + 402 p.

Pressat Roland, 1969. *L'analyse démographique*, 2nd ed. Paris, PUF, 321 p.

Pumain Denise (éd.), 1991. *Analyse spatiale de données biodémographiques.* Paris, INED, John Libbey, 458 p. (Congresses and Colloquia, 6)

Robine Jean-Marie, Vaupel James W., Jeune Bernard and Allard Michel (eds.), 1997. *Longevity: to the limits and beyond.* Berlin, Springer, Fondation IPSEN, 180 p. (Research and perspectives in longevity).

Süssmilch Peter, 1741. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortplflanzung desselben erwiesen*. Berlin, Kultur-Verlag, 360 p. (Reprint of the original edition by J.C. Spener, Berlin, 1741.)

Tooley Michael, 1997. *Time, tense, and causation.* Oxford, Clarendon Press, XVI, 399 p.

Vallin Jacques, 2000. *La fin de la transition démographique: soulagement ou inquiétude?* Paris, INED, 13 p. (Conférence à l'Université de tous les savoirs, Paris, CNAM, 24 Feb. 2000).

Vaupel James and Jeune Bernard, 1994. *The emergence and proliferation of centenarians.* Odense (Denmark), Odense University Press. (Odense Monographs on Population aging 2).

Walford Roy, 1984. *La vie la plus longue*, translated from the English. Paris, Laffont, 250 p.

Walford Roy, 1984. *Maximum life span*. New York, Avon, 230 p.

Wargentin Pehr, 1766 (1983). *Den svenska statistikens fader: en minnesskrift med sju originaluppsatser ur Kungl. Svenska vetenskapsakademiens handlingar för åren 1754, 1755 samt 1766.* Stockholm, Statistiska centralbyråran (1983), 180 p.

Wilmoth John R. and Lundström Hans, 1996. Extreme longevity in five countries: presentation of trends with special attention to issues of data quality, *European Journal of Population/Revue Européenne de Démographie*, vol. 12(1), p. 63–93.

SECTION

I

POPULATION DYNAMICS

Introduction

GRAZIELLA CASELLI AND JACQUES VALLIN

In statistical terms, a population is a set of individuals defined by the same criteria but possessing various values for different attributes. A population can also be considered at all times as consisting of *groups* of individuals possessing the same value for given attributes. The main characterization of a population, therefore, is its size (the number of individuals comprising it) and its composition (the frequencies of the various values attaching to the attributes). On the face of it, this clearly applies to human populations, which are the subject of demography. However, human populations are not frozen in time but are in a state of constant change. Demographic analysis primarily studies the underlying mechanisms of that change (i.e., to understand *population dynamics*).

A first characteristic of the dynamics of a population is its capacity to increase or decrease at a certain rate. For that reason, part one of this section looks at *population growth*. Chapter 1 looks in more detail at what demographers mean by populations and individuals; Chapter 2 considers the dynamic nature of human populations, whereas Chapter 3 examines how population growth depends upon its natural (birth and death) and migratory components. These three short introductory chapters lead on to a problem analyzed in Chapter 4: The fact that the mechanisms of growth cannot be understood independently of the age-and-sex structure of a population. Population movement and structure are inseparable: not only is population structure constantly altered by changing fertility, mortality, or migration patterns, but structural changes themselves influence how these factors of change are reflected in the population growth rate.

To understand this interaction, we must first look more closely at each component of change. The approach taken to do that, the *longitudinal approach*, which is the subject of part two, may initially seem to be leading away from the strict definition of a population, but is nevertheless central to demographic analysis. Each phenomenon (fertility, mortality, migration) will be considered as it is actually experienced by the individuals during their life cycle. So, a benchmark set of individuals born in the same year (a *birth cohort*) will be taken to describe their age-specific mortality from observed deaths in that cohort or a cohort of women to profile their age-specific fertility based on successive childbirths. This somewhat extended section will fall into nine chapters: the first discussing the baseline data (Chapter 5), ways of accounting for them (Chapters 6 and 7), and the theoretical and practical bases of measurement (Chapters 8, 9 and 10). The three final chapters will analyze in detail each of the three components of change: mortality (Chapter 11), fertility (Chapter 12), and migration (Chapter 13).

Although the longitudinal approach can accurately describe the components of change and their underlying trend, it cannot account for year-over-year observable changes in a population. Demographers commonly do this by *period* analysis on the basis of the idea of a *hypothetical cohort*. Instead of describing lifelong mortality, fertility, and migration in a real birth cohort, a hypothetical cohort is designed, and its age-specific death, fertility, and migration rates are calculated for the same single calendar year. This is done at the beginning of part three (Chapter 14). *Standardization methods* will then be used to compare total or synthetic indicators (Chapter 15) and perform age-specific and total differential analyses (Chapter 16). The following two chapters will examine the varying degrees of linkage between period and cohort analyses: *translation models* (Chapter 17), and *APC (age-period-cohort) models* (Chapter 18).

Using this array of instruments to analyze population change, part four *consolidates dynamics and structure*. This is done in two chapters in which *population replacement* (Chapter 19) and *population models* (Chapter 20) are discussed. These will show that, in the final analysis, population structure is entirely determined by the history of its changes, because although structure has a strong duration-specific influence on change, it progressively forgets its past.

In order to analyze the fundamental mechanisms of population dynamics, the realities have had to be oversimplified through two basic working assumptions: (1) all the individuals in a population are, if not identical, at least exposed to the same mortality, fertility, and migration risks, and (2) the probabilities of dying, procreating, and migrating are completely independent events. This is obviously not the case in real life. So the analysis needs to be fine-tuned in order to clarify exactly how the *heterogeneity of a population* can distort the measurement of natural (Chapter 21) and migratory (Chapter 22) changes. We will then consider how to go beyond the independence assumption (Chapter 23). It will quickly be seen that the *life history analysis* used to account for interaction of events has to consider multiple possible levels of aggregation, leading to *multilevel analysis* (Chapter 24). This will be a major step toward understanding demographic phenomena, and the approach to the determinants of change, which will be looked at in Section two of this volume and in Volume II.

PART

I

POPULATION GROWTH

CHAPTER

1

Populations and Individuals

JACQUES VALLIN

Institut national d'études démographiques (INED), Paris, France

To a parent's eyes, a newborn baby, especially a firstborn, looks like no other. Genetic science bears out that conviction: The probability of two (nontwin) human beings being identical is vanishingly small, close to zero. And genetics is not all! What goes for the formation of a person's genotype doubtless goes double for personalities—neither psychologists nor ethnologists would take issue with that. Yet, demography basically invites us to disregard these differences. Individuals can only be counted and turned into statistics by working on the assumption that they are all the same.

I. NUMBER AND COMPOSITION

Essentially, demography is simply the *statistical study of human populations*.[1] In statistical terms, a population is basically a group of individuals who fulfill a definition. The population of France, for example, is everyone living in France, regardless of gender, nationality, religion, skin, or eye color. We can look at this set of people in two different ways: size and composition. In the former, they are simply counted, disregarding any attributes that may differentiate one individual from another. The 1999 census, for example, recorded 58,518,748 people living in Metropolitan France; each of these—regardless of age, size, weight, intelligence, and so forth—counts for one and one only. With typical black humor, the cartoonist Reiser[2] thought the world leper population—estimated at the time by the World Health Organization (WHO) as 14 million—would be more accurately assessed by discounting the hands, arms, feet, or legs rotted off by this dreadful disease, which he calculated would reduce the number of lepers to an estimated 12 million. Not so, however! The size of any population is always and only the total number of its constituent individuals, however great the individual differences between them may be.

From the latter angle—composition—differences between individuals are clearly relevant. And yet, the principle remains unchanged. To determine the distribution of the population according a specific attribute or characteristic, *classes* are created to which all the individuals with the same definition, the same statistical *value*, for that attribute are assigned and a frequency table is constructed for the different values assumed by the *attribute* studied. Suppose that we wish to study eye color in the American population and assign four values to the attribute *eye color*: blue, green, chestnut brown, and black. The inhabitants of the American continent will be assigned to one of four classes according to these values and the number of individuals listed in the corresponding class, disregarding all other attributes, will give the frequency of each value. In fact, what we have done is to distinguish, within the American population, four *subpopu-*

[1] So defined by Achille Guillard, who first coined the word demography for the title of his book: *Éléments de statisque humaine ou démographie comparée* (Paris: Guillaumin, 1855).

[2] Jean-Marc Reiser. La léproserie du dimanche. In: *Les Copines.* Paris: Albin Michel, 1981:20.

lations by reference to eye color and measured their size in exactly the same way as the size of the French population was measured in the 1999 census. There is no essential difference between measuring the frequency of each of the four eye colors in this way and the more usual method of counting the population of each state making up the United States. Only the attribute definition changes.

Clearly, many other attributes of varying degrees of complexity may be observed. The most basic of these is obviously gender, for which there can be only two values (male or female). Others, by contrast (e.g., occupation, birthplace), lend themselves to a vast or even unlimited number of values, where continuous variables like age, weight, size, and so forth are concerned. Distributions that intersect multiple attributes may be studied. Cross-studying sex and age, for example, allows us to produce, as we will see in a subsequent chapter (see Chapter 4), a population pyramid that offers many insights not just into a population's demographic history and predicted development, but also on a host of societal issues such as school, work, health, and so forth. We can take this further and cross-study three, four, five . . . *n* variables. The principle will always be the same—statistically, the individuals will be grouped within a single definition to constitute a population or to distinguish multiple subpopulations within the population. All individuals who satisfy this definition are presumed identical, and so available for being counted and computed, although the heterogeneity of the groups so formed may have to be factored in at a more advanced stage of the analysis.

It is clear that by cross-studying a sufficient number of variables, each containing a sufficient number of classes, we could arrive at a subpopulation of one individual. However, it is equally clear that this will not account for the extreme variations between the individuals who make up humankind (from a purely genetic point of view, the possible number of combinations is infinitely greater than the number of individuals currently alive in the world), but this would have taken us beyond the reach of statistical reasoning (and demographic analysis), which is essentially based on the law of large numbers.

II. SIZE, COMPOSITION, AND DYNAMICS: STATE AND CHANGE

What does demography offer us beyond statistics? Each of the previously described aspects (size and composition) can be approached from two angles: state (What is its size and composition at a given moment in time?) and dynamics (How is it and its components changing and what are the driving forces of those changes?). Demography really starts with the second question.

Admittedly, all literature on methods and materials of demography take as their subject size, composition, and dynamics (state *and* change). But demography would not have been needed simply to measure the distribution of a population between its various constituent categories defined by attributes or characteristics to be studied on the basis of census data or sample surveys. The techniques of descriptive statistics were more than enough for that. Analyzing population change and understanding its mechanisms gave rise to a specific body of methods. That said, the study of change might aim to describe a state in situations where the latter is not directly observable. This, indeed, is the origin of demography. Before censuses became a matter of administrative routine and a reliable source of demographic information, there was no easy way to determine even the total population size of a country or province. Our forefathers believed that, under the dominion of the laws of God and nature, whose purpose was to ensure the proper balance between births and deaths, population sizes varied little if at all. Accordingly, they thought that the tangible embodiment of those laws—baptismal and burial registers—was enough to estimate the population and its principal components. Demography, therefore, originated in the myth of population stability and now enables us to comprehend its dynamics. However, other situations exist in which the analysis of change aims to describe a nonobservable state. For example, attempts to forecast the size and composition of the population 10, 20, or 30 or more years hence. The same process is often applied in a bid to reconstitute past populations. Conversely, as will be seen, the development factors of a population during a given period of its history are closely tied to certain aspects of its initial composition. Without change, therefore, demography, would not exist but state *and* movement are central to the analysis.

This kills two birds with one stone. First, regardless of whether all or each discrete component is concerned, the central overarching question that minimizes individual differences will always be the following: How does the size of a population subject to perpetual change through the entries and exits of the individuals comprising that population increase, decrease, or remain stable? But, because entries and exits may be partly dependent on the population composition, it is not enough to say that what is true of the whole population (of France or America, for example) is also true of its components (the blue-eyed subpopulation, women, senior managers, for example). Not only are entries and exits not necessarily of the same nature among subpopulations as in the whole population (one is *born* a woman, one *becomes* a senior manager), but also change in the whole may be dictated by certain elements of its composition. This means distinguishing those constituent elements that are decisive to the future of the total population, and hence essential to demographic analysis, from those that predicate only that of the subgroups. Thus, it will readily be imagined that gender distribution may play an essential role in the population dynamic because there can be no births (the most ordinary *entry* route into a population) without women. By contrast, the proportion of blue-eyed individuals will have no influence and probably interest the genetic scientist more than the demographer. Between these two extremes, some individual characteristics may play an intermediate role. Arguably, for example, the proportion of senior managers is chiefly of interest only to sociologists and economists. But if senior managers were to have a longer survival rate than others, they would *exit* later. Their share would influence total population development. In other words, we cannot stay at the highest level of aggregation indefinitely.

C H A P T E R

2

Population: Replacement and Change

JACQUES VALLIN

Institut national d'études démographiques (INED), Paris, France

In the previous chapter we defined the field of demographic analysis as the study of a human group made up of people corresponding to the same definition, whether it be according to the criterion most commonly used in identifying a population (people residing in a given territory) or to any other criterion making it possible to distinguish, within such populations, subpopulations made up of people sharing a specific characteristic (blue eyes, male sex, manager, residing in a town, etc.). We have also chosen population *movement* or change as the main subject of this analysis. This does not only concern human demography. The statistical approach, in the more general sense of the word, usually concerns populations made up of transient individuals who are continuously replaced through a process of entries and departures, which may, at the same time, change their dimension or modify their composition. According to the nature of populations, however, two types of processes may come into play: those which ensure replacement in an endogenous manner, without calling upon outside help from a neighbouring population (by bringing in new individuals or by exporting excess numbers of individuals), and those, which on the contrary, rely on exogenous exchanges with other populations.

The most common way of maintaining a supply of objects, whatever their nature, is to draw from outside reserves of new individuals and to reject those which have become undesirable. Thus, the grains in a pile of sand which a mason uses to make his cement will be continuously replaced by regular deliveries from outside to make up for the sand used in his work. In the case of goods produced by human beings, instead of drawing on natural reserves one may produce new objects to increase the collection or replace the items that have become undesirable. The production source, however, is not the collection itself, but an outside manufacturer. This is an example of exogenous replacement.

The living world, to which human populations belong, has the capacity for endogenous replacement which, on the contrary, enables a population to engender new individuals itself. However, the individuals in question are mortal and can also leave the population without adding to the ranks of another population.

One of the fundamental aspects of demography, in the wider sense of the term, is that it has to deal with complex modes of replacement associating endogenous and exogenous processes. More precisely though, demography, which deals with human populations, has developed according to the characteristics of this dual process as it occurs among human beings. In fact, in human societies, as in other forms of societies, these replacement processes are subject to constraints imposed by the environment and biology; they are also, to a great extent, controlled by mankind which has managed to create for itself a certain amount of room to maneuver, allowing the expression and influence of individual or collective decisions. And this is the case in both endogenous and exogenous processes.

I. EXOGENOUS REPLACEMENT: ENTRIES AND EXITS BY MIGRATION

The most common definition concerning a population in demography, and which also gives the most meaning to the discipline, is that of a group of individuals residing within the limits of a territory endowed with a social significance. And among the different types of socially organized territories, national territories organized as States contain populations which, by definition, fall within the realm of demography. Upon this basis, the notion of exogenous replacement through entries into and exits from a territory (international migrations) is clear and relatively easy to define, even if certain clarifications remain necessary. By extension, the same principles may just as easily be applied to any other territorially defined population. However, most of the other populations falling within the field of human demography are also subject to an exogenous replacement process, even if the analogy with the entries and exits through a geographical border may seem a little tenuous.

1. National Borders: Migrations and Simple Visits

To define a population in the usual sense of the term (a group of people living within the boundaries of a socially organized territory) and whether or not entries and exits are exogenous, the notion of "living in a territory" needs to be clearly defined. Two notions are possible at this point. One may rely on the simple observation of a physical presence at a given moment in time and therefore all individuals in this territory at that time are considered as belonging to the population. In this case reference is generally made to a "*de facto population*." This might be, at least in appearance, the simplest solution. It would, however, be entering, in a way, in contradiction with the idea that the territory of reference is only interesting if it has a social significance attached to it, if the individuals in it are socially organized. Would it not be preferable to consider as a part of a population only those individuals who truly maintain social links which organize the population? In that case, visitors who find themselves at a given time in the territory and who have no permanent ties with its "*inhabitants*" and who maintain permanent social relations with another population should be excluded. One must therefore define what is meant by "inhabitant." Usually, reference is made to the notion of residence, of permanent place of residence, which generally relies on criteria of duration, sometimes associated with declarations of intention. But, if one only refers to "*permanent residents*", one must also include in the population being studied, permanent residents momentarily outside the territory. In this case reference is made to "*de jure population*" even if the definition of a permanent resident used for statistical purposes does not generally cover a legal category so to speak.

The evaluation of the exogenous replacement process of a population will of course be made according to the type of definition used for the population. If this concerns the de facto population of a territory, any crossing of the border must be considered to be a migration modifying the population. All entries, even those of outside visitors, will increase the size of the population whereas any exit, even for the shortest period of time, will reduce it. On the contrary, in the case of a de jure population, only changes of permanent residence will play a part in the replacement of the population.

2. Entries and Exits at Other Geographical Levels

The need to make a distinction between de facto and de jure population is also valid, and even more so, when the population is defined by another territory than the national territory, since the problems are then often more complex. The specially strong tie which links inhabitants from the same country usually leaves little room for any ambiguity concerning the notion of permanent residence (or at least, ambiguous cases are quite rare). The smaller the internal administrative area considered, the more migratory exchanges will become frequent, diverse and ambiguous: commuting or seasonal movements for work or studies, main residences and second homes, or even multi-residences, etc. So much so that the definition of de jure population, as well as that of migrations to be taken into account for evaluating exogenous replacement, becomes both difficult and necessary. In most cases, important political, social and legal issues are linked to the total size of the population and its variations. Demographic analysis should therefore be able to deal with de jure population as well as de facto population.

The notion of population as defined by a territory to which a social significance is attached may extend to physical entities that can be considered as part of a territory. A micro-demographic approach will, for example, take into consideration the population of a school, a company, or a hospital. The buildings concerned may then be considered as territories, and populations and migratory movements will be defined in relation to the walls of the buildings . . . But the notion of territory tends to give way to that of the status of the individual person. Individuals belong more to the population considered in terms of the functions they fulfil than in terms of their presence in the place, which in most cases is not permanent but lasts only for the time needed for the person to fulfil his or her function. In the same way, exogenous replacement of a given population may be evaluated more in terms of accession to a function or ending an activity than in terms of observed physical moves. It has more to do with migrations from one category to another than with moving from one place to another.

3. Generalization of the Notion of Migratory Entries and Exits: Extension to Non-territorial Populations

The notion of exogenous replacement through entries and exits extends therefore to the case of populations defined without any reference to a territory.

A first group of populations, the definition of which is largely if not totally independent from the notion of territory, concerns groups of people established on the basis of the family, the clan, the tribe, etc. In many traditional societies, especially in nomadic societies but also often among sedentary populations, the territory is not the main criteria of belonging. It is not even necessarily taken into account. Ties of kinship are the main factors for identifying the limits of a population, since it is on the basis of these ties that the society and its political life are built. It is thus not rare to find situations where—depending on geographical migrations, which in this case do not contribute to replacement—several populations, clearly defined socially and politically, overlap in a very complex manner in the same territory and, in spite of this, are perfectly distinguishable from one another. Such populations, in the same way as territorial populations, fall within the province of demographic research.

The contours of such populations having been established by definition according to ties of kinship, one might think that any idea of exogenous replacement by immigration should be excluded. This is not so. In fact, two factors of exogenous replacement must be taken into account: the exchange of women and processes of adoption. Except for the theoretical hypothesis of an entirely endogamous society totally closed to any outside influence, entry within (or exit from) such a population may occur through the acquisition or the loss of kinship ties. More generally, one may shift from the notion of the clan or of the tribe to that of the ethnic group and, without the notion of territory being primordial, define a population according to its belonging to an ethnic group mutually recognised by the people within it. One will therefore have to take into account for the demographic study of such populations their capacity for filling their ranks from external sources or on the contrary for reducing their numbers through changes in ethnic membership. Therefore, in all of these cases and despite all appearances, the criteria for belonging to a population do not exclude changes in membership of a group and therefore exogenous replacement, even though, of course, endogenous replacement remains prevalent.

The notion of populations based on membership of a category is also to be encountered in what may be termed as "*sector-related sub-populations*". From a demographic point of view, specific groups of individuals may be studied within a whole population corresponding to one of the earlier definitions, according to individual characteristics. Sex, for example, makes it possible to make a distinction between a male and a female population. Marital status, social category, level of education, etc. and, why not . . . the color of the eyes, provide criteria for making distinctions between subpopulations, the size, composition and dynamics of which one may wish to study exactly as if it were a population in its own right. This time, replacement will in most cases be essentially, if not entirely, exogenous.

The question of sex should be left aside as an attribute used in defining a population since it is too closely connected with reproduction to be considered as a case in kind. In all of the other cases, the criterion defining a population being an individual characteristic, one must in fact distinguish two possibilities. Either the characteristic is likely to change (marital status, level of education, socioeconomic category) and therefore the main replacement factor of the population will be a change of the characteristic in question, resulting in the individual moving from one category to another and therefore from one subpopulation to another. In this case therefore it is indeed a migration, which may be considered as similar to a geographical migration (change of status, social mobility, etc.). Or on the contrary, the attribute for a given individual cannot change (the color of the eyes) and the population will not be directly subjected to any exogenous replacement process. However, in every case, these sector-related subpopulations are indirectly subjected to the consequences of replacement flows, whether exogenous or endogenous, affecting the parent-population. Through the latter, the results of the conventional processes mentioned above may be observed. The special criterion used to isolate the subpopulation only adds, in certain cases, to a supplementary replacement process which may just as well be exogenous (marital status, for example) as endogenous if the characteristic is a support for a certain endogamy and may be transmitted to the following generation (nationality, religion, for example).

II. ENDOGENOUS REPLACEMENT: BIRTHS, DEATH

For populations defined by a territory in the strongest sense of the term (country, State) or by a clan, migration usually only plays a secondary role in the replacement process. This is essentially endogenous. Notable exceptions such as the Vatican and the Republic of Mount Athos only confirm the rule, since even if these entities are sovereign states, they more closely resemble subpopulations based on individual criteria rather than on subpopulations residing in a territory or belonging to an ethnic group. To a large degree, demographic reasoning is developed within the framework of "*closed populations*," which, although their use is often well-founded, remain purely theoretical since, as we have just seen, most populations are subject to exogenous replacement processes. In many cases not only does the endogenous replacement process dominate human population dynamics but also, and maybe above all, one of the essential characteristics of these populations—and no doubt the one which establishes demography as a discipline in its own right, distinct from statistics—is their endogenous capacity to reproduce.

1. Reproduction and Death

The capacity living people have of reproducing and giving birth to new living beings, similar to them and endowed with the same abilities, fundamentally distinguishes a population, as defined here, from any other collection of objects the replacement of which can only be ensured from the outside. The pile of sand mentioned earlier can only be supplemented by an outside source which itself depends on the (natural or artificial) transformation of larger rocks into grains of sand. The ability to reproduce, which is a characteristic of the living world, is at the heart of demographic analysis, directly associated with its corollary, death. The ability to procreate has as its counterpart the need to die. And in the same way that procreation is fundamentally different from migration or the modification of a characteristic which introduces a pre-existing individual within the population, death differs from the simple loss of an attribute or from emigration to another population since it makes individuals disappear without them being able to fill the ranks of any other population (except for the fictional one of cemeteries). On the contrary, the grain of sand which as it is worn down becomes too small to be of any use to the mason, thus losing its attribute of belonging to the group being studied, emigrates to another group (finer sand or dust). Thus prevalent in human populations as with all living species, is the endogenous replacement process based on reproduction and death, generating specific entry and exit flows which are generally described as *natural*.

Here again, however, a number of points need to be clarified concerning borders. Where exactly are these natural entries and exits to be situated? Concerning death there is only a little margin for hesitation. Doctors have indeed delayed in adopting a universal definition of death, in choosing a decisive symptom (the stopping of the heart obviously being insufficient), and in the past errors have indeed been made with people being buried alive. On the other hand, the discussion is closed in demographic terms: death is definitely the final event by which an individual exits a population.

On the contrary, can one be so sure that birth constitutes the point of entry? Human beings are mammals, the mother carries her offspring in her womb for nine months and can breastfeed them for two years. What is more, children only become adults after approximately fifteen years of life. When can a new human being thus created be considered to have "entered" the population? At his/her conception? After a certain time in the mother's womb? At his/her birth? When weaned? At puberty? The question may seem trivial since we are used to thinking of birth as the event marking the arrival of a new individual. Yet the point could be discussed. Let us suppose one is studying insects, butterflies, for example; egg, caterpillar, pupa, fully-formed insect: which stage marks the arrival of a new living being? The fully-formed insect is nothing more than a reproductive organ. The caterpillar has done all the work of accumulating the reserves needed to ensure the cycle of life is completed. Can one exclude such active individuals from the population? On the other hand, the pupal state closely resembles a new stage in the gestation of the insect (after that of the egg) making it possible to transform the larvae into a magnificent sexual subject. To which population should the entomologist demographer refer to? Should he exclude the eggs and the pupa because they are only beings in gestation? What strange populations in which each individual could die twice and live two lives!

But one must face the fact that for demographers, life only begins at birth. No doubt because what happens before has more to do with the mother's, or the couple's, sphere of influence than with that of society, although society has long denied the mother any choice in deciding whether or not to continue her pregnancy and the law sometimes attributes rights to the *infans conceptus*. However, let us consider once and for all that for the demographer, the point of entry is birth. Let us draw conclusions from this.

On the one hand, birth must be defined. And on the other hand, one must nevertheless study what happens before.

Although entry within a population occurs at birth, the distinction between who is alive and who is not must be made. A stillborn child can never be considered as having belonged to the population. But the law varies in different countries and may even change with time. In France, for example, only children who were alive at the moment when they were declared to the civil registrar are legally considered as born alive. This declaration must be made within three days following the birth of the child. A child may therefore, theoretically, have lived for nearly three days and never have legally belonged to the French population. The deadline of three days is theoretical since most declarations are made on the first day. However, a substantial number of children born alive and having died after a few hours are considered from a legal standpoint as stillborn. Statisticians and demographers are forced to make corrections in order to base their research on births and deaths.

Moreover, to decide to fix the entry within a population at birth does not in any way dispense the demographer from studying what preceded birth: conception, gestation, fetal morbidity and mortality, late fetal mortality, etc. But none of these aspects will be counted in terms of population. They will only be taken into account as being a part of the reproduction process.

The real future of the individual within the population only starts at birth, at age 0.

2. Age

Because of the very nature of the endogenous replacement process, of life and death, from the outset age presents itself as an unavoidable differentiating characteristic between individuals. One might, in fact, perfectly imagine populations to be in perpetual replacement through external exchanges totally independent from age, for the analysis of which the age reached by the individuals would be of no particular interest. On the contrary, with the endogenous replacement process, in a closed population, entry only occurs at age 0 and from then on starts a cycle of life which will command the expression of the capacity to reproduce and set off the natural process of elimination by death. Age soon appears as an indispensable variable for measuring and analyzing these natural entries and exits which are a part of population dynamics.

Age does not only determine the different phases in the life cycle: maturing of childhood leading to puberty and accession to the ability to reproduce, reproductive activity of the adult ages arriving at its term with the menopause, at least among women, biological ageing and wearing of the body, leading ultimately, naturally, to death, unless death occurs sooner because of external events. Age also determines, in counterpoint, the modes of integration of the individual within society: upbringing and education, productive activity, retirement, to mention only the major stages in life as they operate today in our modern societies. Not only is age the key to natural population dynamics but it is also at the heart of demographic, economic and social phenomena. And finally, simply because of its biological and social connections, age is also at the heart of migratory processes. We will clarify the different meanings of the term in Chapter 6.

3. Sex

Lastly, and we will see further on in this treatise at which point this complicates the analyst's task, human reproduction is sexual and sex is therefore a second fundamental variable. It should even be said that it is just as unavoidable as age. Regarding certain points which remain crucial, however, demographic science has so far only been able to skirt around the problems that this simple reality (two people of opposite sex are needed to create a third person) poses for demographers. This will appear clearly in Chapter 20 concerning population models.

Nevertheless, the numerical balance between the sexes will prove to be a decisive factor in the expression of a population's reproductive capacity. But even more important than the numerical balance is the nature of the psychological, social, economic and political relations between the sexes and, notably, the rules governing the formation and the separation of couples, the respective roles of men and women in the family and society, etc. Without mentioning the fact that, according to their gender and the role they play, individuals have demographic behaviors which differ in terms of reproduction, health, death and migrations.

C H A P T E R

3

Population Increase

GUILLAUME WUNSCH, JACQUES VALLIN, AND GRAZIELLA CASELLI

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

At each instant of time, a population is composed of a certain number of individuals: the population *stock* as economists would say, or the population *prevalence* according to the terminology in epidemiology. This number changes with time under the impact of inputs (births or immigrations) and outputs (deaths or emigrations) (i.e., *flows* in economic terms or *incidences* in epidemiology).

Before distinguishing between different types of flows, first consider the total population increase. Afterward we will distinguish between *natural* flows as they are called (births, deaths) and *migratory* flows[1] (immigrants, emigrants).

I. TOTAL POPULATION INCREASE

Let P_t and P_{t+h} represent population numbers at two different dates, times t and $t + h$. The total population increase between those two dates is as follows:

$$A_{t,t+h} = P_{t+h} - P_t$$

For example, between the census of March 4, 1982, and that of March 5, 1990, population size of France changed from 54.3 million to 56.6 million inhabitants, an absolute increase of 2.3 million. However, the comparison between this increase in absolute numbers with that observed between the last two censuses in China or in Luxembourg would be meaningless. The intercensal periods are not the same. In particular, the population numbers are very different. To compare these increases, one must convert to the same unit of time and population size.

1. The Time Unit

The year is the most common time unit used in demography. However, the study may also cover periods of several years, often rather arbitrarily taken as 5-year periods, or less frequently periods shorter than a year to examine seasonal variations and so forth. In the latter case, months, weeks, or even days can be taken as the unit of time, as for example in the study of infant mortality. On the other hand, few demographic studies use data collected by the hour, minute, or second, although the latter time intervals are often used in other fields, such psychology or physics, which uses time units ranging from a billion years to the nanosecond.[2] The length of the period should be chosen for its relevance to the problem at hand[3]: a 20-year period for the study of mortality would conceal too many important temporal variations, whereas a 20-minute period would not be pertinent for most, if not all, demographic studies, and it would also lead to a very small number of deaths and therefore to a lot of chance fluctuations (or random

[1] More precisely, one should distinguish between the moves themselves (migrations) and the people who actually move (migrants). A same person can move several times during a given period (see Chapter 22, this volume, by Daniel Courgeau).

[2] One billionth of a second.

[3] On this topic, see the book by Janice R. Kelly and Joseph E. McGrath (1988).

noise) in the data. In many studies, the demographer has no control on the length of the time unit, especially if he or she uses data collected by others, such as from administrative sources or existing survey data. For example, data on internal migration might be collected and published by year. In this case, demographers interested in possible seasonal patterns of migration would not be able to obtain needed data. One could interpolate the yearly data to evaluate the number of migrations by month, but this would require knowing the seasonal migration pattern—the object of the study itself!

2. The Rate of Increase

To simplify, consider a 1-year unit of time and suppose one knows the population numbers P_t on January 1, year t, and P_{t+1}, on January 1, year $t + 1$. The rate of increase is obtained by dividing the difference $P_{t+1} - P_t$ by the population size. What denominator should one take, as the size of the population is continuously changing over the year?

One could use the population size at the start of the year, P_t, and compute the yearly *relative increase* as follows:

$$a = \frac{P_{t+1} - P_t}{P_t}.$$

Supposing this relative increase remains constant over time, the population P would increase every year by the quantity aP. Starting from year 0, the population size 1 year later would be:

$$P_1 = P_0(1 + a)$$

and 2 years later:

$$P_2 = P_1(1 + a) = P_0(1 + a)^2$$

Finally, t years later, one would have:

$$P_t = P_0(1 + a)^t$$

Coming back to the intercensal increase of the French population, one can use this formula to compute the average relative increase per year during this period, as one has:

$$a = \sqrt[t]{\frac{P_t}{P_0}} - 1$$

and with the French data of the 1982 and 1990 censuses:

$$a = \sqrt[8]{\frac{56.6}{54.3}} - 1 = 0.005199 \cong 0.5\%$$

Actually, this yearly average increase is not the true measure of the rate demographers wish to estimate. Dividing the absolute increase by the population at the beginning of the year does not take into account the fact that the population is itself the source of events (births, deaths, migrations) that lead to population change. As the size of the population continuously changes during the year, the number of events produced each day by this population also changes. Therefore, at the end of the year, the net population increase (or decrease) does not depend only on the size of the population at the beginning of the year but also of its successive modifications during the year. One should therefore divide the absolute change by an average of all population sizes observed during the year. If the population does not fluctuate too erratically during the year, an estimate of this rate can be based on the simple arithmetic average of the population at the beginning and that at the end of the year. One obtains for the year t the *rate of population increase r*:

$$r = \frac{P_{t+1} - P_t}{(P_t + P_{t+1})/2}$$

If the period of observation is greater than a year, one simply multiplies the denominator by the number of years concerned. For example, the average rate of population increase per year of the French population between the censuses of 1982 and 1990, a period of 8 years, is equal to:

$$r = \frac{56.6 - 54.3}{8(54.3 + 56.6)/2} = 0.005185 \cong 0.5\% \text{ per year}$$

One sees that the difference between this result and the one obtained previously is very small. One can show that for levels of the rate of annual increase of less than 5%, the difference between the results of the two formulas is negligible. Human growth rates for total populations usually remain below 5% per annum, except for certain subpopulations where this rate can be exceeded. For example, one knows that at the beginning of an autoimmune deficiency syndrome (AIDS) epidemic, the population affected by this disease doubles every 2 years on average. In this case, the relative annual increase is 100% but the rate of increase is significantly lower:

$$r = \frac{1}{(1+2)/2} = 67\%$$

When computing the relative increase a one assumes that all the flows depend on the initial population P_t. In the case of the rate of increase r, taken as the ratio of the absolute change to the average of the populations at the beginning and end of the year, one implicitly assumes that the flows going in or out of the population occur at mid period or that they are evenly spread over the period. In other words, one assumes that the population size changes linearly over the year or period. This assumption is not truly correct, as the population is continuously changing and, if the rate of increase is constant, the number of incoming and outgoing events increases (or decreases) relative to the population size.

Thus, as a result of these discrepancies, a precise definition of population growth rate should be based not on discrete units of time but on the continuous values taken by the size of the population over time and the *instantaneous growth rate* occurring during an infinitesimal fraction of time.[4] In other words, we wish to define the rate of increase in the case of a population size changing between times t and $t + \Delta_t$ when Δ_t tends toward 0. If $r(t)$ is the instantaneous rate of increase, one has:

$$r(t) = \lim_{\Delta_t \to 0} \frac{P(t) - P(t+\Delta_t)}{P(t) \cdot \Delta_t}$$

or

$$r(t) = \frac{\mathrm{d}P(t)}{P(t)\mathrm{d}t}$$

which can also be written:

$$\mathrm{d}P(t) = P(t)r(t)\mathrm{d}t$$

Integrating, one obtains:

$$P(t) = P(0)e^{\int_0^t r(\tau)\mathrm{d}(\tau)}$$

Finally, in the finite interval of time $(t, t + h)$, assuming that $r(t)$ is constant over the whole interval, one has:

$$P_{t+h} = P_t \cdot e^{r*h},$$

showing that under this assumption the population grows exponentially during the interval of time $(t, t + h)$.

According to this definition, the rate of increase r^* is thus equal to:

$$r^* = \frac{\ln(P_{t+h}/P_t)}{h}$$

and in the example of the French population, one has:

$$r^* = \frac{\ln(56.6/56.3)}{h} = 0.005186 \cong 0.5\%$$

The difference with r is very small, which is why one usually approximates the growth rate, by convenience, using the average population as denominator.

From one country to another, the growth rate presently varies from –1% per annum to +3.5%. Even within a country the rate can vary greatly for a number of reasons. Crises sometimes lead to dramatic declines in population sizes and long troubled periods of time can lead to durable population reductions. Alternatively, some developing countries recently have experienced periods of very high population growth, with rates above 4% per year. However, total increase rates can be much higher for subnational populations due to migration, which in this case is an important component of growth. This is clearly the case for the rate of increase of the population of urban areas in many less developed countries. For example, the urban population of Tanzania grew at the rate of 6.1% per year during 1990 to 1995, while its rural population grew at the rate of only 2% (United Nations, 1994).

3. Doubling Time of a Population in an Exponential Model

Knowing the annual rate of increase of a population, one can calculate the time needed for a doubling of the population size assuming a constant growth rate. Using the following formula:

$$2P = Pe^{rh},$$

taking logarithms, one obtains:

$$h = \frac{\ln 2}{r} \cong \frac{0.693}{r}.$$

For example, with the average annual growth rate of the French population of the period 1982 to 1990, this population would double every 0.693/0.005 = 138.6 years.

Table 3–1 gives the doubling time of a population for various values of r. Thus, with a rate of growth of 3.5% per annum, a population would double every 20 years. With a population of 10 million in the year 2000, this population would increase to 20 million in the year 2020, to 40 million by the year 2040, to 80 million in 2060, and so forth. A small growth rate can therefore lead to a tremendous increase in population size over a relatively small number of years. Using the growth rate of the Tanzanian urban population as constant, one obtains a doubling time of only 11.4 years. In this case, the population size would be multiplied by 500 in a century!

[4] Though basing the discussion in the present chapter on the number of individuals and events observed during discrete time intervals, it is important to develop at the same time the relationship using calculus, in order to take advantage of the mathematical convenience and power of the latter. Those not familiar with calculus may skip this paragraph if necessary.

TABLE 3–1 Doubling time of a population according to annual growth rate

Annual rate of growth, *r*	Doubling time, y
0.005	139
0.010	69
0.015	46
0.020	35
0.025	28
0.030	23
0.035	20
0.040	17
0.050	14
0.060	12
0.070	10
0.080	9
0.090	8
0.100	7

Exponential population growth of this type led Thomas Robert Malthus (1798) to argue that population has a natural tendency to increase after a geometric (exponential) progression, outstripping the means of sustaining it which grow only following an arithmetic (linear) progression. Although this argument will be examined in Chapter 97, one can already point out that populations never experience constant growth rates for long. Other models of population growth, such as the *logistic* one discussed below, are therefore more suited to actual population trends.

4. The Logistic Model of Population Growth

Some authors propose replacing the exponential growth model with one that supposes that population experiences an accelerated growth over time that progressively decelerates and tends toward 0. This so-called logistic function equation is as follows:

$$P_t = \frac{k}{1+e^{a+bt}} \text{ with } b < 0,$$

where k, a, and b are constants. The Belgian mathematician Pierre-François Verhulst introduced this equation in the 19th century. The equation was then reinvented in the 20th century by the Americans Raymond Pearl and Lowell Reed.[5] If one has three observations in time equally separated from each other, one can compute the values of the constants k, a, and b characterizing the logistic function adjusted to these three time points (Duchêne and Vilquin, 1992). A logistic population increase is probably more suited to sustainable economic development than the exponential model, which leads to implausible results in the long run.

[5] See the interesting discussion by Piero Manfredi (1995).

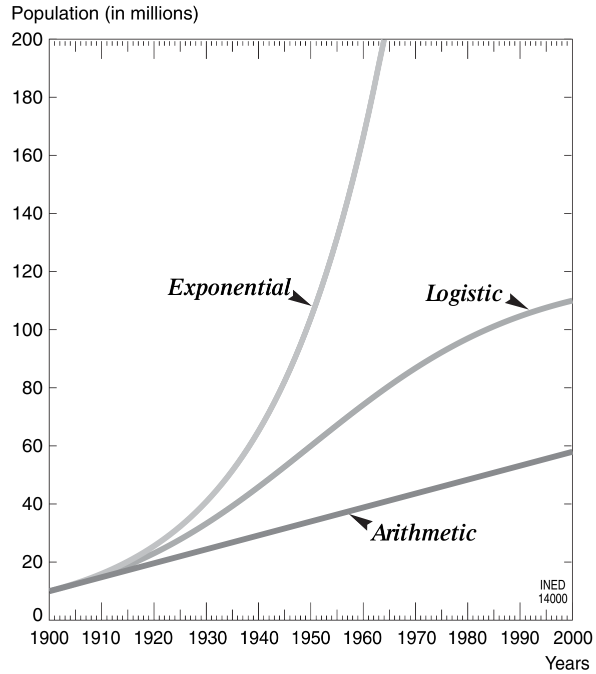


FIGURE 3–1 Evolution over a period of 100 years of a population of 10 million inhabitants with the same initial growth rate of around 5% per annum but following three different models of growth.

Figure 3–1 illustrates the case of a population having a size of 10 million inhabitants in 1900 and a rate of growth of around 5% that year. According to the results of an arithmetic (linear), logistic, or exponential (geometric) growth model, the population would reach in the year 2000 respectively 56 million, 110 million, or 1.1 billion inhabitants.

To conclude this topic, the growth rate is a useful summary indicator of population dynamics, as it shows the differences in population change over time or space. However, population growth results from the combination of various entries into and departures from the population, which must be disentangled to improve one's knowledge of population trends and development of better models of population evolution.

II. NATURAL AND MIGRATORY INCREASE

In the general case of a population of a country or region, there are two types of entries and departures, of inputs and outputs: birth and immigration and death and emigration. During a 1-year period, total population increase can thus be written:

$$A_{t,t+1} = P_{t+1} - P_t = B_{t,t+1} - D_{t,t+1} + I_{t,t+1} - E_{t,t+1}$$

where $B_{t,t+1}$, $D_{t,t+1}$, $I_{t,t+1}$, and $E_{t,t+1}$ represent the number of births, deaths, immigrations, and emigrations, respectively, observed during year t (between January 1, year t, and January 1, year $t + 1$).

One sees that population P increases by the two quantities $(B - D)$ and $(I - E)$. The first is called *natural increase* and results from the two so-called natural components of population growth (i.e., births and deaths), whereas the second is called *net migration* and is simply the balance of immigrations and emigrations. Population changes can therefore be classified into those due to natural increase (or decrease) and those due to net migration, both being the difference between entries and departures. In the case of a *closed population*, without migration, population change results, by definition, from natural increase only. In actual populations, change occurs from both natural increase and net migration, the latter being in some situations more important than the former, such as in the case of many urban populations.

Data on stock or prevalence are usually obtained from censuses or population registers; where they exist, data on flows or incidence such as births, deaths, and migrations come from systems of continuous (e.g., civil [or vital]) registrations or population surveys (see Chapter 121). In countries where the civil registration system works well, current data on births and deaths are routinely collected. If the vital registration system is defective, as in most developing countries, information on births and deaths is obtained from population surveys. It usually is more difficult to obtain reliable data on migrations, especially international migrations. Several countries now have population registers incorporating changes of residence, which yield adequate information on moves inside the country at least. If this type of registration does not exist, the number of migrations can be evaluated thanks to retrospective questions at a census or at a survey (see Chapter 22). In the absence of such data, net migration can be computed as the difference between total intercensal growth and natural increase during the same period. In fact, many developed countries now estimate annually their population size at the beginning or end of each year and the amount of net migration occurring during the year. Some even evaluate their annual flows of immigrations and emigrations.

The relative level of the four flows of entries and departures is measured by *crude rates* of natality,[6] mortality,[7] immigration, and emigration, obtained by dividing the number of events (births, deaths, immigrations, emigrations) during a period by the average population size, following the same approach as for computing the total growth rate of the population. As a first approximation, in discrete terms, one may assume the equal distribution of flows during the year. In that case, the crude birth rate, b, during the year is taken as:

$$b = \frac{B_{t,t+1}}{(P_t + P_{t+1})/2}$$

Replacing the numerator by the number of deaths, D, immigrations, I, or emigrations, E, one obtains the *crude death rate*, d and the crude rates of immigration, i, and emigration e.

Although computed here for a 1-year period (t, t + 1), all these rates can also be derived for longer or shorter periods (t, $t + h$), as was the case for the rate of increase. To bring these rates back to a 1-year period, for comparative purposes, one weights the denominator or average population of the rates by the time interval h expressed in years or fractions of years:

$$b = \frac{B_{t,t+h}}{h(P_t + P_{t+h})/2}$$

The difference between the crude birth and death rates gives the *rate of natural increase*, whereas the difference between the crude rates of immigration and emigration yields the *net migration rate*. One observes, this time at the level of rates, the same decomposition into the two components of change:

$$r = (b - d) + (i - e).$$

Thanks to vital statistics data produced by the civil registration system and to the yearly estimates of population size at the beginning or end of the year, evaluated by the national statistical institutes, one can compute the values of the annual crude birth and death rates (at least in the developed countries) and, taking their difference, that of the rate of natural increase.

In France, for example, in 1989 the crude birth rate was equal to:

$$b = \frac{765{,}473}{(56{,}017{,}000 + 56{,}303{,}000)/2} = 0.01363 \cong 13.6/1{,}000$$

and the crude death rate to:

$$d = \frac{529{,}283}{(56{,}017{,}000 + 56{,}303{,}000)/2} = 9.4/1{,}000$$

[6] Or *crude birth rate*.

[7] Or *crude death rate*.

The rate of natural increase was therefore equal, for the same year 1989, to $b - d = 13.6 - 9.4 = 4.2/1{,}000$.

One cannot usually do the same for migration, in the absence in most countries of a reliable registration system of in- and out-migrations. Often one falls back on estimating intercensal net migration by subtracting natural increase from total intercensal population growth. Similarly, one obtains a rate of net migration from the population growth rate over the period and the crude birth and death rates on the basis of the so-called *growth balance equation*:

$$s = i - e = r - (b - d)$$

One notices that the same net migration value can be produced by very different amounts of flows. For example, a net migration of ten moves can result from the difference 20 immigrants minus ten emigrants or from the difference 10,010 immigrants minus 10,000 emigrants. In fact, as Andrei Rogers (1990) has pointed out, there is no such thing as a *net migrant*: the latter results from the difference between two flows.

Similarly to the rate of increase r, crude rates b, d, i, and e of natality (crude birth rate), mortality (crude death rate), immigration, and emigration, respectively, can also be defined with greater precision using calculus (Keyfitz 1968, Pressat 1995).

In the time interval t to $t + \Delta t$, population increase results from the number of births $B(t,t + \Delta t)$, deaths $D(t,t + \Delta t)$, immigrations $I(t,t + \Delta t)$, and emigrations $E(t,t + \Delta t)$ occurring during the interval:

$$\begin{aligned} P(t + \Delta t) - P(t) = {} & [B(t,t + \Delta t) - D(t,t + \Delta t)] \\ & + [I(t,t + \Delta t) - E(t,t + \Delta t)] \end{aligned}$$

When Δt tends toward 0, one may write for the number of births:

$$B(t) = \lim_{\Delta t \to 0} \frac{B(t,t+\Delta t)}{\Delta t}$$

The number of births occurring during the very small interval of time $(t, t + \mathrm{d}t)$ is therefore $B(t)\mathrm{d}t$ and is dependent on the conditions of time t and the population size $P(t)$ at that instant. This number of births divided by the number of person-years in the interval $(t, t + \mathrm{d}t)$ (i.e., $P[t]\mathrm{d}t$) yields the *instantaneous birth rate* (or *force* of natality) $b(t)$ at times t:

$$b(t) = \frac{B(t)}{P(t)}$$

Assuming that the population size varies exponentially during the finite time interval t, $t + h$, and that the instantaneous birth rate is constant over the same interval, the crude birth rate b^* over the same interval is equal to:

$$b^* = -\ln\left(1 - \frac{B_{t,t+h}}{P_t}\right)$$

Coming back to the French example, one obtains:

$$b^* = -\ln\left(1 - \frac{765{,}473}{56{,}017{,}000}\right) = 0.01376$$

Once again, this result is not very different from the value of b obtained previously and this is the reason why, for convenience, one usually computes b instead of b^*. The smaller the finite interval, the closer the approximation.

Similarly one can compute the instantaneous rates or forces $d(t)$, $i(t)$ and $e(t)$ of mortality, immigration, and emigration under the same assumptions and the corresponding crude rates d^*, i^*, and e^* during a finite interval of time.

The instantaneous rate of growth of the population $r(t)$ can then be shown as follows:

$$\frac{dP(t)}{P(t)dt} = r(t) = [b(t) - d(t)] + [i(t) - e(t)],$$

and therefore

$$r^* = (b^* - d^*) + (i^* - e^*)$$

which is similar to the relation derived previously for the crude rates.

Among national populations, the crude birth rate varies from 10 to 45 per thousand. However, in the recent past, some developing countries have had crude birth rates in excess of 50 per thousand. Yet, birth rates as low as those prevailing today in various developed countries have never been observed in the past. The crude death rate, varying today between 5 and 20 per thousand, has always been much higher. If a crude death rate of more or less 40 per thousand was usual in normal times, it could reach much higher figures in times of crisis. For example, in Finland in 1868, it reached 80 per thousand during a severe famine. In Ukraine in 1933, it nearly reached 100 per thousand for the same reason.

Crude birth, death, and migration rates do not adequately reflect population behaviors as the latter are age dependent and the crude rates are influenced by the age composition of the population. More refined measures of population growth therefore have to take the age structure of the population into account.

References

Duchêne Josianne and Vilquin Éric, 1992. *Mathématiques pour démographes*. Louvain-la-Neuve, Academia, 224 p.

Hillion Alain, 1986. *Les théories mathématiques des populations*. Paris, Presses Universitaires de France, 128 p.

KELLY Janice R. and MCGRATH Joseph E., 1988. *On time and method*. Newbury Park, Sage, 149 p.

KEYFITZ Nathan, 1968. *Introduction to the mathematics of population*. Reading (Massachusetts), Addison-Wesley, 450 p.

MANFREDI Piero, 1995. Some unknown episodes in the history of population dynamics concerning the logistic law: the lesson of the great epidemiologists. *Polish Population Review*, vol. 7, p. 41–52.

PRESSAT Roland, 1995. *Éléments de démographie mathématique*. Paris, AIDELF, 279 p.

ROGERS Andrei, 1990. Requiem for the net migrant. *Geographical Analysis*, vol. 22(4), p. 283–300.

UNITED Nations, 1994. *Urban and rural areas 1994*. New York, United Nations, 27 p.

VALLIN Jacques, 1992. *La démographie*. Paris, La Découverte, 128 p. (Repères, 105).

CHAPTER

4

Population Dynamics: Movement and Structure

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

In the absence of migration, variations in a population's size are the result of the algebraic sum of births and deaths, and its growth rate is equal to the difference between its crude birth rate (CBR) and its crude death rate (CDR).

In such a population, known as a *closed* population, renewal depends on the two components of natural increase, and one might be tempted to consider that CBR and CDR are synthetic indexes that are representative of its propensity to, respectively, procreate new individuals and to lose population. In fact, procreation and mortality, which determine the evolution of a closed population, are the outcome of a complex set of behaviors closely linked to the sex and age of the individuals: to conceive a child, one must be a female of a certain age and have been fertilized by a male; the risks of dying vary across the life course and also between men and women.

Each life history, from birth to death, is written into a biologic time frame that governs both the capacity to procreate and the capacity to stay alive. Even if one considers that all individuals have exactly the same life history, two populations of same size may thus produce different numbers of births and deaths and different birth and death rates, since this will depend on the age–sex structure of the individuals who form the population. The birth rate will be higher when the proportion of women of childbearing age is higher, the death rate will be higher when the proportion of elderly is higher. Thus, in a given population, the propensity to procreate and the risks of dying are, at each moment, translated into births and deaths through the filter of the age–sex structure of the population present at that particular moment. Fertility and mortality dynamics therefore depend, indirectly, on changes in the age–sex structure.

Inversely, changes in the birth and death rates will modify the population's age–sex structure. An increased birth rate will immediately raise the share of infants in the population, and subsequently, as they age, the share of adults then of the elderly. Similarly, a sharp infant mortality decline will produce a similar effect by immediately increasing the share of child survivors, whereas a reduction of mortality risks at older ages will directly increase the share of the elderly. One can also imagine changes in birth and death patterns modifying the population's sex structure. Regarding births, this is no longer pure science fiction, because it is no longer impossible to select offspring on the basis of gender. Regarding deaths, the trends may differ for men and women, not merely in war years, thereby altering the proportions of men and women in the population.

Such modifications in the population's age–sex structure will in turn have an impact on the crude birth and death rates. For instance, the inflated birth cohorts mentioned previously will, given a same reproductive behavior, produce more births when they reach childbearing age than the preceding cohorts, and will, given the same mortality risks, produce more deaths, notably at older ages.

Thus, at each moment, the evolution of a closed population is determined by a continuous process of interactions between the three key elements of its natural dynamics: propensity to procreate, mortality and age–sex structure. The pattern is naturally more

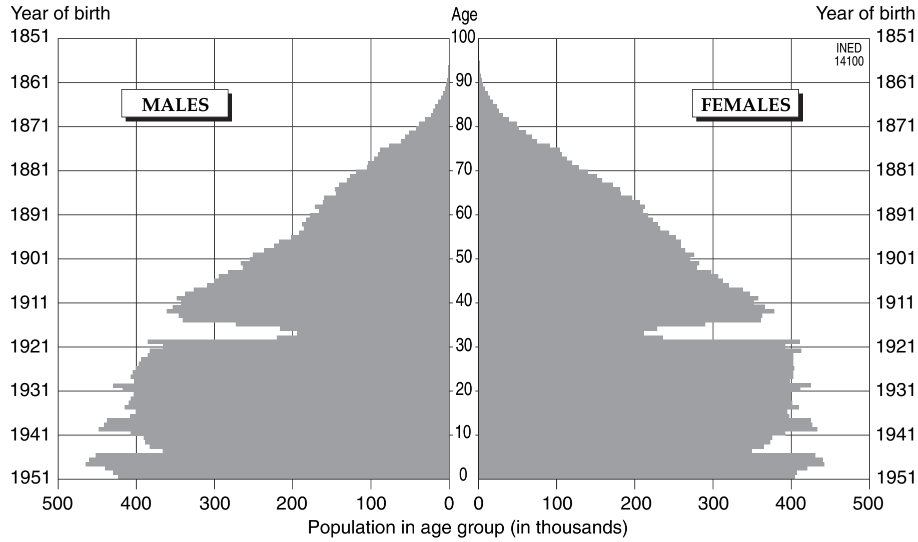


FIGURE 4–1 Population pyramid, Italy, 1951.

complicated when in- and out-migration intervene, but the principle remains the same: Migration also affects the age–sex structure, which in turn acts upon migration.

I. THE AGE–SEX STRUCTURE

To visualize the features of a population's sex and age structure, demographers conventionally construct its *population pyramid*. The definition and utilization of this tool are precious for revealing a number of phenomena and suggesting possible interpretations. We shall therefore make a general presentation of the population pyramid, which will then lead us naturally to the specific issue of the balance between males and females.

1. The Population Pyramid

The population pyramid, whose popularity has spread outside the realm of demography, is constructed by placing side by side the histogram representing the male age structure, generally on the left, and the one representing the female age structure, on the right (Figure 4–1).

a. A Pyramid by Individual Years of Age

For each sex, age is represented on the *y*-axis and the size of the population in the age group on the abscissa. Note that whereas the female pyramid reads normally, from left to right, the male pyramid reads from right to left. Figure 4–1 illustrates the pyramid for the 1951 census population in Italy.

In fact, the Italian 1951 census data were only published by individual years of age from 0 to 79, then by age groups (Istituto Centrale di Statistica, 1956). To construct the complete pyramid, we estimated the population of each individual age beyond 80 by using the year-of-birth distribution published together with the census findings.

b. Unequal Age Groups

Indeed, the complete distribution of the population by sex and individual years of age is frequently lacking, and there is rarely another source of data to fill in the gaps. However, a population pyramid illustrating the structure of the population can be constructed by taking variations in the age scale into consideration. To do so, the population in each age group will be weighted by a coefficient, which is the inverse of the ratio of the size of the age group to the reference age group. Consider the example of the 1951 Italian census data. In Table 4–1 and Figure 4–2, we have grouped them into the unequal age groups commonly used to present census and survey results. In this case, one divides the number in each class by the size of the age group (e.g., the population number in the 1–4 age group will be divided by 4, the number in the 5–9 age group by 5, etc.). For the open age group 85+, we assume that the number of centenarians can be neglected here and divide the number by 15 (i.e., 100 – 85).

c. Proportional Distributions

In absolute numbers, the population pyramid illustrates the age–sex structure while also giving an idea

TABLE 4–1 Italian population by sex and unequal age groups, 1951

Age groups	Population of age groups		Coefficient	Values plotted on the abscissa	
	Male	Female		Male	Female
0	423,509	404,004	1	423,509	404,004
1–4	1,795,191	1,709,419	4	448,798	427,355
5–9	1,981,085	1,892,815	5	396,217	378,563
10–14	2,141,442	2,074,399	5	428,288	414,880
15–19	2,036,794	2,003,583	5	407,359	400,717
20–24	2,062,464	2,038,630	5	412,493	407,726
25–29	1,959,869	2,023,809	5	391,974	404,762
30–34	1,381,717	1,476,727	5	276,343	295,345
35–39	1,675,240	1,757,365	5	335,048	351,473
40–44	1,664,589	1,714,050	5	332,918	342,810
45–49	1,409,080	1,476,825	5	281,816	295,365
50–54	1,182,641	1,327,293	5	236,528	265,459
55–64	1,787,043	2,220,770	10	178,704	222,077
65–74	1,204,453	1,448,840	10	120,445	144,884
75–84	488,117	593,595	10	48,812	59,360
≥85	65,571	94,608	15*	4,371	6,307
Total	23,258,805	24,256,732			

*Since the number of centenarians is negligible here, the group 85+ is taken to extend from 85–100 years.

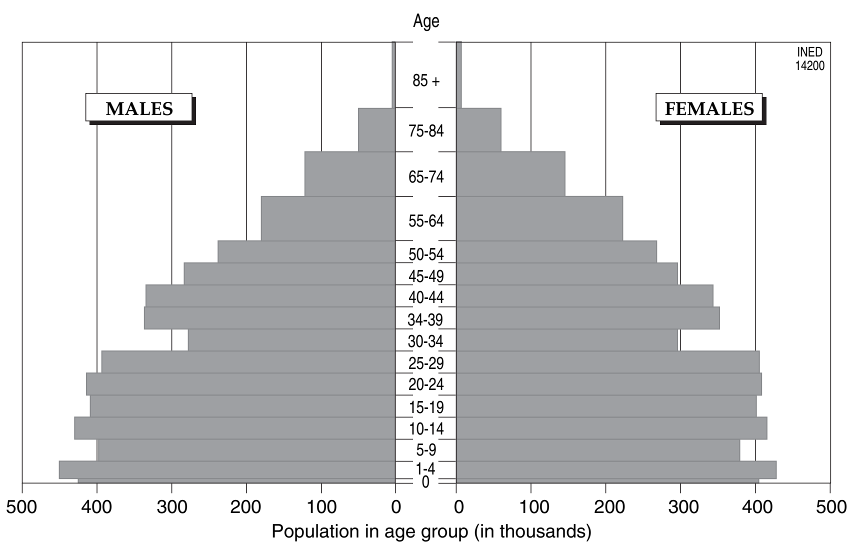


FIGURE 4–2 Population pyramid, Italy, 1951, by unequal age groups.

of the total population size. This may be useful for comparing different populations or the same population across time. However, it may also hamper the comparison if one wishes to focus on differences in the age–sex structure. In this case, a proportional distribution is preferable to absolute numbers (Table 4–2, Figure 4–3A and 3B).

The caveat here is that the population of each age and sex should be related to the total population for both sexes combined. Otherwise, if related to the total population of men or women separately, the result would be two independent histograms and the image of the sex-specific population structure would be lost.

Figure 4–3A and 3B contrasts these two ways of displaying the three sets of country data for comparison. When absolute numbers are used, the differences in population size are so great that this dominates the image, which remains clear even when the three pyramids are superposed (Figure 4–3A). However, the differences in age–sex structure are flattened. By giving a

TABLE 4–2 **Distribution by age groups of the populations of India, Russia and France in 1995**

Age group	Absolute numbers (in thousands)			Proportional distribution (%)	
	Male	Female	Total	Male	Female
	India				
0–4	57,418	53,967	111,385	6.18	5.81
5–9	57,089	53,278	110,367	6.15	5.73
10–14	53,487	49,480	102,967	5.76	5.33
15–19	46,962	43,291	90,253	5.06	4.66
20–24	44,747	40,625	85,372	4.82	4.37
25–29	40,712	36,942	77,654	4.38	3.98
30–34	36,152	32,765	68,917	3.89	3.53
35–39	31,530	28,520	60,050	3.39	3.07
40–44	26,184	23,818	50,002	2.82	2.56
45–49	20,991	19,920	40,911	2.26	2.14
50–54	17,336	17,122	34,458	1.87	1.84
55–59	14,682	14,870	29,552	1.58	1.60
60–64	12,022	12,309	24,331	1.29	1.32
65–69	8,914	9,291	18,205	0.96	1.00
70–74	5,938	6,393	12,331	0.64	0.69
75–79	3,455	3,885	7,340	0.37	0.42
80–84	952	1,114	2,067	0.10	0.12
85–89	595	696	1,292	0.06	0.07
90–94	357	418	775	0.04	0.04
95–99	238	279	517	0.03	0.03
100+	119	139	258	0.01	0.01
Total	479,881	449,122	929,003	51.66	48.34

Age group	Absolute numbers (in thousands)			Proportional distribution (%)	
	Male	Female	Total	Male	Female
	Russia				
0–4	4,205	3,987	8,192	2.84	2.69
5–9	6,040	5,799	11,839	4.08	3.92
10–14	5,956	5,769	11,725	4.03	3.90
15–19	5,463	5,305	10,767	3.69	3.59
20–24	5,234	4,932	10,166	3.54	3.33
25–29	4,850	4,659	9,508	3.28	3.15
30–34	6,043	5,963	12,007	4.09	4.03
35–39	6,366	6,421	12,787	4.30	4.34
40–44	5,733	5,946	11,679	3.87	4.02
45–49	4,245	4,558	8,803	2.87	3.08
50–54	2,874	3,376	6,250	1.94	2.28
55–59	4,285	5,364	9,649	2.90	3.63
60–64	2,959	4,123	7,082	2.00	2.79
65–69	2,813	4,754	7,567	1.90	3.21
70–74	1,140	2,997	4,137	0.77	2.03
75–79	618	1,874	2,492	0.42	1.27
80–84	459	1,688	2,147	0.31	1.14
85–89	156	712	868	0.11	0.48
90–94	39	200	239	0.03	0.13
95–99	4	18	22	0.00	0.01
100+	3	12	15	0.00	0.01
Total	69,487	78,452	147,939	46.97	53.03

Age group	Absolute numbers (in thousands)			Proportional distribution (%)	
	Male	Female	Total	Male	Female
	France				
0–4	1,855	1,771	3,626	3.20	3.05
5–9	1,963	1,876	3,840	3.38	3.23
10–14	2,006	1,915	3,921	3.46	3.30
15–19	1,932	1,852	3,785	3.33	3.19
20–24	2,175	2,123	4,298	3.75	3.66
25–29	2,151	2,130	4,281	3.71	3.67
30–34	2,185	2,196	4,380	3.77	3.78
35–39	2,134	2,161	4,296	3.68	3.72
40–44	2,141	2,145	4,286	3.69	3.70
45–49	2,042	2,010	4,052	3.52	3.46
50–54	1,416	1,401	2,817	2.44	2.41
55–59	1,390	1,444	2,835	2.40	2.49
60–64	1,387	1,534	2,921	2.39	2.64
65–69	1,227	1,476	2,703	2.11	2.54
70–74	1,045	1,403	2,448	1.80	2.42
75–79	445	679	1,124	0.77	1.17
80–84	463	867	1,330	0.80	1.49
85–89	218	525	743	0.38	0.90
90–94	65	212	277	0.11	0.37
95–99	9	44	53	0.02	0.08
100+	1	5	6	0.00	0.01
Total	28,252	29,768	58,020	48.69	51.31

Source: For India and Russia, United Nations 1997; for France, Kerjosse, and Tamby, 1997.

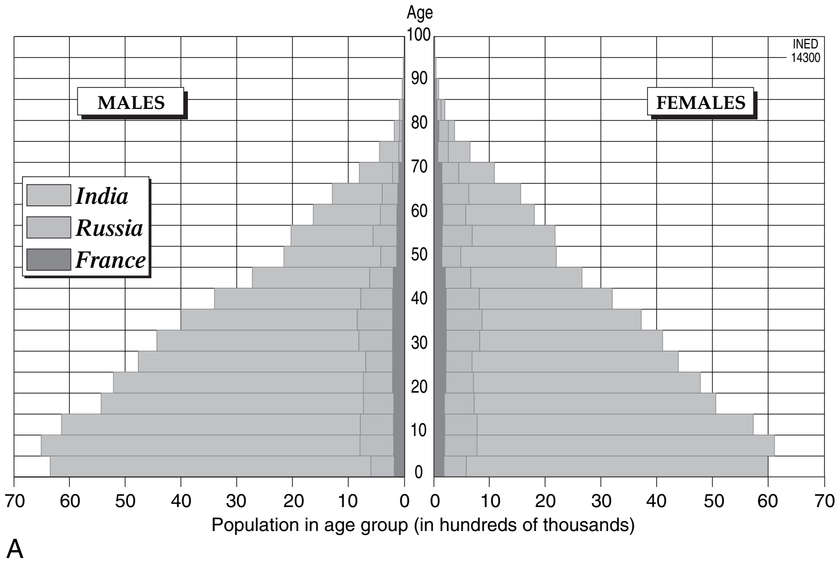


FIGURE 4–3 **A:** Comparison of three population pyramids using absolute numbers (India, Russia, and France, 1995).

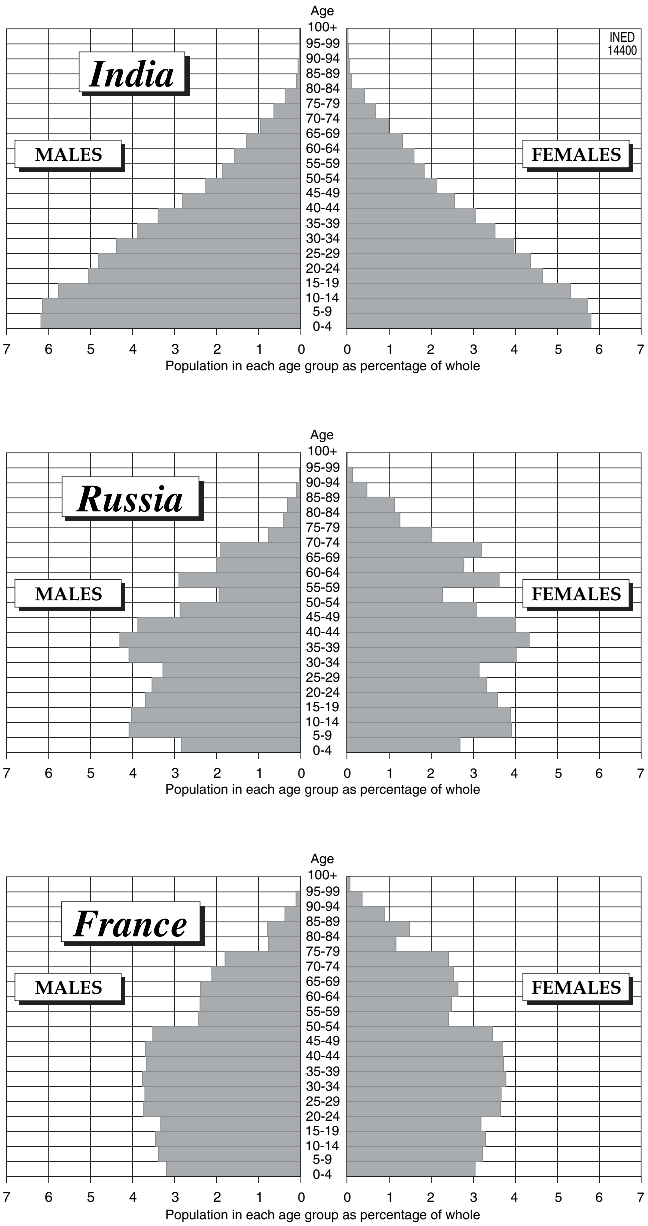
same dimension to all three countries, the proportional distributions are more revealing of these differences (Figure 4–3B).

d. Observation Problems

The population pyramids illustrate survey or census results, but these may be imperfect. Some types of error are immediately visible on the population pyramid, which, for a trained reader, becomes a very useful tool for checking the quality of the data. This is clear from Figure 4–4, which illustrates the Moroccan census of 1960 (Service Central de Statistiques, 1964). All the *round ages* (those ending in zero) are protuberant. The numbers at these ages are overestimated by an effect known as *age heaping*. This problem is common among populations that have no reliable vital registration system and age is not known with accuracy. In this case, enumerators or respondents tend to report the age as the most probable round digit. In some cases, other digits, those ending in five in particular, attract the imprecise cases (Roger *et al*., 1981; Clairin and Lohlé-Tart, 1988; Tabutin, 1995). The instructions given to interviewers to remedy this type of error may be so strict that they have the opposite effect and produce a marked deficit at the round ages: this occurred with the 1960 census in Niger (Roger *et al*., 1981). The heaping effect may also concern years of birth, when the questionnaire focuses on date of birth rather than on age. Thus, cohorts born in years ending in zero (or five) will be overestimated, as was the case in the Zambian census of 1969 (Ohadiké and Tesfaghiorghis, 1975).

In this case, the heaping effect will naturally appear on the pyramid at ages other than multiples of five or 10, unless the census or survey was held in a round year. Finally, if a historical calendar is used to help respondents to locate themselves in time, the heaping will be determined by major events in the calendar and will concern ages that are neither round nor resulting from a round year of birth (Waltisperger, 1988). Age heaping, so blatant on the Moroccan population pyramid for 1960, continued to mar most European censuses throughout the 19th century. It is clearly visible on the French population pyramid for 1851, for instance (Tabah, 1947; Meslé and Vallin, 1989).

The population pyramid may also reveal other forms of error. In some developing countries, for instance, the numbers of nubile girls are often underestimated, because their age tends to be inflated when they are married and understated when they are not (Simonet, 1974; Roger *et al*., 1981). This is quite clear in the case of Morocco (Figure 4–4). Another common example of age misreporting that can show up on the population pyramid concerns the first year of age. Indeed, age 0 is a very abstract notion, and infants who have not yet reached their first birthday may well be reported as being "age 1." The result is a sharp indentation at the base of the pyramid for those younger than 1 year. Another sign of underestimation detectable on the pyramid concerns elderly women, who often escape the enumeration when they are widowed or divorced (Négadi *et al*., 1974), or men in the armed forces, who are systematically excluded. However, it is important not to confuse these



B

FIGURE 4–3 (*Continued*) **B:** Comparison of three population pyramids using proportional distributions (India, Russia, and France, 1995).

systematic errors due to age misstatement with actual perturbations resulting from specific historical events. It is not because the population studied is a developing one that every unusual feature of the pyramid is a sign of age misreporting (Picouët, 1969).

2. How to Read the Population Pyramid

Not only does the population pyramid reflect an image of the current composition of the population by age, with all the implications for expression of the needs of the different categories of population (preschool and school-aged children, working population, pensioners, balance of the sexes, etc.), it also, by placing side by side 100 successive birth cohorts, traces the history of these different generations. Indeed, the current population of each individual age or age group is the combined result of an initial number of births in the cohort and the deaths or entries/exits by migration that have occurred since. The interpretation of the pyramid calls on two kinds of historical events that have each left their mark. First, the bumps and hollows trace the memory of the hazards of history experienced by specific birth cohorts. Second, the general shape of the pyramid reflects the major trends that have shaped the course of fertility, mortality, and migration.

a. The Hazards of History

In the last decades, some countries have lived through a terribly tormented history, so tormented that the distortions produced on the population pyramid are staggering. Three of the most remarkable examples to date are shown in Figures 4–5, 4–6, and 4–7. Let us just consider the spectacular impact on specific birth cohorts of three major events: World War II in Russia, the ban on abortion in Romania, China's "great leap forward."

World War II caused almost 20 million deaths in Russia[1] and the crude death rate suddenly soared from 20.1/1000 in 1939 to almost 60/1000 on average in the period 1941 to 1944.[2] This trauma concerned many birth cohorts simultaneously, which makes its impact less visible on the population pyramid (Figure 4–5). The trace it leaves is less a clearcut break in the dimension of successive age groups than a break in the balance between the male and female populations—although the imbalance produced by the war tends to blur with time, mixing into the structural effect of the very high excess male mortality which characterizes the Russian population. Nonetheless, on the 1997 pyramid, the hollowing out of the male population aged 69 to 78 is a permanent scarring caused by the tragedy.

The most visible impact of the war, however, is the indentation of the population at ages 46 to 49, men and

[1] On the basis of recent studies (Andreev *et al.*, 1990), Alain Blum (1994) estimates that World War II produced 26 to 27 million more deaths than there would have been in peacetime. According to Alexandre Avdeev (personal oral communication May 1999), 20 million of these war deaths concern Russia alone.

[2] Supposing the majority of the excess mortality (20 million deaths) can be attributed to the years 1941 to 1944 and that, without the war, the number of deaths per year would have been much the same as in 1939.

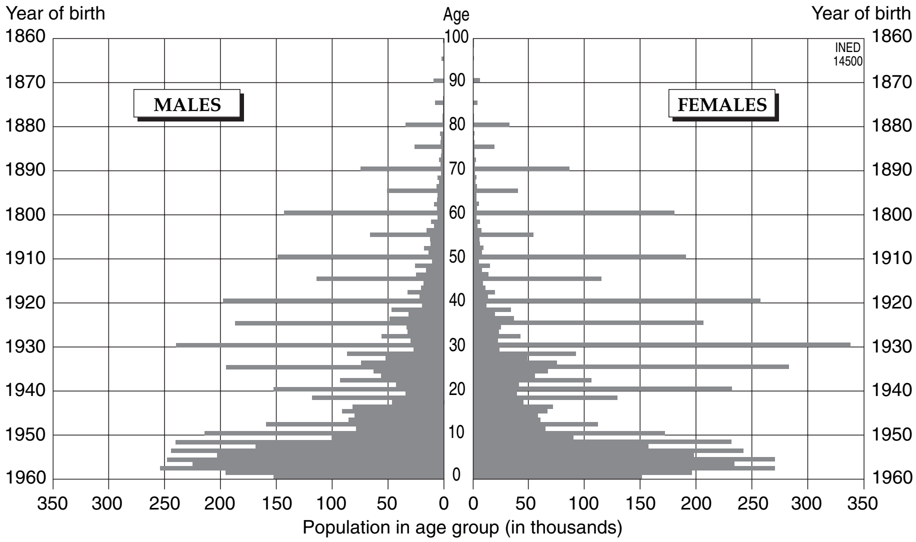


FIGURE 4–4 Moroccan population pyramid based on the 1960 census data.

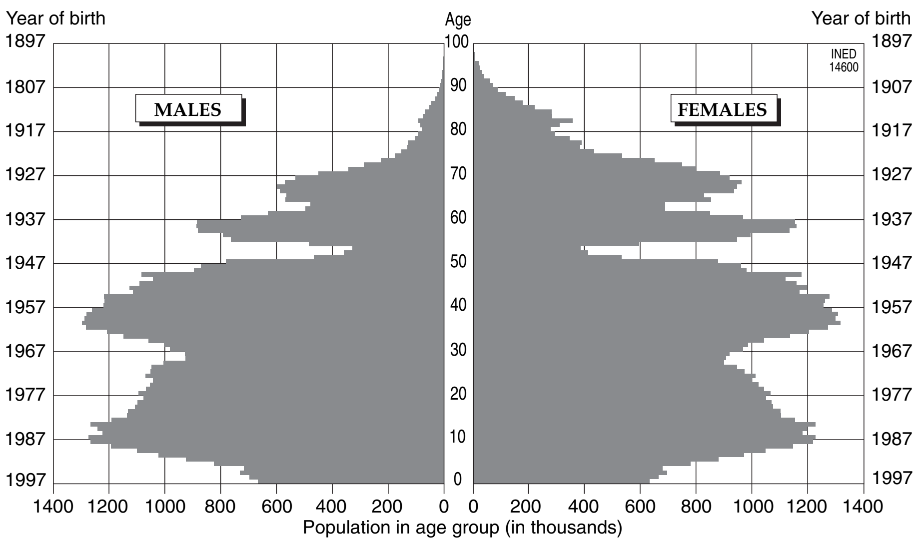


FIGURE 4–5 Population pyramid of Russia, January 1, 1997. (Source: estimated by Goscomstat, data kindly communicated by Alexandre Avdeev.)

women alike. These are the cohorts born during the war years, depleted compared with the earlier and later cohorts because so many couples were separated by the war. Indeed, the annual birth rate plummeted from 37/1000 in 1939 to 13/1000 in 1943.[3] This *deficit of births*, which relates to a precise number of cohorts, leaves a more striking mark on the pyramid than the more devastating, but more diffuse, effect of the 20 million war victims. The impact of the deficit was such that a second indentation can be observed on the pyramid marking the moment when the cohorts born during the war years entered the reproductive ages: In turn, these depleted cohorts produced depleted cohorts, in the mid 1970s. This is a striking example of the interaction between the structure of a population and a component of its natural dynamics, births. First, in the 1940s, a sudden change in the birth rates profoundly modified the age structure, compared with what it would have been had the previous birth rate level been maintained. Then, one fourth of a century later, because of this anomaly in the age structure, the

[3] These estimates refer to the USSR, but they are probably not so different for Russia. The births estimated by Serge Adametz *et al.* (1994), 7.3 million in 1939 and 2.6 in 1943, were related to a total population of about 195 million.

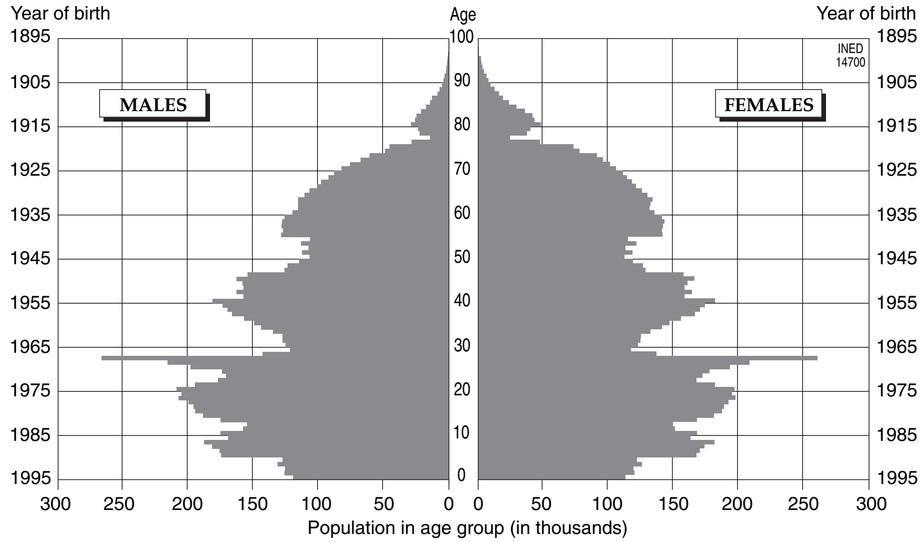


FIGURE 4–6 Population pyramid of Romania, July 1, 1995. (Source: United Nations, 1997.)

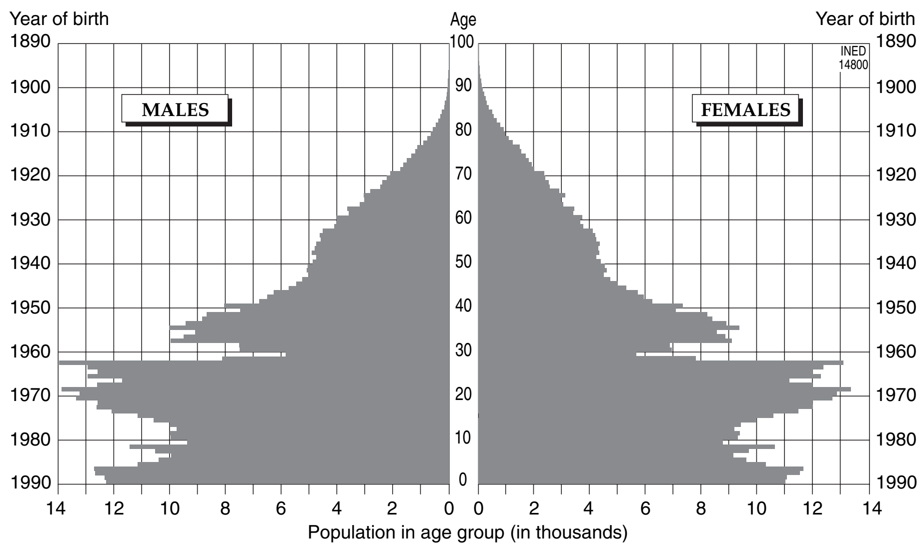


FIGURE 4–7 Population pyramid of China based on the 1990 census (Source: PCO, 1993.)

birth rates fell, producing a new indentation of the age structure. And so on . . . because another quarter century on, another sudden drop in fertility is observed, just emerging at the base of the 1997 pyramid, which will produce another set of depleted cohorts. This new phenomenon is not, however, merely the backlash of history: The erosion at the base of the pyramid, observed since the end of the 1980s, is enhanced by an actual decline of the propensity to reproduce. The cyclic effect of the shortfall of births during the war tends, on the contrary, to become diluted, as the births in the years concerned are not the offspring of depleted cohorts only, but also of neighboring nondepleted cohorts who are also of childbearing age. Yet, in the present case, the impact on the Russian population pyramid has been so great that it will be a long time before the slate is wiped clean.

The population pyramid for Romania (Figure 4–6) is marked by the same kind of cyclic perturbations related to the war (even the impact of World War I, which was more devastating in Romania than in Russia, remains clearly visible).

However, we have chosen this example to illustrate an inverse effect. In 1966, the Romanian government suddenly decided to ban abortion, which was at the time unrestricted and widely used as the principle method of birth control. As a result of this sudden decision, many women who were then pregnant had no choice but to keep their baby. The outcome was a sharp *excess of births* in the following year (Gethau, 1970, 1983; Lévy, 1990; Muresan, 1996). The birth rate leaped from 14/1000 in 1966 to an extraordinary 27/1000 in 1967, thereby doubling the size of the 1967 birth cohort compared with the preceding one, and creating a lasting and unprecedented anomaly in the Romanian population pyramid. This was also a one-off increase, since in the following years women resorted to alternative methods of birth control (illegal abortion or more systematic contraception) and the birth rate fell back to below 20/1000 (18.2 in 1973). Despite the extent of this upturn, the fact that the excess of births was limited to the years 1967 to 1968 means the cyclic shockwave has been slight compared with the Russian deficit of births during World War II. Indeed, 20 years after this extraordinary baby boom, only a slight increase in births is observed.

Looking at China's population pyramid for 1990 (Figure 4–7), we find an example of birth deficit, not a wartime one in this case, but the consequence of the terrible crisis engendered by the Great Leap Forward decreed by Mao Tse Tung in 1957. This brutal political change not only caused 15 to 30 million deaths between 1957 and 1961[4] (Banister, 1997, p. 85), but also a sudden downturn in birth rates, owing to the forced separation of many couples and to the fact that the others were unwilling to have children who were doomed to famine. The birth deficit was made all the more acute as the rise in mortality targeted infants in particular. Thus, the population pyramid constructed from the 1990 census data shows a marked indentation at ages 29 to 32 years. Once more, these depleted cohorts produce, in turn, another anomaly in the pyramid 20 years later: The number of children born around 1980, aged 9 to 13 in 1990, was visibly reduced, even if this particular deficit is combined with the effects of the one-child family policy.

b. The Shaping of the Pyramid: The Role of Births, Deaths, and Migration

Such specific historical events that have a brutal impact on the population pyramid are but one aspect (and no doubt not the most important one) of the interaction between population dynamics and population structure. In a much more general and fundamental way, the long-term trends in births, deaths, and migration give their overall shape to the pyramid. This explains why the population pyramids of different countries display such a great variety of shapes. Figure 4–8 shows five examples of the most classic situations.

In ancient Europe, but also until quite recently in southern Europe, population pyramids were almost perfectly triangular in shape, which explains why this representation is called a "pyramid." This situation, illustrated by India in 1951, corresponds to a demographic regime of very high mortality and fertility, but relatively slow population growth. The population is very young, with very few elderly: in India in 1951, 47.5% of the population were younger than 20 years, 48.9% were adults aged 20 to 65 years, and 3.6% were older than 65 years.

In comparison, in Kenya in 1969, the pyramid is still triangular but its base is broadened and the sides are concave. This difference is because birth rates increased in Kenya during the 1950s and 1960s because of a reduction of breast-feeding and postpartum abstinence, and at the same time infant and child mortality began to decline. Consequently, the population growth rate accelerated rapidly, which greatly broadened the base of the pyramid. Compared with India in 1951, the proportion of those younger than 20 years is even higher (58.6%), and the proportions of adults and elderly are lower (37.9% aged 20–65 years and 3.5% 65 years or older).

The population in France in 1911 illustrates a situation in which birth rates were falling for a long time (see Figure 4–8). Despite the concomitant decline in infant and child mortality, which has slowed down the process, the base of the pyramid has narrowed to the extent that the cohorts of young adults are barely smaller than the cohorts of children. The proportion of those younger than 20 years is much lower than in the previous examples (33.3%) and those of adults and elderly are higher (57.5% and 9.2%, respectively). The French pyramid is typical of a population in which the decline of birth rates has constantly offset the mortality decline and maintained a very low rate of population growth.

The base of the Italian pyramid for 1995 is narrower than the French. Indeed, for the last two decades Italy has experienced an unprecedented fertility decline. So, the crude birth rate is exceptionally low, in the region of 10/1000, which is even lower than the crude death rate and means that natural increase is negative. Consequently, the proportion of those younger than 20 years is even lower than that in France in 1911. The share of those older than 65 years is, on the contrary, higher: It even exceeds the share of those younger than 15 years (16% versus 14%).

[4] According to Judith Banister (1987), the excess mortality due to the crisis is about 15 million deaths if the death rate recorded in 1957 (10.8/1000) is taken to be correct. She considers, however, that there must have been substantial underregistration of deaths and that the excess mortality may be as high as 30 million.

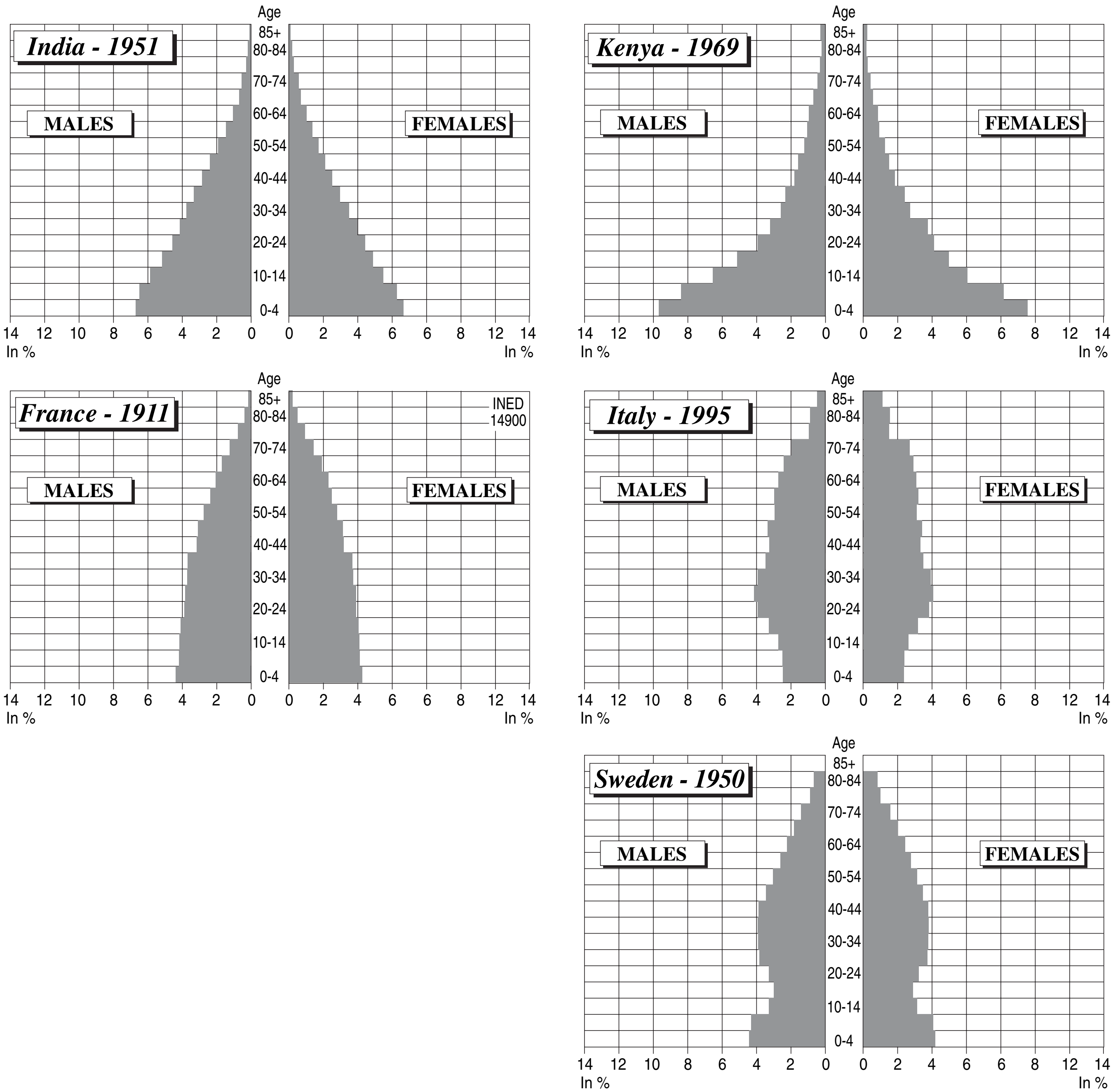


FIGURE 4–8 Population pyramids of India in 1951, Kenya in 1969, France in 1911, Italy in 1995, and Sweden in 1950. (Source: Population censuses)

In the case of Sweden in 1950, we observe a trend that led, as in France and Italy, to a narrowing of the pyramid, but which was then followed by a baby boom in the 1940s and 1950s, which expanded the base again (Figure 4–8). The result is a characteristic ace-of-spades form.

In Figure 4–9, we show two even more pronounced patterns illustrating the last two situations: Emilia-Romagna and Catalonia. These two regions are known for the extent of their recent fertility decline and the very low levels reached at the end of the 1980s, particularly in Emilia-Romagna. Here, the proportion of those younger than 5 years had fallen well below 2% in 1991.

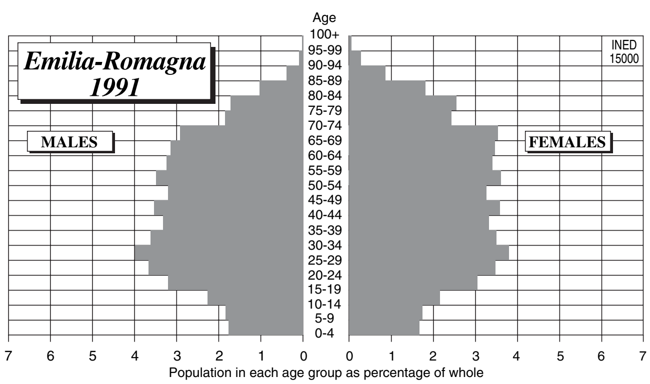
The two regions differ, however, in the general aspect of their population pyramids. In Emilia-

Romagna, the population had already aged considerably before the recent fertility decline, which explains the pyramid's barrel shape. In Catalonia, on the contrary, the decline has modified the base of a more classic pyramid, giving a Christmas-tree shape.

Migration generally plays a lesser role in shaping the population pyramid than fertility and mortality, which explain most of the population growth. However, in specific countries and specific periods, international migration has been so strong, and so selective by age and sex, that it has had a notable impact. Figure 4–10 shows three such examples: Italy in 1911, Algeria in 1974, and Kuwait in 1980.

Because of very high emigration, the Italian population had a clear deficit of young adults at the turn of the century. In 1911, the pyramid shows a clear discontinuity around the 20th birthday for women as well as men, a sign that emigration concerned both in roughly equal proportions. The situation is different for Algeria in 1974, where emigration visibly affected men more than women. Lastly, Kuwait illustrates the impact of particularly high immigration of (mostly male) adults.

The impact of migration on the population pyramid is more pronounced at the local level, when very specific internal migration flows leave their mark. The French city of Toulouse, for instance, like many towns whose economy is based on an active university life and modern, dynamic industry, has recently attracted a flow of young adults who have swollen the central part of the pyramid, which elsewhere is rather monotonous since the age groups below 35 and above 60, up to about 85, are of comparable dimension (Figure 4–11). The contrary is noted in an extreme example,



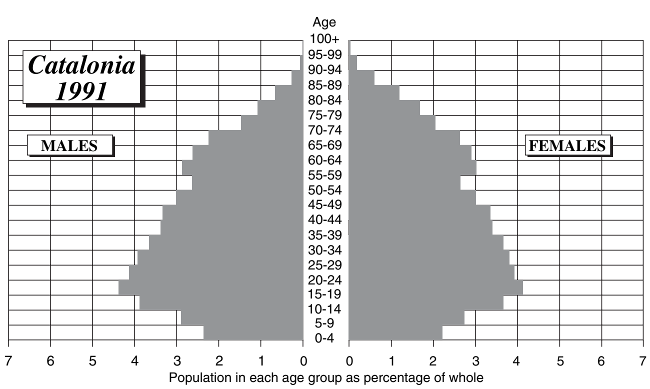
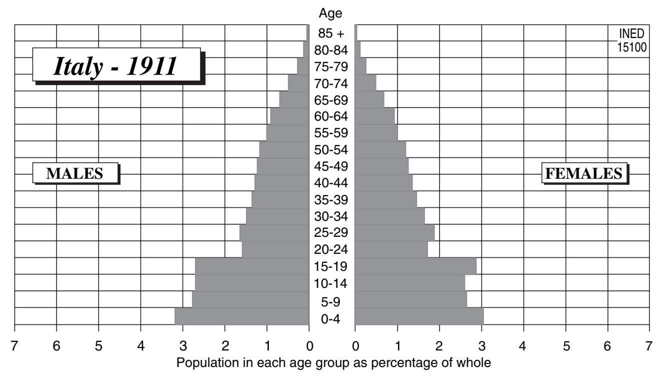
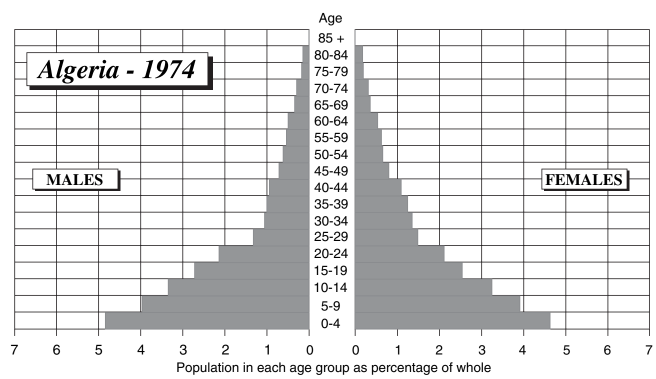


FIGURE 4–9 Population pyramids of Emilia-Romagna and Catalonia in 1991. (Source: Emilia-Romagna: ISTAT, 1995b; Catalonia: INE, 1993.)





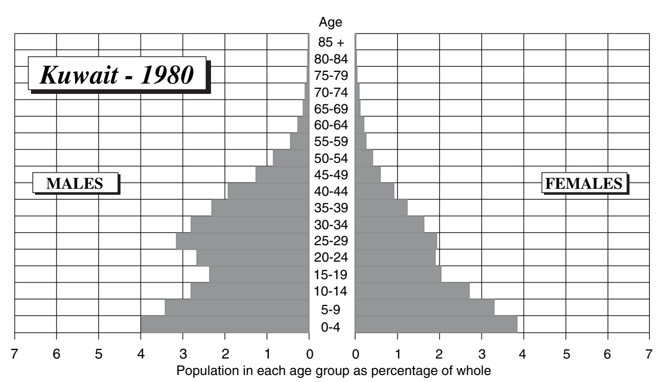


FIGURE 4–10 Population pyramids of Italy in 1911, Algeria in 1974, and Kuwait in 1980. (Sources: Algeria: ONS 1975; Italy: DGSL, 1914; Kuwait: 1994.)

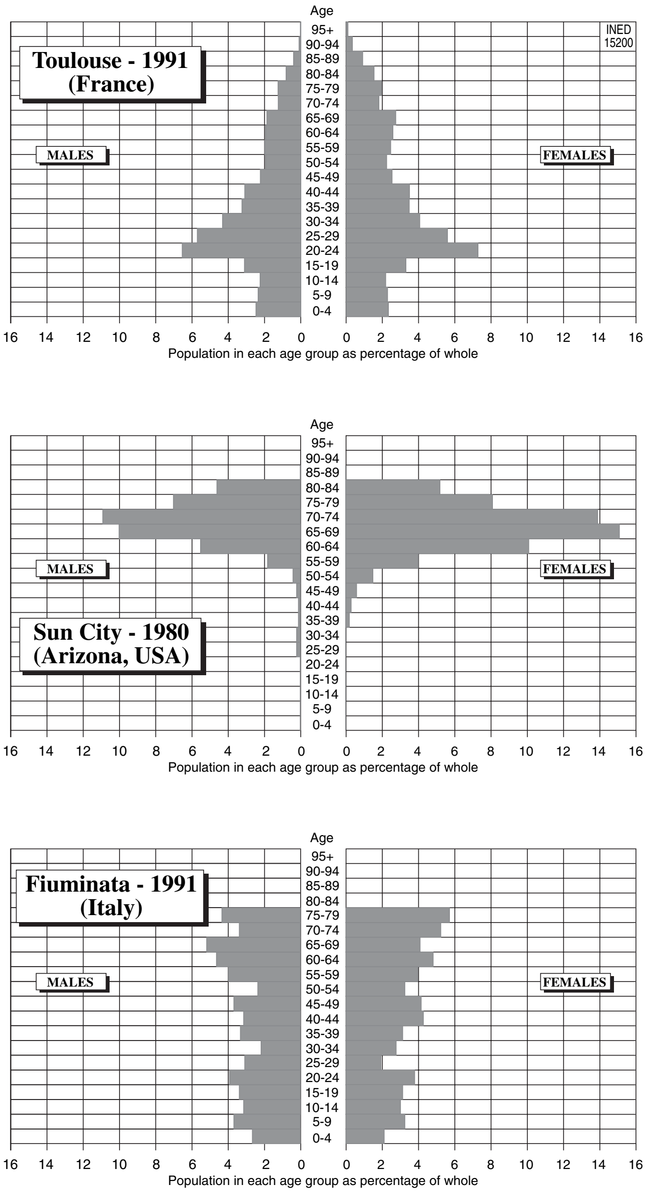


FIGURE 4–11 Population pyramids of Toulouse (France), Sun City (Arizona, US). and Fiuminata (Italy). (Sources: Toulouse: INSEE, 1992; Sun City: taken from Leridon, 1997; Fiuminata: ISTAT, 1995.)

Sun City in the United States, with an almost exclusively pensioner-based economy, which offers the hallucinating image of an inversed population pyramid, floating unanchored without a single inhabitant below the age of 25 (Figure 4–11). Finally, Fiuminata, a small rural town in the Italian Apennines, has been marked by an emigration of adults, which, combined with decades of low fertility, gives the upside-down pattern of a pyramid that is broader towards the top than at the base (Figure 4–11).

c. Three-dimensional Pyramids

It may be of interest to distinguish, within a population pyramid, different categories of population by introducing, together with age and sex, another variable such as economic activity, marital status, nationality, and so forth. Figure 4–12 presents an example: the Italian census population in 1991, broken down by sex, age, and marital status. The distribution of the single, married, divorced, and widowed members of the population within the age–sex groups is thus shown. The pyramid is massively single for those younger than age 20 for men and 17 for women. The married group then rapidly predominates, for both sexes, confining the single to a narrow central band. The divorced occupy an outer fringe between the ages 30 and 60, very small in size because of the low frequency of divorce and the effect of remarriage. From age 50 on, widowhood emerges, with a varying impact on men and women. Among the latter, widows quite rapidly gain ground to become the majority group, whereas among men, widowers remain constantly marginal. This reflects the fact that men die earlier than women. In contrast, in the upper half of the pyramid, there are constantly more married men than married women. This is because men marry women younger than themselves, and more numerous. This imbalance is particularly marked above age 70, when there are three times as many married men as there are married women.

In Figure 4–13, the population pyramid again demonstrates its efficiency for summarizing at a glance how a population is structured according to three variables: here, the research and teaching staff of the Italian Faculty by age, sex, and faculty status. Not only is the gender imbalance striking, it is clear that it is closely associated with age and status. The highest status category (professor) is totally dominated by men, with only a very slight fringe of women, while the lowest status category (researcher) has an important proportion of women. But, for a same status category, women are on average younger than men, which suggests that women have embraced this career more recently than men and that their numbers will in the future draw closer to those of men.

3. The Sex Balance

Because the balance between males and females conditions couple formation and reproductivity, this general presentation given by the population pyramid can usefully be completed by a specific analysis. Indeed, globally, the numbers of males and females are rather well balanced, but in contrast the ratio of men to women varies substantially from age to age.

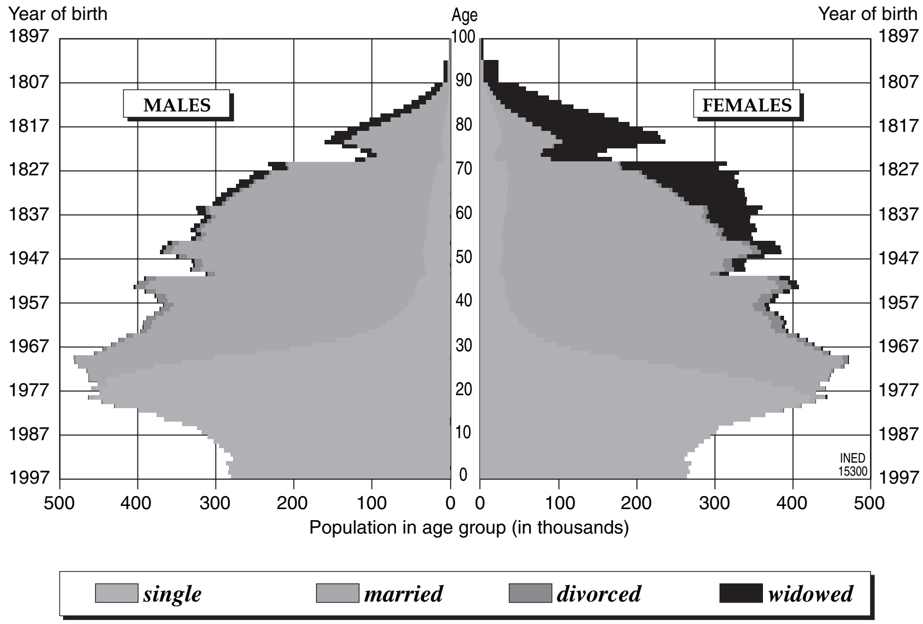


FIGURE 4–12 Population pyramid of Italy, 1991, by marital status. (Source: ISTAT, 1995a.)

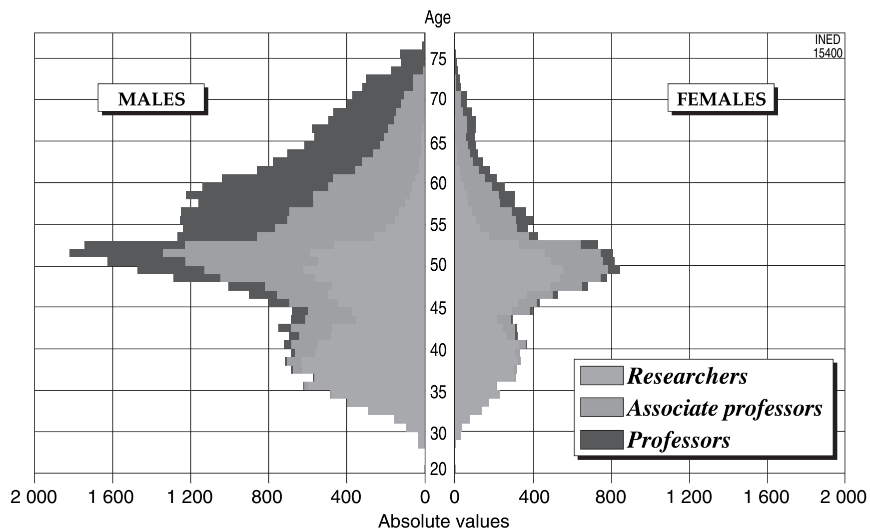


FIGURE 4–13 Population pyramid of the research and teaching staff of the Italian Faculty in 1997, by faculty status. (Derived from data published on the Internet by the Ministero dell'Università e della Ricerca Scientifica e Tecnologica, MURST. Consulted in May 1999.)

a. The Global Balance Between Men and Women

At the national level, in most countries worldwide, the *sex ratio* of the population (the number of males per 100 females) is close to 100 (Figure 4–14). There are, however, some interesting variations. Many industrialized countries (but also others) have values slightly below 100 whereas some southern countries are well above 100. In fact, the global balance between the sexes that is generally observed results (in the absence of substantial migration) from two phenomena that offset each other: Slightly more baby boys are born than baby girls, but they die earlier than the baby girls. If some populations move away from this balance, it is a result of the mortality gap between the sexes has grown or the discriminatory effect of substantial migration flows. For instance, the sex ratio is particularly low in Russia, where excess male mortality is breaking all known records, but it is higher than 100 in India where female mortality was until quite recently higher than male mortality. When Algeria in 1974 is compared with Kuwait in 1980, the sex ratio in

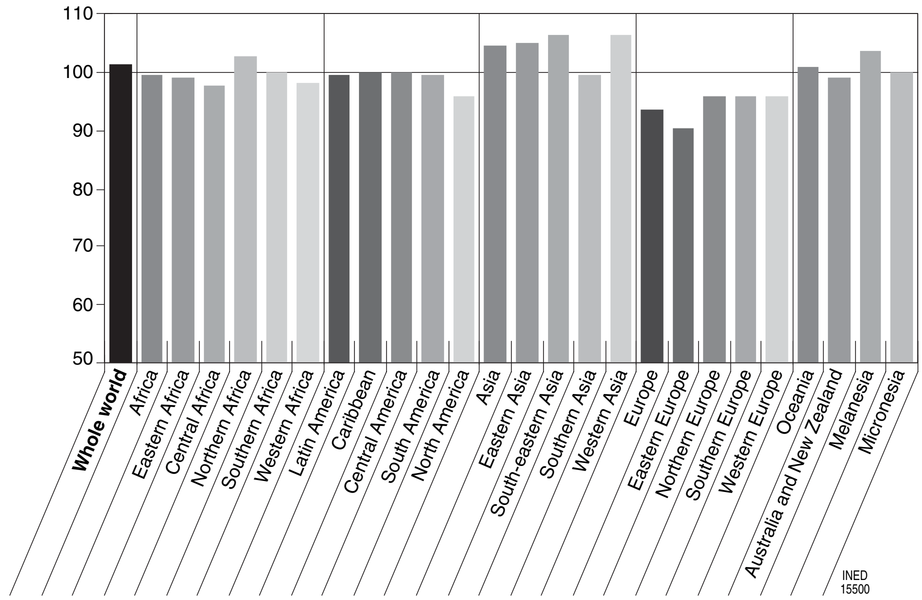


FIGURE 4–14 Distribution of countries by global sex ratio circa 1995. (Source: United Nations, 1997.)

the former is 91 because of the high emigration of males and 134 in the latter owing to high immigration of males.

Naturally, however, the effect of birth, death and migration flows on the balance between the sexes operates differently at the different ages. Accordingly, the population's age structure acts on its sex structure: the more elderly there are in a population, the more the impact of excess male mortality affects the sex structure.

b. Age-specific Sex Ratio

In the absence of migration, the *age-specific sex ratio* (the number of men of a given age to the number of women of same age) follows a classic pattern: around 105 at birth, then with equal balance between men and women at ages 40 to 50, before falling far below 100 at the older ages.

The starting point is the fact that nature produces more baby boys than baby girls. The *sex ratio at birth* (number of male births per 100 female births) is generally very close to 105. This is one of the very rare demographic parameters that is virtually constant. Without any voluntary intervention on the sex of the future offspring (such interventions being limited for the moment),[5] the variations observed around this level of 105 are rare and slight, to the extent that any deviation alerts the expert demographer to the risk of birth registration deficiencies (Vallin, 1973).[6]

This slight headway for boys then dwindles gradually (in the absence of migration) as a result of excess male mortality, which may vary a lot across countries. Figure 4–15 displays the age-specific sex ratios for India and Russia in 1995. The trace of international migration is almost negligible (except for Russia around age 25 years, owing to the recent waves of immigration of young adults, mostly men). Excess male mortality is, on the other hand, much higher in Russia than in India, a country where excess female mortality from childhood to puberty increases the sex

[5] Social preferences for boys or girls (usually for boys) are very strong in some societies, but could until now have little impact on the sex ratio at birth, since sex could not be determined prenatally. The preference for boys has sometimes led to infanticide of baby girls (not only in China), but that affects the ratio of male to female death rates and not the sex ratio at birth. In recent years, however, the development of prenatal scans and amniocentesis has come to permit the prenatal diagnosis of sex, thus opening the door to possible selective abortion. Apparently the latter is beginning to spread in some Asian countries. In certain regions of India and China, the sex ratio at birth in the early 1990s reached values as high as 110 or even 120, and the level increases very significantly with rank of birth (Miller, 1996). At an earlier stage, now we have *in vitro* procreation, the child's sex can theoretically be chosen at time of conception. However, this procedure is rare and statistically negligible.

[6] When registration is incomplete, more baby girls than boys are excluded from the statistics. This is traditionally a case of negligence: it is less important to report the birth of a daughter than of a son. It may also, more seldom, mean some female births have been concealed: in China, for instance, the one-child family policy leads some families to hide the fact they have had a daughter in the hope of having a son later (Miller, 1996).

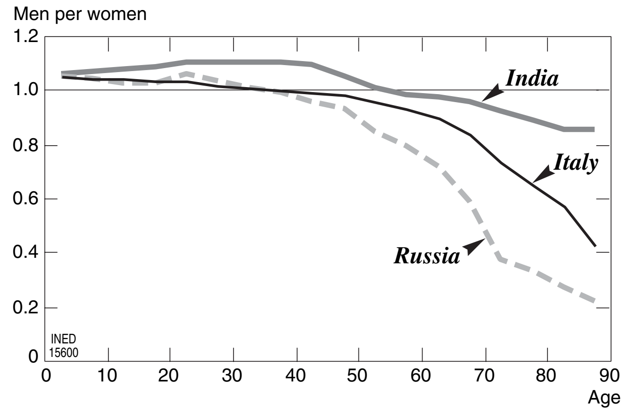


FIGURE 4–15 Age-specific sex ratio in Russia and India in 1995 (Source: United Nations, 1997.)

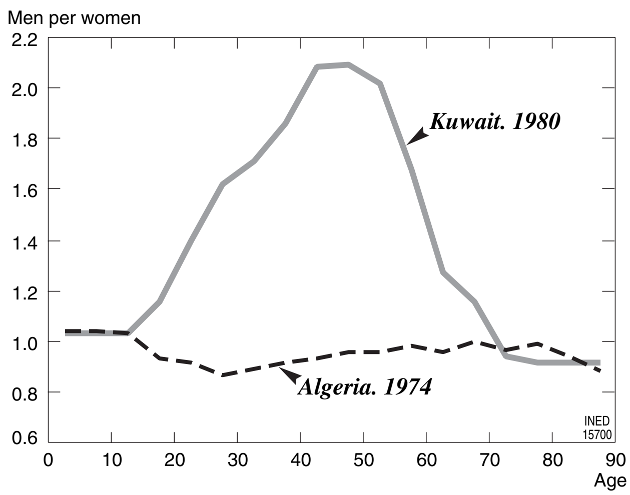


FIGURE 4–16 Age-specific sex ratio in Algeria (1974) and Kuwait (1980). (Source: Algeria: ONS, 1975; Kuwait: CSO, 1995.)

ratio until about the age of 55. In Russia, by contrast, there is excess male mortality at all ages and, leaving aside the anomaly observed at age 25, the sex ratio declines age-wise from birth on, with acceleration after age 50; by the age of 90, there are no more than 20 men for 100 women. We have plotted the curve for Italy in 1995 between these two extreme cases; it is a country with no excess female mortality (unlike India) but with more moderate excess male mortality than in Russia.

Figure 4–16 shows Algeria in 1975 and Kuwait in 1980, two countries whose recent history is marked by migration flows, mostly of men, but in opposite directions. In Algeria, the mass emigration of workers to Europe, France in particular, during the 1960s and 1970s has produced an exceptional deficit of young adult men, whereas in Kuwait, the mass immigration of (often skilled) labor as a consequence of the oil boom has brought the sex ratio to exceptionally high levels at the active ages. Thus, variations in the age-specific sex ratio bring into sharper focus the phenomena that were already clearly detectable on the population pyramid.

Once again, we see to what extent a population's age–sex structure varies as a result of birth, death, and migration trends. Nonetheless, we can also understand the interest of refining the indicators we have used so far (the crude rates) to measure these three factors of population dynamics. Indeed, the variety of the structures means that the crude rates of birth, death, and migration do not measure exactly the phenomena at stake (fertility, age at death, and migration at different ages), but only their combination with the age–sex structure prevailing at a given moment in time.

To go further in the analysis, it is necessary to study fertility, mortality, and migration by sex and age.

II. CRUDE RATES AND AGE-SPECIFIC RATES

Just as the crude death rate is estimated by relating the events occurring in a period to the mean population in that period[7], so the age-specific rates are calculated by relating the events occurring at a given age during the reference period to the mean population of this age during the period.

Like the crude rates, the age-specific rates are generally calculated from annual observations. They can also be calculated on the basis of shorter or longer periods of observation. In this case, to obtain the annual dimension that is universally used for purposes of comparison, the denominator must simply be multiplied by the number of years (or the fraction of a year) that gives the dimension of the reference period.

1. Age-specific Mortality

The *death rate* $m_{(x,t)}$ at age x in year t is obtained by relating the deaths $D_{(x,t)}$ occurring that year among persons of that age to the mean population of same age (generally estimated by the simple average of the

[7] In the preceding chapter, we have seen that a crude rate for all ages can be calculated by relating the number of events to the mean population. For instance, the crude death rate m_t in year t is equal to:

$$m_t = -\log\left(1 - \frac{D_t}{P_t}\right) \cong \frac{D_t}{(P_t + P_{(t+1)})/2}$$

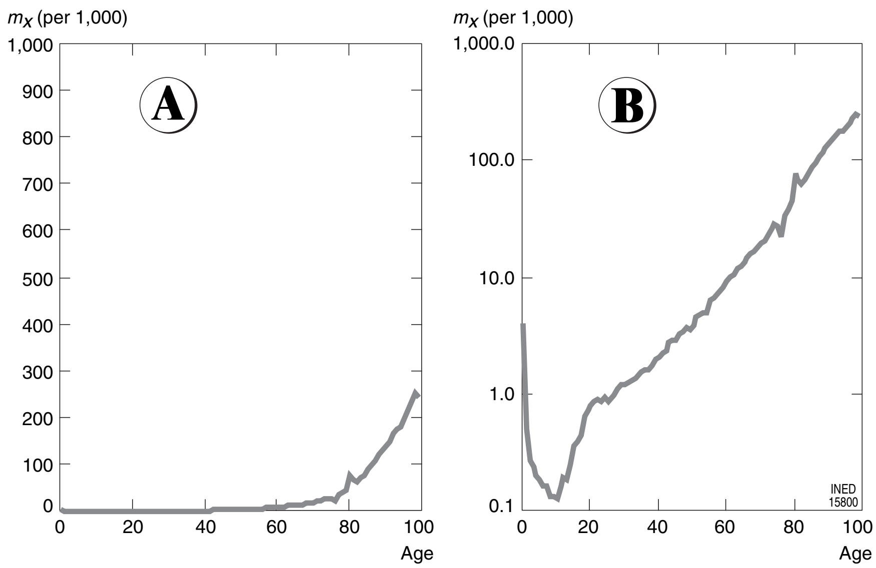


FIGURE 4–17 Age-specific death rates in France, 1995: arithmetic and semilogarithmic representations.

populations $P_{(x,t)}$ and $P_{(x,t+1)}$ on the two 1 January surrounding t)[8]:

$$m_{(x,t)} = \frac{D_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}$$

Thus, in France, the *Statistique du mouvement de la population* (Demographic Yearbook) for 1995 informs us that 774 people died that year at the age of 22. According to Institut National de la Statistique et des Etudes Economiques, the population of this age was 861,362 on January 1, 1995 and 808,621 on January 1, 1996. The death rate at age 22 is therefore:

$$m_{22} = \frac{774}{(861,362 + 808,621)/2} = 0.93\text{‰}$$

The rate can be calculated in this way for each individual age or by age groups, in particular when the distribution of deaths (or of population) is known only by age groups:

$$m_{((x,x+a),t)} = \frac{D_{((x,x+a),t)}}{(P_{((x,x+a),t)} + P_{((x,x+a),t+1)})/2}$$

[8] This approximation follows on directly from the preceding footnote:

$$m_{(x,t)} = -\log\left(1 - \frac{D_{(x,t)}}{P_{(x,t)}}\right) \cong \frac{D_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}$$

We shall see in Chapter 9 that to estimate mortality by this rate supposes, in addition, either the absence of all competing factors (here, migration) or independence between the competing factor(s) (here, migration) and the studied phenomenon (here, mortality). Finally, it supposes of course that the population is homogeneous (Chapters 9 and 14).

Figure 4–17 illustrates the results for France in 1995. On the left side, with an arithmetic scale, the rates are very low until age 50 to 60 years, then rise rapidly with age, in a near-exponential manner. That explains why, conventionally, age-specific mortality rates are plotted using a log scale (right-side graph): the variations at young ages become more visible and the near-exponential growth at older ages can be verified.

This age-wise mortality increase can be interpreted as the statistical translation of the process of biological aging of the human body. The older one becomes, the more the vital functions become impaired and the greater the risk of dying. The regularity of the age-wise increase in the risk of dying is striking, even if causes other than aging can bring about death. There are two exceptions to this regular progression. One, observed everywhere, concerns the first years of life; the other is around age 20 and is more specific to the example shown here. First, at the youngest ages, mortality begins by diminishing with age, until approximately age of 10. This is the effect of another biologic process: birth and the first days of life are particularly perilous (obstetric trauma, environmental aggressions with which the newborn cannot yet cope, elimination of serious congenital malformations spared by intrauterine mortality, etc.), and these dangers disappear rapidly as the infant ages. At around age 10, this process is essentially completed and the process of biologic aging starts to take over, thus inverting the age-specific mortality trend. Around age 20, however, another exception is observed: a strange hump in the curve that has no biologic explanation, but reflects the excessive importance of traffic accidents in mortality at these ages. Indeed, within the framework imposed by the biologic process, an actual mortality level depends also on social behaviors and on the health system.

2. Age-specific Fertility

The shift from crude birth rate to age-specific rates is more complex. It is obviously not the newborn who must be differentiated by age but the parents, whose propensity to procreate must be measured. To mark this difference, the conventional terminology distinguishes between *fertility*, which is measured by age, and *natality*, which refers to the global frequency of births. Furthermore, fertility depends on three elements: women's reproductive capacity, that of men, and the meeting of these two aptitudes through couple formation, with all three elements depending on age. There are therefore at least three ways to measure a population's fertility, according to whether it is approached from the angle of women, men, or couples. The latter is the most complex approach. It requires knowledge of couple formation and dissolution by duration of union. But couple formation is closely linked to reproductive capacity and it is this (therefore the age of the partners) that first and foremost governs a population's fertility; thus, the idea of approaching fertility via the individual. However, a choice must be made: via men or via women? The convention is to measure women's fertility,[9] for at least two reasons, a theoretical one and a practical one. Biologically, it is the woman and not the man who sets the limits of a population's fertility: a woman generally brings to maturity only one ovum per cycle, when fecunded she then has 9 months of pregnancy, and once the child is born there will be a lapse of time (that may be lengthened if she breast-feeds) before ovulation is resumed. A man emits millions of spermatozoa and is subjected to none of the constraints of pregnancy. In a population, therefore, women's fertility could theoretically reach its biologic maximum with a very small proportion of men. By contrast, it is difficult to imagine how many women would be needed to bring just one man's fertility to a biologic maximum. This argument is theoretical, however, insofar as humans are essentially monogamous. More concretely, maternity is more visible and easier to verify than paternity and its observation is simpler. It is more rare for a birth to be registered as "of mother unknown" than "of father unknown." In a survey, it is easier to enumerate a woman's offspring than a man's, and a man may even have children whose existence is unknown to him.

Having opted for this course of action, the age-specific *fertility rates* $f_{(x,t)}$ can be calculated in the same way as the age-specific death rates, but putting in the denominator only the number of women: at age x in year t, the births $N_{(x,t)}$ to mothers aged x in year t are related to the average of the numbers $F_{(x,t)}$ and $F_{(x,t+1)}$ of women aged x on the first of the two Januarys surrounding t.

$$f_{(x,t)} = \frac{B_{(x,t)}}{(F_{(x,t)} + F_{(x,t+1)})/2}$$

So, in France in 1995, 25,766 children were born to mothers aged 22 and the number of women of this age was 424,430 on January 1, 1995 and 398,774 on January 1, 1996, which gives a fertility rate at age 22 of:

$$f_{22} = \frac{25,766}{(424,430 + 398,774)/2} = 62.6\%$$

This rate can be calculated for each age or age group. It is obviously null before puberty and after the menopause; the usable age span is approximately from age 15 to 50.

Reading the results for France in 1995 (Figure 4–18), the fertility rate is very low before age 18, then increases rapidly, peaking at ages 26 to 27 at a level of 150/1000 before decreasing, more slowly, and becoming almost negligible after age 40. Like the relation of the mortality curve to aging, this curve is closely related to a biologic process: fecundity. Fecundity begins at puberty and ends normally with the menopause, but may also, for a variety of reasons, continue for longer or stop earlier, so that even between ages 15 and 50, the proportion of fecund women is

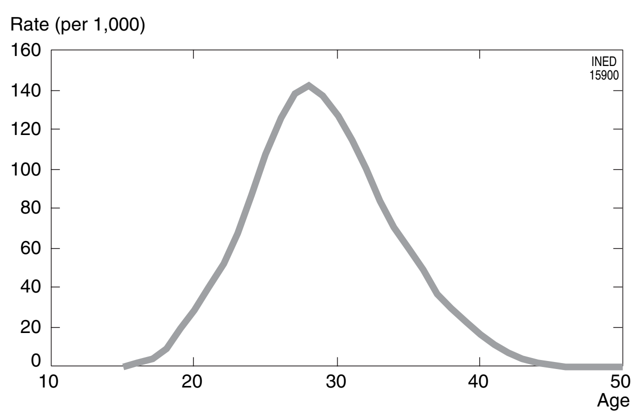


FIGURE 4–18 Age-specific fertility rates in France, 1995.

[9] Men's fertility is sometimes measured, and the results can be quite different, with a notably higher level of male than female fertility, in particular in cases of polygamy and large age gaps between spouses (Brouard, 1977, 1989; Maffioli, 1992; Pison, 1982).

never 100% and varies with age, peaking, once more, at around the age of 25. However, as in the case of mortality, within the limits set by this biologic constraint, fertility is filtered by social rules and individual attitudes, which organize sexual relations (notably through couple formation) and modulate their results.

Even more than for mortality, this fertility curve allows us to perceive to what extent the total number of births observed in a population (and so the crude birth rates) can depend on its age–sex structure. An immigrant population, for instance, can have high fertility but low birth rate, if the immigrants are mostly men. Similarly, when depleted cohorts, such as those born during the war, enter the childbearing ages, the birth rate falls all the more sharply as the fertility curve is concentrated on a small number of ages.

3. Age-specific Migrations

In the same way as for mortality and fertility, it is possible to calculate at each age the rates of *immigration*[10] $i_{(x,t)}$, *emigration* $e_{(x,t)}$ and *net migration* (or *balance of migration*) $nm_{(x,t)}$, by relating the numbers of immigrants $I_{(x,t)}$, emigrants $E_{(x,t)}$ and the net number of migrants ($I_{(x,t)} - E_{(x,t)}$) observed at each age in a given period, to the mean population of same age in the period. Thus, at age x in year t, we have:

$$i_{(x,t)} = \frac{I_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2},$$

$$e_{(x,t)} = \frac{E_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}$$

and

$$nm_{(x,t)} = \frac{I_{(x,t)} - E_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}$$

The problem is that the data required to calculate these rates are rarely available. Indeed, few countries routinely record migrations, and even when this is the case, registration is even more rarely complete. In most cases, the data are from retrospective questions asked during a census. Even if they are of good quality, that only gives, for international migration, numbers of arrivals but not of departures. For provinces or regions within a country, however, it is possible to calculate rates of in-migration, out-migration, and net migration by age.

The calculation of these rates will differ from the mortality and fertility rates. The data generally refer to a period of more than year and the rates calculated directly will not be annual rates; the population at end of period is known, but not the population at beginning of period, so the mean population cannot be obtained as above.

Let us take the example of the Italian census of 1991, which contained a question on place of residence 5 years earlier. For a given region—say, Lazio—we therefore have the number of residents in 1991 who were not residents in 1986 (immigrants) and the number of residents in 1986 living elsewhere in 1991 (emigrants). We note here a third problem: among the "immigrants" we have in-migrants from other regions of Italy as well as immigrants from other countries, whereas among the "emigrants" we have only out-migrants to other regions of Italy and not emigrants to other countries, who are obviously not covered by the census. The symmetry between immigration and emigration rates is therefore not perfect. This is acceptable, however, when international emigration is negligible compared to mobility within the country. This is roughly the case for Italy that became a country of immigration in the 1980s.

To calculate the mean population for the denominator, an initial population must be calculated by adding to the census population of each age in 1991 the deaths that have occurred during the five preceding years in each of the corresponding birth cohorts and net migration must be subtracted.

We then obtain rates of immigration, emigration, and net migration by age (here, in fact, we have only 10-year age groups) for the period 1986 to 1991. For instance, at ages 25 to 34, the immigration rate is:

$$i_{(25-34)} = \frac{I_{(25-34)}}{P_{25-34,1991} + (D_g - I_g + E_g)/2}$$

where D_g, I_g and E_g are the numbers of deaths, immigrants, and emigrants since 1986 in the birth cohorts aged 25–34 in 1991. To estimate E_g, of course, international emigration that escapes the census must also be estimated. Using a number of hypotheses that we lack space to develop here, Istituto Nazionale di Statistica has reconstructed the populations by region and age groups in 1986. On the basis of these estimations, having now a population at beginning and end of period, we can proceed as follows:

[10] Leaving aside the conceptual problem raised by the notion of immigration rate, mentioned in the previous chapter, given that immigration is an event that affects individuals who are external to the study population. The immigration rate is in fact a risk of emigrating for individuals external to the population, which cannot be measured because the denominator is unknown. By calculating immigration rates, it is considered that the host population (and not the individuals forming it) is globally subjected to a risk of immigration. The nature of this risk is obviously not the same as the risk of emigration, even if the two are apparently perfectly symmetrical.

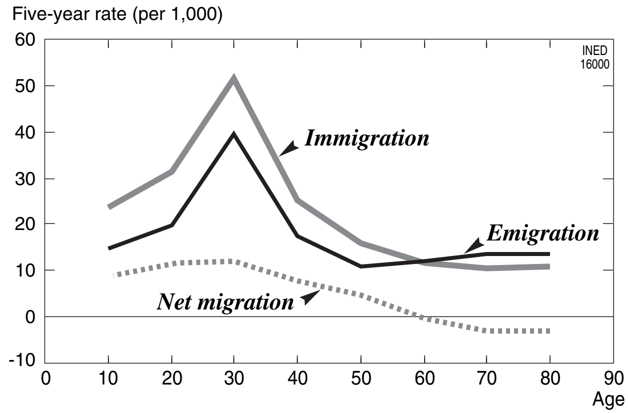


FIGURE 4–19 Age-specific immigration, emigration, and net migration rates for the Italian region of Lazio during the 5 years before the 1991 census.

$$i_{(25-34)} = \frac{I_{(25-34)}}{(P_{25-34,1986} + P_{25-34,1991})/2}$$
$$= \frac{41,817}{(788,721+830,524)/2} = 0.052 = 52\%$$

In the same way:

$$e_{(25-34)} = \frac{32,133}{(788,721+830,524)/2} = 40\%$$

and

$$nm_{(25-34)} = \frac{41,817-32,133}{(788,721+830,524)/2} = 12\%$$

Figure 4–19 shows the variations in these different rates by 10-year age groups. The immigration and emigration curves are similar and classic, peaking sharply around age 30. Net migration on the other hand is biased, as we have mentioned, by the fact that international emigration is not taken into account.

These results, we have said, concern a 5-year observation period. To give them an annual dimension, for comparison with the mortality and fertility rates calculated previously, we can as a first approximation divide them by five.[11]

4. The Relationship Between Crude Rates and Age-specific Rates

The crude rates measure the frequency with which the studied events (deaths, births) occur in a population, but they do not inform us about actual mortality and fertility, which are expressed more accurately by the mortality and fertility functions by age described by the age-specific rates. The crude rate is in fact nothing but a weighted average of the age-specific rates by the proportions of population[12] of each age. So, for the crude death rate d we have:

$$d = \sum_{x=0}^{\omega} m_x p_x \Big/ \sum_{x=0}^{\omega} p_x$$

and for the crude birth rate b:

$$b = \sum_{x=0}^{\omega} f_x p_{f,x} \Big/ \sum_{x=0}^{\omega} p_x$$

where $p_{f,x}$ represent the proportion of women by age and p_x for both sexes by age.

With the same mortality and fertility rates, two populations with different age–sex structures will have different crude death and birth rates. Two examples will amply illustrate the practical importance of this difference between crude rates and age-specific rates.

Let us suppose, for example, that the population of Kenya in 1969, whose population pyramid we have shown in Figure 4–8, had the same mortality as France that year. Using the French age-specific death rates, we calculate the crude death rate that Kenya would have had that year and compare it with the actual French one. To do so, we simply multiply, age by age, the French rates by the Kenyan proportions of population (Table 4–3).

The same mortality gives a crude rate of 11.6/1000 with the French age structure and only 6.9/1000 with the Kenyan age structure. This is obviously explained by the fact that the population of Kenya is much younger than in France, where the high death rates among the elderly concern much larger numbers of the population and weigh more heavily on the crude rate than when these rates are applied to the Kenyan population. We can see how important a population's age structure is for the expression of one of the components of its natural dynamics, mortality.

The same holds, more or less, for fertility, although with a slight difference in that the balance between men and women intervenes more directly and that, with respect to age, it is the childbearing ages that weigh most heavily on the expression of the second component of natural dynamics. Let us apply, for instance, the French age-specific fertility rates in 1980 to the age–sex structure of Algeria in 1974 and China in 1995. With the same fertility rate, we obtain a crude birth rate of 13/1000 in Algeria and 18/1000 in China instead of 15/1000 in France.

[11] By so doing, we suppose that events are distributed equally over the 5-year period. Another hypothesis would be to suppose the annual rates were constant over the 5 years.

[12] We refer here to the average population observed during the considered interval of time (the mid-year population for annual rates).

TABLE 4–3 Comparison of the crude death rate in France in 1969 and the crude death rate Kenya would have had that year with French age-specific death rates

Age	Age-specific rates	Age structure of population		Weight of age-specific rates	
	France, 1969	France	Kenya	France	Kenya
0–4	0.0046	8.45	19.23	0.0004	0.0009
5–9	0.0004	8.60	16.54	0.0000	0.0001
10–14	0.0004	8.41	12.60	0.0000	0.0000
15–19	0.0009	8.39	10.10	0.0001	0.0001
20–24	0.0011	8.43	8.02	0.0001	0.0001
25–29	0.0012	5.96	6.95	0.0001	0.0001
30–34	0.0015	6.16	5.30	0.0001	0.0001
35–39	0.0023	6.71	4.72	0.0002	0.0001
40–44	0.0034	6.79	3.62	0.0002	0.0001
45–49	0.0053	6.47	3.07	0.0003	0.0002
50–54	0.0078	3.80	2.48	0.0003	0.0002
55–59	0.0121	5.65	1.98	0.0007	0.0002
60–64	0.0192	5.22	1.80	0.0010	0.0003
65–69	0.0305	4.52	1.26	0.0014	0.0004
70–74	0.0518	3.07	0.86	0.0016	0.0004
75–79	0.0840	2.03	0.54	0.0017	0.0005
80–84	0.1718	0.97	0.40	0.0017	0.0007
85+	0.4757	0.37	0.51	0.0018	0.0024
Total		100.00	100.00	0.0116	0.0069

It may seem surprising that the birth rate obtained for Algeria is lower than that in France although the relative proportion of women in the Algerian population is augmented by the mass exodus of male workers (Figure 4–16). In fact, this phenomenon is more than offset by the very large numbers at the base of the population pyramid (Figure 4–10), which, on the contrary, reduce the weight of women of childbearing ages. By contrast, in China, the proportion of women of childbearing ages is swollen by the facts that the younger cohorts are reduced by the recent fertility decline, the cohorts aged 50 to 60 years are reduced by the birth deficits due to the Sino-Japanese and second world wars, and the older cohorts are still very small; this explains why the crude birth rate is higher than in France and far higher than in Algeria. We can see how the exceptional Chinese age structure promotes high birth rates, so that, despite the considerable headway already achieved, the fertility level continues to preoccupy the authorities.

The same relationship naturally exists between the crude rates and the age-specific rates of migration. Given the same propensity to migrate at all ages, the crude rate will depend on the population's age structure.

TABLE 4–4 Comparison of the French crude birth rate in 1980 and the crude birth rates 1974 Algeria and 1995 China would have had given the French age-specific rates

Age	Age-specific rates	Proportion (%) of total population			Weighting of age-specific rates		
	France 1980	France 1980	Algeria 1974	China 1995	France 1980	Algeria 1974	China 1995
				Female			
0–14	0.0000	10.69	24.32	12.63	0.0000	0.0000	0.0000
15–19	0.0166	4.04	5.10	3.58	0.0007	0.0008	0.0006
20–24	0.1230	3.97	4.23	4.41	0.0049	0.0052	0.0054
25–29	0.1450	3.96	2.98	5.11	0.0057	0.0043	0.0074
30–34	0.0752	4.07	2.71	4.40	0.0031	0.0020	0.0033
35–39	0.0264	2.94	2.50	3.40	0.0008	0.0007	0.0009
40–44	0.0054	2.74	2.19	3.65	0.0001	0.0001	0.0002
45–49	0.0004	2.99	1.60	2.71	0.0000	0.0000	0.0000
50+	0.0000	16.57	5.20	9.21	0.0000	0.0000	0.0000
				Male			
All ages	0.0000	48.87	50.69	50.90	0.0000	0.0000	0.0000
				Total			
All ages	—	100.00	100.00	100.00	0.0153	0.0132	0.0178

III. THE RELATIONSHIP BETWEEN AGE-SPECIFIC RATES, AGE STRUCTURE, AND POPULATION GROWTH

We have now seen how the two elements of natural population dynamics (fertility and mortality) affect the age structure and how, reciprocally, the latter affects the translation of population dynamics in terms of crude birth and death rates and, therefore, natural increase. To show this, we have focused on the diversity of particular situations often inherited from specific events. Let us now consider globally the historical transformation of age structures resulting from the revolution of demographic behaviors that began in Europe in the 18th century, and more recently in the rest of the world, and how this structural change has affected population growth.

Starting in the mid 18th century in Europe, a long period of in-depth economic and social transformations led to a decisive mortality decline, among other consequences. In a first stage, this mortality decline created a gap with fertility, which remained unchanged, thereby opening up an era of exceptional population growth. In a second stage, fertility also began to decline and population growth decelerated. The birth rate then drew closer again to the death rate until a quasi-balance was once more struck between the two components of natural increase, this time at a very low level for both. These mortality and fertility declines radically transformed the population pyramid, which, in turn, contributed to translating these declines in terms of population growth.

1. The Historical Transformation of Two Population Pyramids: France and England

This same transformation affected France and England differently. In England, the pattern of change was a classic one, whereas in France where the mortality and fertility declines began early and simultaneously, population growth was modest and the aging of the pyramid occurred sooner (Figure 4–20).

In England[13], between 1750 and 1850, the mortality decline, in the absence of fertility decline, produced an unprecedented population growth and a broadening of the base of the pyramid. The total population soared from 5.8 million to 18 million in the space of a century, whereas the proportion of those younger than 15 years rose from 32.9% to 36.0%. With fertility essentially constant[14], the total population growth was due essentially to the mortality decline, but this also made the population younger, since the gain was in infant mortality and this produced a result comparable to an increase in birth rates. By contrast, during the same 100 years in France, the population growth was moderate, from 25 to 36 million, and the proportion of those younger than 15 years fell from 32% to 27%. The reason is that the fertility decline, begun concomitantly with the mortality decline as early as the mid 18th century, so reduced the birth rates (from 40 to 27 per 1000) that the absolute annual number of births stopped rising and even fell slightly, despite the population growth and an age structure that was increasingly favorable to the proportion of women of childbearing ages. On the other hand, as in England, the mortality decline focused on infant mortality, which to some extent thwarted population aging by slowing down the reduction of the proportion of those younger than 15 years.

Half a century later, with the continued reduction of both fertility and mortality in France, the population has virtually stopped growing (from 36 to 38 million) but the proportion of children has continued to fall (from 27.3% to 25.6%). In England, by contrast, the population has continued to progress rapidly (from 18 to 33 million). However, it is also in the middle of this period that the long-term fertility decline finally emerged in England, more than a century after France, and the pyramid can be seen to start narrowing at the base as the proportion of those younger than 15 years falls from 36% to 32%.

Let us jump forward another 50 years to 1951. We find a total population that has barely increased in France (43 million), a proportion of those younger than 15 years that has continued to shrink (23%), and a very irregular population pyramid. In fact, this situation is the combined result of several factors working in opposite directions. On one side, World War I severely reduced the male cohorts born in 1880 to 1898 and produced a major deficit of births in 1914 to 1918. These same depleted cohorts were then hit by World War II and produced new undersized birth cohorts in 1940 to 1944.

On the other side, after 40 more years of fertility decline, the behavioral changes that emerged during World War II suddenly brought about a rise in birth rates after the Liberation, which by far exceeds a simple catching-up of postponed births, and marks the beginning of a new period of population growth in

[13] For England, we reconstructed the historical population pyramids from unpublished data elaborated for the preparation of *The population history of England 1541–1871* (Wrigley and Schofield, 1981) and kindly communicated by Edward Wrigley.

[14] In fact, it increased during the second half of the 18th century, then fell in the early 19th century, so that the birth rate recovered in 1830 to 1850 the level of 35/1000 it had had in 1750 to 1760 (Vallin, 1994).

France. Without this invigoration of fertility, which suddenly widens the base of the pyramid, the proportion of those younger than 15 years would have fallen below 20% and the French population would have been no more than 40 million. These 40 million would not have been reached without the net migration that was positive almost continuously from the end of the last century (Figure 4–20).[15]

During the same interval, England's population continued its strong growth, from 33 to 44 million, but it was now its turn to enter fully into the phase of population aging as a result of the fertility decline. This evolution is accentuated by the birth deficits in the two world wars (even if the impact of the first is much slighter than in France). Finally, like France, it is marked by the baby boom of the late 1940s, but that is not enough to prevent the proportion of those younger than 15 years falling from 32% to 22%, the same level as in France.

At this stage, especially in France but also in England, the changing shape of the population pyramid had a profound impact on the conditions of expression of mortality and fertility in terms of crude rates and consequently of natural increase. In fact, apart from the fluctuations produced by the aging of the depleted wartime cohorts, these changes do not affect so much the birth rates, insofar as there is a gradual swing between the reduction of the share of young people and the increase of the proportion of elderly, leaving the adult (and childbearing) ages relatively stable. By contrast, regarding mortality, the population aging tends to slow down considerably the reduction of the crude death rate as the age-specific rates decline. So, in 1950 to 1951, the crude death rate in France was 12.8/1000, but with the age structure of 1750, it would be no more than 7.5/1000. The crude birth rate, instead of the observed 19/1000, would be 22, whereas the growth rate, instead of 0.6% would be 1.5%, almost three times more. We can see to what extent change in population structure, under the effect of change in fertility and mortality, modifies in turn the resulting population dynamics.

Since 1950, there have been many further changes in France and England. With slight variations, after the exceptional rejuvenation of the pyramids linked to the baby boom, both countries continue their route of rapid population aging under the dual effect of a resumption of the secular fertility decline and a change in the nature of the mortality decline. Indeed, if fertility remained at the relatively high level of the 1950s until around 1970, it then began a new downward trend, the end of which is difficult to forecast today. Also in the 1970s, a new mortality decline emerged, in which cardiovascular diseases took over from infectious diseases as the focus of the victory over death. The result is an unprecedented mortality decline at older ages, while the progress made in infant mortality, already at very low levels, has stagnated. Previously a factor of rejuvenation, the mortality decline thus becomes per se a factor of population aging.

At the dawn of the 21st century, therefore, France and England, with some 60 million inhabitants each, present similar age structures: already filled out towards the top (20% of 60+ in France in 1997 and 21% in England in 1990) and narrowed at the base (19% of under-15s in both countries), but having for the moment exceptionally high proportions of adults aged 20 to 50 years. The latter presently delay the reduction of birth rates, but this reduction may be all the more severe in the future when the following cohorts, who are much less numerous, reach the childbearing ages; at the same time, the currently exceptional proportions of adults will accelerate the aging of the population as they progress further up the pyramid. Thus, a third factor of demographic aging, added to the specific effects of mortality and fertility change, is already written into the age structure of these two populations.

2. Fifty Years of Change in China

Figure 4–21 illustrates the case of China, whose population pyramid has undergone the same combined effects of interactions between structure and dynamics, as a result of the same kind of evolution, but within a much shorter lapse of time. Indeed, China has covered in a few decades virtually the same path as France in two centuries.

In 1953, the Chinese pyramid was almost exactly triangular above age 8 or 9, as a result of long-term near-constant fertility and mortality, which produced a fairly modest rate of growth. However, the base of the pyramid is very broad owing to the rapid but recent decline of infant mortality, a sign of the considerable health progress made as from the late 1940s.

In 1964, the pyramid above age 10 is in large part simply a moving upward of the pyramid for 1953. We note just a perceptible hollowing for the cohorts born around 1945, who were still children during the Great Leap Forward and suffered especially from the famine in 1960 to 1961. It is this period of crisis that cuts deeply into the pyramid at age 4, most acutely, but also at ages 5, 6, and 3, signing the remarkable deficit of births we have previously discussed (Figure 4–7). The

[15] Apart from the 1930s when the economic crisis reversed the trend.

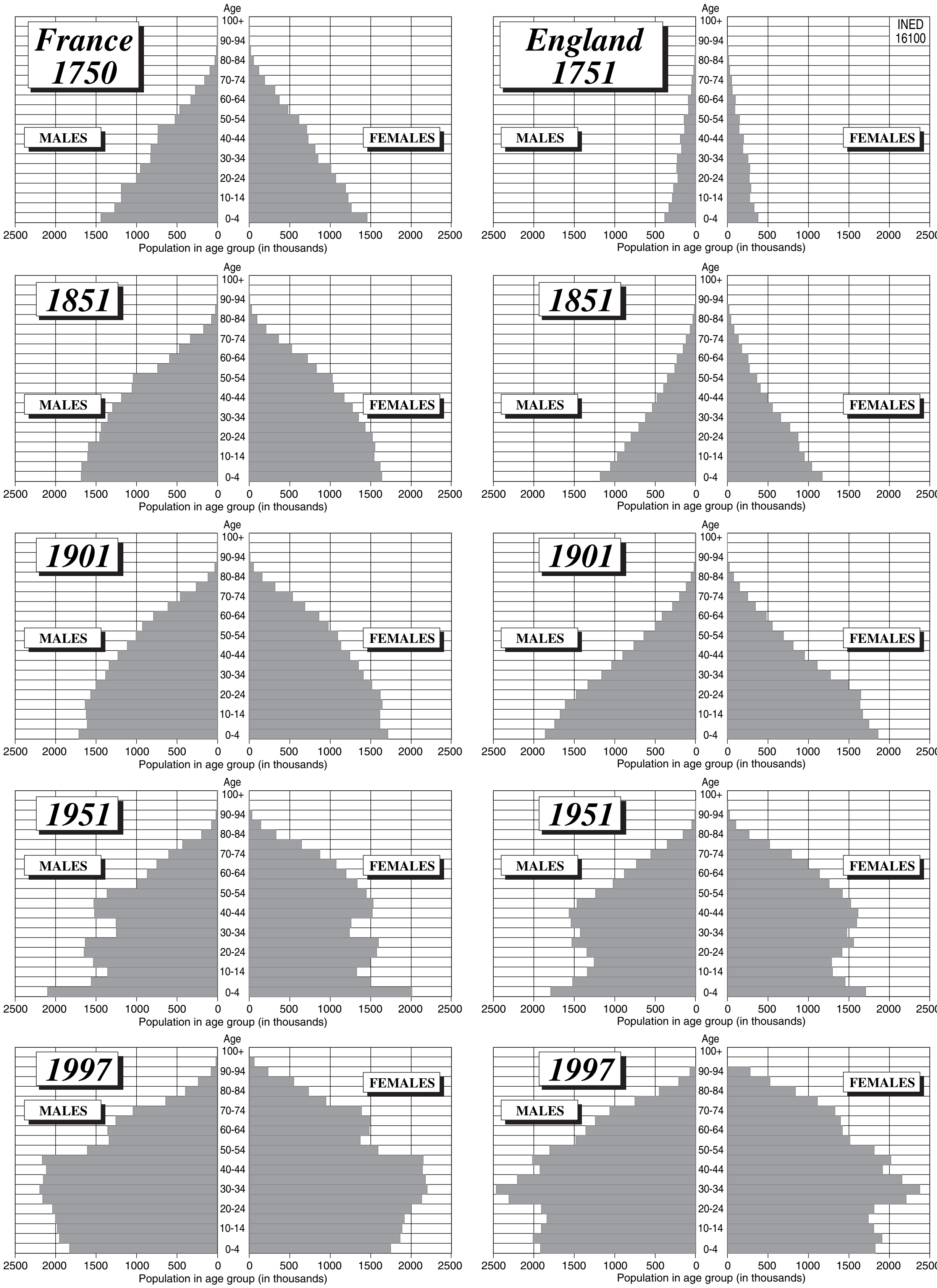


FIGURE 4–20 Evolution of the population pyramids of France and England since 1750.

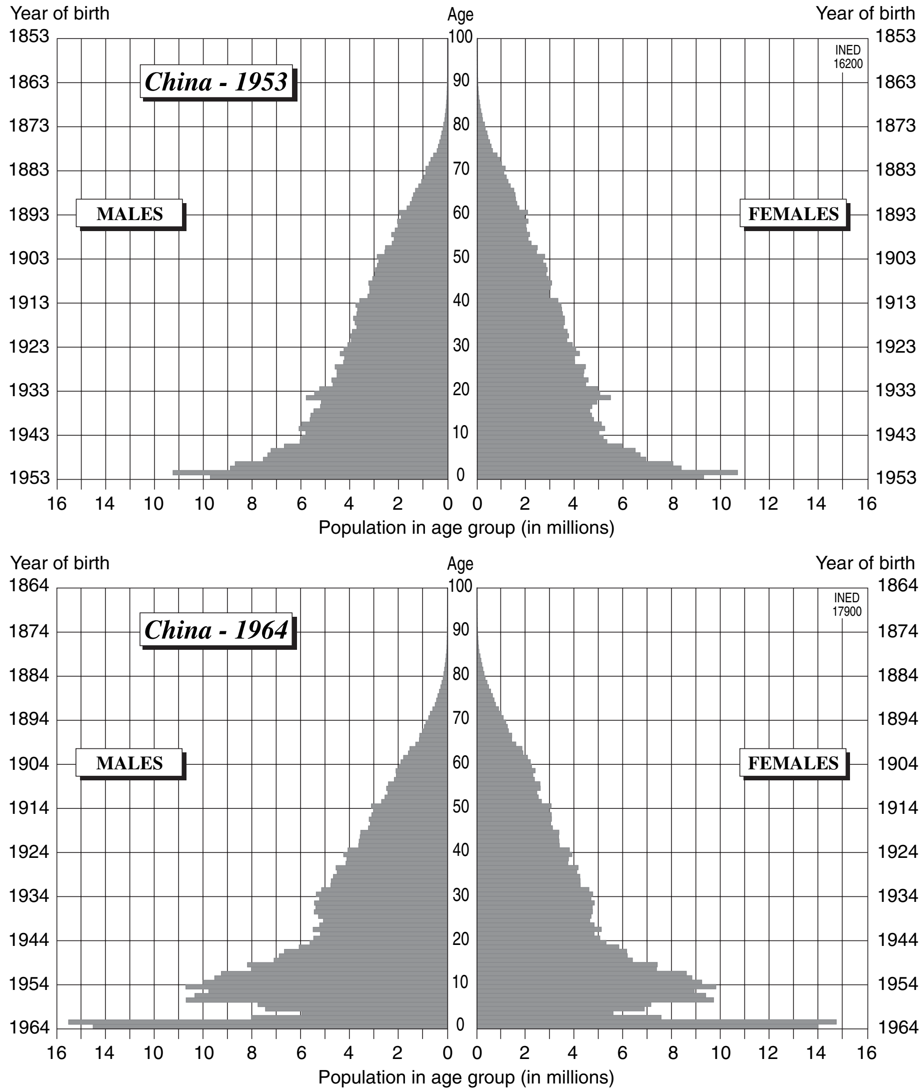


FIGURE 4–21 Evolution of China's population pyramid, 1953–1990.

exceptionally large populations of ages 0 and 1 express the revival of births that then followed this dramatic crisis.

These events can also be traced in the 1982 population pyramid, moved on 18 years from that of 1964. But the base of the pyramid has again changed. After the relatively high fertility of the post–Great Leap Forward years, which produced 10 inflated, although fluctuating, birth cohorts from 1963 to 1973, a new change can be noted. Starting in 1970, the birth control policy became more severe, which downsized the subsequent cohorts. The impact was all the greater as the couples concerned included the depleted cohorts of the Great Leap Forward.

Once again, the population in 1990 bears the marks of these events, some ages higher up, but once again the base of the pyramid is changing. The very large cohorts born after 1962 have now reached the childbearing ages and, as a ripple effect, engender a second baby boom, although a minor one compared to the years 1963 to 1972. Contrary to what was often suggested at the time, the upturn is indeed due more to this structural effect than to a real failure of China's birth control policy. The very base of the pyramid shows, in fact, a reduction of birth rates after 1988 and a new cutback in the size of the last two cohorts.

These interactions between population structure and population dynamics, which we have approached intuitively by examining the historical evolution of a number of population pyramids, will find formal expression in Lotka's equations and in population models (see Chapter 20). Before exploring these, however, we will need to delve further into mortality and fertility analysis (Chapters 11 and 12).

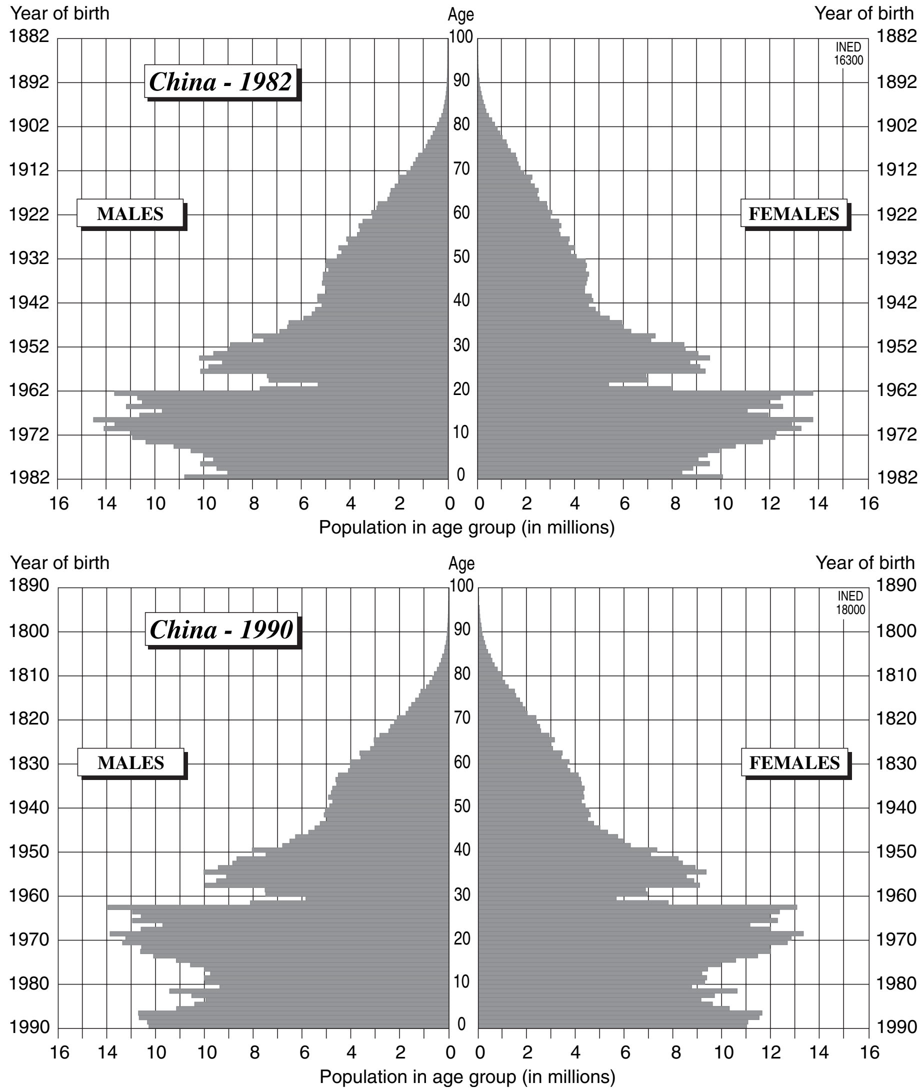


FIGURE 4–21 (*Continued*)

References

Adametz Serge, Blum Alain, and Zakharov Serge, 1994. *Disparités et variabilité des catastrophes démographiques en URSS*. Paris, INED, 100 p. (Dossiers et Recherches, no. 42).

Andreev Evgenij, Darskij Leonid, and Kharkova Tatiana, 1993. *Naselenie sovetskogo Sojuza 1922–1991* [The population of the Soviet Union, 1922–1991]. Moscow, Nauka, 140 p.

Banister Judith, 1987. *China's changing population*. Palo Alto, California, Stanford University Press, 488 p.

Blum Alain, 1991. Redécouverte de l'histoire de l'URSS (1930–1945). *Population et Sociétés*, no. 253, p. 1–4.

Blum Alain, 1994. *Naître, vivre et mourir en URSS 1917–1991*. Paris, Plon, 274 p.

Brouard Nicolas, 1977. Évolution de la fécondité masculine depuis le début du siècle. *Population*, vol. 32(6), p. 1123–1158.

Brouard Nicolas, 1989. *Mouvements et modèles de population*. Yaoundé, IFORD, 184 p.

Chidambaran V. C., Cleland John G. and Verma Vijey, 1980. *Some aspects of WFS data quality: a preliminary assessement*. Voorburg, The Hague, International Statistical Institute 1980, 35 p. (WFS Comparative Studies no. 16).

Clairin Rémy and Lohle-tart Louis, 1988. la collecte des données sur la fécondité, *in:* Louis Lohle-Tart Louis and Rémy Clairin (eds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 225–237. Paris, CEPED, UIESP, IFORD, 330 p.

Committee on Population and Demography Ncr, 1981. *Collecting data for the estimation of fertility and mortality*, National Academy Press, Washington, DC.

CSO, 1995. *Annual Statistical Abstract 1994*. Kuwait, State of Kuwait, CSO (Central Statistical Office), 452 p.

DGSL (Direzione Generale della Statistica del Lavoro), 1914. *Censimento della popolazione del Regno al 10 giugno 1911*, vol. 2. Rome, Ministero di Agricultura, Industria e Commercio, 638 p.

François Michel, 1988. L'état civil, *in:* Louis Lohle-Tart and Rémy Clairin (eds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 91–111, Paris, CEPED, UIESP, IFORD, 330 p.

Ghetau Vasile, 1970. Le redressement spectaculaire de la natalité en Roumanie, *Economie et Statistique*, no. 12, p. 51–55. Paris, INSEE.

Ghetau Vasile, 1983. Évolution de la fécondité en Roumanie: une approche longitudinale, *Population*, vol. 38(2), p. 274–265.

Guzman José-Miguel, 1989. Some new approaches to the collection of demographic data in Latin America, *International Population Conference New Delhi, September 20–27 1989 / Congrès international de la population, 20–27 September 1989*, vol. 2, p. 3–16. Liège, Belgium, IUSSP, 378 p.

INE, 1993. *Censos de poblacion y viviendas 1991*. Madrid, Instituto Nacional de Estadistica (INE), vol. 1, 525 p.

INSEE, 1992. *Recensement général de la population de 1990. Population-activité-ménages*: *le département et ses principales communes, 31 Haute-Garonne*. Paris, Institut national de la statistique et des études économiques, 132 p.

ISTAT (Instituto Centrale di Statistica), 1956. *IX Censimento generale della popolazione, 4 novembre 1951, vol. III*: *Sesso, età, stato civile, luogo di nascità*. Rome, Soc. Abete, 440 p.

ISTAT (Istituto Nazionale di Statistica), 1972. *XI Censimento generale della popolazione, 24 ottobre 1971. Popolazione e abitazioni. Fascicolo della provincia di Macerata, no. 55.* Rome, Società Abete, X+49 p.

ISTAT (Istituto Nazionale di Statistica), 1995a. *Popolazione e abitazioni. Fascicolo nazionale. Italia. 13e Censimento generale della popolazione e delle abitazione 20 ottobre 1991.* Rome, Istituto Nazionale di Statistica, 515 p. +annexes.

ISTAT (Istituto Nazionale di Statistica), 1995b. *Popolazione e abitazioni. Fascicolo regionale Emilia Romagna. 13e Censimento generale della popolazione e delle abitazione 20 ottobre 1991*. Rome, Istituto Nazionale di Statistica, 263 p. +annexes.

Kerjosse Roselyne and Tamby Irène, 1997. *La situation démographique en 1995*: *mouvement de la population.* Paris, INSEE, 264 p. (INSEE résultats, no. 584–585, Démographie-société, no. 65–66)

Leridon Henri and Toulemon Laurent, 1997. *Démographie. Approche statistique et dynamique des populations* Paris, Economica, 440 p. (Économie et statistiques avancées).

Lévy Michel-Louis, 1990. À propos de la Roumanie. *Population et Sociétés*, no. 243, p. 4.

Maffioli Dionisia, 1982. Alcuni risultati relativi ai mariti: fecondità, atteggiamento riproduttivo e contraccezione dal punto di vista maschile, *in:* Paolo De Sandre (ed.), *Indagine sulla fecondità in Italia. Rapporto generale. Volume I. Metodologia e analisi*, p. 361–401. Padoue, Istituto di Statistica, Florence, Dipartimento di Statistico, and Rome, Istituto di Demografia, 446 p.

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIXe siècle. *Population*, vol. 44(6), p. 1121–1158. (Also published in English in *Population: An English Selection* vol. 46(3), 1991, p. 33–62).

Miller Barbara D., 1996. Changing equality: the politics of sex selective abortion in Asia, *in:* IUSSP Committee on Anthropological Demography and Centre for Development Studies (ed.), *Seminar on socio-cultural and political aspects of abortion in a changing world, Kovalam, Trivandrum, India, 25–28 March 1996*, 15 p. Liège, IUSSP, multiple pages.

Mitchell B. R., 1988. *British Historical Statistics.* Cambridge, Cambridge University Press, 836 p.

Muresan Cornelia, 1996. L'évolution démographique en Roumanie: tendances passées (1948–1994) et perspectives d'avenir (1995–2030). *Population*, vol. 51(4–5), p. 813–844.

Negadi Gourari, Tabutin Dominique, and Vallin Jacques, 1974. Situation démographique de l'Algérie, *in:* Ahmed Bahri (ed.), *La population de l'Algérie*, p. 1–62. Paris, CICRED, 174 p.

Ohadike Patrick O. and Tesfaghioghis Habtemariam, 1975. *The population of Zambia.* Paris, CICRED, 184 p.

ONS, 1976. *Annuaire statistique 1975* Alger, Office national de la statistique, 373 p.

ONS, 1994. *Annuaire statistique de l'Algérie*: *résultats 1991.* Alger, Office national de la statistique, 326 p.

PCO, 1993. *1990 China's 4th national population census. Tabulation on the 1990 census.* Beijing, Population Census Office, People's Republic of China, 3681 p.

Picouët Michel 1969. Étude de la structure par âge de la population tunisienne à partir de l'examen des pyramides d'âge, *Revue Tunisienne des Sciences Sociales*, no. 17–18, p. 177–214.

Pison Gilles, 1979. Âge déclaré et âge réel: une mesure des erreurs sur l'âge en l'absence d'état civil. *Population*, vol. 34(3), p. 637–648.

Pison Gilles, 1982. *Dynamique d'une population traditionnelle, les Peuls Bandé, Sénégal oriental.* Paris, INED-PUF, 278 p. (Coll. Travaux et documents, Cahier no. 99).

Roger Gilles, Waltisperger Dominique, and Corbille-Guitton Christine, 1981. *Les structures par sexe et âge en Afrique* Paris, Groupe de démographie africaine, IDP, INED, INSEE, MINCOOP, ORSTOM, 556 p.

SCS, 1964. *Résultats du recensement de 1960, vol. I, Nationalité, sexe, âge.* Rabat, Royaume du Maroc, Service central de statistiques, 160 p.

Simonet Michel, 1974. *Questions de démographie africaine.* Abidjan, Ministère de l économie et des finances, Levallois, Cabinet Olivier, 125 p.

Tabah Léon, 1947. La répartition par âge de la population française en 1851. *Population*, vol. 2(1), p. 349–354.

Tabutin Dominique, 1995. *Analyse des données démographiques déficientes. Méthodes d'évaluation et d'analyse des données démographiques pour les pays en développement.* Louvain-la-Neuve, Diffusion universitaire CIACO, 236 p. (Demo 3255).

United Nations, 1997. *The sex and age distribution of the world population. The 1996 Revision.* New York, Department of Economic and Social Affairs, 884 p. (Document ST/ESA/SER.A/162)

Vallin Jacques, 1974. Rapport de masculinité à la naissance et qualité de l'observation dans l'enquête fécondité algérienne 1970. *Bulletin de liaison du Groupe de démographie africaine*, spécial, no. 7, September 1974, p. 149–172. (Paper presented at the International Population Conference, IUSSP, Liège, 1973).

Vallin Jacques, 1994. *La population française.* Paris, La Découverte, 128 p. (Coll. Repères).

Waltisperger Dominique, 1988. Les structures par âge, *in:* Louis Lohle-Tart and Rémy Clairin (eds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 203–216. Paris, CEPED, UIESP, IFORD, 330 p.

Wrigley Edward A. and Schofield Roger S., 1981. *The Population History of England 1541–1871.* London, Edward Arnold, 780 p.

PART

II

THE LONGITUDINAL APPROACH

Description of Demographic Processes in a Birth Cohort

GRAZIELLA CASELLI AND JACQUES VALLIN

So far, the natural and migratory components of population dynamics have been considered through a period analysis of specific situations. Thus, the period risks of individuals of a given age were approached by defining age-specific fertility, mortality, and migration rates. These age-specific risks clearly display between-population variability and between-period variability within the same population. A closer analysis means clarifying, for each component of change, the nature of the relations between these age-specific risks, to combine, then incorporate them in population models, to give a fuller picture of the mechanics of population dynamics.

The most straightforward way to understand what these age-specific rate series mean and how they can be combined is to stay as close as possible to the real-life experiences and events of the individuals who make up the population. Instead of looking at the risk exposure at each age during a single year, we will look at the risk exposure at each age of their lives by a universe of individuals born the same year. This is what demographers call the *cohort approach*, as opposed to the *period approach*, based on an analysis of the risks in a specified period.

Essentially, this is a purely descriptive approach to demographic phenomena, considered discretely. It is an essential first step before moving on to a closer analysis that involves the way they interact. This step obviously rests on two fundamental working assumptions: The population is *internally homogeneous* and the phenomena studied are *independent*. Obviously, these assumptions will almost never be perfectly true at all times, so they must be discarded at a later stage to analyze in more specific detail how these different phenomena *interact* in *heterogeneous populations*.

We first briefly describe an age-specific rate matrix by which to distinguish more precisely between the cohort and period approaches (Chapter 5), and then specify the different time breakdowns usable in demographic analysis (Chapter 6) and methods of accounting for and combining cohort and period approaches (Chapter 7). We will then consider the conceptual differences between different measures of incidence relations—frequencies, rates, and probabilities (Chapter 8)—and look more closely at the assumptions of independence between events and population homogeneity (Chapter 9). We will discuss observation methods specific to the cohort analysis of fertility, mortality, and migration (Chapter 10), preparatory to constructing a cohort mortality table (Chapter 11), defining the various longitudinal fertility measures (Chapter 12) and introducing the cohort analysis of migration (Chapter 13).

C H A P T E R

5

Variation Through Time of Age-Specific Rates

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

Table 5–1 shows a few elements making it possible to find one's way in the matrix of age-specific death[1] rates observed during two centuries in France. In fact, it is rare to have access to such series of age-specific rates. To calculate them, one needs annual statistics concerning age-specific deaths and age-specific population estimates for January 1 of each year. In France even the age-specific death series is only available from 1899 and INSEE, the French Statistical Institute, has only been publishing regular annual estimates of the population on January 1 of each year since 1950. The death rates used here therefore are based essentially on historical reconstructions.[2]

The type of table shown in Table 5–1 can be read from different angles. Each column gives a series of age-specific death rates calculated for a given calendar year, whereas each row represents the annual variations in age-specific mortality. Therefore, Figure 5–1, which covers 10 annual age-specific death rates in columns, illustrates the evolution, in 20-year intervals over two centuries, of the age-specific curve, whereas Figure 5–2, which covers 10 annual series of rates represented in rows, illustrates the evolution of mortality at the ages of 0, 10, 20, . . ., 80, and 90 years, since 1806. These are two illustrations of the same reality seen, on the one hand, from the angle of transformations that health improvements have imposed on the age-specific structure of mortality and, on the other hand, from the angle of specific changes in causes of death at different ages.

This matrix can however, also be interpreted from a third angle, by following people's biographies from their birth until their death. All that is needed for this is to start from age 0 in a given year, to pass to age 1 of the following year, to age 2 of the year after that, and so on until the limit imposed by the available data. If one starts at age 0 in 1881, for example, one reaches the age of 100 in 1981. In this way we follow, according to age, the mortality that affects approximately[3] all of the individuals born during the same calendar year, which demographers commonly refer to as a *birth cohort*. Thus, Figure 5–3 illustrates the changes in the age-specific mortality structure of birth cohorts, by 20-year birth intervals, since 1806.

The differences between Figures 5–1 and 5–3 illustrate the difference between the period and cohort approaches to mortality. One example of the passage from one to the other is indicated on each of the two figures by points in bold.[4] In Figure 5–1, these points make it possible to follow observations of the 1926 birth cohort. The difference between these two illustrations of the same database is obviously due to the fact that mortality has changed over the centuries with improvements in health status. A given birth cohort

[1] The rates used here have been obtained directly, on the basis of deaths by age and on population estimates by age for the period from 1899 to 1996 (Vallin and Meslé), but indirectly, for the period from 1806 to 1898, on the basis of estimated annual age-specific probabilities of dying from historical reconstructions of complete annual life tables (Meslé and Vallin, 1989). The corresponding definitions and calculation methods are described in subsequent sections.

[2] For the period from 1899 to 1950 (Vallin, 1971) and for the 19th century (Meslé and Vallin, 1989).

[3] We will see in subsequent sections that a birth cohort can only be rigorously followed if special conditions are respected in terms of computation.

[4] Birth cohort 1926 in Figure 5–1 and year 1946 in Figure 5–3.

TABLE 5–1 Elements of a complete matrix of annual age-specific death rates for France from 1806 to 1995 (rates per thousand)

Age	Calendar year												
	1806	1807	1808	. . .	1850	. . .	1900	. . .	1950	. . .	1994	1995	1996
0	188.97	199.18	193.95		145.16		188.78		53.54		5.99	5.00	4.80
1	80.98	84.04	75.82		59.09		34.20		5.28		0.49	0.51	0.45
2	37.73	39.16	35.33		27.54		18.73		1.84		0.28	0.27	0.28
3	25.01	25.96	23.42		18.25		11.83		1.26		0.23	0.23	0.19
4	20.36	21.05	19.18		14.69		9.08		0.98		0.18	0.20	0.19
5	16.71	17.19	15.85		11.89		6.52		0.82		0.19	0.18	0.15
. . .													
50	22.29	22.55	21.58		15.94		16.68		8.59		4.62	4.58	4.24
. . .													
90	274.70	269.48	271.05		323.84		357.68		317.38		171.43	172.79	160.00
91	291.67	285.63	287.70		350.54		417.11		292.92		187.34	188.85	177.69
92	309.66	302.73	305.33		379.34		399.07		312.71		206.08	212.54	192.12
93	328.72	320.81	324.01		410.38		448.01		309.79		229.75	230.50	211.11
94	348.90	339.93	343.78		443.79		363.17		321.10		253.94	253.65	225.48

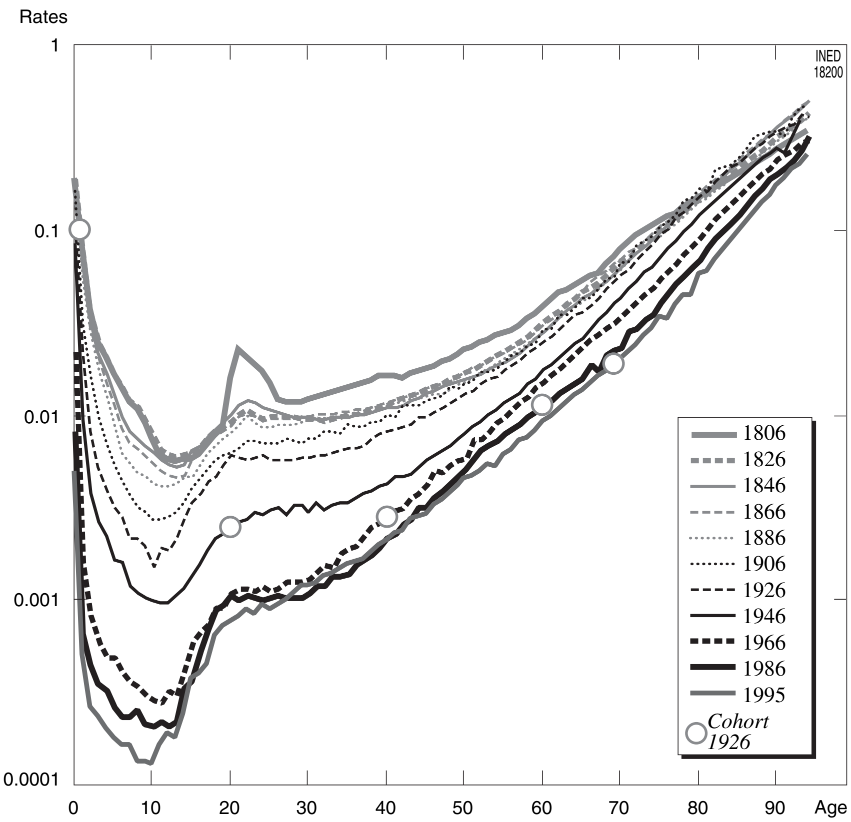


FIGURE 5–1 Year of observation: Age-specific death rates for different years (in 20-year intervals from 1806 to 1995) in France.

benefits from health conditions that change as it grows older. If, for example, one compares the mortality of a birth cohort with that of the year of its birth, mortality at the first year of age will of course, be largely shared by both series, but gradually the mortality conditions of the birth cohort will grow further away from those of the year of birth and become closer at the older ages to those of the year at which the birth cohort will die out. On the whole, since the change is a decline in the death rate, conditions at the end of life will be better than those prevailing in the year of birth. However, more temporary differences may appear in the

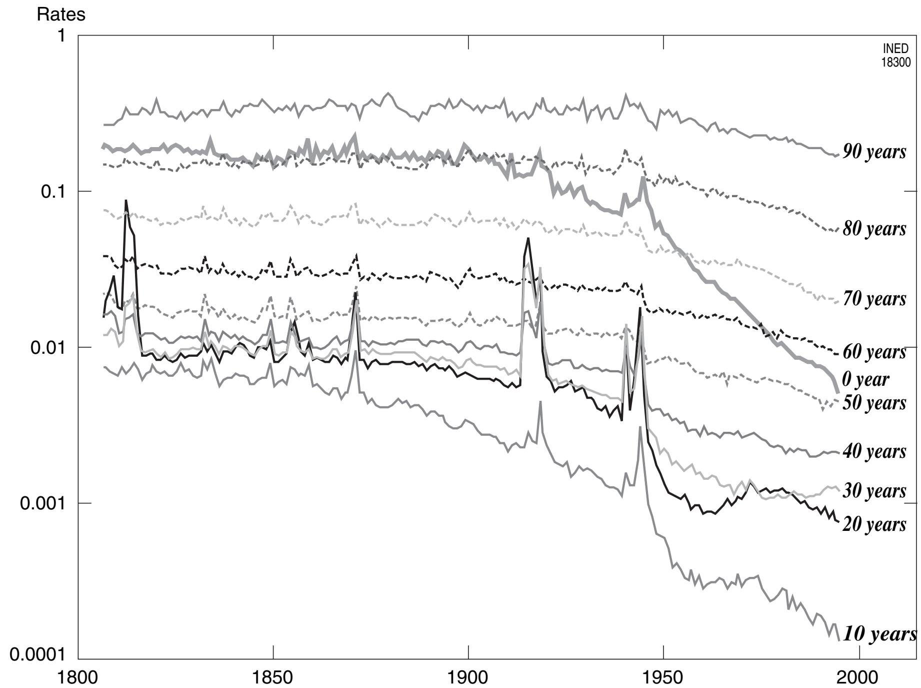


FIGURE 5–2 Annual change (1806–1996) of certain age-specific death rates (0, 10, 20, . . ., 80, 90 years) in France.

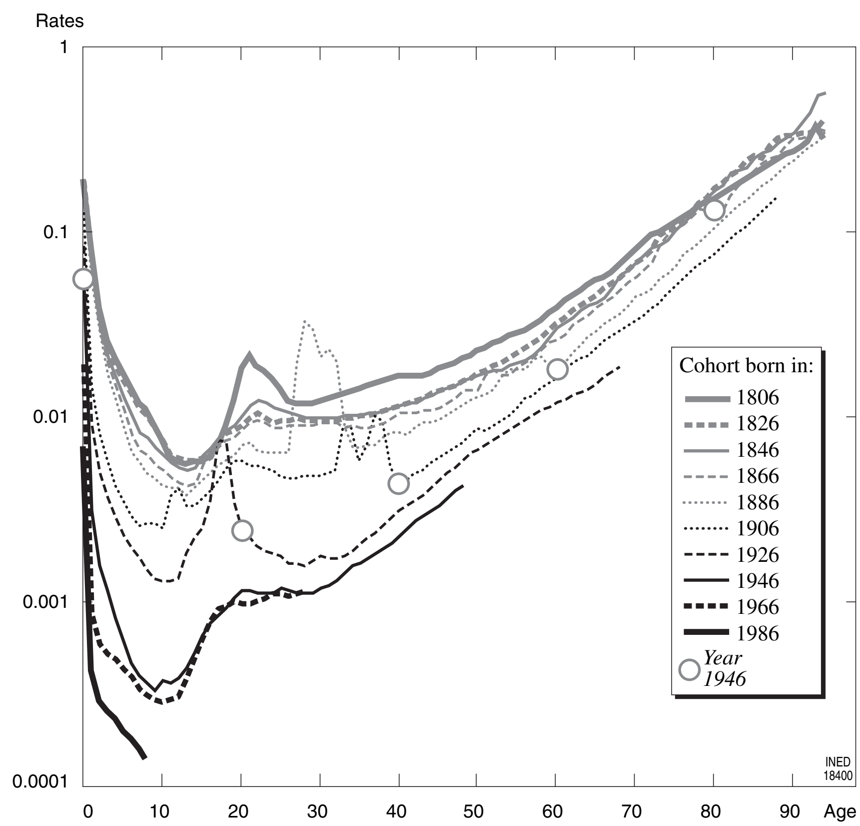


FIGURE 5–3 Age-specific death rates for cohorts born at 20-year intervals from 1806 to 1986 in France.

opposite direction. Figure 5–3, for example, clearly indicates that certain birth cohorts have experienced, because of the two world wars, specific mortality risks at certain ages that make the age-specific mortality curves completely atypical.

It should immediately be pointed out that death rates by year of age and by year of observation calculated by period, such as those in Table 5–1, are not an exact reflection, when assembled diagonally, of the mortality of birth cohorts. In fact, at each age they measure the mortality of a given calendar year, whereas at each age the mortality of a birth cohort must necessarily be spread out over two calendar years. This is what we must now explain by presenting more clearly how demographic events may be located in time and how they should be grouped together in order to measure their frequency.

References

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIX[e] siècle. *Population*, vol.44(6), p.1121–1158. (Also published in English in *Population, English selection* vol. 46(3), 1991, p.33–62).

Vallin Jacques, 1973. *La mortalité par génération en France depuis 1899*. Paris, INED, PUF, 484 p. (Travaux et Documents, Cahier 63).

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX[e] et XX[e] siècles et projections pour le XXI[e] siècle*. Paris, INED, 102 p. + CD-rom. (Données statistiques, no. 4, 2001.)

C H A P T E R

6

From Situating Events in Time to the Lexis Diagram and the Computing of Rates

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

I. TIME, AGE, COHORT

1. Absolute Time and Relative Time: Dates and Ages (or Durations)

Any demographic event may be situated in time in two ways, according to whether one refers to *absolute time*, as marked out in the official calendar (its *date of occurrence*) or in *relative time*, determined by the time elapsed between the event in question and an initial event directly concerning the individual. If this initial event is the person's birth, this relative time will be the *exact age* the person will have reached when the event studied occurs. But the initial event may, in certain cases, also be of another nature, for example, arrival in the country, starting school, marriage, a first pregnancy, and so forth. We are then dealing with specific concepts of *duration of a state* or of the time elapsed since a given event. It is, in fact, possible to distinguish two types of relative time according to whether the initial event and the event being studied concern the same person:

- Relative time between different events in the life of the same person (length of time between first marriage and birth of the first child, time elapsed between two successive births, time elapsed between one migration and the following migration, etc.);
- Relative time between different events in the lives of different people (time elapsed between the death of a young child and the birth of the following child, time elapsed between the death of a person and that of his/her spouse, etc.).

According to the objective of the analysis, a choice will have to be made between one relative time and another. If, for example, one wishes to measure fertility in marriage without making the distinction of birth order, reference must be made to all the births of the same woman at the time elapsed since marriage. If on the contrary one wishes to analyze independently the occurrences of successive births, for each birth the time elapsed since the previous birth will be used, with reference to the marriage (or first sexual intercourse) only for the first birth. In the same way, one may analyze professional careers either by constantly referring to the time elapsed since entry into employment or by referring for each employment to the time elapsed since obtaining the previous employment.

Last, when combining in the same analysis the impact of several phenomena, for example fertility and migration, one may have to introduce a multidimensional time, associating elapsed time between births and elapsed time between successive migrations.

2. Exact and Complete Age (or Duration)

Time is constantly passing, of course, and events occur selectively, each one at a precise moment, a *date*, if absolute time is concerned, or an *elapsed period of time*

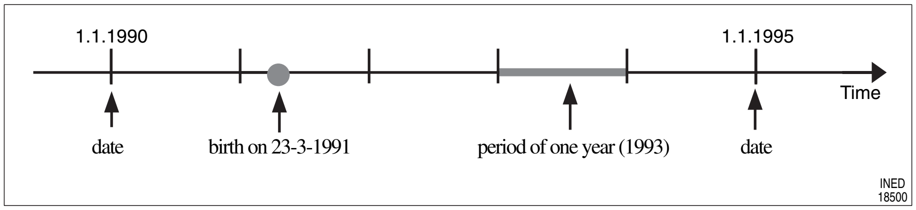


FIGURE 6–1 Absolute time elapsed, date, period, and situation of an event on the basis of the calendar year.

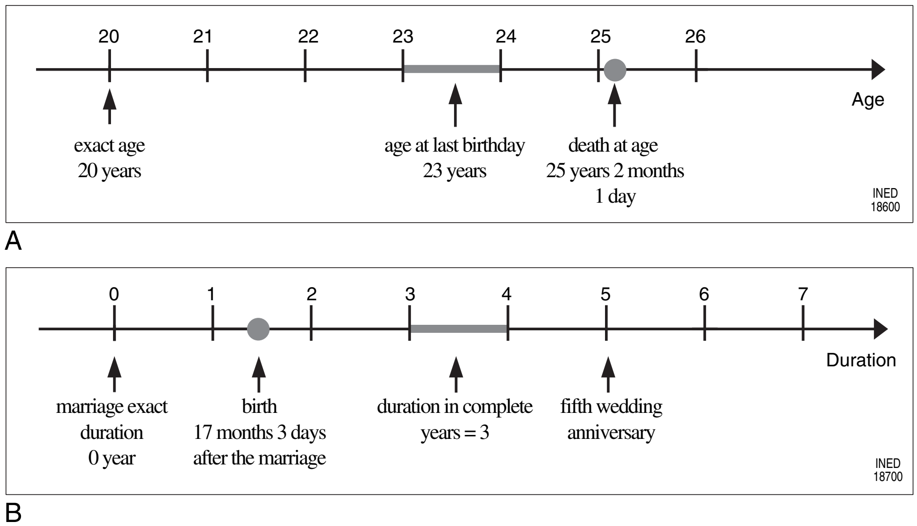


FIGURE 6–2 **A:** Relative time elapsed since the birth: birthday, exact age at the time of the event, and age in complete years. **B:** Relative time elapsed since a given event, for example a wedding: wedding anniversary, exact duration of the marriage at the time of the event and complete duration of the marriage.

(*exact* age, *exact* length of stay, *exact* length of marriage, etc.) if relative time is concerned. But between these events there are discontinuities, and for calculating rates one must bring together those events that occur during the same lapse of time. This is what we have done until now, without any greater degree of precision, by adopting the calendar year as a unit of time, which is the reference generally used in demography. A line representing absolute time in Figure 6–1 localizes the first of January, which marks the beginning of the calendar years, the date of an event (in this case a birth), and a period of one calendar year during which events may be observed. Figure 6–2A situates in the same way, but this time on a line representing age, a number of birthdays, the exact age of an event (in this case a death), and a length of time elapsed between two birthdays, which is usually referred to as *age in complete years* or age at last birthday, whereas Figure 6–2B represents the exact duration of a marriage at the time of the birth of a child as well as different wedding anniversaries and the complete duration of the marriage.

3. Complete Age and Age Reached During the Year

This distinction between exact age and age in complete years may seem naive. However, it is fundamental for the analysis. To be correctly used, a statistic by age (or time elapsed since a previous event) must have been clearly established according to a correct definition of the age (or of the duration) at the time when the data are gathered. When, in a census, a survey or a declaration made to the civil registrar, the age of a person is registered, the way the age of the person is calculated must be agreed upon. This will rarely be the exact age because it would take an unnecessarily long time to detail the age in years, months, days, hours, and so forth, without attaining total accuracy. An approximate solution that is often used consists of obtaining the date of birth as precisely as possible (day, month, year) and to then deduce the age by deducting the difference from the date of registration. If the date of birth is not used, there are possible choices between several age concepts. The age in complete years, the one we defined in the prior section, is customarily fixed at the level reached at the last birthday and remains constant until the following birthday. This must not be confused with the *age reached during the year*[1] (or in period differences), which remains constant during the whole calendar year in terms of the birthday that will be that year. Nor should it be confused with the *individual year of age*. Therefore, on March 31, 1999, a baby born July 1, 1998, will be *aged 0 in complete years* but it will be 1 *year old in the year* and in its first *year of life*. But 6 months later, on September 30, the infant will be aged 1 *complete year*, 1 *year in the year*, and will be in its second *year of life*.

In fact, although saying "I am in my x[th] year of life" is quite common in certain cultures, this notion of age is almost never used in the production of statistical tables. On the other hand, the two other solutions of grouping together exact ages into *age groups* are used and it is essential for the analysis to know which type of age group is concerned: standard *age in complete years* or *age reached during the year*.

4. Birth Cohort and Cohorts in General

From the moment that an event is dated in a precise manner according to the calendar (date) and the age (exact age), a third parameter may be deduced, which sometimes resembles a third dimension of time: the fact of belonging to a *birth cohort*, which is an essential tool in longitudinal analysis in demography.[2] A birth cohort is composed of all the people born during the same calendar year. This notion of cohort can in fact be generalized, as explained in a previous section, to the whole time elapsed since a prior event (migration, marriage, entry into employment, etc.), and reference is thus made to *marriage cohorts* or cohorts in general. For practical reasons until the end of this chapter we will deal essentially with age, but the same approach is valid for any time elapsed since a given event.

II. THE LEXIS DIAGRAM

A little over a century ago, Wilhelm Lexis (1875) proposed a method, which has since become famous, for illustrating demographic events through the use of the dual time dimension (calendar and age) and representing the different useful classifications for analysis, according to age, period, and birth cohort.[3] At the cost of a transformation rendering it even more practical, this mode of representation was popularized at the end of the 1950s among French-speaking demographers, and by Roland Pressat, on the occasion of the publication of the first edition of his manual of demographic analysis (1961).

1. Point-events, Lifelines, and Birth Cohorts

The principle is very simple. It is a Cartesian system in which the first axis measures absolute time (the calendar) and the second, relative time (age). It is then very easy to situate within it any event (birth, marriage, migration, maternity, death, and so forth) according to these two coordinates. By linking the events concerning the same person, a *lifeline* is obtained, which is automatically situated on a parallel to the bisector of the two axes, since age increases at the same pace as time elapses. Figure 6–3, for example, represents the life of a woman born on January 31, 1902 who migrated on March 16, 1916, married on May 22, 1926, gave birth to a child on May 8, 1927, and died on July 7, 1939. At each date of each event, there is a corresponding vertical line that cuts the lifeline at a point situated on a horizontal line indicating the exact age at which this event occurs.[4]

[1] The age reached during the year is the difference between the year of observation and the year of birth. Therefore, during the whole of 1998, a person born in 1956 is 42 years of age reached during the year.

[2] The *birth cohort*, a group of people born in the same year, is indeed a special example of the more general notion of the *cohort* (group of people having experienced the same event during the same unit of time).

[3] There is a brief summary of the origins of this type of representation in an article written by Graziella Caselli and Enzo Lombardo (1990) and in an article by Christophe Vanderschrick (1992). In fact the idea originally came from Georg Knapp (1868, 1869) before being adopted by Karl Becker (1874) and clarified by Wilhelm Lexis (1875), but also and at the same time, by Abraham Verweij (1874), alias Abraham Verwey (1875).

[4] In fact, this definition of what is known as the *Lexis Diagram*, proposed by Roland Pressat, is different from the one developed by its inventor. Wilhelm Lexis followed the lives of people on a vertical to the time line, by tracing *life durations* and not the *lifelines* drawn by Roland Pressat. Thus, he followed, in counter-diagonal lines, *isochrones* (groups of events occurring at the same date) which, in Pressat's diagram, are read vertically. It should be noted that this currently standard presentation of the Lexis diagram is however, not universal. The Anglo-Saxons, who use it much less than French-speaking demographers, sometimes draw an inverted diagram in which the age is presented on the x-axis and time on the y-axis, from the top downwards. The lifelines are therefore descending diagonals. However, Anglo-Saxons also use the presentation proposed by Pressat. Nathan Keyfitz, for example, used the same year the Pressat-type presentation for his *Applied Mathematical Demography* (Keyfitz, 1977b) and the inverted presentation for the revised edition of his *Introduction to the Mathematics of Population* (Keyfitz, 1977a), the first edition of which dated from 1968.

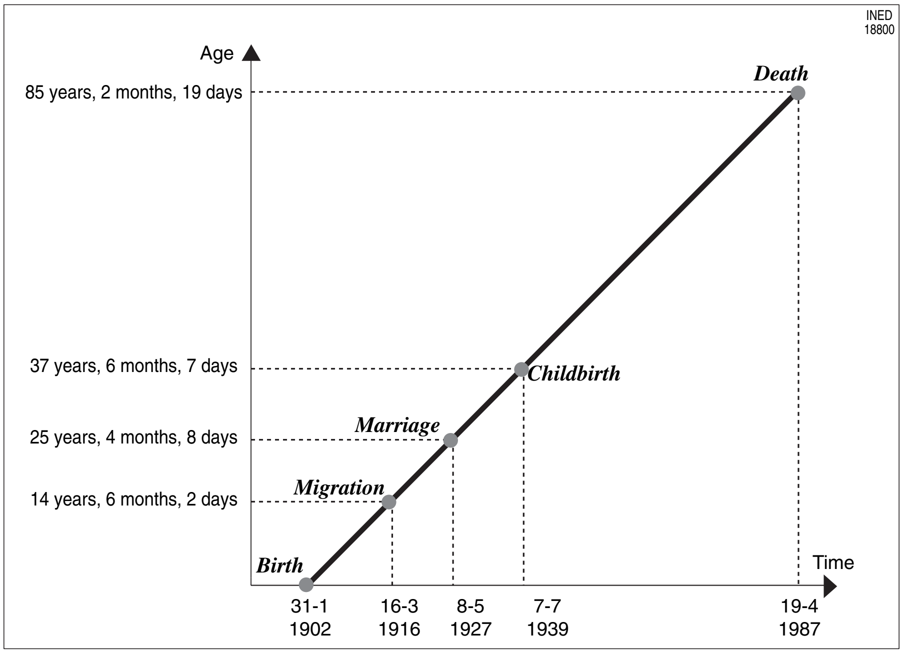


FIGURE 6–3 Representation, on a Lexis diagram of a lifeline, with an event point (migration, marriage) and a terminal death point.

2. Populations and Events

The Lexis diagram makes it possible to make useful classifications for the analysis of events as well as of populations. Concerning populations (Figure 6–4) lifelines may be grouped together in columns of varying thickness according to the time elapsed between the births of the first lifeline from that of the last lifeline. If one groups together the lifelines of people born the same year, framed within diagonals starting January 1 and December 31 of the year in question, one identifies a birth cohort. Our graph represents, therefore, a beginning of a column of lifelines composing birth cohort *C*, born in 1991. The people still living, when the column of lifelines of a same birth cohort cross a vertical line in the diagram, make up the population of this birth cohort at the date corresponding to this vertical line. If the date is January 1, these people are all of the same age in complete years and they make up the population of this age at this date.[5] Thus, *P* represents the population aged 2 on January 1, 1991, composed of people from the birth cohort of 1988 still alive at that date.

Concerning events, the sets of births, occurring by definition at the exact age of 0, obviously always appear as segments on the time axis.[6] Therefore, on Figure 6–4, *B* represents the births of 1993. But all the other events (marriages, migrations, deaths, etc.), occurring at different ages, will be represented in specific surfaces of the Lexis diagram according to the way they have been identified by age and time.[7] Only very special events such as *birthdays* will also be assembled on the segments of lines (parallel to the time line). Therefore, segment *V* in Figure 6–4 illustrates the

[5] In certain statistical publications one may find populations on January 1 and in others, populations on December 31. In fact, there is no difference between a given population on December 31 of one year and the same population on January 1 of the following year. In fact, in the first case the estimate is made at the end of the year, December 31 at midnight, and in the second case it is made at the beginning of the year, January 1 at 0 hour.

[6] At least in this general case of a diagram representing the life of individuals within a population. In such a case births must be represented at age 0. However, these births are in fact the sum of all the *maternity* points that appear, in the course of a year, among women of different ages. These maternity points, as we shall see in the case of fertility analysis (Chapter 12) will be grouped together within surfaces of the Lexis diagram according to the age or the cohort of the mother. The births of a mother will be considered as events whereas all of the births of one year appear here as the initial size of the birth cohort.

[7] Note, however, that the Lexis diagram may be used while replacing the age by the time elapsed since a given event. It is then this reference event, which will be represented by the segments on the time line. Thus, if one studies fertility in marriage and refers to the time elapsed since marriage and not to the mother's age, the lifelines will start with the marriage and will be grouped together in *marriage cohorts*. The marriages of each cohort will occupy a segment of the time line but the maternities (along with any other events such as migration, death, etc.) will be identified among the different surfaces of the Lexis diagram.

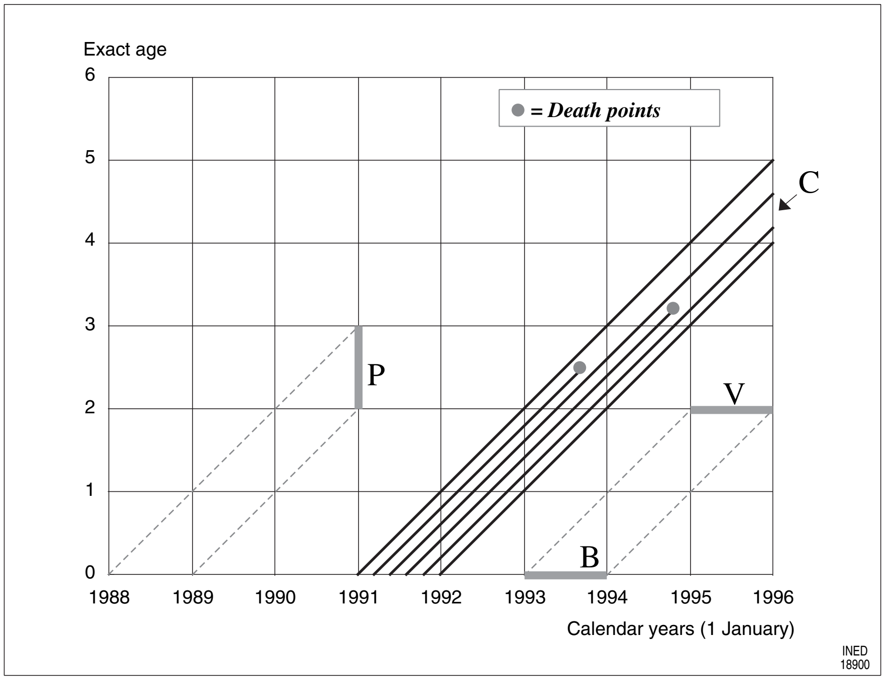


FIGURE 6–4 Lexis grid with the births of 1 year, the population by age, and the column of lifelines of a birth cohort.

second birthday celebrated in 1995 by individuals born in 1993. This number may also be interpreted as being that of the people belonging to the birth cohort of 1993 in the population having the exact age of 2 years.

3. Different Ways to Classify Events by Year of Age

Figure 6–5 illustrates the three standard ways of classifying events according to age, time and birth cohort, which correspond to the three ways that will allow us to locate events classified in the usual statistical publications. Parallelogram E_2 groups together events having occurred during a given year (in this case year A_4) within a given birth cohort (in this case, the cohort born in year A_0). This is how events classified by *calendar year and year of birth* or by *calendar year and age reached during the year* are situated. It becomes immediately apparent that these events occur at the exact age of $x + 4$, but that they do not all occur around the same age in complete years. Some occur before $x + 4$ and others afterward. They therefore concern two age groups in complete years, groups $(x + 3, x + 4)$ and $(x + 4, x + 5)$, without, for that matter, grouping together all the events that may be observed that year in these two age brackets.

On the contrary, parallelogram E_1 groups together events having occurred within the same birth cohort (born in year A_0) at the same age in complete years, the age $(x + 1, x + 2)$. It represents the ages classified by *age and year of birth* or even *age and cohort*. This time it should be noted, however, that they occur before and after a January 1 and are distributed therefore over 2 years (here, years A_1 and A_2).

The third way of classifying events according to age and time consists of grouping by *age and calendar year* or, more precisely, by *age in complete years and calendar year*. This classification corresponds to square E_3 in Figure 6–5. This time, we are indeed dealing with a group of events all occurring at the same age in complete years, the age bracket $(x + 1, x + 2)$, during the same calendar year (here, year A_4). On the other hand, they do not all concern the same birth cohort, since some of them occur among people born in year A_2 and others, the following year, A_3.

Each of these classification methods leads to different perceptions of reality that may seem very similar but that may also, in certain cases, have significant consequences for the analysis. The Lexis diagram makes it possible to avoid sometimes major errors of interpretation.

Otherwise, each one of these three computation methods presents different advantages depending on the type of analysis being made. Only the first method of classification, by calendar year and year of birth, may be used indifferently for either *cross-sectional analysis* (by calendar year) or *longitudinal analysis* (usually by cohort). On the other hand, it presents the

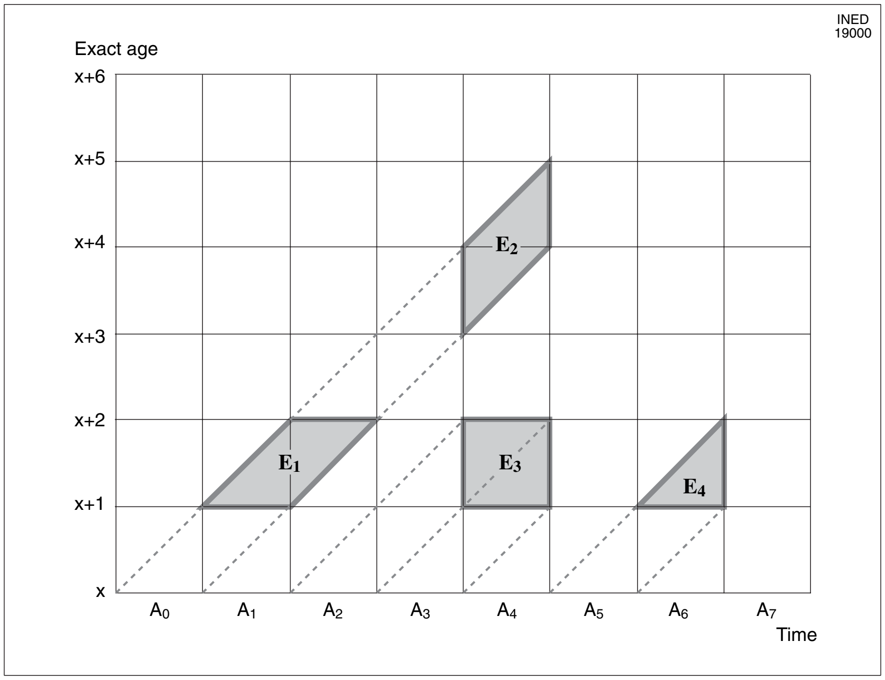


FIGURE 6–5 Representation of events on the Lexis diagram according to different possible annual classification modes.

disadvantage of not clearly identifying the age, because the events are spread out before and after the birthday, whereas age is more naturally perceived as an interval between two birthdays. For cross-sectional analysis, distribution of death by age in complete years and calendar years (the squares of type E_3) is preferred, whereas the ideal for longitudinal analysis is to have deaths distributed according to age in complete years and by cohort (parallelograms of type E_1).

Increasingly, however, current statistics are delivered according to what is customarily called *double classification*, which makes it possible to reconstruct the three cases of the previous figures indifferently. The events of a given calendar year, thus classified *by age and by year of birth* or *by age and by cohort*, then appear in the triangles of the Lexis diagram, such as triangle E_4 in Figure 6–5, which represents the events occurred during the year A_6 in the cohort born in the year A_5 and therefore, with an age in complete years (x + 1, x + 2). By appropriately combining these triangles two by two it is, of course, possible to reconstruct the parallelograms and the square described here.

4. Pluri-annual or Infra-annual Classifications

Conversely, statistics may only provide large classifications, by grouping together ages, calendar years, or cohorts. Here again, the Lexis diagram is very useful for rapid visualization of the type of data one is dealing with. Figure 6–6 provides a few examples. Parallelogram *ABCD*, for example, represents a case of events observed during the same calendar year (year A_5) but classified *by quinquennial groups of birth years* (cohorts relating to the years A_0 to A_4). The individuals concerned have, in the year A_5, exact ages between x and $x + 6$ years, but some events having occurred that year at the ages in complete years of x or x + 5 are excluded, because they either concern the cohort that is too old, relating to A_{-1} or the cohort that is too recent, relating to A_5 (the two triangles next to the extremities of parallelogram *ABCD*).

In the same way, rectangle *EFGH* represents the case of events having occurred in the course of the same calendar year (year A_7) classified by *quinquennial completed age groups*. It is apparent that these events concern not five but six cohorts (those relating to the years A_2 to A_7).

Lastly, parallelogram *FIJK* represents the case of events observed, for example, during a retrospective survey, classified by *quinquennial age groups within the same cohort*. In this case, the observation covers six calendar years (A_8 to A_{13}).

These types of data may also be classified by calendar year within the same cohort (parallelogram *LMNO* in Figure 6–6). The observation within the same birth

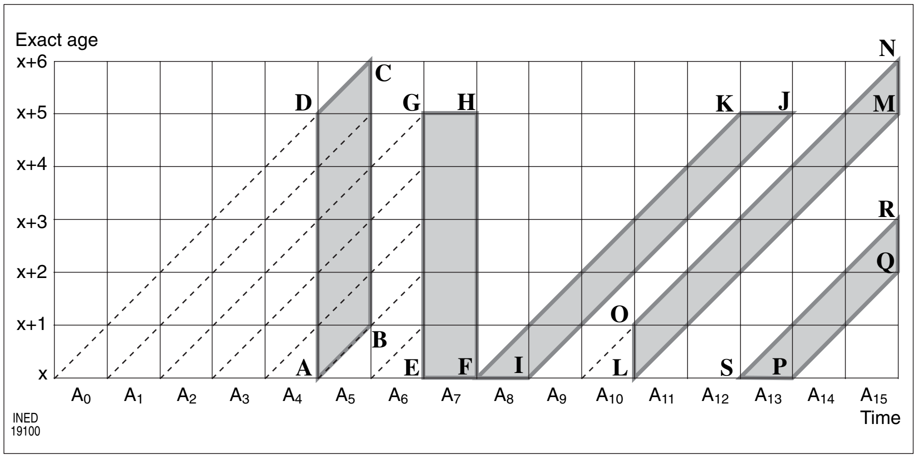


FIGURE 6–6 Representation of events on the Lexis diagram, according to different age groups.

cohort therefore covers five full calendar years but the extreme ages at last birthday are incomplete.

It is even more obvious in this case that the classification of events by calendar year and by quinquennial age group (rectangle *EFGH*) applies perfectly to cross-sectional analysis, whereas classification by quinquennial age groups within the same cohort (parallelogram *FIJK* in Figure 6–6) naturally lends itself to cohort analysis. If one accepts a slight difference in the situation of ages this is also respectively the case for the classification of events in the year by quinquennial groups of cohorts (parallelogram *ABCD*) for cross-sectional analysis, and for events concerning a cohort classified by groups of calendar years (parallelogram *LMNO*), for cohort analysis.

Sometimes classifications of events require changes in criteria. Notably, the observation of data from retrospective surveys, which are usually provided by cohort, find themselves, however, interrupted by the date of the survey and some groups of events are of the same type as that represented in Figure 6–6 by the trapezium *PQRS*, which represents events having occurred in a cohort relating to year A_{13} during the 3 years (A_{13} to A_{15}) before the survey. Here again, we have chosen a favorable, hypothetical case in which the survey took place on January 1.

Finally, data may often be supplied by groups of years of observation. Figure 6–7 situates on the Lexis diagram the three types of classifications in the case of decennial age groups and groups of five cohorts or of five calendar years: *completed age groups and calendar year groups* (rectangle *ABCD*), *intervals of exact age and groups of cohorts* (parallelogram *ABEG*), and *completed age groups and groups of cohorts* (parallelogram *BEFH*).

Of course, in addition to helping to visualize different pluri-annual groups, the Lexis diagram also provides valuable help in illustrating the classification of events according to age and calendar intervals that are smaller than a year. All that needs to be done in such cases is to change the scale. It is possible to work in months, weeks, days, or in any other unit of time adapted to the data being analyzed.

III. ELEMENTS FOR COMPUTING A RATE, ACCORDING TO THE CLASSIFICATION MODE OF EVENTS

1. Rates for a Year of Age

Previously, we defined the rates by age for fertility, mortality, and migrations (Chapter 4). In the case of mortality, for example, we have seen that the rate for year t at age x is equal to:[8]

$$m_{(x,t)} = \frac{D_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}$$

In fact, the significance of parameters x and t varies according to the data used for the calculation. The Lexis diagram can be a great help in identifying, according to these variations, the elements that are

[8] It should be noted that this is an approximation based on the hypothesis of a linear change in the size of the population:

$$m_{(x,t)} = -\log\left(1 - \frac{D_{(x,t)}}{P_{(x,t)}}\right) \cong \frac{D_{(x,t)}}{(P_{(x,t)} + P_{(x,t+1)})/2}.$$

We will see in Chapter 9 that the evaluation of mortality with such a rate supposes the absence of any competing phenomena (in this case migration) or independence between the competing phenomena (in this case migration) and the phenomenon studied (in this case mortality). In addition, it supposes that the population is homogeneous (Chapters 9 and 14).

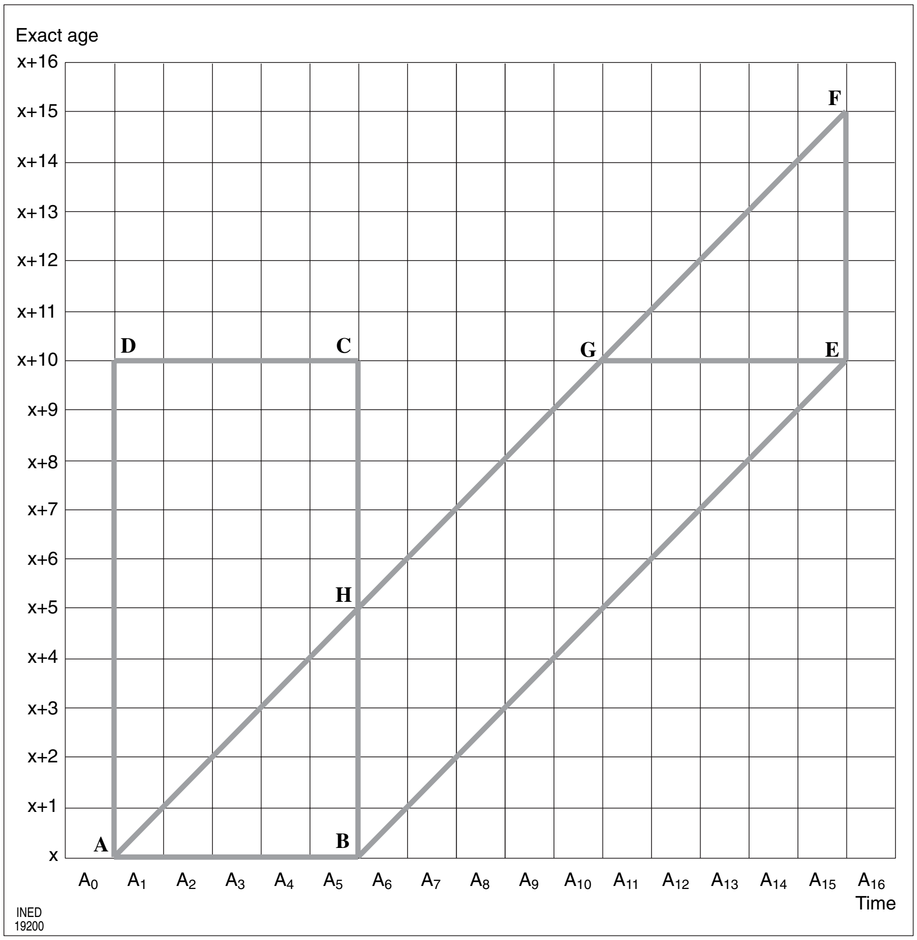


FIGURE 6–7 Representation of events on the Lexis diagram, according to different pluri-annual classification modes.

really needed for calculating rates in the different possible situations (Figure 6–8).

If the deaths are classified by year of age and calendar year (square *ABCD*), deaths having occurred during the year considered t (between January 1 t and $t + 1$) in the age class (x, $x + 1$) will be related to the mean population of age (x, $x + 1$) during the year. It is generally possible to estimate this average by the half sum of the populations observed at the beginning and at the end of the year. A first, most common, rate at age x is thus obtained, which we will call type 1 rate:

$$m'_{(x,t)} = \frac{D^t_{(x,x+1)}}{\left(P^t_{(x,x+1)} + P^{t+1}_{(x,x+1)}\right)/2}$$

These are rates by year of age and calendar year that have been used to construct the second part (1899–1995) of the complete matrix (1806–1995) a few elements of which are given as examples in the beginning of Chapter 5 (Table 5–1). Because of a lack of precise data, the first part (1806–1898) of this matrix has been built indirectly based on an estimate of *probabilities of dying* by age which will be dealt with further on (Chapter 11).

If the deaths are classified by year of birth and calendar year (parallelogram *EFGH* in Figure 6–8), the deaths having occurred during the year considered t in cohort C ($t - x$) reaching the exact age x during this year will be divided by half the sum of populations of age ($x - 1$, x) on January 1 t_3 and of age (x, $x + 1$) on January first t_4. A second rate at age x (type 2) is obtained, which is sometimes referred to as a prospective rate:

$$m''_{(x,t)} = \frac{D^{C(t-x)}_{(x-1,x+1)}}{\left(P^t_{(x-1,x)} + P^{t+1}_{(x,x+1)}\right)/2}$$

When deaths are grouped by age in complete years and year of birth or cohort (parallelogram *IJKL* in Figure 6–8), it is necessary to proceed in the same way as above, to divide the deaths occurred in this birth cohort between the exact ages of x and $x + 1$ by half

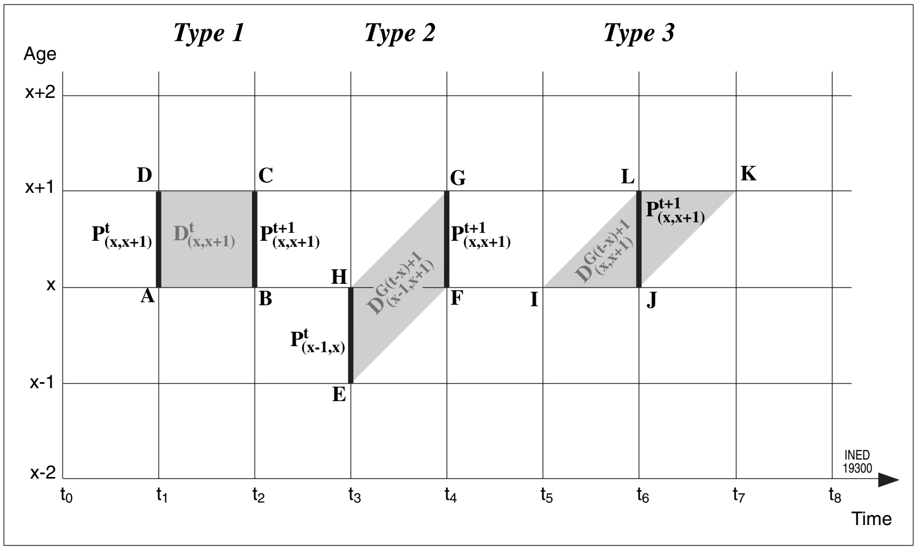


FIGURE 6–8 Representation of the elements required for computing a death rate by age according to the classification mode of the deaths.

the sum of the numbers of individuals of the cohort present at the time of their birthdays x and $x + 1$, but these data do not feature in any of the usual statistics. Because the deaths concerned are distributed more or less equally between the triangle *IJL*, belonging to the year t and the triangle *JKL* belonging to the year $t + 1$, we will take the denominator of the population at complete age $(x, x + 1)$ on January 1 of the year $t + 1$, which is generally a good estimator of the mid-year population. A third rate at age x (type 3) is obtained, which can truly be defined only if one has a double classification of deaths by age and by year of birth:

$$m'''_{(x,t)} = \frac{D^{C(t-x)}_{(x,x+1)}}{P^{t+1}_{(x,x+1)}}$$

2. Differences Among the Three Rate Types

The *double classification* of the deaths of each calendar year by age and by year of birth, which in France for example has been in existence since 1907, will enable us to illustrate our point with an example with figures, demonstrating that in certain cases such subtleties in computations are far from superfluous.

Table 6–1 provides an extract of these French mortality statistics for the years 1993, 1994, and 1995, relating to a number of age groups, accompanied by population estimates at the same ages on January 1st of 1994 and 1995. Figure 6–9 places on the Lexis diagram the elements of this table that are appropriate for computing death rates by age for the year 1994, according to the three usual approaches mentioned previously.

For instance, at the *completed age* of 74, the death rate (type 1, Figure 6–9) is equal to:

$$m'_{74} = \frac{2{,}889 + 4{,}275}{(121{,}312 + 202{,}140)/2} = 0.04430$$

On the other hand, at the age of 74 *reached during the year*, the rate (type 2, Figure 6–8) is equal to:

$$m''_{74} = \frac{4{,}142 + 4{,}275}{(210{,}506 + 202{,}140)/2} = 0.04080$$

But this rate, in fact, concerns an age that is half a year less than the type 1 rate. To make a valid comparison with the previous rate, the average of the type 2 rates at 74 and 75 years has to be made, i.e., from Table 6–1:

$$\overline{m''_{74}} = \frac{m''_{74} + m''_{75}}{2} = \frac{0.04080 + 0.04348}{2} = 0.04214$$

Lastly, the rate computed between the exact ages of 74 and 75 in the cohort celebrating its 74th birthday in 1994 is equal to:

$$m'''_{74}(C_{1920}) = \frac{4{,}275 + 4{,}084}{202{,}140} = 0.04135$$

This result must be compared with the rate of the same type obtained in the previous cohort celebrating its 74th birthday in 1993 and which also crosses the square of the Lexis diagram concerned by the type 1 rate:

$$m'''_{74}(C_{1919}) = \frac{2{,}318 + 2{,}889}{121{,}312} = 0.04292$$

TABLE 6–1 Extract of mortality statistics by age for 1994 and corresponding populations[1] on January 1 of 1994 and 1995

Age (x)	Deaths						Populations		Rates (for 1000) in 1994					
	1993		1994		1995				Type 1	Type 2		Type 3		
	C	C-1	C	C-1	C	C-1	1994	1995	(Lexis square)	x	$(x + x + 1)/2$	1993–94	1994–95	Average
65	2,788	2,598	2,575	2,662	2,845	2,637	259,605	258,546	20.21			20.99	20.16	20.58
66	2,818	2,828	2,850	2,719	2,866	2,752	249,347	254,288	22.12	21.45	22.14	22.21	22.03	22.12
67	2,948	2,908	2,911	2,914	2,981	2,899	244,661	243,768	23.85	22.83	23.69	23.96	23.83	23.90
68	3,206	3,043	3,020	3,080	3,114	3,136	237,909	238,791	25.59	24.55	25.66	26.42	25.78	26.10
69	3,300	3,081	3,205	3,184	3,279	3,197	224,448	231,630	28.02	26.77	27.90	28.89	27.64	28.26
70	3,600	3,321	3,240	3,208	3,457	3,291	219,548	218,205	29.46	29.02	30.02	31.01	29.93	30.47
71	3,619	3,557	3,498	3,344	3,425	3,349	213,348	212,922	32.10	31.01	32.33	32.64	32.16	32.40
72	3,801	3,961	3,719	3,648	3,613	3,596	213,184	206,391	35.12	33.65	34.79	34.94	35.44	35.19
73	4,183	2,858	3,878	4,142	3,925	3,954	210,506	205,693	38.54	35.93	38.36	39.55	38.08	38.81
74	2,318	2,386	4,275	2,889	4,280	4,084	121,312	202,140	44.30	40.80	42.14	42.92	41.35	42.14
75	2,393	2,118	2,275	2,417	4,646	3,095	100,716	116,196	43.26	43.48	46.44	47.76	46.22	46.99
76	2,163	2,058	2,438	2,238	2,376	2,456	84,490	95,863	51.85	49.39	51.28	52.09	51.05	51.57
77	2,259	2,222	2,137	2,005	2,667	2,235	74,641	80,091	53.54	53.17	55.86	57.13	54.59	55.86
78	3,326	4,030	2,242	2,326	2,376	2,238	87,657	70,442	57.79	58.55	62.13	64.48	63.60	64.04
79	4,294	4,305	3,249	4,108	2,345	2,322	120,679	82,035	72.59	65.71	68.37	69.62	67.91	68.77
80	4,560	4,492	4,171	4,325	3,444	4,321	113,402	112,397	75.25	71.04	75.53	78.35	75.55	76.95
81	4,738	4,256	4,400	4,465	4,308	4,313	104,048	104,663	84.95	80.02	84.56	88.45	83.25	85.85
82	4,397	4,360	4,410	4,161	4,554	4,477	87,060	95,172	94.07	89.10	94.55	98.30	93.38	95.84
83	4,522	4,271	4,131	4,300	4,472	4,151	80,816	78,770	105.66	100.01	105.29	109.16	105.14	107.15
84	4,268	4,163	4,168	4,072	4,178	4,175	69,063	72,348	116.54	110.57	117.14	120.76	115.32	118.04
85	4,134	3,815	3,974	3,898	4,132	3,996	60,008	61,017	130.09	123.71		133.85	130.62	132.23

[1] The population numbers are estimates by INSEE, on each January 1, based on the numbers obtained at the censuses, taking into account births, deaths and migration balance.

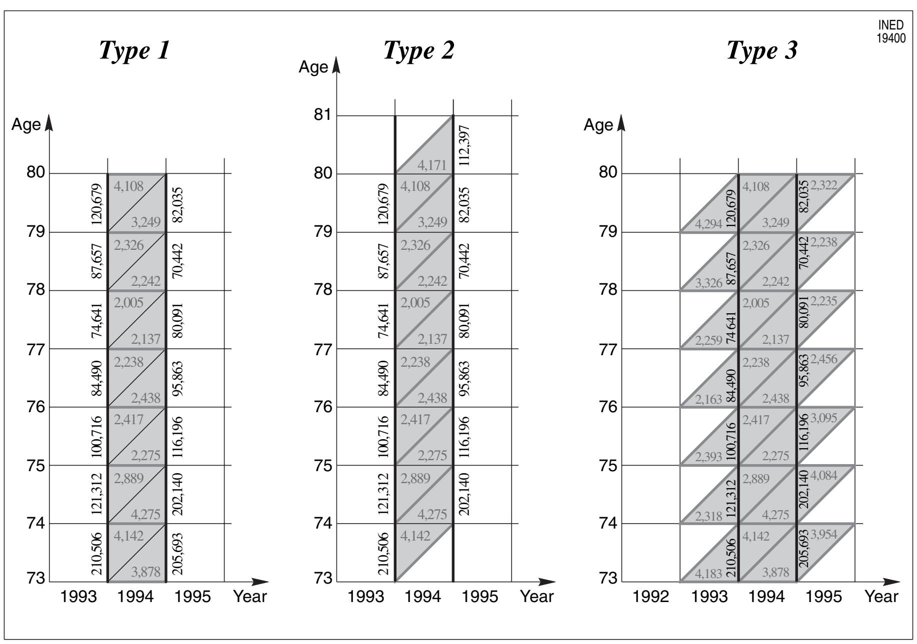


FIGURE 6–9 Elements for computing different types of death rates between the ages of 65 and 85 years in France (men).

For comparing with the previous rates one must, here again, compute the average of these two results to focus the measure on 1994. Thus the rate of 0.04214 is obtained, which is very close to the result obtained with the type 2 rate.

Thus, we now have three death rates established for the same calendar year at the same age but opposing a higher value obtained with the type 1 rate (44/1000) to the two other lower values provided by types 2 and 3, which are much closer to one another (42/1000).

In fact, the differences observed here correspond to an exceptional situation. Indeed most of the time, there is practically no difference between the three types of rates. This supposes, however, that the sizes of the neighboring cohorts that come into the calculation are distributed in a similar manner according to age. In most cases this is fine, and the method used for calculating the rate is of little importance. In the example taken here, the sizes of successive cohorts varies suddenly. In fact, this concerns cohorts born in 1919 and 1920, just after World War I, which, as we know created a sharp rise in births. Therefore, on January 1, 1994, the number of men suddenly declined from 210,506 at age 73 to 121,312 at age 74 (Figure 6–9). The 1919 cohort represents those in the last year in the deficit of births, whereas those born in 1920 belong to the first full cohort after the War. It is this fluctuation that, by suddenly modifying the distribution by age of individuals within cohorts, is at the origin of the discordance observed here between type 1 rates and the others (Figure 6–10).

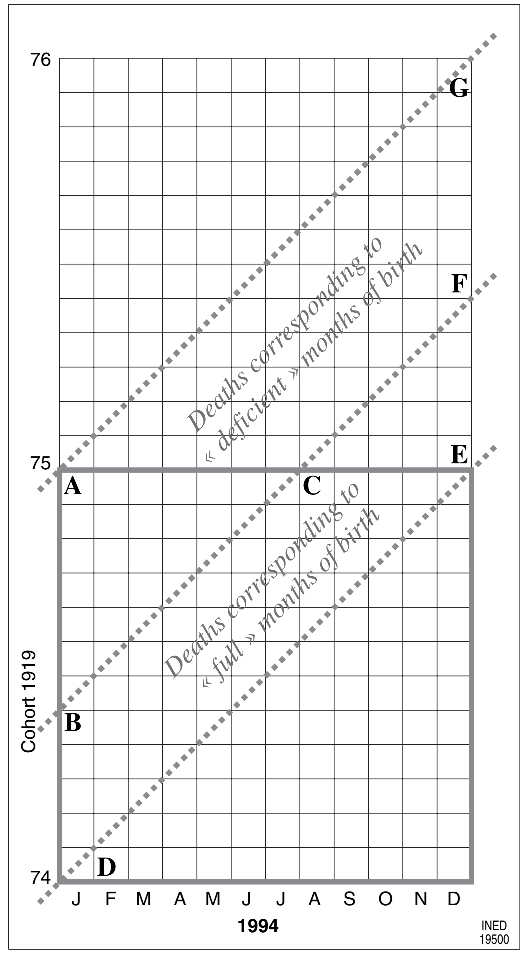


FIGURE 6–10 Representation on the Lexis diagram of two portions of the cohort of 1919, from the months with low birth rates and the months with high birth rates, during the year of 1994.

The sudden change in the number of births that causes the difference in the sizes of the cohorts did not, in fact, happen exactly at the passing of 1919 to 1920 but during 1919, precisely at the end of summer, 9 months after the Armistice and demobilization. Because of this, among the deaths in 1994 that occurred in the 1919 cohort at 74 years of age, some of the deaths correspond to the 8 or 9 first months with a low number of births and others to the 3 or 4 last months with a high number. However, for the first part of this cohort there is less time between January 1, 1994 and the 75th birthday (triangle *ABC* in Figure 6–10) whereas for the second part there is much more time in the same interval (trapezium *BCDE*). Therefore, the number of deaths at that age in this cohort in this year depends more on the people born during the months with a high number of births than on those born in the months with low numbers; divided, as if nothing had happened, by the population size, without any distinction in terms of month of birth, this will overestimate the death rate at that age. Hence the abnormally high level of the death rate at 74 in 1994. Inversely, the same year, at 75, the rate is underestimated because the contribution of the months with a high number of births of the cohort of 1919 to mortality is very short (triangle *CEF* in Figure 6–10) whereas that of the months with low numbers is much longer (trapezium *ACFG*).

The same phenomenon occurs in the opposite direction with the passage of the cohorts born in 1914 (the last full year before the war) and 1915 (first year with a deficit in births), which reached 80 and 79 years of age, respectively, in 1994. This time, it is within the cohort of 1915 that monthly births suddenly declined in the spring. The same causes producing the same effects; therefore, in 1994, one obtains a death rate of type 1, which is abnormally low at 78 and abnormally high at 79.

Type 2 and 3 rates, calculated strictly within the same cohort, are, on the contrary, practically unaffected by this statistical artifact, whether it be before or

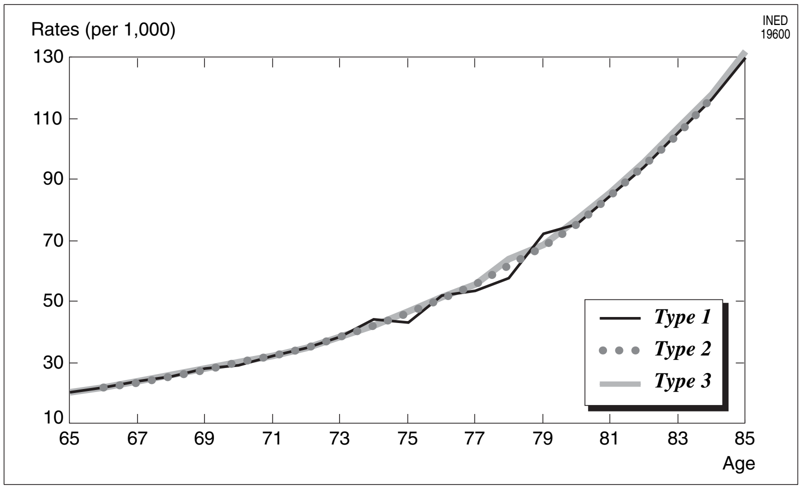


FIGURE 6–11 Death rate by age between 65 and 85 years in France, in 1994, according to method of calculation.

after the average of the two contiguous rates has been obtained to center the result on the same complete age and the same calendar year as the type 1 rate.

In any case, this type of problem only arises when there are strong variations in sizes between contiguous birth cohorts. Figure 6–11 clearly shows that apart from the ages reached by cohorts affected by the war, there is practically no difference between the three types of rates by age.

To account for this problem due to the computation of rates in the Lexis squares, Gérard Calot proposed a weighting method according to the monthly distribution of births forming each cohort. All that is needed is to apply to the deaths in each Lexis triangle a correction factor calculated according to this monthly distribution (Calot, 1984).

3. Rates by Age Groups

In Chapter 4, we discussed the calculation of rates by age groups following the general formula:

$$m_{((x,x+a),t)} = \frac{D_{((x,x+a),t)}}{\left(P_{((x,x+a),t)} + P_{((x,x+a),t+1)}\right)/2}$$

As with the rates by individual year of age, different computation methods are possible according to the way the available data is presented. Figure 6–12 illustrates the data that may be used for computing death rates by age groups according to possible classification modes.

If, for example, deaths are classified by calendar years and quinquennial age groups (rectangle *ABCD* in Figure 6–12), we will divide deaths having occurred between the exact ages x and $x + 5$ by the average of the populations of the same ages on each January 1:

$$m'_{((x,x+5),t)} = \frac{D^{t}_{(x,x+5)}}{\left(P^{t}_{(x,x+5)} + P^{(t+1)}_{(x,x+5)}\right)/2}$$

In the similar case in which the deaths of the calendar year are grouped according to year of birth (parallelogram *EFGH* in Figure 6–12), one obtains:

$$m''_{((x,x+5),t)} = \frac{D^{(t)}_{G(t-(x+5),t-x)}}{\left(P^{t}_{(x,x+5)} + P^{(t+1)}_{(x+1,x+6)}\right)/2}$$

Note that these two types of rates according to age groups are annual rates, because, although they group together the deaths of five age brackets, they relate these deaths to equivalent groups of population size by age. This is true whatever the number of age groups. These rates are directly comparable to the rate by individual year of age mentioned here, the first being obviously similar in nature to type 1 and the second to type 2. To do so, place the rate by age group in relation with the rate by age corresponding to the central age of the age group. Thus, the first quinquennial rate mentioned here will be related to a rate at complete age $x + 2$, or on average at exact age $x + 2.5$. The second will be compared with a rate at exact age $x + 3$. It is, of course, impossible to compute a similar type 3 rate.

On the other hand, if the deaths are classified by birth cohort and group of calendar years, it is possible to obtain in cohorts a type 2 rate and a type 3 rate, of pluri-annual dimension, which will not be directly comparable with any type 1 rate.

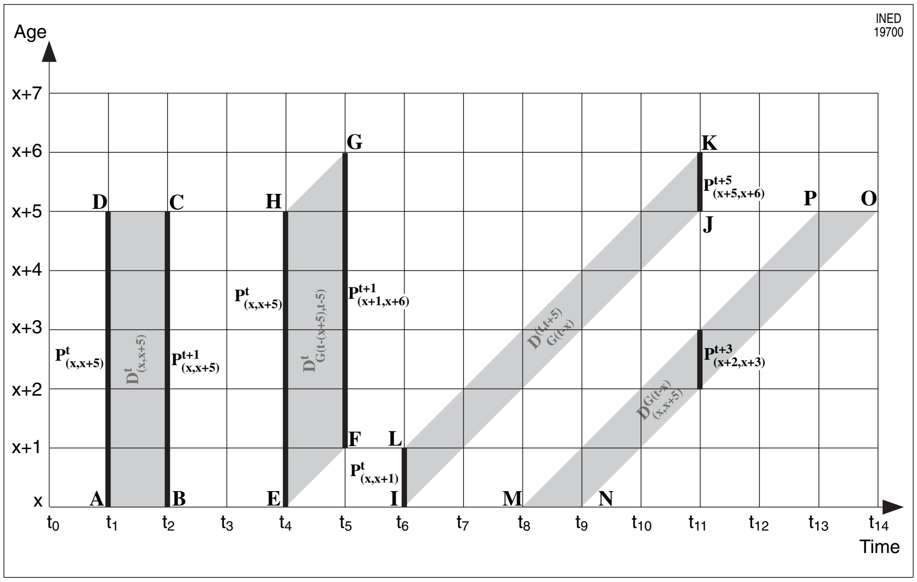


FIGURE 6–12 Representation of the elements needed for the calculation of death rates by age according to how deaths are classified.

If, for example, deaths are classified by birth cohort and groups of five calendar years (parallelogram *IJKL* in Figure 6–12), a type 2 rate can be calculated as follows:

$$m''_{((x,x+5),(t,t+5))} = \frac{D^{(t,t+5)}_{C(t-x)}}{\left(P^{t}_{(x,x+1)} + P^{t+5}_{(x+5,x+6)}\right)/2}$$

In the same way, if the deaths are classified by cohort and quinquennial age intervals (parallelogram *MNOP* in Figure 6–12), one will take as an estimator of the average size of the actual population at the x^{th} and $(x + 5)^{\text{th}}$ birthdays of the cohort born in the year $t - x$, the population aged $(x + 2, x + 3)$ on January 1 of the central year t_{+3}. One then obtains:

$$m'''_{((x,x+5),(t+5))} = \frac{D^{C(t-x)}_{(x,x+5)}}{P^{t+3}_{(x+2,x+3)}}$$

In both of these cases, the rates are no longer on an annual dimension scale since the deaths of five age groups are related to the mean size of a single birth cohort. One may of course obtain this annual dimension by dividing the result by five (or, more generally, by the number of age groups concerned). It is only on such a condition that one may compare these results to those of rates per individual year of age by referring to the central age of the group.

Finally, note that in most cases age groups are identified by ages in complete years rather than by exact ages. Therefore, one notes in all these formulas, x instead of $(x, x + 1)$, $(x, x + 1)$ instead of $(x, x + 2)$, . . ., $(x, x + 4)$, instead of $(x, x + 5)$, and so forth., thus replacing an interval between two extremes of exact ages by a series of ages in complete years.

4. Pluri-annual or Infra-annual Rates

The reader will easily extrapolate to the cases where deaths are classified by age groups as well as by groups of birth cohorts or calendar years. Moreover, we have only used as an example here the most standard cases. In fact, the data available are much more varied, but the principle remains the same: The simple fact of correctly entering on a Lexis diagram a few elements of the death statistics used will certainly facilitate the choice of population to be entered in the denominator of the rates.

5. A Special Case: The Infant Mortality Rate

In addition to the three possible types of rates mentioned so far, mortality at under 1 year of age is very often measured by an *ad hoc* indicator, called *the infant mortality rate*, but which is very different from the rates we have just defined. Instead of relating deaths of less than 1 year to an average population of less than 1 year as indicated above for any age x, these deaths are related to births:

$$mi_{(t)} = \frac{D_{(0,t)}}{B_t}$$

We shall see in Chapter 8 that this "rate," used extremely often, is similar to a *probability* such as we will define later on. Given the title of the chapter, we could not end without evoking its existence.

References

Beker Karl, 1874. *Zur Berechnung von schterbetafeln an die Bevölkerungsstatistik zu stellende anforderungen*.—Berlin, 76 p.

Calot Gérard, 1984. *La mesure des taux en démographie. Âge en années révolues ou âge atteint dans l'année; incidence du choix de la définition; application à la fécondité générale (France, 1946–1981).* Paris, INED, PUF, 322 p. (Travaux et Documents, Cahier 104).

Caselli Graziella and Lombardo Enzo, 1990. Graphiques et analyse démographique. *Population*, vol. 45(2), p. 399–413.

Keyfitz Nathan, 1977a. *Introduction to the mathematics of population with revisions*. Reading (MA), Addison-Wesley Company, 481 p.

Keyfitz Nathan, 1977b. *Applied mathematical demography*. New York, John Wiley and Sons, 388 p. (Reedited in 1985 by Springer-Verlag, New York, 442 p).

Knapp Georg, 1868. *Die Ermittlung der Sterblichkeits.* Leipzig.

Knapp Georg, 1869. *Die Sterblichkeit in Sachsen.* Leipzig, Dunker & Humbolt, 216 p.

Lexis Wilhelm, 1875. *Einleitung in die Theorie der Bevölkerungsstatistik*. Strasbourg, Trübner, 137 p.

Pressat Roland, 1961. *L'analyse démographique*, 1st ed. Paris, PUF, xii + 402 p.

Vanderschrick Christophe, 1992. Le diagramme de Lexis revisité. *Population*, vol. 47(5), p. 1241–1262.

Verweij Abraham, 1874. *Die warrnemingen der bevölkerungsstatistik.* Deventer, Rustering en Vermandel (Academisch proefschrift, ter verkrijging van den graad van doctor in de wis—en natuurkunde, aan de Hoogschool te Utrecht, te verdedigen op vrijdag den18 december 1874, Des namiddags te 3 uren).

Verwey Abraham, 1875. Principles of vital statistics, *Journal of the Statistical Society*, vol. XXXVIII, p. 487–513.

C H A P T E R

7

Frequency Surfaces and Isofrequency Lines

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

To examine the evolution of any demographic phenomenon—the structure of a population, or the intensity of the components of its dynamics, such as fertility, mortality, or migration—we need to take at least three dimensions into account: time, age, and the intensity of the phenomenon. This is the case, for example, with the matrix of mortality rates by age and calendar year shown in Table 5–1 of Chapter 5: The matrix provides an indicator of the force of mortality for about 100 age-years and nearly 200 years of observation. This would also be true of a matrix that assembled fertility rates or population size by age for a period of one or two centuries. Unless we could develop three-dimensional graphic media,[1] such matrices would pose the classic problem of how to display a three-dimensional phenomenon on a two-dimensional sheet of paper.

In fact, there are only two solutions: We can try to depict the three dimensions in identical fashion, using perspective-creating devices of the kind first used by Renaissance painters,[2] or we project one of the three dimensions on the plane formed by the two others. Naturally, the second solution is more common. In demography, however, it is used in only two of the three possible ways: Intensity is always chosen as one of the two axes of the projection plane. Indeed, intensity curves (for population size, fertility rates, mortality rates, etc.) are very often shown as a function of age, or of time, by projecting on this Cartesian plane the situation characterizing a specific period or age, respectively. The third dimension is materialized simply by juxtaposing several curves on this projection plane. These two ultraclassic types of projection are illustrated in Figures 5–1 and 5–2 of Chapter 5. In the far less common projections of the third kind, intensity is projected on the plane formed by age and time. Yet it is this third solution, well known to geographers, that undoubtedly offers the best substitute for the impossible "perfect" display of the three dimensions.

[1] A few decades ago, for example, Institut National d'Etudes Démographiques exhibited a three-dimensional volume structure showing the change in the French age pyramid over a century.

[2] Notably Mantegna in the Sala degli Sposi (Bridal Chamber) in Mantua.

I. ATTEMPTS AT THREE-DIMENSIONAL REPRESENTATION

The problem is identical to the one that geographers have faced for centuries in mapping relief. Unless they can produce actual *relief maps*—as cartography institutes are now doing ever more routinely—geographers need to represent, on a plane, a nonplane surface each of whose points is situated in a space defined by three dimensions: longitude, latitude, and altitude. The earliest attempts typically consisted in using *trompe-l'œil* images of hills and mountains to provide an ornamental identification of places mapped in the two dimensions regarded as the most important: longitude and latitude. This practice of pictorial representation later led to stylized graphic devices such as cross-hatching or shading. In fact, however, the standards now required for these devices depend on a more modern invention: *contour lines*. The first demographers made their own attempts to put into perspective the three main dimensions of the phenomena they were studying (see Caselli and Lombardo, 1990).

For example, Luigi Perozzo (1880) used an earlier diagram by Berg (1860) as a model for a perspective representation of the change in the Swedish population that was exhibited at the 1875 Geographic Congress and the 1878 Paris Exhibition. The diagram comprised a line plotting the number of births over more than a century (1750–1875). Above that line, Perozzo placed a perspective view of the age pyramids observed at each census, at 5-year intervals (Figure 7–1).

A few years later, the same author (Perozzo, 1883) published a new chart with a twin perspective view of the probabilities of a man aged *x* marrying a woman aged *y* and the probabilities of a woman aged *x* marrying a man aged *y*. The probabilities were shown as a function of the age of each spouse (Figure 7–2). Each view featured a splendid "witch's hat" illustrating the high concentration of marriages in a small number of spouse–age combinations.

With modern computer technology, this formerly tedious operation has become much easier. It is, for example, commonly used in architecture. However, it is not widely used in demography. The advantage of computer-aided perspective design is the possibility of switching perspective angles instantly, on screen. One can thus take decisions on the final appearance of a construction by examining the various angles from which it can be viewed once it has been built.

Similar attempts have been made by Thomas Pullum (1980), for example, concerning fertility in the United States (Figure 7–3). By presenting the 1920 to 1970 matrix of age-specific fertility rates from different angles, he shows that the visual impression varies considerably with the angle chosen.

This finding effectively pinpoints the major drawback of the stratagem as applied to demography. Unlike the architect, the demographer is not attempting to choose the most handsome structure or variation of age-specific fertility or mortality, but to describe observed reality as faithfully as possible. It may be

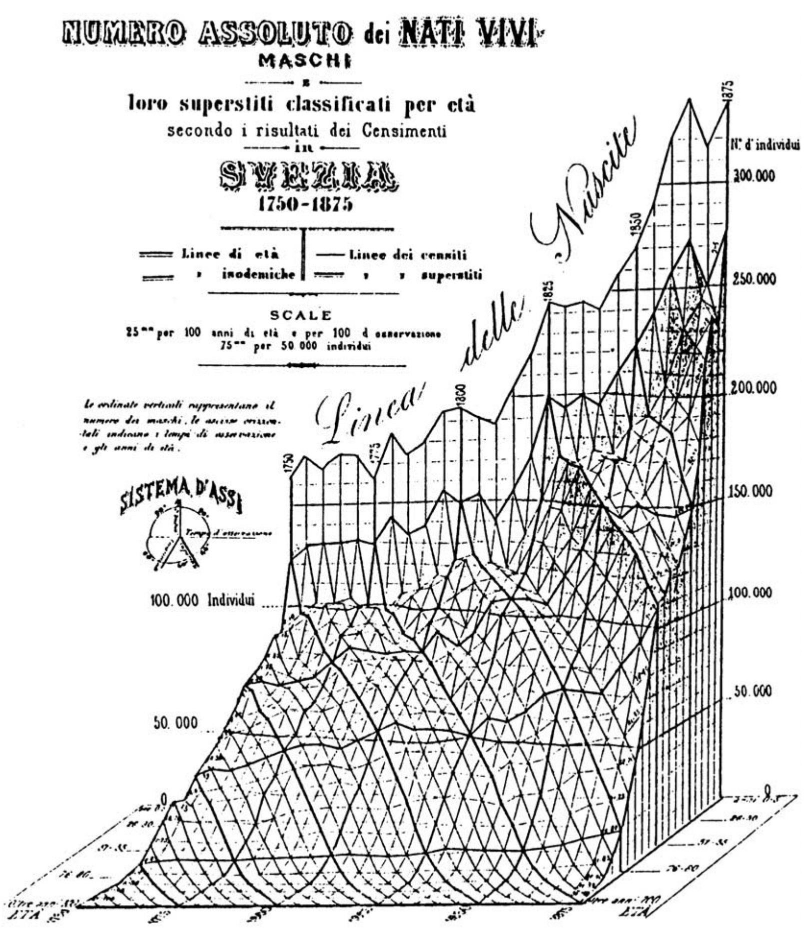


FIGURE 7–1 Perspective view of change in Swedish age pyramid compared with birth curve, by Luigi Perozzo (1880).

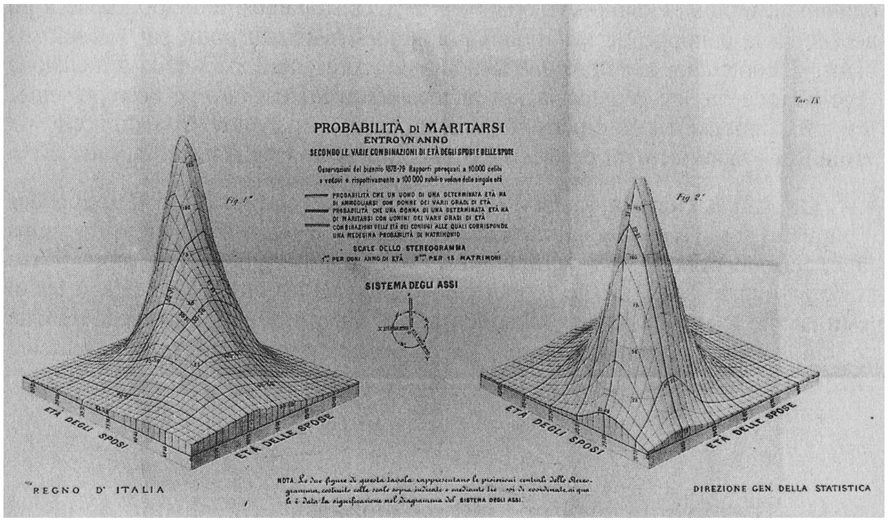


FIGURE 7–2 Representation by Luigi Perozzo (1883) of the probability, for a man aged x, of marrying a woman aged y (left-hand chart) and, for a woman aged y, of marrying a man aged x (right-hand chart).

useful, for this purpose, to systematically examine on a computer screen all the possible facets of a perspective view, but at the time of publication the demographer must choose which chart to print, and there is no objective reason to pick one angle instead of another.

Figure 7–4 shows the large 1806 to 1995 matrix of French age-specific male mortality, under what we regard as the most expressive angle, in terms of (1) the radical fall in mortality over the long term and (2) the severe disruptions caused by the Napoleonic wars, the Franco-Prussian War of 1870, and the two world wars. However, our deliberate choice to highlight these major historical events promotes the legibility of some aspects to the detriment of others that are, perhaps, no less important. For example, we can very easily track the change in infant mortality, but it is much harder to glimpse mortality at age 60. Likewise, it is fairly easy to distinguish the overall profile of age-specific mortality and its considerable change over two centuries, but we have trouble making out the details of this transformation.

From this dual standpoint, perspective adds nothing to the classic projection on a Cartesian plane in which age or time is plotted against intensity. These two classic approaches—which require a choice between age and time to define the projection plane—are, however, relatively powerless in themselves to depict a fundamental demographic reality: the life of generations. In fact, from this third angle, we need to define age and calendar as the two axes of the projection plane—as Wilhelm Lexis did when designing his diagram (see Chapter 6). Age and calendar thus become the two major references for locating an event in time, just as longitude and latitude are the geographer's two preferred references for locating a spot on the Earth's surface. What we now need to do is to follow the geographers' example and adopt their most rigorous technique for mapping relief: contour lines. This will enable us to visualize the intensity of demographic phenomena on the projection plane.

II. CONTOUR LINES

It is no coincidence that the first demographer to use contour lines, Pierre Delaporte, is also the first to have attempted to plot the mortality of French birth cohorts. The problem to be solved here is the way in which the data are compiled. We know the number of inhabitants in a given year, or the mortality or fertility rate at a given age, or, conversely, the values of these indicators in a given calendar year for a given age. By contrast, we do not know the age or the instant to which the rates 10, 15, 20 . . . per thousand apply, rates that could be used to determine the contour lines. To plot contour lines showing the century-long change in age-specific mortality, we must therefore begin by constructing what Delaporte calls *mortality surfaces* from the available values for age-specific and period mortality rates.

For this construction task, Pierre Delaporte actually made the debatable choice of smoothing as much as possible the variations in age-specific mortality and in

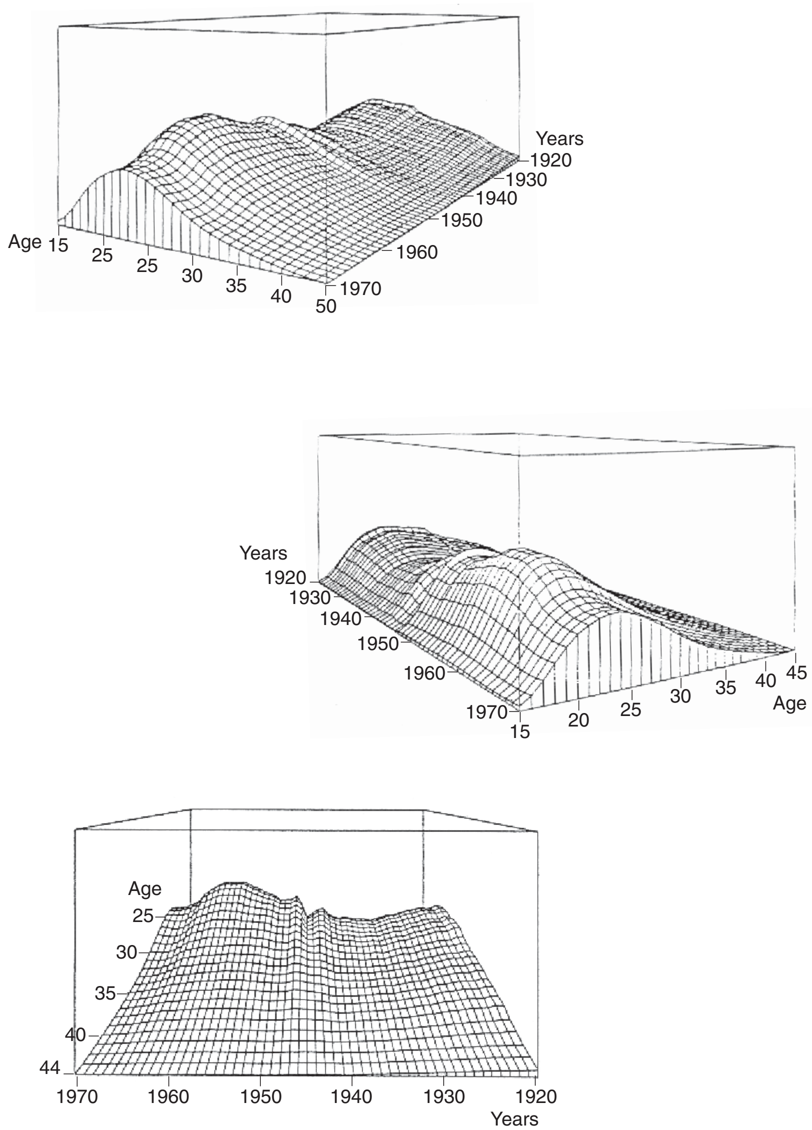


FIGURE 7–3 Alternative perspective views of changes in age-specific fertility of white U.S. females, 1920–1970, prepared by Thomas Pullum (1980).

mortality over time.[3] He therefore gave preference to rates for age groups covering several years, and interpolated between the central ages and the median dates of these rates. After examining different mathematical interpolation models, he opted for simple graphic interpolations. He started, of course, by interpolating the age-specific rates of each period, followed by the change over time of each age-specific rate. This enabled him to determine, for each period, the exact age at which the rate reached a particular, noteworthy

[3] Admittedly, in some cases, smoothing is entirely justified as a way to minimize the effect of observation errors. One can thus readily agree with Pierre Delaporte when he writes, "We have eliminated oscillations from the Demonferrand tables above age 50, for they seem solely due to attraction phenomena for round ages" (p.28). But we can hardly concur with his subsequent comments, "We have used slightly lower rates for the adjustments than the rates of the Bertillon tables between ages 0 and 40, as the deaths in these tables include those of the 1848 Revolution (France registered 973,000 deaths in 1848 compared with 836,000 in 1847), the Crimean War (cholera epidemic and siege of Sebastopol, which reportedly caused the death of 80,000 French), the Italian war, and the occupation of Algeria. France also experienced two cholera epidemics in 1849 and 1854. We classify cholera among the catastrophic events that could usefully be eliminated from the analysis. Having made its first appearance in Europe in 1832, it should not be included in the natural variation of the mortality surface." This deliberate removal of all "bumps" judged to be abnormal is both an illusion (resting on the misguided quest for a "natural" law of evolution) and a serious impediment to the goal pursued (i.e., the study of mortality by cohort). Such cross-sectional smoothing is likely to eliminate phenomena that would be interesting to observe longitudinally.

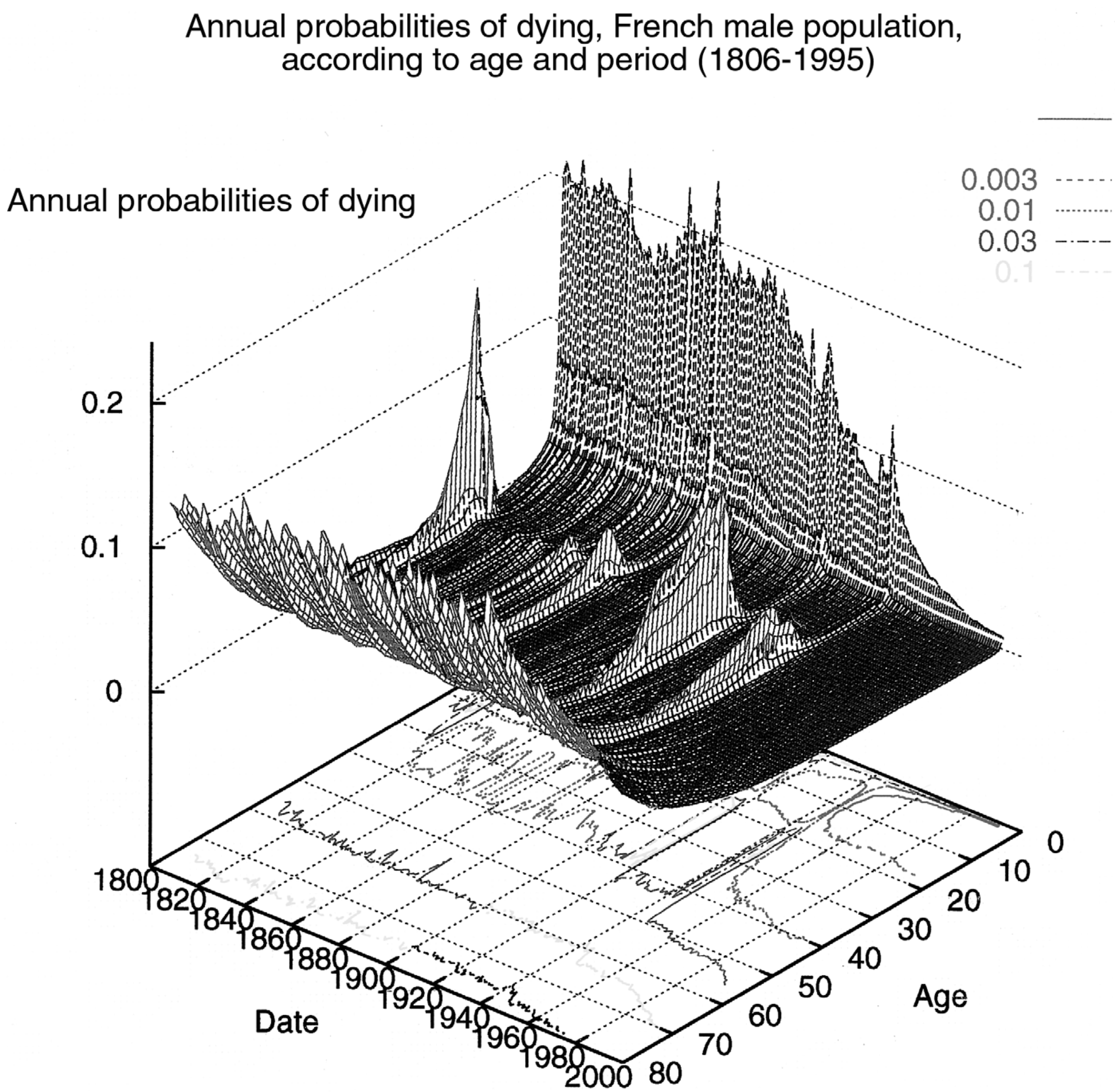


FIGURE 7–4 Perspective view of 1806–1995 matrix of French age-specific male mortality (chart from Nicolas Brouard [1997]). (Figure 7–4 also reproduced in color plate section.)

value, and for each age the exact date at which that same value was reached. From this, he deduced several points of the cross-section of the mortality surface by level. A final smoothing turned the broken lines thus obtained into *adjusted contour lines* (Figure 7–5).

Forty years later, at a time when cohort analysis enjoyed a revival of interest thanks to the rise of APC (age, period, cohort) models[4] (Hobcraft *et al.*, 1982; Caselli and Capocaccia, 1989), James Vaupel, Bradley Gambill, and Anatoli Yashin developed a software program at IIASA[5] that allowed the automatic projec-

[4] See John Wilmoth's Chapter 18.

[5] International Institute for Applied Systems Analysis, Laxenburg, Austria.

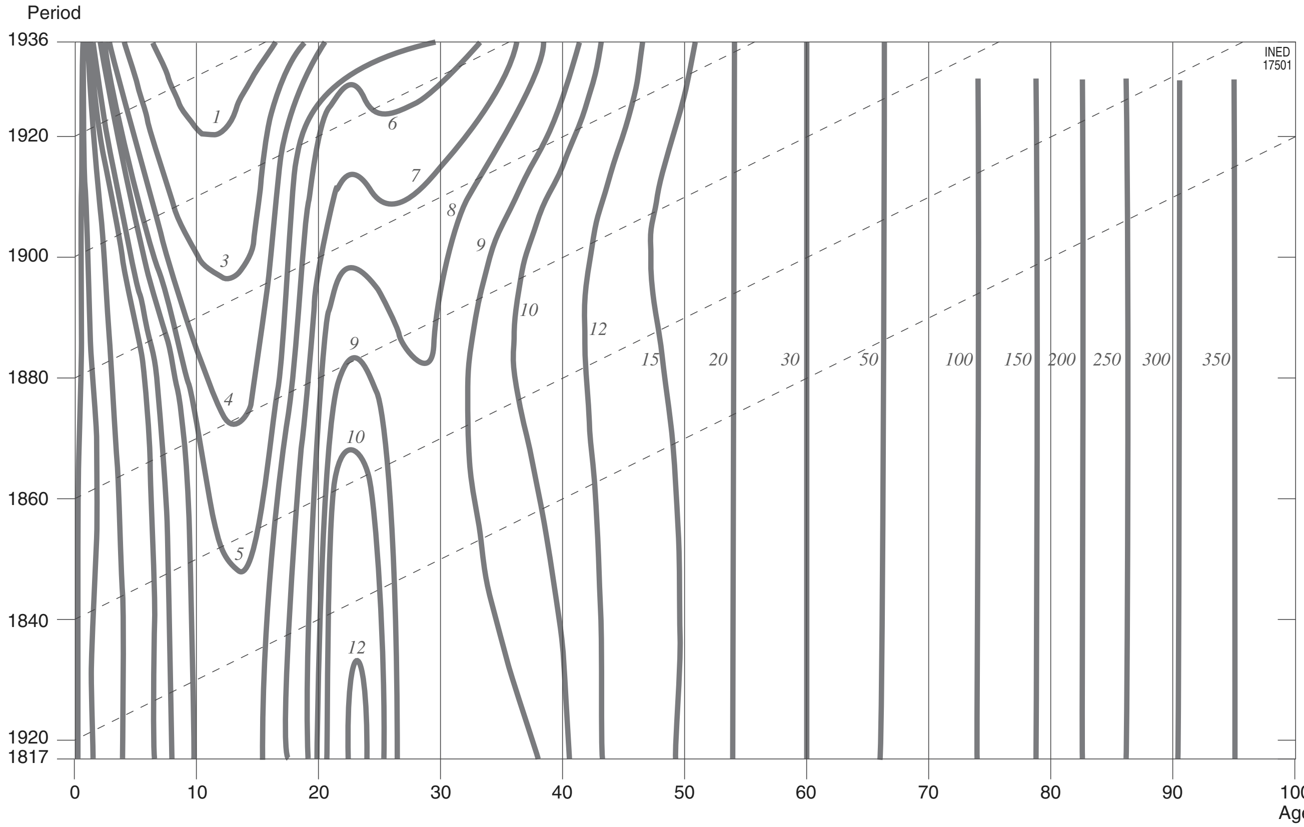


FIGURE 7–5 Contour-line projection, on age/time plane, of French male mortality surface, 1817–1936, by Pierre Delaporte (1941).

tion, on the age/period plane, of a contour-line display of the mortality surface generated by a matrix of age-specific and period rates. The principle is identical to Pierre Delaporte's, but the interpolations needed to locate the exact dates and ages for a given value of the mortality rate are automated (using a linear interpolation of the rates' logs).

Figure 7–6 illustrates the result obtained with the matrix of the probabilities of French males dying in the period of 1806 to 1995.[6] The surfaces are colored to show the mortality levels to which they correspond. This graphic display enables us to read the change over time in the age-specific risks of death and the age-related variations of those risks. Following the horizontal lines, we can see the mortality rate for a given age decrease with the color shift. Likewise, along the vertical lines, the color shifts give us the variation in age-specific mortality for a given observation year. The combination of these two variations causes the blue zone of the lowest mortality rates to gradually widen over time: in the 19th century, it was narrowly confined to adolescent ages; in the 20th century, it widened rapidly to embrace all ages between 1 and 40 by the 1990s. The chart also vividly illustrates the war-related crises of excess male mortality, from the Napoleonic wars to the Franco-Prussian war of 1870 and the two world wars. The crises are shown by vertical bars whose colors contrast with those displayed by the same ages in the surrounding years. We can also read this type of chart diagonally to follow the individual histories of selected birth cohorts. French male mortality is not deeply marked by generation effects.

By contrast, Italian male mortality offers a good example of generation effects. Figure 7–7 uses the same method to display the matrix of Italian male probabilities of dying for the period of 1887 to 1992. Some birth cohorts stand out by forming diagonal lines of distinctive colors. They notably include two groups of cohorts: children born during World War I, and individuals aged 10 to 15 years during that war.

In conclusion, to broaden the range of examples, we have used the contour-line method in Figure 7–8 to plot the change in the age distribution of the French male population over a century, from 1899 to 1996. The

[6] In Chapters 8 and 11, we elaborate on the definition of probabilities of dying and their differences with the rates used until now. In the present context, probabilities and rates can be used interchangeably.

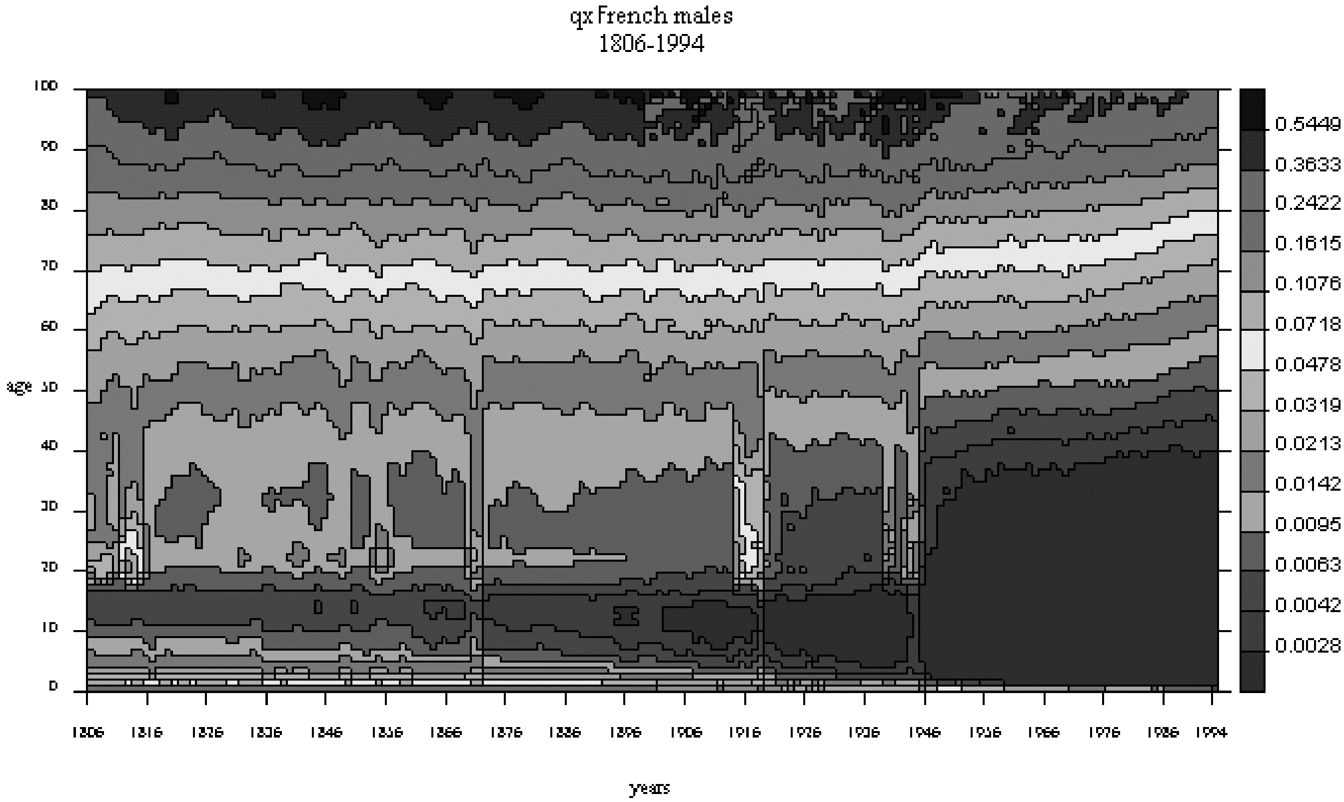


FIGURE 7–6 Contour-line projection, on age/time plane, of probabilities of dying of French males, 1806–1995 using method of James Vaupel *et al*. (1985). (Figure 7–6 also reproduced in color plate section.)

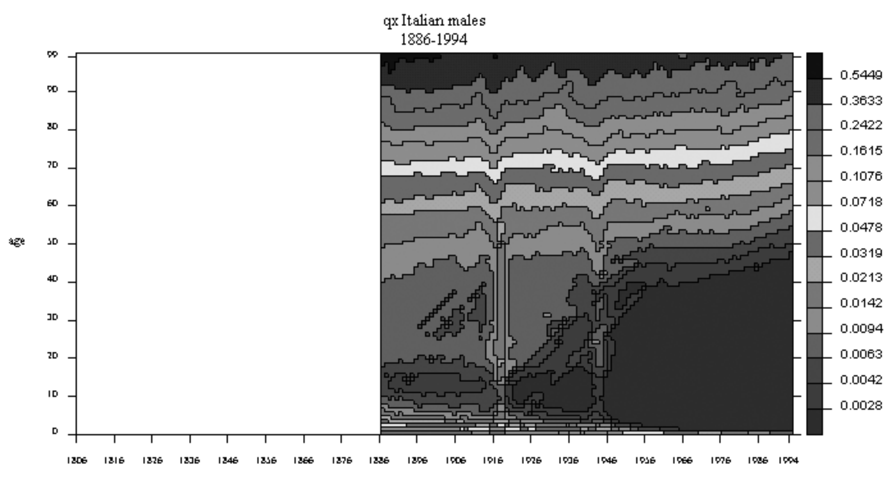


FIGURE 7–7 Contour-line projection, on age/time plane, of Italian male probabilities of dying, 1887–1992, using method of James Vaupel *et al*. (1985). (Figure 7–7 also reproduced in color plate section.)

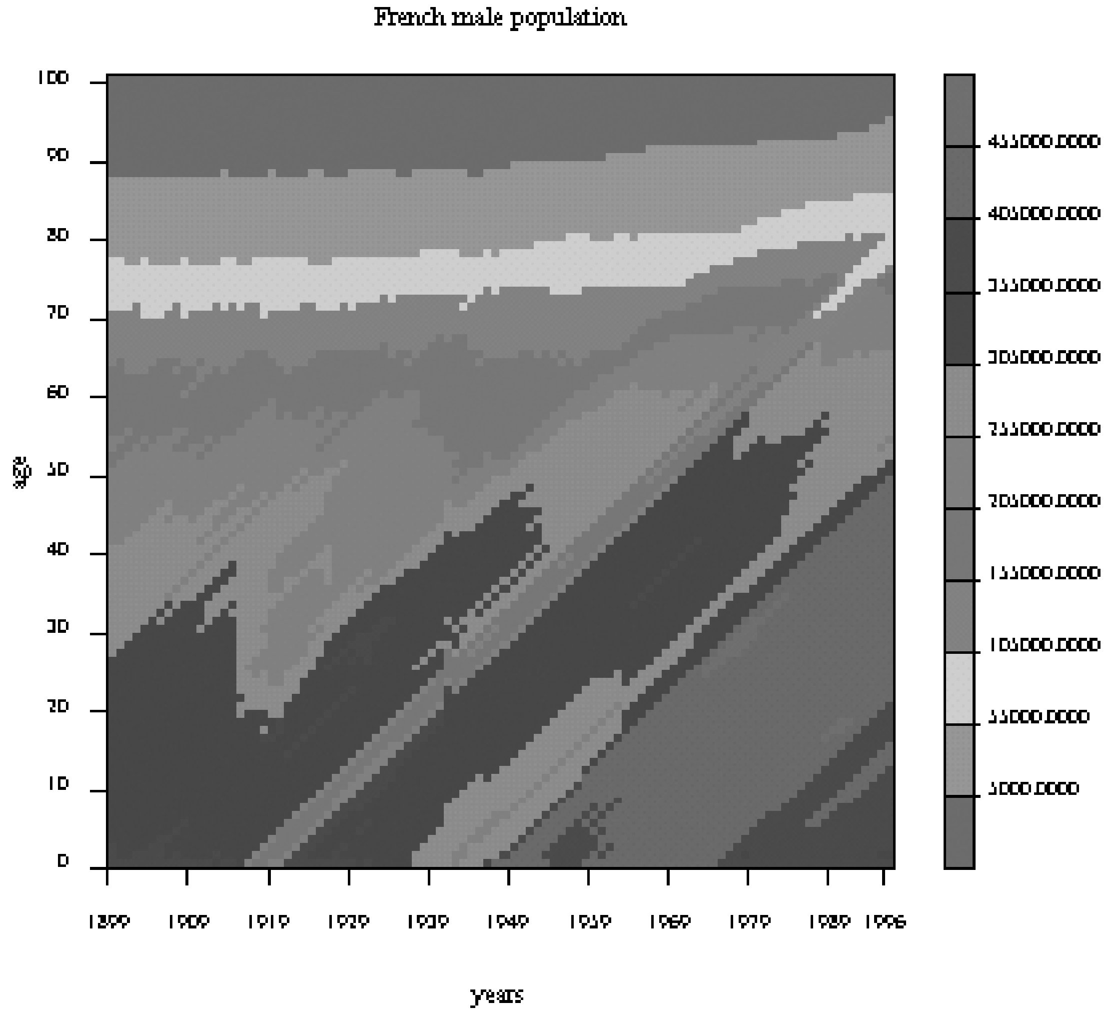


FIGURE 7–8 Contour-line projection, on age/time plane, of French male population size, 1899–1995, using method of James Vaupel *et al*. (1985). (Figure 7–8 also reproduced in color plate section.)

cohort effects are trivial, since it is obvious that the current population sizes depend directly, at any given moment, on the number of births at the origin of each generation.

References

Berg Gerard J. (van den), 1860. Befolknings statistik *in: Underdåniga berättelse för åren 1856–1860*, ny följd II, 3. Stockolm, Statistika Central-Byråns.

Brouard Nicolas, 1997. *Health expectancy: multi-state method, incidence based versus prevalence based*, Paris, INED, 8 p. (Tutorial on health expectancy calculations, Third Euro-REVES meeting, Fontevrault and Paris, April 1997).

Caselli Graziella and Capocaccia Riccardo, 1989. Age, period, cohort and early mortality: an analysis of adult mortality in Italy. *Population Studies*, vol. 43(1), p. 133–153.

Caselli Graziella and Lombardo Enzo, 1990. Graphiques et analyse démographique. Translated from Italian by Jacques Vallin. *Population*, vol. 45(2), p. 399–413.

Caselli Graziella, Vallin Jacques, Vaupel James, and Yashin Anatoli, 1986. *L'évolution de la structure par âge de la mortalité en Italie et en France depuis 1900: effets de période et effets de générations*. Paris, INED, 28 p. (Dossiers et recherches, no. 4).

Caselli Graziella, Vallin Jacques, Vaupel James, and Yashin Anatoli, 1987. Age-specific mortality trends in France and Italy since 1900: period and cohort effects, *European Journal of Population/Revue européenne de démographie*, no. 3, p. 33–60.

Caselli Graziella, Vaupel James, and Yashin Anatoli, 1985. Mortality in Italy: contours of a century of evolution, *Genus*, vol. 41(1–2), p. 39–55.

Delaporte Pierre, 1941. *Évolution de la mortalité en Europe depuis l'origine des statistiques de l'état civil (tables de mortalité de générations)*. Paris, Statistique générale de la France, Imprimerie nationale, 155 p. (Études démographiques no. 2).

Hobcraft John, Menken Jane, and Preston Samuel H., 1982. Age, period, and cohort effects in demography: a review. *Population Index*, vol. 48(1), p. 4–43.

Perozzo Luigi, 1880. Della rappresentazione grafica di una collettività di individui nella successione del tempo, e particolari dei diagrammi a tre coordinate, *Annali di Statistica*, series 2, vol. 12, p. 1–16.13 + charts.

Perozzo Luigi, 1883. Nuove applicazioni del calcolo delle probabilità allo studio del fenomeni statistichi e distribuzione dei medesimi secondo l'età degli sposi, *Annali di Statistica,* series 3, vol. 5, p. 175–203 + tables.

Pullum Thomas W., 1980. Separating age, period and cohort effects in white US fertility, 1920–1970. *Social Science Research*, vol. 9(3), p. 227–244. New York, Academic Press.

Vaupel James W., Gambill Bradley A., and Yashin Anatoli I., 1985. *Contour maps of population surfaces*. Laxenburg (Austria), IIASA, 84 p. (Working Paper no. WP-85-47).

Zeng Yi, Vaupel James, and Yashin Anatoli, 1985. *Marriage and fertility in China: a Lexis-surface analysis*, Laxenburg (Austria), IIASA, 18 p. (Working Paper no. WP-85-70).

C H A P T E R

8

Rates, Frequencies, and Probabilities

GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Up to now, various types of rates have been computed by dividing the number of events by an average population, without taking account of the differences in substance between births, deaths, and migrations, nor the types of situations one encounters when computing these rates. The purpose of this chapter is therefore to distinguish between the different types of events and rates when evaluating the incidence of a particular demographic phenomenon such as fertility or mortality.

I. TYPES OF EVENTS AND POPULATION STATES

1. Repeatable and Nonrepeatable Events

Death is a unique event, in all the meanings of the word, whereas most women and men experience giving birth or fathering more than one child. Many people also migrate several times during their lives. Death is therefore a *nonrepeatable* (or nonrecurrent) event whereas birth and migration are *repeatable* (or recurrent) events, in the sense that individuals can experience these events several times during their lifetime. Repeatable events can however be made nonrepeatable if one gives them a *rank order*. Births can thus be ordered according to rank one (first birth), rank two (second birth), and so forth. In a similar way, one can distinguish the first migration from the second, and so on. Death always has rank one! One simple way to describe a demographic phenomenon is then to classify events according to their rank order, as in a table of births by rank. In a low-fertility country, most births are located in the rank one and two categories; in high-fertility countries, numerous births occur at higher ranks too. Birth certificates should therefore always include the order of the birth in the lifetime of the person, mother, or father.[1]

Even though birth is a repeatable event, it must be stressed that each birth is nevertheless a unique individual experience. Giving birth to one's first daughter Sally is a different life experience from giving birth to one's fourth child Billy. The reasons people do or do not have a first child can therefore be quite different from those explaining why couples have had a fourth child or not. The same holds true for migration: Migrating as a child depends on causes other than migrating as a young adult or after retirement. Once again, although migration is a repeatable event, each move is a unique experience based on a specific set of circumstances, such as marriage, change of occupation, need of a nursing home, and so forth. A detailed analysis of migration has to take these various reasons into account. Death is nonrepeatable, but the process leading to death might entail a series of repeatable issues, such as a succession of illnesses. Once again, each sequence should be considered separately from each other, in a life course perspective.

2. Ranks and States

One can establish a one-to-one correspondence between the ranks of the events one experiences and

[1] The two are not necessarily the same, due to polygamy or to multiple marriages in one's past.

the states one is in. In the case of death, one obviously moves from the state of being alive to the state of deceased. To give another example, first marriage changes the state of the female or male from single to married. In the field of fertility, having a first child modifies the *parity* of the woman concerned: she has moved from parity zero to parity one. If she has a second child, she will move on to parity two, and so on. Parity k is therefore the state one is in when one has had k children during one's past life. A childless woman or man is thus of parity 0. Note that as one moves from parity k to parity $k + 1$, the number of persons involved usually decreases, as not all individuals of parity k will have a child of rank $k + 1$. People of parity $k + 1$ are therefore a subset of those of parity k. Counting either nonrepeatable events or changes of states lead both therefore to the same numerical result. Similarly, estimating the probability of having a fourth child, for those exposed to risk, is equal to determining the probability of moving from parity three to parity four. As one produces a table on births by rank order based on the vital registration system, it is also useful to draw up a table of women by age according to parity, based on data from a census or a survey. Once again, in a low-fertility country, most women at the end of their reproductive period will be of low parities, while in a high-fertility country a large proportion of women will be in the higher parity groups.

3. Absorbing and Transient States

The state of deceased is an *absorbing* state, in the terminology of Markov chains: once entered, it is never left. On the other hand, parity three and parity four are states that *communicate* with each other: One may move from parity three to parity four, and then on possibly to parity five or more, although backward transition (from four to three or five to four) is impossible. In this case, there is only a one-way transition between parity state three and parity state four. In other situations, such as in the study of health and morbidity, one can observe on the contrary two-way transitions between *transient* states, moving from the state of good health to the state of being ill, and then back again to the state of good health if one recovers from the disease. Explaining demographic trends implies describing as a first step these changes between successive states. *Multistate* demographic models follow individuals from one state to the other and compute the probabilities of going out, coming back, or staying in each state during a specified period (see e.g., Halli and Rao, 1992, Section 2.5). The data one deals with should therefore distinguish the order of the events concerned and the corresponding states one is in.

4. Possible and Fatal Events

Not only does death differ from migration or maternity because it is nonrepeatable, but contrary to the other two events it is also *fatal*. Although a woman may spend her whole life without bearing a child, or a person may never cross the borders of his or her country, no human being can escape the fact that he or she will eventually die. This difference will obviously be reflected in the measures of final *intensity* (i.e., the *quantum* or number of events per person) of the phenomenon. At the scale of one's lifetime, the intensity of mortality is always 100%, meaning one death per person. On the contrary, in the fields of fertility and migration, these measures can take values other than one, such as bearing two children over one's lifetime or experiencing five migrations or changes of residence between birth and death.

5. Censoring

An event such as death or international out-migration excludes those who experience it from observation. The observation is *censored*, as one says, by death or by migration. On the contrary, having a birth does not exclude the female from observation; it changes her parity state, as we have seen, but we can still follow her after the child is born. Internal migration does not censor the data either, if the person can be followed in the country from one place of residence to the other thanks, for example, to a population register. Once again, internal migration changes the characteristics of the person (from a one-time migrant to a two-time migrant for example) but he or she can still be followed in the country one is dealing with. Data on fertility can therefore be obtained by a retrospective survey, for those who have not experienced international out-migration or death; the same is true for internal migration or for morbidity. In all these cases, the person who is interviewed should adequately recall his or her past life[2] and produce the information required on the number of births, migrations, or illnesses he or she has experienced. On the contrary, due to censoring, international out-migration and especially death cannot be observed by retrospective interviews. These people are therefore lost to follow-up, and a major question is to what extent those who have died or migrated out of the country differ from those who have survived and

[2] As cognitive abilities decrease with age, the quality of retrospective data often deteriorates at older ages, even excluding extreme cases of, e.g., persons affected with Alzheimer disease. Errors are more important for certain types of information than for others, depending on the type of memory involved.

stayed. This important issue will be discussed in Chapter 9.

II. REPEATABLE EVENTS: RATES AND FREQUENCIES

Consider a retrospective cohort study where women, interviewed at the end of their reproductive period (e.g., 55 years of age), are asked how many live births they have had during their lifetime. Let F_{55} stand for the number of women interviewed. If B_x represents the number of births occurring in this cohort of women at completed age x, then B_x/F_{55} is the relative number of births at age x per woman. This ratio is called the *age-specific fertility "rate"*[3] or more rarely the *age-specific fertility frequency* at age x. The sum Σ_x (B_x/F_{55}) over all x yields the average number of children[4] born during the reproductive period (i.e., a measure of the *intensity or quantum* of fertility). The denominator F_{55} concerns all women, having had or not a child during their lifetime.

As only survivors (to death and migration) can be interviewed at the end of the reproductive period, the measure of fertility quantum described above is net of the effects of mortality and migration. Furthermore, if mortality and migration do not select women with higher or lower fertility than those who survive until the end of their reproductive period, this measure of quantum is identical to that which would have been obtained if there had been no mortality nor migration at all during the reproductive period, except for larger confidence intervals due to the smaller number of cases involved. If censoring such as death or migration does, on the contrary, select women with different fertility levels, then the measures obtained by retrospective questions do not yield the same result as in the absence of censoring. In this case, birth, death, or migration are not *independent* from each other due to known or unknown common causes. For example, if the population is *heterogeneous* such that women of lower socioeconomic status have higher fertility and also a higher chance of dying than low fertility women of higher social status, censoring will select women with potentially higher fertility and progressively lead to a cohort of survivors of lower fertility (see Chapter 21). Social status is here a *common cause* of both fertility and mortality, and the two phenomena are therefore not independent. In this case, fertility measures will differ in the presence or in the absence of censoring.

Suppose now that births by age of mother are recorded each year by the vital registration system, giving for the same cohort births B'_x at completed age x. If F_x stands for the number of women surviving (to death and to migration) at that age, the ratio B'_x/F_x will be equal to the ratio B_x/F_{55} if women surviving until age 55 are not a selected group with respect to their fertility. Once again, women with higher fertility might be exposed to higher risks of dying or of migrating; in this case, B_x/F_{55} would actually be lower than B'_x/F_x due to the selection effect, as F_x also includes here women with higher fertility who will die or migrate eventually before reaching age 55. Once again if there is no selection (i.e., if there is *independence* between fertility on the one hand and mortality and migration on the other hand), the age-specific fertility "rate" or frequency B'_x/F_x will be equal to B_x/F_{55} at each age x, and therefore one can also characterize the average intensity or quantum of fertility by the sum of frequencies or "rates" $\Sigma\, B'_x/F_x$.

In practice, the denominator F_x is more or less equal to $F_{x+0.5}$ at exact age x + 0.5, if births are recorded by completed age x (see Chapter 6). The fertility "rate" can then be written $B'_x/F_{x+0.5;}$ the extension to intervals x to x + n is immediate, as the "rate" or frequency is then more or less equal to ${}_nB'_x/F_{x+n/2}$. Once again, the average number of children born for this cohort of women is obtained by summing the number of "rates" or frequencies by age intervals x to $x + n$ over all x. If the fertility "rate" is computed on a period basis for a particular year, by dividing the number of births ${}_nB'_x$ recorded during the year between exact ages x to $x + n$, by the mid-year population ${}_nF_x$ aged x to $x + n$, then the period rate should be multiplied by the age interval n for comparison with its cohort equivalent, as ${}_nF_x \cong {}_nF_{x+n/2}$ under the hypothesis of identical cohort fertility "rates." As an alternative, one can consider that the denominator of the "rate" is $1/n^{\text{th}}$ of the *person-years* lived in the age range x to $x + n$, i.e. roughly ${}_nF_x/n$.

Examples can be found in most demographic textbooks (such as Pressat, 1983; Preston, Heuveline, and Guillot, 2001; or Wunsch and Termote, 1978). As an illustration, the Belgian female cohort born in 1959 has given birth to 7494 children during the year 1990, subdivided into 3841 births at completed age 30 and 3653 at completed age 31. The corresponding mid-year estimated female population is 78,663. The age-specific fertility "rate" or frequency is thus 7494/78,663 = 0.09527 at exact age 31 on average. Similarly, if the data on births and on population numbers are available each year, one can compute the fertility "rate" at each

[3] In French, it is called *taux de fécondité* or sometimes *naisssances réduites* in Louis Henry's terminology. As the term *rate* has various meanings in demography, the term is put here between inverted commas in view of distinguishing it from the *exposure rate* which will be defined further on.

[4] In demography, fertility is usually restricted to *live births* only.

age during the whole reproductive period for birth cohort 1959. The sum of these rates will be equal to the average number of children born per woman of this cohort in the absence of mortality and migration, if the latter do not select women with lower or higher fertility than those females who remain in the population.

III. NONREPEATABLE EVENTS, PROBABILITIES, AND OCCURRENCE/EXPOSURE RATES

The approach developed in the previous section applies to all *repeatable and nonrepeatable events*, but only for those demographic processes such as fertility or nuptiality (see Chapter 25) that do not exclude the population from observation. In the case of *nonrepeatable* events, such as births by birth order, the same procedure can be used (see Chapter 12 for an example). Once again, to obtain measures of fertility by birth order net of the effects of censoring by death or migration, independence between fertility and the censoring phenomena is required although in reality, selection effects do occur due to the heterogeneity of the population. In particular, as we are now dealing with nonrepeatable effects—there should be no difference in mortality and migration between women of different parities, at the same age. This assumption has been called the condition of *continuity* in the demographic literature (Henry, 1972, Chapter 4). If this condition does not hold, as for example in the case where having a child increases one's risk of death, censoring will once again select women with higher fertility and the actual data will not reflect what would have been the quantum of fertility in the absence of censoring. These conditions of independence and of continuity will be more thoroughly discussed in Chapter 9.

This methodology cannot be used in the field of mortality because mortality excludes people from observation; it is therefore impossible to collect mortality data by retrospective surveys[5] as "dead men tell no tales." As "rates" or frequencies (as defined above) cannot be used in this situation, the analysis of mortality in the absence of migration is based instead on the *probabilities* of dying by age estimated in the absence of censoring by migration. The same approach can in fact be applied to all nonrepeatable events, such as births or emigrations by rank order.

[5] One can however obtain indirect estimations of mortality at retrospective surveys by obtaining e.g. data from mothers on the proportion of their children dead or by asking the household head how many deaths have occurred in the household during a given period. These indirect methods are used when vital registration does not exist or is of poor quality (see Volume 4).

1. Nonrepeatable Events and Probabilities

Suppose that the exact times of death are known and that deaths are registered per day. This is the case if the death certificate registers the time of death by day, month, and year; for example, it is known that Mr. George Brown has died on February 21, 1998. If a cohort is followed day by day, it is possible to compute the risks of dying for this cohort during each day j by the method developed by E. L. Kaplan and P. Meier;[6] (see e.g., Estève *et al*. 1993, Chapter 4, for an application). The probability of dying in the absence of migration, during day j, is taken equal to $q_j = D_j/P_j$, where D_j represents the number of deaths which have occurred during day j and P_j stands for the population exposed to risk at the beginning of the day, for example those having survived to death *and* out-migration during their lifetime until day j. If E_j out-migrants are observed during the same period j, one supposes that all D_j deaths have occurred before the E_j migrations; the population P_j exposed to death is therefore not affected by the occurrence of E_j. The population P_{j+1} exposed to the risk of dying during the next day $j + 1$ will then be equal to $P_j - D_j - E_j$, and $q_{j+1} = D_{j+1}/P_{j+1}$.

The only assumption the method requires is that death occurs during each period before its competitor, migration in this case. The method also easily takes care of immigration. If I_j immigrants are observed during day j, one assumes once again that they occur after deaths D_j. The formula for q_j therefore remains the same. The population exposed to risk at time $j + 1$ then becomes equal to $P_j - D_j - E_j + I_j$. Having computed the daily probabilities of dying q_j in the absence of migration, it is easy to derive the yearly probability of dying q_t during year t, in the absence of migration, using the multiplicative property of probabilities of survival. One has

$$q_t = 1 - \Pi_j(1 - q_j)$$

for j varying from the first day of the year (day 1) to the last day of the year (day 365); as usual, Π_j is the symbol of multiplication over j.

Aalen (1978; see also Courgeau and Lelièvre, 1992) has generalized this approach to take account of dependent risks in multiple decrement models. His approach is based on martingale counting process theory, leading to a multiplicative intensity model. An advantage of this model is that the changes in the risk set can depend quite arbitrarily on past, though not on

[6] Kaplan and Meier (1958). Assuming that deaths occurring during an *i*th time interval are considered as a binomial variable, E.L. Kaplan and P. Meier propose that each death occupy an interval by itself. See e.g. Harris and Albert (1991) for further reading, including confidence bands for the Kaplan-Meier survival curves.

future, failures. The Aalen estimator is a cumulative hazard rate, the values of which can be plotted and compared between multiple samples. The theory also allows the construction of confidence bands for the empirical cumulative intensities.

The *Kaplan-Meier* method is recommended as long as the units of time remain short; ideally, the estimator should be computed for each death separately when it occurs, except if several deaths happen at the same time. In practice, one depends upon the unit of time adopted in the data set (e.g., the hour or the day). If deaths are recorded per year, the estimator $q_t = D_t/P_t$ would not be adequate, as the assumption that competing events all occur after all deaths have been observed becomes less and less plausible as the unit of time increases. If out-migration had not occurred, the number of deaths during year t would have been greater than the number D_t actually observed, as some of the would-be migrants would have died during the year. Let K_t stand for this unknown number. The probability of dying in the absence of migration would then be

$$q_t = \frac{(D_t + K_t)}{P_t}$$

The numerator takes account of the number of deaths D_t actually observed, and the number K_t one would have observed in addition if there had been no migration. The problem is then estimating K_t. If all potential out-migrants die during the period, $K_t = E_t$; if all survive, $K_t = 0$. The probability would therefore be at the most $(D_t + E_t)/P_t$ and at the least $(D_t)/P_t$. The real situation is somewhere in-between.

Let Q_t $(0 \le Q_t \le 1)$ stand for the probability of dying of the potential migrants. Assuming rather arbitrarily that they would have been subjected to half the risk during the period, if they have left the country more or less uniformly over the period on average,[7] the probability in the absence of migration is then

$$q_t = \frac{(D_t + 0.5E_tQ_t)}{P_t}$$

Q_t is unknown and can be less than, equal to, or greater than q_t, depending on the fact that out-migrants are healthier than others, similar to the nonmovers, or in a less good condition. In the absence of information on the difference between Q_t and q_t, one usually takes $Q_t = q_t$ under the rather strong assumption that migrants are a perfect random sample of the population and therefore do not differ on average from those who remain. This is equivalent to assuming that death and migration are *independent* events (i.e., that there are no common causes or causal links between mortality and migration). Mortality and migration would in this case be two independent competing risks for the survivors (to death, to migration) at the beginning of the time interval considered. Under this assumption, the probability of dying is then estimated by what is sometimes called the approximate *Berkson formula* for the probability of dying, which is frequently used in the construction of life tables (see Chapter 11):

$$q_t = \frac{(D_t + 0.5E_tq_t)}{P_t}$$

or:

$$q_t = \frac{D_t}{P_t - 0.5E_t}$$

In the case of immigration, the formula simply becomes:

$$q_t = \frac{D_t}{P_t - 0.5E_t + 0.5I_t}$$

For example, having recourse once again to the Belgian data for the 1959 female birth cohort observed during the year 1990, one knows that 44 deaths have occurred in this cohort during the year. There have also been 789 immigrations and 615 emigrations of females during the same period at this age. The population of the cohort exposed to risk at the beginning of the period is 78,599. Without correction for migration, the probability of dying would therefore be equal to:

$$q_{1990} = \frac{44}{78{,}599} = 0.00055980 \cong 0.6 \text{ per thousand}$$

The *Berkson* formula for the corrected probability is in this case equal to:

$$q_{1990} = \frac{44}{78{,}599 - 0.5 \times 615 + 0.5 \times 789}$$
$$= 0.00055918 \cong 0.6 \text{ per thousand}$$

The correction is not significant in this example, but this is not always the case in practice. Having computed the probabilities of dying at all ages for this cohort, one can set up a cohort *life table* as discussed in Chapter 11. Cohort mortality studies are usually restricted to segments of people's lives, such as between, for example, 55 and 75 years of age if one is particularly interested in the mortality experience of this age group.[8]

[7] A mathematically more precise demonstration is given e.g. in Pressat (1995, p. 83–88).

[8] This would be the case for example if the study focuses on particular diseases of older ages, such as ischemic heart disease or cancer.

It must be stressed that the probability of dying at each age is in fact an average measure applied to all the members of the cohort surviving until that duration. If the population is *heterogeneous* with regard to mortality, and this will be the usual case, persons with a higher probability of dying will on average experience death earlier than those with a lower risk of dying. As age increases, the population of survivors will proportionately be composed more and more of individuals with lower risks, as the others have dropped out of observation due to death. For example, in the developing countries, it is well known that measles kills the weakest children, especially those with malnutrition. In this case, the survivors would consist of the strongest children having a lower average probability of dying. In this example, there could however be another selection effect. Survivors could possibly consist of weakened children through the detrimental effect of measles on the state of nutrition. The latter children would therefore have a probability of dying greater during the following months than the one the population would have experienced if it had not been exposed to measles, as observed in the Kasongo for example.

In each actual situation, one should thus be aware of possible selection effects and try to point them out. Attempts have been made to correct the impact of unobserved heterogeneity assuming subgroups with different risks among the population (see Chapter 21). Results are however sensitive to the choice of the mixture model. A good solution is trying to find out the causes of the underlying heterogeneity, in view of delimiting more homogeneous subgroups, as pointed out a long time ago by, for example, Henry in demography, MacMahon and Pugh in epidemiology, or Salmon in the philosophy of science. Better still, if the data are available, one can take account of individual life histories and apply the techniques of event-history analysis to this data set. Finally, one can put together individual data and contextual variables in a multilevel model, thus avoiding in the analysis both the atomistic and the ecological fallacies (Courgeau, 2004).

Probabilities may also be computed in the field of fertility for example, if one is dealing with births by rank order (see Chapter 12). Over very small periods of time however, such as t to $t + \delta$, one can nevertheless evaluate the probability for a woman to give birth in the interval $t; t + \delta$ regardless of birth order, assuming that the occurrence of birth at the micro level is generated by an underlying Poisson process or better still by a Yule or a Polya process (Chiang, 1968); the result will remain between 0 and 1. The probability that an event occurs during the time interval $(t, t + \delta)$ can be written $\lambda\delta + o(\delta)$ where $o(\delta)$ represents the probability that more than one event occurs in $(t, t + \delta)$. The notation $o(\delta)$ stands for any function of δ, which tends to 0 faster than δ. Different assumptions concerning λ yield different models of fertility.

2. Occurrence/Exposure Rates

In the case of *nonrepeatable* events, one may also compute *rates* of exposure in addition to or instead of *probabilities.* For example, with mortality rates instead of dividing the number of deaths observed during a period by the initial population of survivors exposed to risk, possibly corrected for competing events such as migration, take the ratio of the number of deaths (occurrences) to the exposure time or *person-years* of exposure observed during the period. Using the same symbols as before, the number of person-years of exposure during year t is equal to:

$$P_{t+1} + 0.5\ D_t + 0.5\ E_t - 0.5\ I_t$$

One assumes that deaths D_t, in- and out-migrations I_t, and E_t all occur on average at the middle of the year as, for example, in the case of an uniform distribution of events over the year or a linear survivorship function over the interval. Survivors P_{t+1} at exact age $t + 1$ have obviously lived 1 whole year[9] between t and $t + 1$, while those experiencing death or migration have been exposed on average one half-year. The *occurrence/exposure rate* or *central rate* of death is then equal to

$$r_t = \frac{D_t}{P_{t+1} + 0.5D_t + 0.5E_t - 0.5I_t}$$

or

$$r_t = \frac{D_t}{0.5(P_t + P_{t+1})}$$

The last formula is obtained by replacing P_{t+1} in the denominator of the rate by its equivalent $P_t - D_t - E_t + I_t$, yielding the average of P_t and P_{t+1} assuming an uniform distribution of events over the interval. The formula is easily extended to a period of n years, from t to $t + n$:

$${}_n r_t = \frac{{}_n D_t}{0.5 \cdot n \cdot (P_t + P_{t+n})}$$

${}_nD_t$ representing the number of deaths during the n-year period. For example, the Belgian data on mortality would yield a rate of

[9] Except those who have immigrated, on average at mid period, during the year; this is the reason why 0.5 I_t are subtracted from the sum.

$$r_{1990} = \frac{44}{0.5(78,599 + 78,729)}$$
$$= 0.00055934 \cong 0.6 \text{ per thousand}$$

the numerator being the number of deaths observed in the 1959 female birth cohort during the year 1990, and the denominator being the female population surviving at mid-period, an estimate of the person-years of exposure to risk during the time-period.

As for probabilities over discrete intervals, occurrence/exposure rates can be computed for all nonrepeatable events such as death, birth by rank, migration by order, and so forth. The denominator is always the *person-time* of exposure to risk by those who have not experienced censoring either by the event studied (e.g., death) or the competing events (e.g., migration).

Rates have the advantage over probabilities that they do not depend upon the length of the time interval considered. For example, one may not compare the probability of dying between 20 and 21 years of age with that between 20 and 25, as the longer the interval the more deaths there will be. These probabilities cannot be put, for example, on the same graph. On the contrary, one may compare the occurrence/exposure rates between these two age groups, as rates are expressed per person and per unit of time. Rates do not depend therefore on the length of the interval. To give another example in the field of infant mortality, one may not compare the neonatal probability of dying to the postneonatal probability, as the former covers the first month of life and the latter the last 11 months of the year. The time interval of the latter is thus 11 times that of the former. One may however compare the neonatal infant mortality rate[10] to the postneonatal rate.

3. Frequencies and Rates

One should distinguish between two types of fertility or emigration *rates*. In the case of fertility for example, one rate is equal to the age-specific fertility "rate" or frequency obtained by dividing births (all birth orders or by birth order) by the population surviving to competing events (death and migration), while the other rate is an age-specific occurrence/exposure rate obtained by dividing the number of births by order by the population which has neither experienced the event studied (birth of a certain order) nor the competing events (death and migration). As we will see in Chapter 14, the latter type of rate is much to be preferred to the former. The terminology actually used in the literature can be confusing, as pointed out in footnote 10. In French, one usually distinguishes between frequencies ("reduced events" or "second category rates"), occurrence/exposure rates ("first category rates"), and probabilities ("quotients"). In English, the terminology actually used is not always that clear, the term *rate* often being used in the sense of a probability, of an exposure rate, or of a frequency as defined above! The terms *central rate* and *occurrence/exposure rate* are synonyms.

4. Relations Between Probabilities and Occurrence/Exposure Rates

As seen above, probabilities and exposure rates are quite similar indicators, but they differ radically from frequencies (as defined previously). Probabilities and rates both apply to nonrepeatable events. The numerators of the two are identical, but they differ in their denominator. In the case of probabilities, the denominator is the population exposed at the beginning of the interval, while in the case of rates it is equal to the person-years of exposure. The latter denominator can be approximated by the average population exposed at the beginning and end of the year, assuming a linear survivorship function, if the time interval is 1 year. It is therefore easy to transform rates into probabilities and vice versa. A general relation would be:

$$_nq_t = 2n\ _nr_t / (2 + 2(1 - {_na_t})n\ _nr_t)$$

where n is the age interval t to $t + n$ and $_n\mathrm{a}_t$ is the fraction of time spent alive in the age interval t to $t + n$ by those who die during this interval (Chiang, 1968). If deaths are uniformly distributed over the age interval, for example, if the survivorship function is linear, $_n\mathrm{a}_t = 0.5$ yielding the formulas described below. Assuming a uniform distribution of events over the year, one thus obtains:

$$q_t = \frac{2r_t}{2 + r_t}$$

To demonstrate, replace the rate r_t by its value given above $D_t\ /\ (P_{t+1} + 0.5\ D_t + 0.5\ E_t - 0.5\ I_t)$; one obtains the probability $q_t = D_t / (P_{t+1} + D_t + 0.5\ E_t - 0.5\ I_t)$. Replacing P_{t+1} by its value $P_t - D_t - E_t + I_t$ yields $q_t = D_t / (P_t - 0.5\ E_t + 0.5\ I_t)$, which is in fact the approximate *Berkson* formula obtained above for the net probability of dying. The formula can easily be extended to n-year intervals:

$$_nq_t = \frac{2n \cdot {_nr_t}}{2 + n \cdot {_nr_t}}$$

[10] In current practice, the term *infant mortality rate* is often used for *infant mortality probability*, an unfortunate linguistic confusion between rate (probability) and rate (exposure rate)!

The inverse relationship can also be obtained:

$$ {}_n r_t = \frac{2\,{}_n q_t}{n(2 - {}_n q_t)} $$

These formulas can also be adapted to take into account the possible nonuniform distribution of events over the interval, if one knows this distribution. For example, an exponential survivorship function would yield:

$$ {}_n q_t = 1 - e^{-n \cdot {}_n r_t} $$

For example, if the Belgian male occurrence/exposure death rate in the 80 to 85 age group is 0.12420, the corresponding probability of dying will be

$$ {}_5 q_{80} = \frac{2 \times 5 \times 0.1242}{2 + 5 \times 0.1242} = 0.47386 $$

supposing a linear survivorship function, and

$$ {}_5 q_{80} = 1 - e^{-5 \cdot 0.1242} = 0.46259 $$

assuming an exponential one. The difference is not totally negligible as, at that age, mortality is relatively high and the form of the survivorship function in this 5-year age interval must be taken into account.

IV. SINGLE AND MULTIPLE DECREMENT

Up to now, probabilities of dying for example have been corrected to obtain a measure of the probability of dying *net* of migration, in what is called the *single decrement* approach. One wishes to know in this case the probability of dying in the absence of other censoring events. In the actual world, the population is affected by both events, death and migration, and this has to be taken into account in population projections. In this case, one may wish to compute in a *multiple decrement* approach a *gross* probability of dying and a *gross* probability of emigrating with no correction for competing risks. One would then have, respectively:

$$ \kappa_t = \frac{D_t}{P_t} \quad \text{and} \quad \varepsilon_t = \frac{E_t}{P_t} $$

the probability of attrition by both death and out-migration being in this case:

$$ s_t = \frac{D_t + E_t}{P_t} = \kappa_t + \varepsilon_t $$

which is equal to the sum of the gross probabilities. Gross probabilities are therefore additive, which would not be the case with net probabilities as it is now shown. Let q_t and e_t stand respectively for the *net* probabilities of dying and out-migration obtained by the Berkson formula:

$$ q_t = \frac{D_t}{P_t - 0.5E_t} \quad \text{and} \quad e_t = \frac{E_t}{P_t - 0.5D_t} $$

In this case, out-migration is competing for censoring with death and vice-versa. It is easy to check that the probability of exit by either death or out-migration is not equal here to the sum of q_t and e_t. However, the probability of neither dying nor emigrating is equal to the product of their complements, if the probabilities are independent. The global probability of attrition due to death and migration, in the *single decrement* approach, is therefore equal to:

$$ s_t = 1 - (1 - q_t)(1 - e_t) = q_t + e_t - q_t e_t $$

For example, if the male *net* probability of dying in the 80 to 85 age group is 0.46259, as obtained above, and the net probability of out-migrating from the area of residence is 0.18926, the total probability of exit due to both mortality or out-migration would be:

$$ {}_5 s_{80} = 0.46259 + 0.18926 - 0.46259 \times 0.18926 = 0.56430 $$

If they were *gross* probabilities, one would have instead:

$$ {}_5 s_{80} = 0.46259 + 0.18926 = 0.65185 $$

As the difference can be significant, as one sees from this example, it is important to know if the probabilities one is using are gross or net; not all authors explain how they have computed the probabilities they have obtained, and the results can therefore be misleading in comparative studies.

The single decrement approach based on net probabilities is often used when one is interested in one particular event, such as death in the absence of migration or, to give another example, death from ischemic heart disease in the absence of all other causes of death. On the other hand, the multiple decrement approach based on gross probabilities is adopted when one is examining the multiple causes of exit, such as in the case of the various causes of death. Examples and further discussion can be found in Preston *et al*. (1972, 2001). Note that computing gross probabilities requires no assumption, whereas establishing net probabilities usually assumes the independence between these probabilities. Without making any assumption on the relationship between risks, one may therefore compute gross probabilities of internal migration after marriage or vice versa, on the basis of survey data, as long as one does not raise the question of what these risks would have been in the absence of mortality and international migration.

V. THE FORCE OF ATTRITION

In the previous sections, it has been stated that demographic measures can be estimated for very small intervals of time. However, the advantage of greater precision is compensated by the disadvantage of dealing with small numbers of events and individuals. As the time interval becomes smaller, one reaches at the limit what is called the *force* or the *instantaneous rate* already developed in Chapter 3 for the crude rate. This section therefore develops briefly the concept of *force*, which is used in the actuarial literature and in mathematical demography, mortality being taken here as an example. Consider once again the number of survivors P_x at exact age x, and suppose there is no migration. The force of mortality is then defined as:

$$\mu_x = \lim_{\Delta x \to 0} \frac{P_x - P_{x+\Delta_x}}{\Delta_x P_x} \quad \text{for} \quad \Delta_x \geq 0,$$

or:

$$\mu_x = -\frac{\mathrm{d}P_x}{P_x \mathrm{d}x}$$

The minus sign is due to the fact that the survivorship function P_x decreases as x increases; its derivative is therefore negative. As the force of occurrence must be positive, one must multiply the derivative by – 1. One notices that, in fact, the force of mortality is an instantaneous death *rate*, with deaths as the numerator and person-time as the denominator. The discrete death rate centered on age x, as presented before, is therefore usually a very good approximation of the force of mortality at age x. The above expression μ_x is also equal to

$$\mu_x = -\frac{\mathrm{d}(\ln P_x)}{\mathrm{d}x}$$

where *ln* stands for the natural logarithm.

Inversely, one can obtain P_x by integration. Writing $\mathrm{d}(\ln P_x) = -\mu_x \mathrm{d}x$ integration gives

$$\ln P_x = -\int \mu(a)\mathrm{d}a + c$$

where c represents an integration constant. The former expression is equal to

$$P_x = c^* e^{-\int \mu(a)\mathrm{d}a} \quad \text{with} \quad \ln c^* = c$$

As the number of survivors at duration x depends upon the forces of death from 0 to x, one writes:

$$P_x = c^* e^{-\int_0^x \mu(a)\mathrm{d}a}$$

For $x = 0$, $P_0 = c^*$ so that finally

$$P_x = P_0 e^{-\int_0^x \mu(a)\mathrm{d}a}$$

The discrete probability of dying can then be obtained by the complement of the probability of survival over the interval, for example:

$$q_x = 1 - \frac{P_{x+1}}{P_x} = 1 - e^{\int_x^{x+1} \mu(a)\mathrm{d}a}$$

Applications can be found in Keyfitz (1977). This approach will be discussed again in the context of the life table (Chapter 11).

References

AALEN Odd O., 1978. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, vol. 6(4), p. 701–726.

CHIANG Chin Long, 1968. *Introduction to stochastic processes in biostatistics*. New York, Wiley, 313 p.

COURGEAU Daniel, 2004. Du groupe à l'individu. Synthèse multiniveau. Paris, Editions de l'INED, 242 p.

COURGEAU Daniel and LELIÈVRE Èva, 1992. *Event history analysis in demography.* Oxford, Clarendon Press, 226 p.

ESTÈVE, Jacques, BENHAMOU Ellen, and RAYMOND Luc, 1993. *Méthodes statistiques en épidémiologie descriptive.* Paris, INSERM, 307 p.

HALLI Shiva S. and RAO Vaninadha, 1992. *Advanced techniques of population analysis*. New York, Plenum, 226 p.

HARRIS Eugène K. and ALBERT Adelin, 1991. *Survivorship analysis for clinical studies*. New York, Marcel Dekker, 200 p.

HENRY Louis, 1972. *Demographie: analyse et modèles.* Paris, Larousse, 341 p.

KAPLAN E. L. and Meier P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457–481.

KEYFITZ Nathan, 1977. *Applied mathematical demography*. New York, John Wiley and Sons, 388 p. (reedited in1985 by Springer-Verlag, New York, 442 p).

PRESSAT Roland, 1966. *Pratique de la démographie*. Paris, Dunod, 301 p.

PRESSAT Roland, 1983. *L'analyse démographique: Concepts, Méthodes, Résultats*, 4th ed. Paris, Presses Universitaires de France, 295 p.

PRESSAT Roland, 1995. *Éléments de démographie mathématique.* Paris, AIDELF, 279 p.

PRESTON Samuel H., HEUVELINE Patrick, and GUILLOT Michel, 2001. *Demography: measuring and modeling population processes.* Oxford, Blackwell Publishers, 291 p.

PRESTON Samuel H., KEYFITZ Nathan and SCHOEN Robert, 1972. *Causes of Death: Life Tables for National Populations*. New York, Seminar Press, 787 p.

WUNSCH Guillaume J. and TERMOTE Marc G., 1978. *Introduction to demographic analysis*. New York, Plenum Press, 274 p.

C H A P T E R

9

Competing Risks, Independence, and Continuity

GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

As seen in Chapter 8, measures of the components of population change must address the problem of censoring and of competing risks. Censoring occurs when the observation is truncated and the persons are lost to follow-up. For example, in a prospective study on mortality, where survivors are followed up over a period of several years, censoring might occur as a result of out-migration. To give another example, in the field of fertility, in a study on the use-effectiveness of a new contraceptive, where users are followed up for several months, censoring might happen due to death or out-migration, but also because of withdrawing from the survey to switch to another contraceptive method or because they wish to conceive. These exits from follow-up disturb the observation, but they also hamper, for example, the evaluation of the use-effectiveness of the new contraceptive if censoring or loss to follow-up selects persons with higher or lower use-effectiveness than those who remain in the study. Similarly, in the study of mortality, out-migrants might have risks of dying different from those who do not move; probabilities of dying drawn from the study do not therefore reflect the actual risks in the total population. As this heterogeneity of risks seems likely, between those who stay in the study and those who do not, the question is what would have been the result if there had been no censoring effects (i.e., no disturbances or competing risks)?

I. THE CONDITION OF INDEPENDENCE

The first problem concerns the measurement of the risk over a specific time interval in the absence of other causes of exit. As seen in Chapter 8, if individual exact times of occurrence are available, the *Kaplan-Meier* estimator of this probability is highly recommended, because it does not assume any particular relationship between risks. This is also the case of the *Aalen* estimator (Courgeau and Lelièvre, 1992, Chapter 4). The only assumption concerns the occurrence of more than one event during the interval; in that case, one assumes that the event of interest (e.g., death) happens before its competitor (e.g., migration). If individual times of occurrence are not available, such as when deaths and migrations are recorded by 1-year intervals, formulas have been developed under the assumption that the *forces* of mortality and migration, or the events themselves, vary over the interval in a specific way, and that the risks are independent in the sense that, for example, the probability of dying is not influenced by the presence of migration. If competing *forces* are assumed to be proportional over the interval, one obtains the so-called *Elveback formula* for the net probability of dying, whereas if deaths and migrations at each age are considered to be proportional over the interval, one obtains the *Berkson formula* of which an approximation was given in Chapter 8. The reader is referred to Le Bras and

Artzrouni (1980) for a thorough discussion of these two approaches, and to Chiang (1968) for a general discussion of competing risks theory.

As an example, suppose that survivors at ages x and $x + 1$ are equal to P_x and P_{x+1}, and that D_x deaths and E_x emigrations occur during the interval x to $x + 1$. One has $P_{x+1} = P_x - D_x - E_x$. The *Elveback* formula would estimate the probability by taking:

$$q_x = 1 - \left(\frac{P_{x+1}}{P_x}\right)^{\frac{D_x}{D_x+E_x}}$$

while the more common *approximate Berkson* formula would lead, as seen in Chapter 8, to:

$$q_x = \frac{D_x}{P_x - 0.5E_x}$$

In practice, both formulas give close results, especially if the *exact Berkson* or *Schwartz and Lazar* (1961) formula is used instead of its approximation:

$$q_x = \left[1 + 0.5\left(\frac{D_x}{P_x} - \frac{E_x}{P_x}\right)\right] - \sqrt{\left[1 + 0.5\left(\frac{D_x}{P_x} - \frac{E_x}{P_x}\right)\right]^2 - 2\frac{D_x}{P_x}}$$

For example, if for a cohort of males aged 75 in completed years (i.e., between exact ages 75 and 76 years) one observes 4732 deaths, 3294 out-migrations, and 78,231 survivors at exact age 75, the *Elveback formula* would give the following result:

$$q_{75} = 1 - \left(\frac{70,205}{78,231}\right)^{\frac{4732}{4732+3294}} = 0.06183$$

The *approximate Berkson formula* would lead to:

$$q_{75} = \frac{4,732}{78,231 - 0.5 \times 3,294} = 0.06179$$

Finally, the *Schwartz and Lazar* or *exact Berkson formula* would yield:

$$q_{75} = \left[1 + 0.5\left(\frac{4,732}{78,231} - \frac{3,294}{78,231}\right)\right] - \sqrt{\left[1 + 0.5\left(\frac{4,732}{78,231} - \frac{3,294}{78,231}\right)\right]^2 - 2\frac{3,294}{78,231}} = 0.06183$$

One sees that the *Elveback* and *Schwartz and Lazar* formulas give in this case identical results, while the *approximate Berkson* formula yields a lower estimate. There are no major reasons for choosing one or the other estimate, except on the basis of the underlying models and therefore of one's assumptions relating to the two competing risks. The *approximate Berkson formula* is simple and easily takes immigration into account;[1] it could therefore be preferred for these reasons.

As a second problem, consider the condition of *independence*, which is at the basis of the methods developed in the preceding section. For example,[2] are death and migration independent events? Although mutually exclusive, these events might be dependent. For example, migration might select healthier persons than those staying in the country of origin, leading to biased probabilities of dying in this country, in the sense that if no migration had occurred probabilities of dying would have been lower than those actually computed. Usually, and in the absence of contrary evidence, the event of interest and its competitor(s) are considered independent from each other, but this is only a stopgap. In the presence of information on possible common causes of competing events, one's approach should control for the common factor(s) (see e.g., S. Prati, 1995). Oftentimes, lack of suitable data will leave these factors uncontrolled. Assuming the absence of selection effects, as one does under the independence assumption, is not satisfactory in this case; procedures requiring independence should then be considered only as a model of the underlying process, meaning here a simplified representation of the true but unknown situation.

Some demographers have tried to waive the condition of independence by incorporating some form of dependency between competing risks (for an overview, see J. Duchêne, 2002). Without outside information this solution is inadequate, as it has been shown[3] that to each competing risks model with dependent risks, one may associate another model with independent risks, which is observationally equivalent in the sense that data will never be able to distinguish between the two models. The best way out is therefore to obtain more knowledge on the possible common causes of competing events.[4] Confounders can then be controlled as seen above, if one has collected the necessary data. If data on individual life courses are available, this set can be analyzed using event-history techniques taking into account each individual life course, under the assumption that one's past conditions determine one's present situation. Data, analysis, and synthesis or explanation, cannot be dissociated in practice, as all scientists know.

[1] The Schwartz and Lazar formula can lead to aberrant results in the case of immigration.

[2] A more detailed discussion is given in Chapter 44, on dependencies between causes of death.

[3] For a recent demonstration, see Mouchart and Rolin, 2002, theorem 4.2.

[4] See for example Kalbfleisch and Prentice, 1980, Chapter 7.

II. THE CONDITION OF CONTINUITY

In the field of fertility, it was seen in Chapter 8 that two conditions were to be satisfied in order to use rates or frequencies, the condition of *independence* and the condition of *continuity* in Louis Henry's terms. In that case, frequencies or rates could be used instead of probabilities, the advantage being that rates require less data than probabilities because only the population surviving the disturbances (death, migration) is concerned, whereas probabilities also require information on the population that has not (yet) experienced the event under study (e.g., the move from parity 3 to parity 4). However, the advantage of using rates is minimized by the supplementary condition of continuity, which must be satisfied in this case. Both approaches require that the condition of independence be satisfied. Independence is probably not guaranteed in most cases, and the same is probably true for continuity too.

From a causal perspective (see G. Wunsch, 1988), two events A and B are *causally independent* if there are on the one hand no *common causes* to both such that C causes A and B, and on the other hand if there is no *causal relationship* between A and B such that A causes B or B causes A. The *independence* assumption in demography covers the first part of the requirement, while the *continuity* assumption addresses the second part. The absence of common causes is needed for all measures, rates, probabilities, and frequencies, except if one can control for all of the common causes—a doubtful exercise indeed. Some form of unknown selection probably operates in most cases, and the independence model stands as a significant simplification of reality in this case. *Continuity*, on the other hand, supposes no causal relation between competing processes, such as fertility changing one's probability of survival. If this happens, relating births by order to the total population escaping disturbances does not lead to the true ratio in the absence of disturbances, as the parity distribution would not be the same in both cases. For example, females without children would have a higher chance of surviving than females with children, if fertility lowers one's probability of survival. As fertility rates and probabilities take parity into account, continuity is not required in this case.

As stressed in Chapter 8, the problem of dependent or independent risks arises when people are lost to follow-up and one wishes to know, in this case, what would have been the risk in the absence of censoring. If there is no censoring (i.e., if all persons remain in observation), there is no major reason to adopt this viewpoint. One can study whether marriage influences occupation and internal migration, or vice versa, without having to assume independence between these factors. On the contrary, if marriage does influence occupational status, there is by definition in this case a causal dependence between both factors, as marriage is then a cause of occupational change. Assuming independence in this situation would be absurd! The same is true in the case of nonmarital cohabitation and marriage, as the duration of cohabitation increases, most likely, the probability of marriage. Once again, there is a causal dependency here between the length of cohabitation and marriage, and the hypothesis of independence would be incorrect.

To conclude, one may say that the usual measures demographers use, probabilities, rates, or frequencies, require *causal independence* between competing events, if one wants to answer the "what" question of the single decrement approach: what would have been the rate-probability-frequency in the absence of censoring and competing events? The condition of causal independence is most probably not satisfied in practice, and the measures one obtains using the traditional formulas should therefore be considered with some degree of skepticism. One's efforts should then be geared to testing the independence assumption by checking if the variables considered are associated by common causes or causal links. This implies in practice moving from description to explanation, and having recourse to more powerful methods based on statistical analysis, such as covariance structure modeling, multistate transition modeling, or event-history analysis if individual data are available. The ideal is probably measuring all links between the variables one is interested in, both known . . . and unknown. In this situation, there would be no need anymore to assume independence and continuity! This ideal can probably never be attained,[5] but it is good scientific practice to try to reach it over and over again.

References

Chiang Chin Long, 1968. *Introduction to stochastic processes in biostatistics*, New York, Wiley, 313 p.

Courgeau Daniel and Lelièvre Èva, 1992. *Event history analysis in demography*, Oxford, Clarendon Press, 226 p.

Duchêne Josianne (2002). The measurement of mortality by cause and of morbidity, *in*: Guillaume Wunsch, Michel Mouchart, and Josianne Duchêne (eds.), *The life table: modelling survival and death*. Dordrecht, Kluwer, 306 p.

Kalbfleisch John D. and Prentice Ross L., 1980. *The statistical analysis of failure time data*, New York, John Wiley & Sons, 322 p.

Le bras Hervé and Artzrouni Marc, 1980. Interférence, indifférence, indépendance. *Population*, vol. 35(6), p. 1123–1144.

[5] Hopefully, because if we knew everything, scientists would soon be out of a job!

Mouchart, Michel and Rolin Jean-Marie, 2002. Competing risks models: problems of modelling and of identification, *in*: Guillaume Wunsch, Michel Mouchart and Josianne Duchêne (eds.), *The life table: modelling survival and death*. Dordrecht, Kluwer, 306 p.

Prati, Sabrina, 1995. Une méthode d'analyse et d'interprétation des risques concurrents de mortalité par cause. *Population*, vol. 50(4–5), p. 1013–1030.

Schwartz Daniel and Lazar Philippe, 1961. Taux de mortalité pour une cause donnée de décès en tenant compte des autres causes de décès ou de disparition. *Revue de l'Institut International de Statistique*, vol. 29, p. 44–56.

Wunsch Guillaume, 1988. *Causal theory and causal modeling*. Leuven, Leuven University Press, 200 p.

CHAPTER

10

The Longitudinal Approach

GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

As pointed out several times in the previous chapters, there are two different temporal ways to look at data in demography. One determines the characteristics of a population and its behaviors relating to fertility, mortality, and migration at a particular point in time, and possibly compare these characteristics from year to year. This approach is known as a *cross-sectional* or *synchronic study*, also called a *period* observation and analysis.[1] This type of study can therefore be considered as a *photograph* of the population, or of a sample of the population, at point of time *t* or at several points of time. The second way to look at data entails following up people who have had, for example, a heart attack (myocardial infarction), determine their probability of dying each year, and compute the average span of time between their infarct and their eventual death. We can also relate each individual's survival or death, at a specific duration elapsed since his or her heart attack, to his or her age, gender, way of life, and other possible explanatory factors. This approach is known as a *longitudinal study*.[2] As we will see in this chapter, from the viewpoint of data collection, longitudinal observations can be conducted prospectively or retrospectively.

During the last decades, the longitudinal approach has become a major means of study, due to the search for explanations and causal factors or determinants. In this case, one must indeed time-order the succession of events, as causes must always precede effects in time.

As longitudinal analysis is required for the understanding of the internal dynamics of demographic phenomena (i.e., their intensity and tempo), we have preferred to develop the method here, although its usefulness will be made much clearer in Section II and subsequent volumes.

I. INDIVIDUAL AND COHORT DATA[3]

Longitudinal analysis implies following the units of observation or cases over time, to study their changes of states in temporal succession and the durations of the interval between two successive transitions. For example, in the case of fertility, one may be interested in knowing what proportion of women move from parity zero to parity one, from parity one to parity two, and so forth, and in computing the average waiting times between these changes of states. *Time* is therefore a key issue in longitudinal research. One must adequately define the duration of the study (e.g., a study of female fertility should follow women from menar-

[1] Strictly speaking, a period study refers to the incidence of events occurring during a specified period of time in the area considered, such as deaths occurring during the year 2004 in Belgium. A cross-sectional study, such as the estimation of the point-prevalence of disability in the Belgian population, refers on the other hand to a specific point in time (e.g., June 30, 2004). A cross-sectional (synonym: synchronic) study shows the characteristics of people surviving at that time, and the latter may differ however from those who died before the study.

[2] The terms *longitudinal*, *follow-up*, *life course*, *prospective* and *cohort* are often used as synonyms. In this section, longitudinal is more general than prospective (synonym: follow-up) and cohort.

[3] This section is partly based on a chapter by G. Wunsch (1993).

che to menopause), decide on the right unit of time for the study (a time unit of 1 year is too long if one is interested in seasonal variations), and time order correctly the succession of events if we are involved in causal analysis (causes always precede effects in time).

From the viewpoint of data collection, longitudinal observations can be conducted prospectively or retrospectively. *Prospective longitudinal research* implies following the cases forward in time, according to the *arrow of time* as philosophers say. This is the procedure adopted for patients having suffered a heart attack for example; their survival or death is assessed from medical records, registration data, or multiround surveys.[4] A significant advantage of a prospective study, contrary to a retrospective one, is that the characteristics of those who drop out from observation are known for those variables that have been recorded.[5] One can therefore determine if censoring selects individuals with different characteristics than those who remain in the study. Moreover, prospective studies are not affected by recall bias as in the case of retrospective or case-control studies. Prospective longitudinal studies have several drawbacks, however. They are very expensive to carry out, are subject to attrition or loss to follow-up, and require large initial populations if events are rare.[6] Due to sorting (selection) effects, the group probably becomes more and more biased as time passes: People who drop out of the study are often different from those who remain. Once again, heterogeneity of risks among individuals has to be taken into account.

If the phenomenon studied does not exclude individuals from observation, a much cheaper and easier way of conducting a longitudinal research is through *retrospective observation* (also called a *retrospective panel design*). For example, in fertility research, females who have reached the end of their reproductive period are questioned on their past fertility experience. Reproductive histories taking account the number of live births and the date of each birth, among others, can therefore be constructed for each woman interviewed at the end of her reproductive life. However, retrospective longitudinal research is influenced by recall lapses (some events may be forgotten), inadequate recollection of time patterns (dates are often poorly remembered), selective recall (the respondent interprets the relationship between events), and changes through time in the social representation of events (such as abortion or cohabitation). Finally, retrospective analysis may suffer from biased group composition; for example, females with higher fertility may have had higher risks of dying before their menopause and will therefore not be correctly represented in the survey. Moreover, retrospective analysis is inadequate (as noted previously) for phenomena leading to population exclusion, such as death or emigration: Unfortunately, we cannot interview the deceased on their cause of death! For a good discussion of the respective advantages and disadvantages of retrospective and prospective analysis, see Henry (1959).

In longitudinal research, prospective designs (also called *follow-up studies* or *prospective panel designs*) follow the *same* individuals forward in time. Individual *life courses* or trajectories, such as one's reproductive career for example, can then be constructed for all cases that are followed over time. However, as we have seen, this research design will lead to sample attrition or loss to follow-up. Another approach, which avoids this problem to a large extent, is to collect longitudinal data on *nonidentical but comparable* individuals (or cases) over time, as in a *repeated cross-sectional design.* In the latter approach, probability samples are drawn from the same population and the characteristics of the cases are compared from sample to sample. Samples of the same or similar sizes can be compared, but individual life courses cannot be constructed, as the same individuals are not observed over time (see e.g., Jones, 1996, Chapter 6) and therefore techniques such as event–history analysis cannot be used. In this approach, comparisons are therefore possible on the aggregate level, but not on the individual level. The same situation arises if we have recourse to anonymous vital registration data.

Up to now, cases have been followed either forward or backward in time (i.e., prospectively or retrospectively), but the entry of the cases into observation did not refer to a particular time period; furthermore, one did not take into account the periods of time (month, year) during which the observations were made. For example, all patients having a heart attack (the onset of the process studied, called the *event origin*[7] were followed, regardless of the date of their heart attack. This date was not taken into consideration, as the duration since the onset of the process was the only time variable deemed to be of interest. The patient's *age* at onset was obviously considered, but the patients could be grouped together according to their age at onset, whatever their year of birth. In this design, aging is the prime temporal variable, and no particular considera-

[4] Also called *panel surveys* in the social sciences.

[5] The two groups (those who drop out and those who remain in the study) may differ on unrecorded variables.

[6] In this situation, *case-control* studies are probably preferable as they are much cheaper and easier to carry out (see Gray, 1988). For a thorough discussion of the case-control approach, see Bernard and Lapointe (1991).

[7] The term was coined by Henry (1959).

tion is given either to the year of birth of the patient or to the periods (calendar months or years) of observation. The duration elapsed since the heart attack is therefore an indicator or proxy for biologic aging, given the chronologic age of the patient at onset. One could replace age, in these circumstances, by a series of biologic indicators that would reflect the aging process itself.

Consider now the situation where we suspect that the environmental conditions of the times we live in may have an impact on heart disease, due for example to the incidence of more stressful events.[8] We may then take the succession of *calendar years* as a proxy for the changes in environmental conditions (e.g., marital conditions, political conditions, socioeconomic conditions, etc.) over time. History, as represented by the succession of calendar years, would then be our temporal variable in this case, if we neglect the *duration* component. In this situation, one could compute the time series of deaths occurring after a heart attack, whatever the age of the patients, and study the association[9] between this time series and one or several time series of environmental indicators, such as income per capita or the percentage of unemployed.

Finally, suppose that a new anticoagulant drug has an impact on tertiary prevention if delivered in the days after a heart attack. Some of the patients in the sample have had their infarct at a time when the drug was still unknown. Others, on the other hand, were born later and therefore had the opportunity to receive the drug.[10] In this case, another temporal variable—the year of birth of the patient—must be taken into account: For a given age at onset, some patients were born too long ago to receive the drug and others on the contrary benefited from the discovery. The *birth cohort* can be taken in this case as the temporal factor of interest; it is in fact a proxy for a dummy variable (such as 0 and 1) indicating if one has or has not taken the drug.[11]

Since the seminal work of Whelpton (1954), Henry (1959, 1966a and b), and Ryder (1965), the cohort approach has become a common way of structuring demographic data. As Ryder (1965, p. 845) has stressed, members of a birth cohort share "a common historical location" as they have lived through similar experiences. Therefore, in Ryder's terms, "each cohort has a distinctive composition and character reflecting the circumstances of its unique origination and history; [. . .] the community of date equips each cohort with its own expanse of time, its own style, and its own truth." For example, wars or epidemics may leave a negative imprint on the future health of individuals who have experienced these hardships at a similar age, or unfavorable health conditions early in life might leave the strongest alive and lead to reduced cohort mortality later in life. These and other possible cohort effects are discussed in Hobcraft *et al.* (1982). Cohort effects seem to be less common in the field of fertility, where period effects predominate (Bhrolchain, 1993). Therefore, as Calot (1993) has pointed out, even if the cohort approach might seem more natural than the period approach, in the sense that one is following a group of individuals according to the *arrow of time*, this does not mean that cohorts are necessarily a better temporal frame of reference than periods. The choice between a period or cohort approach thus depends on the phenomenon one is studying and on the questions raised.

Consider now the risk of dying at a certain age x. This risk is age dependent, as mortality varies greatly according to age-related biologic factors. It is also period dependent, as the conditions of the period, in the social, economic, and public health fields for instance, have an impact on mortality. It is finally cohort dependent, as the similar history of the members of the cohorts in such matters as vaccination, availability of medicinal drugs, smoking habits, and so forth have an impact on their eventual mortality. One can therefore consider that this risk of dying is subjected to *age effects*, *period effects*, *and cohort effects*, and we might then wish to try to disentangle these effects. In a causal perspective, this means incorporating the various underlying causes into an explanatory mechanism or theory and then applying a suitable causal model such as covariance structure analysis.[12] Statistical problems can crop up however in *APC (age-period-cohort) models* if we want to use age, period, and cohort as proxies of the latent unknown causes, as the three variables are not linearly independent: Knowing two of these variables yields the third by simple addition or subtraction (see Chapter 18).

Generally speaking, cohort analysis refers to what was called above a repeated cross-sectional design, in the sense that the experience of a *group* of persons—and not individual trajectories—is followed through time. As Ryder (1965) has noted, "the cohort record, as macro-biography, is the aggregate analogue of the individual life history." This is typically the case when

[8] Such as the collapse of the Soviet Union and its impact on the quality of life and health of the Russian population.

[9] Such as the Granger co-integration method.

[10] The same holds for the introduction of techniques such as coronarography.

[11] Care must be taken however to distinguish between *actual* (or real) cohorts and *synthetic* (hypothetical or fictitious) cohorts, which rely solely on the data of a period.

[12] Such as LISREL or Mplus.

data are provided by a vital registration system: Individual records are usually not available because the data are grouped over individuals and time, but one may study, for example, the nuptiality or mortality of birth cohorts over time if marriages or deaths are classified each year by age and year of birth. French-speaking demographers therefore make the useful distinction between *données continues* (aggregate longitudinal data) and *données suivies* (individual longitudinal data); see for instance Henry (1959).

If *individual* longitudinal data are available from a survey or a register (such as a cancer register), for example, it is always possible to group the data by cohort according to a specific event origin, for example by birth cohort. Computation is made easier in this case but we lose the valuable information on individual differences *within* the cohort, that is on *intracohort heterogeneity*. The converse is obviously not possible: If one has only aggregate data, it is impossible to break the latter down into individual life courses. Greater effort should therefore be made to collect individual histories (life courses) rather than aggregate cohort data; this point was already stressed more than 40 years ago by Louis Henry (1959). It is thus strongly recommended to use the full information on individual life histories, instead of aggregating the data by cohort, and to apply *event history analysis* (such as the Cox hazard models) to the data on individual trajectories. This is possible thanks to the availability of adequate computer software. Cohorts can be distinguished in this approach by including in the statistical model the year of birth (or the calendar period defining the cohort) among the individual explanatory variables, similarly to such categoric variables as race or place of birth. One could also conduct the analysis at a double level, that of the individual and that of the cohort. *Multilevel* statistical methods (see Chapter 24) can be used for this purpose (Snijders and Bosker, 1999).

II. THE LONGITUDINAL APPROACH: DESCRIPTIVE MEASURES

1. Fertility

In the field of *fertility*, longitudinal data for individuals are very often collected by surveys[13] incorporating retrospective questions on the number of *children born alive* and on the date of birth of the children or the age of the mother or father at birth of the child. Most surveys collect data for women, although male

TABLE 10–1 Number and proportion of women according to parity

Indicator	Parity					
	0	1	2	3	. . . *i*	*k*
No. women	P_0	P_1	P_2	P_3	P_i	P_k
Proportion of women:	p_0	p_1	p_2	p_3	p_i	P_k

fertility can be studied, too. As the number of children depends, among others, on the age of the woman, young women usually having borne fewer children than older women because of a shorter length of exposure time, fertility measures are usually computed for women of the same age. If one takes women aged 50 and over for example, *completed fertility* or *ultimate family size* is obtained in the sense that these women have terminated their fertility history. The *quantum* or *intensity* of fertility for each female in this group is simply the total number of children born during her reproductive life, while her *tempo* of fertility is the succession of ages at which she has given birth to these children. Moreover, the difference between dates of birth of two successive children yields the length of the *birth interval* between these two events. If a woman has borne no child during her reproductive lifetime, she is defined as infertile. *Infertility* can be either due to biologic reasons (it is then called *infecundity* or *sterility*) or to the couple's decision to avoid childbearing (see the chapters in Section II on the biologic and social determinants of fertility, including Chapter 34 by Henri Leridon). A fecund woman is therefore not necessarily fertile,[14] but all fertile women were fecund at least until the time of their last child.

Some simple summary measures of the quantum (or intensity) and tempo of fertility can now be computed. A more detailed presentation and an example are given later on in this volume, in Chapter 12. A first descriptive measure is the distribution of the population of females according to the number of children they have borne. Write P_i for the number of women having borne *i* children during their life. One can then draw Table 10–1, *k* being the maximum number of children born.

Women having had *i* children during their lives are said to be of total *parity i*. To bear *i* children during their lifetime, they have had children of respectively birth *order*[15] 1, 2, . . . *i*. Each proportion p_i is simply obtained by taking the ratio $P_i/\Sigma_i P_i$. *The average number*

[13] Censuses can collect information on fertility using retrospective questions.

[14] Be careful! The French term for *fertility* is *fécondité*, and for *fecundity* it is *fertilité*. This confusion is unfortunate and must be kept in mind when reading papers in French.

[15] Multiple births are given different birth orders.

of children born, a measure of average intensity often used in fertility research, is then equal to

$$\Sigma_i iP_i/\Sigma_i P_i \text{ or } \Sigma_i ip_i, \text{ as } \Sigma_i p_i = 1,$$

with the sum extending from 0 to k. The number p_0 is equal to the proportion of infertile or childless women, as defined above in this section. One can also compute the variance or the standard deviation of the fertility distribution—measures of *dispersion* of total parity around the mean. Other descriptive measures such as the median fertility and quartile deviation can also be obtained.

One can also compute an *average age at childbearing*, a summary measure of the tempo of fertility, by taking the average of all ages of mothers at birth of their children.[16] Similarly, one can compute the average age of mothers at the birth of order one, of order two, and so forth. Once again, measures of dispersion around the mean may be computed. One may also compute the *probability* for females of a given age x to bear a child during a certain period, for example 1 year between ages x and $x + 1$, yielding a schedule of fertility probabilities by age. Other tempo measures can be obtained by taking the average of the *birth* intervals[17] between birth orders one and two, between birth orders two and three, and so forth. These measures and others are developed in Chapter 12, where examples are given of the various measures of quantum and tempo of fertility.

Up to now, all women past their childbearing age have been considered. Similar measures can be computed for women who are still in their reproductive period, once again controlling for their age. For example, consider females age 30. One can compute for this group the average number of children born up to this age and the average age at childbearing, keeping in mind that the data are *censored* by the date of the survey (i.e., these women can still have other children in the future after age 30). Birth intervals can also be computed between births of successive orders. As the data are censored, *open birth intervals* might however be closed in the future if another birth occurs. As intervals *straddling* the survey date are on average longer (for the same birth orders) than those intervals that are already closed, care must be taken if average birth intervals are computed in this way. These and other problems related to censored data are examined in Chapter 12.

For the open age group 50+, it must be stressed that fertility measures are usually not computed for the whole group, even if the survey or census addresses women of all ages.[18] As fertility has most probably varied in the past, it makes sense to take account of the various ages (i.e., of the various *birth cohorts*) involved. For example, one may consider separately all women aged 50 to 55, 55 to 60, and so forth, and compute for each of these age groups the descriptive measures developed above. For example, if the survey is held at the end of year 1995, fertility measures could be computed separately for women born in 1941 to 1945, in 1936 to 1940, and so on. The choice of the age interval depends on on the precision with which one wants to study past fertility and on the number of females in the sample. If few women are involved, the data must be grouped together to reduce random *noise*.

At high ages, e.g., above 80, the number of women is usually very low and these women might be selected[19] for their higher (or lower) fertility. For example, women having many children are more exposed to maternal mortality. In this case, women surviving at the time of the interview will have had, on average, a lower fertility at a given age than those who have died. Migration can also be a selection factor: If women with low parities tend to migrate (i.e., leave the country) more than women of higher parities, the sample will be biased in favor of the latter. In general, if the phenomenon studied is not *independent* of the processes that exclude individuals from observation, retrospective questions will lead to biased response, as only those *surviving* these processes (mortality, out-migration) will be interviewed. In the case of fertility, it is therefore possibly preferable to avoid dealing with the higher age groups and to restrict the analysis to women younger than 70 years. The rational choice of a cut-off age must consider the factors discussed previously.

Age is obviously not the only factor having an impact on fertility. The probability for a woman of giving birth during a certain period, a year for example, is dependent on other factors such as the number of previous births, her marital or partnership status, her economic condition (activity, household income), use or nonuse of contraception, her values and those of her partner, and so forth. More and more surveys therefore include questions on the current and past conditions and behaviors of the spouses or partners. In this case, fertility can be analyzed by taking account of the women in each of the various categories

[16] If data are aggregated by age, one can determine at each age x during the reproductive period the *age-specific fertility rate* (ASFR) at age x by dividing the number of births at that age by the average number of women of the same age. This yields a fertility schedule of *rates* by age of mother; see Chapter 12.

[17] Usually excluding multiple births in this case.

[18] Very often, fertility surveys only consider females aged 15 to 50 or 55. In this case, past trends in completed fertility cannot be observed.

[19] As usual, the term *selection* is not used here in its Darwinian meaning.

considered. For example, the probability of giving birth at age 30 can be computed for women who are married and economically active, who already have two children, who are practicing Roman Catholics, and so forth. The population of females is in this case classified into the various possible categories taking into account the variables at hand, and fertility measures of quantum and tempo can be computed for each of these categories and compared between categories. As usual in data analysis, *intragroup* variance can be compared to *intergroup* variance, and the relevance of the various categories can thus be assessed. This is possible because individual data are used rather than just group means.

As the cells in the multidimensional table resulting from this decomposition often contain only a small number of women and results are therefore subjected to a lot of random noise, another procedure is to have recourse to statistical methods of *event history analysis.* In this case, the probability of transition from one parity state to another is related to the various determinants of fertility (or covariates) as contained in the interview questionnaire, by way of some specific statistical function. Following the path-breaking work of Cox, it is now possible to estimate the impact of these determinants using exponential statistical models with time-dependent covariates. These covariates can be quantitative or qualitative, and computer programs (such as TDA, Transition Data Analysis) can estimate the parameters of the regression equation. Blossfeld and Rohwer (1995) present a good overview of this approach.

Up to now, only individual data on fertility have been considered. The advantage of this approach is that one has information on the individual's fertility behavior during her lifetime and on the possible determinants of this behavior drawn from retrospective surveys concerning the individual and, in some cases, the context or environment in which she has lived. Data on fertility are also recorded in all the developed countries by the *vital registration system.* Usually these data are aggregated and made available in tables or on magnetic or optical support. Longitudinal fertility measures can be computed from this source too, keeping in mind however that individual data are usually not available, for privacy reasons, and few data on the characteristics of the individual and a fortiori on the possible determinants of fertility are collected in the vital registration system. Chapter 12 discusses this type of data and shows how longitudinal summary measures of quantum and tempo of fertility can nevertheless be computed for aggregates such as birth cohorts.

The situation is drastically improved if data from the vital registration system are linked to the census data through a *personal identification number.* In this case, the characteristics of the individual are known, at least those recorded on the census form. Individual census forms can also be linked from one census to another, yielding individual life histories *sampled* at each census time. The problem however is the temporal *sampling* frequency: intercensal intervals are usually long (around 10 years) and, therefore, changes that may have occurred during the intercensal periods remain unknown. Nevertheless, the information thus obtained can be treated as longitudinal individual data, even if the actual names of the individuals remain unknown for privacy reasons.[20] The approaches developed previously for individual data can be used in this case.

2. Mortality

Longitudinal individual data in *mortality* research can be provided by *prospective surveys.* The intensity of mortality being always equal to one (everybody dies!), mortality research focuses instead on the tempo of mortality (i.e., on the age at which one dies). The average of ages at death is called the *expectation of life at birth* for the persons concerned (see Chapter 11). The variance of ages at death can easily be obtained from the individual ages at death. Aggregating individual experiences, *probabilities* of dying can also be computed by age and sex, to obtain the age- and sex-specific mortality schedules. Medical *causes of death* are also examined, and studied over time in a longitudinal perspective (see Volume 2). Finally, the characteristics of the deceased can be used to explain their age and cause of death, using various types of regression techniques (see e.g., Kahn and Sempos, 1989).

Contrary to fertility research, *retrospective* questions are not very useful in the field of mortality: The dead do not adequately answer the questionnaire! Relatives of the deceased person can however be asked retrospective questions, and this approach is used in *case-control studies*, for example. However, information thus obtained, even from persons close to the deceased, are often not very reliable.[21] For example, few widows know the actual occupation (or income) of their former husband or his sexual behavior if he has died from venereal disease! However, retrospective data on child

[20] Who is interested in the name of the person anyway? Good research and privacy are therefore compatible if each individual can be followed through time thanks to a personal code or identification number.

[21] Even a question on deaths occurring in the household during the past 12 months does not yield very good information, as surveys conducted in the developing countries have shown.

mortality have been collected in many surveys in the developing countries, such as the World Fertility Survey (WFS) and the Demographic and Health Survey (DHS) by asking the mother questions on the survival or death of her children and their age at or date of death (see the chapter on data collection in Volume 4, Chapter 121). Questions on the survival of siblings are also asked to evaluate adult mortality.

In demography, no large *prospective survey* on mortality has ever been carried out, contrary to the field of epidemiology. This source requires large samples and the cost is indeed enormous. Prospective surveys are also hampered by *loss to follow-up* by people eventually dropping out of the survey and therefore reducing the sample size and possibly biasing its composition, as those who drop out do not necessarily have the same characteristics as those who remain in the sample. Longitudinal research with individual data in mortality can be based on the *record linkage* of census and vital registration data. For example, in the case of Norway, a longitudinal study (Wunsch *et al*.,1996) has been conducted on males and females who were followed from one census to the next since 1960. The fact of dying or surviving at each age has been related to the present and past characteristics of the persons deceased or alive using logit regression techniques—these characteristics being derived from the individual census forms.

Contrary to prospective surveys, very large samples or even the whole population can thus be followed. On the other hand, only the characteristics recorded at census are available, although the situation can be further improved if the individual data provided by other registers (e.g., cancer or cardiovascular registers) or other sources (social security, for example) are also linked to the census and vital registration data. The use of the same personal identification number in all these sources makes linkage at least theoretically possible. Record linkage is slowly becoming one of the major sources of individual longitudinal mortality research in developed countries. The method requires an excellent information system and a good coverage; thus, it is not suitable for developing countries. However, even in developed countries, the method has been criticized by those who want to protect individual privacy, although the use of personal codes (no names and addresses) essentially guarantees the confidentiality of information.

If record linkage is not available, one must resort to aggregate mortality data provided by the vital registration system. In this case, longitudinal research can only be performed on aggregates. Mortality rates or probabilities can be computed by age and sex and birth cohort. Cohort age- and sex-specific mortality schedules can therefore be derived, and *cohort life tables*[22] can be set up (see Chapter 11). Intercohort differences in the timing of death and possibly in the medical causes of death can then be observed longitudinally, but intracohort differences are unknown as individual data are not available. Furthermore, care must be taken to achieve comparable definitions of medical cause of death over time if the latter are used (see Vallin and Meslé, 1988).

3. Migration

Finally, in *migration* research (see Chapters 13 and 22), longitudinal analysis can also rely on *retrospective* surveys for all types of mobility except international emigration; in the latter case, the emigrants have left the country and cannot be interviewed. Otherwise, the approach is quite similar to that used in fertility studies. One can ask the persons sampled when they have moved during their lifetime, where they came from, and where they went. Measures of intensity (number of moves) and of tempo (age at migration) can be obtained. Furthermore, the interval between moves can be computed, and the geographical trajectory assessed, as migration involves *time* and *space.* One can moreover collect background information on possible determinants of migration, and relate these possible determinants to the moves. For example, Courgeau and Lelièvre (1992) have used data from a triple biographic survey on migration, occupation, and family to study the interrelations between these factors. They have also developed statistical methods and computer programs to analyze this type of data set.[23] In fact this type of data can be modelled by parametric, semiparametric, and nonparametric *event history models*, according to the applicable assumptions. More methods and computer programs are being developed in this area, and individual longitudinal data analysis should take advantage of this relatively recent methodology. Individual data on migration can also be obtained by longitudinal prospective studies, although these are rare in this field.

If individual data on migration are not available, longitudinal cohort data can be analyzed in countries, such as Belgium, that have a *population register* incorporating a registration of changes of residence. This register can yield individual data if each person in the register can be followed over life by his/her personal identification number. Compared with retrospective surveys, however, the population register includes few

[22] For an example, see Mamelund and Borgan (1996).
[23] Courgeau and Lelièvre (1992).

covariates that can be used to explain the quantum and tempo of migration. However, the register covers the whole population and takes account of all migration flows occurring in the country during any period of time. Ideally, one should have both a population register for measuring flows and retrospective surveys for explanatory studies, possibly linking the two sources together. One could also link the population register data on moves to the characteristics of the movers obtained from the census forms, if both sources are linked. The advantage here is that both sources cover the whole population.

In all longitudinal studies discussed in previous sections, whatever the method, the main advantage is to locate events in *time*. The temporal sequence of events can indicate possible causal patterns, because causes always have to precede effects in time for reasons which will not be discussed here (Wunsch, 1988). For example, the probability of marriage for males having started their occupational career in agriculture is much higher after leaving agriculture; on the other hand, their change of activity is not dependent on their marital status (Courgeau and Lelièvre, 1992). The temporal order of the two events (leaving agriculture and marriage) is therefore highly important if one wants to explain the interrelation between the two events. To give another example, one observes that female activity and fertility are correlated, women with more children being less active than women with few children. This might be due to the fact that women with few or no children have more opportunities to look for jobs or to the fact that women who work want fewer children. A single-round survey giving only the *prevalence* of women according to parity and activity cannot discriminate between these two plausible explanations. One needs again the life history of women to determine the temporal order of the two variables. One might then see that a woman decides not to work once she has had a child or that her occupation leads to postponement of childbearing. In fact, the same woman can experience the two situations at different stages of her life.

In the field of longitudinal mortality research, as Duchêne and Wunsch have initially suggested (Caselli *et al*., 1989; Duchêne and Wunsch, 1991; Salhi *et al*., 1995), there are advantages in taking account of the order of events and the type of states one has experienced during one's lifetime, a state being defined by the various characteristics of an individual at a point in time, and finally of the time spent in each state. Such a *state–time–order transition model* could be used in other fields. Once again, individual longitudinal data are required in this case. Longitudinal individual data enable one to distinguish the time order of events and eliminate incorrect explanations. This cannot be achieved with cross-sectional or longitudinal aggregate data. Good demographic data collection and analysis should thus address individual life histories even if there is a price to pay; prospective or retrospective studies are required, and record linkage should be achieved.

Finally, longitudinal research can move beyond the individual and consider what happens to the individual's family, environment, and so forth. A family or a household is difficult to follow over time, but other procedures are being developed to link the individual to the group, such as using the concept of *contact circle* or having recourse to multilevel modelling (Lelièvre *et al*., 1998; Courgeau and Baccaïni, 1998; see also Chapter 24). The behavior of an individual also depends on the context in which he or she lives, and demographic explanations must take these multilevel strategies increasingly into account.

References

Bernard Paul-Marie and Lapointe Claude, 1991. *Mesures statistiques en épidémiologie*. Presses de l'Université du Québec, Sillery, 314 p.

Blossfeld Hans-Peter and Rohwer Götz, 1995. *Techniques of event history modeling*. Mahwah, Lawrence Erlbaum, 294 p.

Calot Gérard, 1993. Synthèse des taux et synthèse des quotients. *Population*, vol. 48(2), p. 405–418.

Caselli Graziella, Duchêne Josianne and Wunsch Guillaume, 1989. L'apport de la démographie à l'explication de la mortalité différentielle, *in:* Josianne Duchêne, Guillaume Wunsch and Éric Vilquin (eds.), *L'explication en sciences sociales: la recherche des causes en démographie*, p. 41–50. Brussels, CIACO/ARTEL, 478 p. (Chaire Quetelet, 1987).

Courgeau Daniel and Baccaïni Brigitte, 1998. Multilevel analysis in the social sciences, *in:* Daniel Courgeau (ed.), New methodological approaches in the social sciences, *Population*, *English Selection*, vol. 10(1), p. 39–71.

Courgeau Daniel and Lelièvre Èva, 1992. *Event history analysis in demography*, Oxford, Clarendon Press, 1992, 226 p.

Duchêne Josianne and Wunsch Guillaume, 1991. Life histories: a new approach to the analysis of mortality, *in:* Janina Jozwiak and Irena E. Kotowska (eds.), *Usefulness of demographic modelling*, p.111–120. Warsaw, Warsaw School of Economics, 212 p.

Gray Ronald, 1988. L'intérêt des méthodes de l'épidémiologie et des études *cas témoin* pour la mesure de la mortalité et de la morbidité, *in:* Jacques Vallin, Stan D'Souza and Alberto Palloni, *Mesure et analyse de la mortalité*, p. 63–84. Paris, INED, Presses Universitaires de France, 460 p. (Coll. Travaux et Documents, Cahier no. 119).

Henry Louis, 1959. D'un problème fondamental de l'analyse démographique. *Population*, vol. 14(1), p. 9–32.

Henry Louis, 1966a. Perturbations de la nuptialité résultant de la guerre 1914–1918, *Population* (2), p. 273–332.

Henry Louis, 1966b. Analyse et mesure des phénomènes démographiques par cohortes, *Population* (3), p. 465–482.

Hobcraft John, Menken Jane and Preston Samuel H., 1982. Age, period, and cohort effects in demography: a review. *Population Index*, vol. 48(1), p. 4–43.

Jones Russell A. 1996. *Research methods in the social and behavioural sciences*, 2nd Ed. Sunderland, Sinauer Associates, 369 p.

Kahn Harold A. and Sempos Christopher T., 1989. *Statistical methods in epidemiology*. New York, Oxford University Press, 212 p.

Lelièvre Èva, Bonvalet Catherine and Bry Xavier, 1998. Event history analysis of groups, *in:* Daniel Courgeau (ed.), New methodological approaches in the social sciences, *Population English Selection*, vol. 10(1), p. 11–37.

Mamelund Svenn-Erik and Borgan Jens-Kristian, 1996. *Cohort and Period Mortality in Norway 1846–1994*. Oslo, Statistics Norway, 236 p.

Ni Bhrolchain Maire, 1993. Histoire passée, indices synthétiques de fécondité du moment. *Population*, vol. 48(2), p. 427–431.

Ryder Norman B., 1965. The Cohort as a concept in the study of social change. *American Sociological Review*, vol. 30(6), p. 843–861.

Salhi Mohamed, Caselli Graziella, Duchêne Josianne, Egidi Viviana, Santini Antonio, Thiltgès Évelyne and Wunsch Guillaume, 1995. Assessing mortality differentials using life history: a method and applications, *in:* Alan Lopez, Graziella Caselli, and Tapani Valkonen (eds.), *Adult mortality in developed countries: from description to explanation*, p. 57–79. Oxford, Clarendon Press, 361 p. (International Studies in Demography).

Snijders Tom and Bosker Roel, 1999. *Multilevel analysis*. London, Sage, 266 p.

Vallin Jacques and Meslé, France, 1988. *Les causes de décès en France de 1925 à 1978*. Paris, INED, Presses Universitaires de France, 608 p. (Travaux et Documents, Cahier no. 115, with seven annexes published previously).

Whelpton Pascal, 1954. *Cohort fertility: native white women in the United States*. Princeton, Princeton University Press, 492 p.

Wunsch Guillaume, 1988. *Causal theory and causal modeling*. Leuven, Leuven University Press, 200 p.

Wunsch Guillaume, 1993. Cohort Analysis of Vital Processes, *in:* Donald J. Bogue, Eduardo E. Arriaga and Douglas L. Anderton (eds.), *Readings in population research methodology*, vol. 5, p. 12–38. Chicago, Social Development Center, 93 p.

Wunsch Guillaume, Duchêne Josianne, Thiltgès Évelyne and Salhi Mohamed, 1996. Socio-economic differences in mortality. A life course approach. *European Journal of Population*, vol. 12(2), p. 167–185.

CHAPTER

11

Cohort Life Table

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

As death is a *nonrenewable event* at the individual level, mortality can be directly subjected to the probabilistic risk measurement defined earlier (Chapter 8). This approach will enable us to follow, from its birth, the gradual extinction of a cohort whose initial population is reduced by death from one age to the next, until its total disappearance. Indeed, as long as humans have not discovered the secret of immortality, this total disappearance is inevitable. In demography, death is termed a *fatal event*. Probabilistic analysis thus has a simpler aim when applied to mortality than in its general applications: Here, we do not have to measure *final intensity*, which necessarily equals unity. It is sufficient to measure the *tempo*, in other words, the age at death, absent any competing risks. That is the purpose of life-table construction.

In this chapter, therefore, we will begin by spelling out the various practical alternatives for determining the age-specific *probabilities of dying* in a cohort. Next, we construct the life table, describing its main functions and their interrelationships. In the final section, we discuss the difficulties involved in computing mortality at the oldest ages and in closing the table, as well as the special treatment customarily reserved for infant mortality.

I. CALCULATING A COHORT'S PROBABILITIES OF DYING

Monitoring a birth cohort directly up to its complete extinction is no easy task. Imagine taking a sample of births observed in a given year and documenting the children's survival, year after year, until the last one dies . . . 100 or 120 years later! No demographer could, of course, survive such an endeavor. And any attempts to base a health care policy on such a long-term approach would very likely be soon discouraged. The notion is not, however, totally ridiculous. In fact, it underlies several intensive studies on selected populations that lack any reliable civil registration system. In these regions, demographers have set up *population observatories* to monitor, for the longest possible period, the population of a very specific geographic area starting from an initial census and from each year's new births.[1] The data gathered can at least be used to begin series of cohort-specific probabilities of dying, even if the probabilities are confined to the life interval between the initial births and the most recent observations. This method is thus chiefly of use for analyzing childhood mortality. To compute the probabilities directly would require two conditions, both of them highly improbable: Either the observed population must not lose any of its members through emigration, or a supplementary survey must be carried out to track the emigrants in their new localities. In theory, the data collected in such population laboratories should also allow the compilation of partial probability series for the cohorts born before the initial census (and hence an analysis of adult mortality) on the basis of the population enumerated at each age. There are several problems, however. First, the series would be cut off

[1] For more details, see Chapters 121 and 124 in Volume 4 of this Treatise (see also Das Gupta *et al.*, 1997).

or censored at both ends: before the initial census and after the most recent observation. Second, as a rule, the series are undermined by the unreliability of information on age in these populations without civil records; hence the risk of incorrectly identifying the series' starting-points.

Some retrospective surveys could also yield suitable data for the direct computation of cohort-specific probabilities of dying.[2] One particularly relevant source is the surveys aimed at reconstructing the complete maternity history of women. These surveys must collect, with sufficient accuracy, the two vital items of information on child survival: date of birth and date of death. Maternity surveys have been conducted in many countries in recent decades, not only in the developed countries to study the determinants of fertility, but also—and even most commonly—in a growing number of developing countries, for which this type of survey is often the only reliable way to measure fertility and infant and child mortality (before even analyzing their causes). Two large-scale international data-gathering operations have been carried out for these purposes: the World Fertility Survey (WFS) in the 1970s by the International Statistical Institute (ISI),[3] and the Demographic and Health Survey (DHS) by Macro International Inc. from the 1980s onward.[4] Again, this type of observation confines the analysis to child mortality. It also has two weak spots. First—and, no doubt, most important—the data gathered depend on the quality of responses by the women surveyed. Respondents may fail to mention certain events, either through forgetfulness or to avoid painful memories. In either case, this can have a major impact on infant and child mortality statistics, as neither the births nor the deaths will be declared. Second, even if these events are actually mentioned, there may be much imprecision about the dates of their occurrence, which will considerably impede the analysis. In addition to this critical data collection problem, there is a methodologic issue. Naturally such surveys can only serve to explore mortality among the offspring of women still alive and present in the survey locality at the time of the survey. We therefore assume that the mortality of the children of these women does not differ from that of the children of the women whom the survey was unable to cover, either because they too were deceased, or because they had left the locality where they resided at the time of the preparation of the sampling frame from which the surveyed sample was drawn. In other words, it is assumed—in all likelihood, wrongly—that there is no link between the children's mortality and their mothers' mortality or migration. Yet, this serious theoretical problem is less important than the observation problem.

In any event, it is wishful thinking to hope we can ever construct a complete series of probabilities of dying that would provide a full record of a cohort's mortality from its birth to its extinction, from data directly gathered on a longitudinal base. The task is practically beyond the reach of population laboratories and strictly excluded from the field of retrospective surveys on women's maternity history. Ultimately, the only way to arrive at such a result is to use the current statistics compiled from civil records and population censuses, in countries where these information sources—in particular, civil registration—have been properly maintained for at least a century.

1. Starting a Series of Cohort-specific Probabilities of Dying from a Retrospective Survey

The data gathered from retrospective surveys on the maternity history of a sample of women theoretically provide—subject to the reservations mentioned previously—an ideal starting point for calculating the cohort-specific probabilities of dying. However, the data gathered must be processed to the necessary level of detail, which is seldom the case.

Let us take, for example, the Egyptian DHS survey of 1992 (El-Zanaty *et al*., 1993). As in all DHS surveys, module 2 of the questionnaire, *reproduction*, asks each woman to provide a complete list of the children to whom she has given birth; for each child, she is asked to indicate the sex, month, and year of birth; survival at time of survey; and, if the child has died, age at death. Using such information, we can obviously group the children together by birth year and, for each cohort thus constructed, obtain the number of deaths occurring in the cohort at successive ages, regardless of whether the children have left the mother's residence or not. For these children, we are operating strictly within the framework of a closed population (i.e., without migration). The annual probability of dying at age x, ${}_1q_x$, can be obtained directly by dividing the D_x deaths observed at the age reached during the year x to the B births of the cohort minus the deaths occurring up to the exact age x:

[2] See Dominique Tabutin's Chapter 121 in Volume 4.

[3] Between 1973 and 1981, 44 national surveys were carried out under the World Fertility Survey program: 13 in Africa, 16 in Asia/Pacific, 14 in Latin America, and one in Europe (ISI, 1986; Cleland and Scott, 1987).

[4] From 1986 to now June, 2002, 153 surveys were carried out in 68 countries under the Demographic and Health Survey program: six in Central Asia, 32 in Latin America and the Caribbean, 25 in South/Southeast Asia, 20 in North Africa/West Asia/Europe, and 70 in Sub-Saharan Africa (http://www.measuredhs.com).

$$_{1}q_{x}=\frac{D_{x}}{B-D_{(0,x)}}$$

Unfortunately, the data from this type of survey are rarely processed with the necessary level of detail. In particular, none of the reports published on DHS data supplies the table that would allow a computation of these annual probabilities. In fact, these reports only give the total deaths observed between birth and the survey date for each birth cohort. For Egypt, for example, we have the following table (Table 11–1), whose elements are plotted on a Lexis diagram in Figure 11–1.

TABLE 11–1 Births and deaths by cohort, from 1992 Egyptian demographic and health survey

Year of birth	Births	Death before survey date	Survivors at time of survey
1992	1,511	51	1,459
1991	1,707	120	1,587
1990	1,694	125	1,568
1989	1,774	149	1,626
1988	1,964	158	1,806
1987	1,884	152	1,732
1986	2,008	244	1,763
1985	1,751	239	1,512
1984	1,766	248	1,518
1983	1,902	265	1,637
1988–1992	8,649	603	8,046
1983–1987	9,312	1,149	8,163
1978–1982	8,376	1,288	7,088
1973–1977	6,070	1,240	4,830
before 1973	5,802	1,521	4,281
Total	38,209	5,801	32,408

Note: This table, reproduced from the original as published, includes a few minor discrepancies between certain sums and the summed elements.

Source: El Zanaty Fatma H., Sayed Hussein, Zaky Hassan H., and Way Ann A., 1993. *Egypt Demographic and Health Survey 1992*. Cairo, National Population Council and Calverton, Macro International Inc., 317 p.

As the survey was conducted in November 1992, 1992 is a year of incomplete observation, which is why the number of births (1511) is considerably smaller than those of preceding years. Likewise, the Lexis trapezoids containing the deaths are missing a month on the right. That is not the main problem, however. The missing element is the distribution by years of age (by Lexis parallelograms) within each trapezoid. To obtain the distribution would require a supplementary processing of the survey.

The authors of the report have avoided doing so, preferring to compute probabilities of dying by period (and not by cohort): We will comment on these later when discussing cross-sectional analysis and the use of the fictitious cohort as a device (Chapter 14).

Reprocessing the basic data of the Egyptian survey,[5] we obtain the following age distribution of deaths occurring among the 6070 children born in the 5-year period from 1973 to 1977 (Table 11–2). By grouping births in 5-year periods, we reduce the stochastic effect inherent in small numbers; by examining this group of years of birth—which is fairly remote from the survey date—we can track the mortality of these children over a 15-year period.

The probabilities are very simple to calculate in these conditions. To obtain the annual probability at age 0, $_{1}q_{0}$, we divide the deaths observed between 0 and 1 years of age (first line of second column of Table 11–2) by births (first line of third column).

$$_{1}q_{0}=\frac{D_{(0,1)}}{B}=\frac{796}{6,070}=0.13114$$

We then subtract the deaths at less than 1 year from births to obtain the number of cohort survivors at the first birthday (second line of third column); we divide the deaths between ages 1 and 2 by that survivor figure to obtain the annual probability of dying, $_{1}q_{1}$, at age 1, and so on, until the bottom of the table.

$$_{1}q_{1}=\frac{D_{(1,2)}}{B-D_{(0,1)}}=\frac{207}{6,070-796}=0.03925$$

These probabilities are annual in age terms, which is why, by convention, the index 1 is placed to the left of the symbol q. However, the probabilities represent the average mortality of a 5-year group of birth cohorts.

We can also determine probabilities with different age dimensions. For example, to measure the mortality of children older than 1 year, it is common to calculate the probability $_{4}q_{1}$, between 1 and 5 years, and the probability $_{5}q_{5}$, between 5 and 10 years:

$$_{4}q_{1}=\frac{D_{(1,5)}}{B-D_{(0,1)}}=\frac{375}{6,070-796}=0.07110$$

and

$$_{5}q_{5}=\frac{D_{(5,10)}}{B-D_{(0,5)}}=\frac{45}{6,070-796-375}=0.00919$$

The procedure is simple, but the main drawback is that it is confined to a very narrow portion of the life of the cohorts studied.

[5] We thank Zaia Bedidi for having carried out this special processing of the Egyptian DHS data.

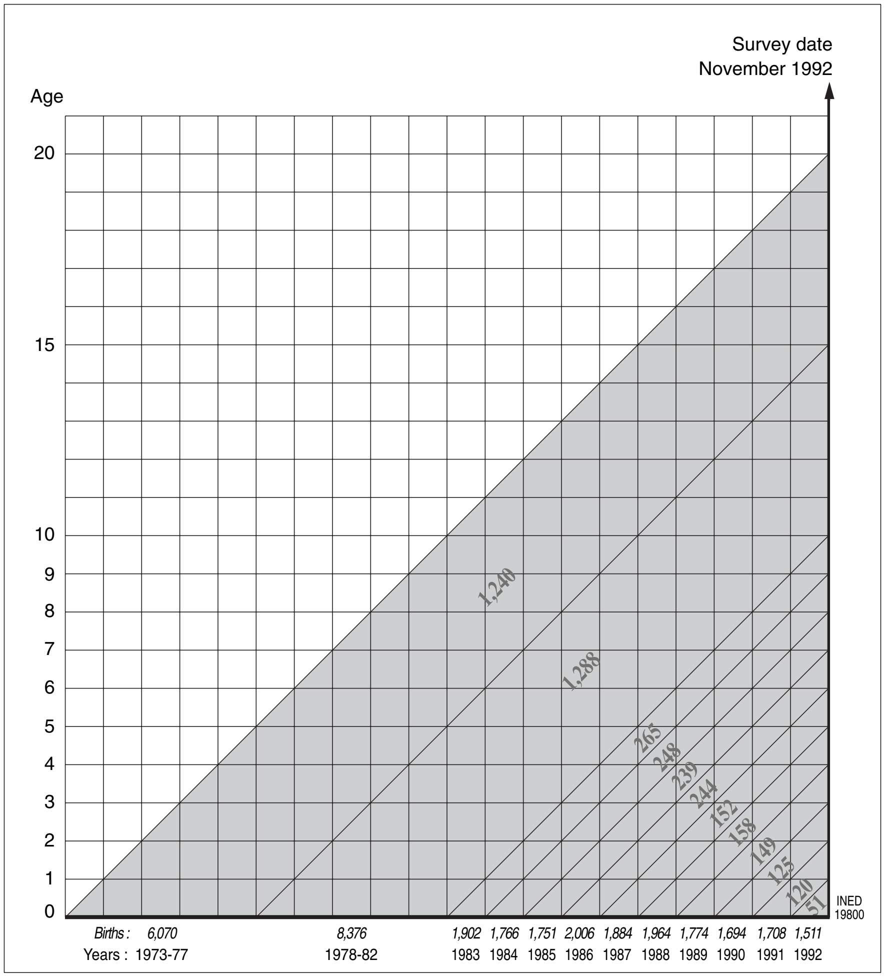


FIGURE 11–1 Representation on a Lexis diagram of child-survival data published in 1992 Egyptian Demographic and Health Survey report.

2. Calculating a Complete Series of Cohort-specific Probabilities of Dying from Current Statistics

To extend this calculation to all ages of life, we must use the long data series available in certain countries, with a sufficiently long tradition of current civil registration statistics and censuses. Let us go back to the French data from which we computed the series of age-specific proportions from 1806 to 1995, used in Chapters 5 and 6 and updated to 1997 (Vallin and Meslé, 2001). This data set is actually divided into two subsets, 1806 to 1898 and 1899 to 1995. For the second—thanks to the data published periodically by INSEE (the French national statistical institute) since World War II and a reconstruction of series for earlier years (Vallin, 1973)—we now have complete annual series of births, population by year of age on January 1, and deaths by year of age and year of birth (the basic triangles of the Lexis diagram). For the first subset, which covers the 19th century, the series are a historical reconstruction from far less detailed data that led to the estimation of annual series of probabilities of dying by year of age (Meslé and Vallin, 1989).

The second data subset (1899–1995) offers an example of favorable conditions for the construction of complete series of cohort-specific probabilities of dying from current statistics. As we shall see, it is possible to calculate these probabilities by making acceptable hypotheses on migrations and their relations with mortality.

a. Classic Probability, Between Two Birthdays

Let B_t be the births occurring in year t, $P_{t,(x,x+1)}$ the population of age $(x, x+1)$ on January 1, t, $D^i_{t,(x,x+1)}$ the deaths in year t at age $(x, x+1)$ in cohort i, and $D^{i-1}_{t,(x,x+1)}$ the deaths occurring in the same year t at

TABLE 11–2 Age distribution of deaths occurring among children born in 1973–1977 according to 1992 Egyptian demographic and health survey and calculation of probabilities of dying

Exact age x	Deaths between ages x and $x+1$	Remaining population at age x	Probability of dying between ages x and $x+1$
0	796	6070	0.1311
1	207	5274	0.0392
2	96	5067	0.0189
3	52	4971	0.0105
4	20	4919	0.0041
5	11	4899	0.0022
6	16	4888	0.0033
7	8	4872	0.0016
8	7	4864	0.0014
9	3	4857	0.0006
10	3	4854	0.0006
11	3	4851	0.0006
12	4	4848	0.0008
13	1	4844	0.0002
14	4	4843	0.0008
15		4839	

the same age $(x, x+1)$ in the preceding cohort $(i-1)$.

With no Migration

Let us start with the B_t births of cohort i born in year t. Absent any migrations, the probability of dying at age 0 would be equal to:

$$ {}_1q_0 = \frac{D^i_{t,(0,1)} + D^i_{(t+1),(0,1)}}{B_t} = \frac{D^i_{(t,t+1),(0,1)}}{B_t} $$

where i is the cohort born in year t. More simply, if we abandon the indexing by calendar year, which is redundant with the indexing by cohort:

$$ {}_1q_0 = \frac{D^i_{0,1}}{B^i} $$

Then, subtracting from the births the deaths occurring in the cohort before age 1, we would obtain the denominator by which to divide the deaths occurring in the same cohort between ages 1 and 2 to determine the probability of dying at age 1:

$$ {}_1q_1 = \frac{D^i_{1,2}}{B^i - D^i_{0,1}} $$

And finally, by generalizing, we could determine the probability of dying at any age $x > 0$:

$$ {}_1q_x = \frac{D^i_{x,x+1}}{B^i - \sum_1^x D^i_{x-1,x}} = \frac{D^i_{x,x+1}}{B^i - D^i_{0,x}} $$

With Migrations: Problems and Solutions

However, like most real populations, the French population is not *closed*. It experiences inflows and outflows due to migration. The application of the formulas above in this section would bias the mortality calculation for two reasons: (1) deaths of immigrants, who do not belong to the initial birth cohort whose mortality we seek to measure, are included in the deaths that form the numerator of the probabilities; (2) deaths of emigrants, who, unlike immigrants, do belong to the initial cohort studied, are not counted. If we use the general formula given for x, this bias worsens with age because the greater the distance from the birth year, the wider the interval disturbed by migrations. To analyze populations open to migrations, we must therefore introduce a correction into the initial formulas.

For simplicity's sake, let us consider the net migration (immigrants minus emigrants) and use $M^i_{x,x+1}$ to denote this balance for cohort i at age $(x, x+1)$. We now need to subtract the deaths of these net immigrants from the deaths forming the numerator. Unfortunately, there are no statistics available on the subject. We must therefore resort to a *first hypothesis*. Let us admit, again for the sake of simplicity, that the mortality of migrants is identical—at least in a first approximation—to that of the entire population studied. We can then estimate the number of deaths to be subtracted from the numerator by assigning to net immigrants the probability of dying that we are seeking to determine. However, there is a second practical problem: The migration inflows and outflows are spread over time between the two birthdays concerned. If they had all occurred at the exact age x, we would obviously have had to assume that all the net migrants were exposed to the risk of dying in the population studied for the entire duration of age $(x, x+1)$, and the deaths to be subtracted from the numerator for age $(x, x+1)$ would be estimated by $(M^i_{x,x+1} \cdot {}_1q_x)$. Conversely, if they had all occurred at the exact age $(x+1)$, none should be subtracted. In most cases, it is fairly safe to assume that, on average, these movements occur midway on the time line, and we can thus subtract the average of the two quantities—for example, $(0.5 \cdot M^i_{x,x+1} \cdot {}_1q_x)$, from the numerator. Combining this *second hypothesis* with the first, we can write the probability of dying at age 0 as follows:

$$ {}_1q_0 = \frac{D^i_{0,1} - 0.5 \cdot M^i_{0,1} \cdot {}_1q_0}{B^i} = \frac{D^i_{0,1}}{B^i + 0.5 \cdot M^i_{0,1}} $$

and, more generally, for the other ages x:

$$ {}_1q_x = \frac{D^i_{x,x+1}}{B^i - (D^i_{0,x} - M^i_{0,x}) + 0.5 \cdot M^i_{x,x+1}} $$

Calculation without Migration Data

The third problem is that France keeps no reliable record of migration movements comparable to its civil registration of births and deaths, so we do not know the $M^i_{x,x+1}$ values. On the other hand, France does publish estimates of the population by year of age on January 1 of each year. We can therefore attempt to estimate the population at birthday x from this far more recent information rather than from the initial number of births. Let $M^i_{t,(x,x+1)}$ be net immigrants of year t at age (x, $x + 1$) in cohort i (first triangle of Lexis diagram) and $M^i_{(t+1),(x,x+1)}$ the net immigrants of the following year (t + 1) at the same age (x, $x + 1$) in the same cohort i (second Lexis triangle). We can replace the previous formula by:

$$ {}_1q_x = \frac{D^i_{x,x+1}}{P^i_{t,(x,x+1)} + D^i_{t,(x,x+1)} - M^i_{t,(x,x+1)} + 0.5 \cdot M^i_{x,x+1}} $$

In theory, this equation assumes an even more detailed knowledge of migrations because it refers to a dual classification of migrations by age and year of birth. In fact, however, it shows that we can calculate the probabilities by eliminating all explicit references to migrations. For this, we need only accept a *third hypothesis*, which is simple and in most cases close to reality—namely, that the number of net immigrants in a cohort at a given age is evenly distributed between the two triangles of the Lexis diagram concerned. This gives:

$$ M^i_{t,(x,x+1)} = 0.5 \cdot M^i_{x,x+1} $$

and, in final form:

$$ {}_1q_x = \frac{D^i_{x,x+1}}{P^i_{t,(x,x+1)} + D^i_{t,(x,x+1)}} $$

We can therefore easily disregard the migration issue if we have annual statistics for (1) deaths by year of age and year of birth and (2) populations by age on January 1. However, we must bear in mind that the estimates of population by year of age at January 1 taken here for granted rely on intercensal reconstructions. These, in turn, assume, if not full data on migration flows (rarely available), at least an estimate of intercensal flows and a model for determining their annual distribution (see Chapter 13).

More concretely, let us go back to the example taken from the French data set and used in Table 6–1 of Chapter 6 to compute the male mortality rate at age 74 in 1994. Let us now determine the probability of dying of the 1920 cohort, which turned 74 in 1994. Obviously, we need to use the items required for calculating type 3 rates, as in Chapter 6. Figure 11–2 plots their positions on the Lexis diagram (parallelogram *ABCD*).

The probability is equal to:

$$ {}_1q_{74} = \frac{D^{1920}_{74,75}}{P^{1920}_{1995,(74,75)} + D^{1920}_{(1994),(74,75)}} = \frac{4,275 + 4,084}{202,140 + 4,275} = 0.04050 $$

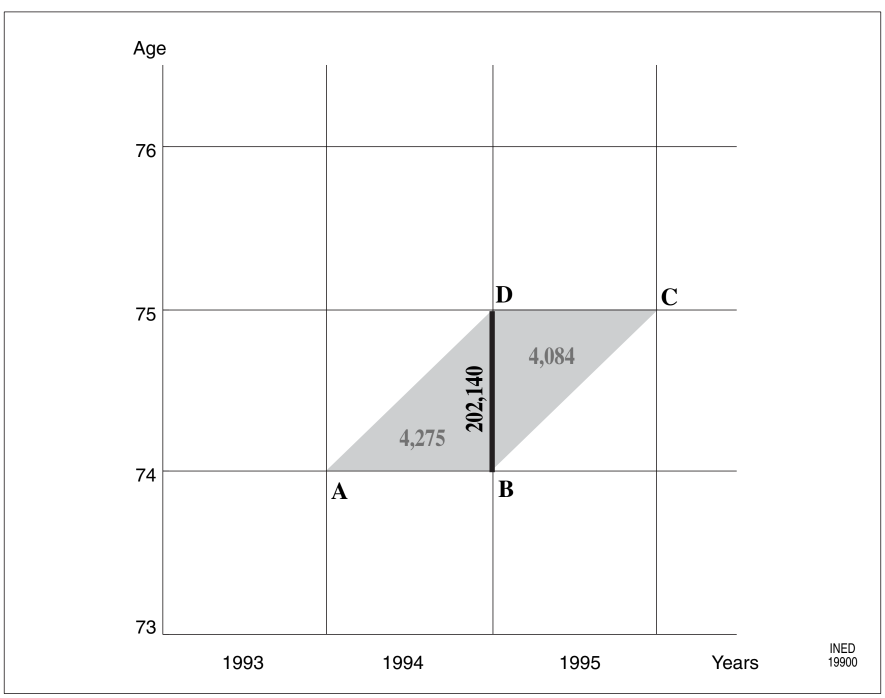


FIGURE 11–2 Representation on a Lexis diagram of items needed to compute probability of dying at age 74 for 1920 French male cohort.

where the year shown in the index corresponds to January 1 of that year.

Rates and Probabilities

It will be recalled that the type 3 mortality rate, computed with these same data in Chapter 6, was 0.04135. The difference is small. This is generally the case as long as we are dealing with relatively low death risks.

However, by definition, the difference between the rate and the probability increases with the degree of risk.[6] We can see this in summary form in Table 11–3, which compares the theoretical results that we would obtain in a closed population and a uniform distribution of deaths between two birthdays. The difference is minimal when the probability is below 10/1000. But it rises very quickly. If the probability reaches 100/1000, the rate is 5% higher than the probability. For a probability of 500/1000, the rate exceeds the probability by one third; at the top end of the scale, whereas the probability cannot—by definition—exceed unity, the rate peaks at twice that value.

In a population such as the French, the age-specific probability of dying rarely reaches values high enough to diverge substantially from the rates determined with the same data. In the male cohort born in 1899, however, we note sizable differences for infant mortality and for mortality at the oldest ages (Figure 11–3). In this cohort, the probability of dying before age 1 is 185/1000, and the rate exceeds the probability by nearly 15%. The difference stays below 3% until

[6] Using the rate-to-probability conversion formula given in Chapter 8, we effectively obtain:

$$_nq = \frac{n \cdot {}_nm}{1+\frac{n}{2}\cdot {}_nm}$$

in other words:

$$\frac{_nm}{_nq} = \frac{_nm}{n \cdot {}_nm}\cdot\left(1+\frac{n}{2}\cdot {}_nm\right) = \frac{1+\frac{n}{2}\cdot {}_nm}{n} = \frac{1}{n}+\frac{_nm}{2}$$

and, in the end, if $n = 1$:

$$\frac{m}{q} = 1+\frac{m}{2}$$

TABLE 11–3 Theoretical comparison between rates and probabilities

Deaths	Initial population	Final population	Probability	Rate	Ratio
1	100,000	99,999	0.000010	0.000010	1.000005
10	100,000	99,990	0.000100	0.000100	1.000050
100	100,000	99,900	0.001000	0.001001	1.000500
1,000	100,000	99,000	0.010000	0.010050	1.005025
10,000	100,000	90,000	0.100000	0.105263	1.052632
30,000	100,000	70,000	0.300000	0.352941	1.176471
50,000	100,000	50,000	0.500000	0.666667	1.333333
70,000	100,000	30,000	0.700000	1.076923	1.538462
90,000	100,000	10,000	0.900000	1.636364	1.818182
100,000	100,000	0	1.000000	2.000000	2.000000

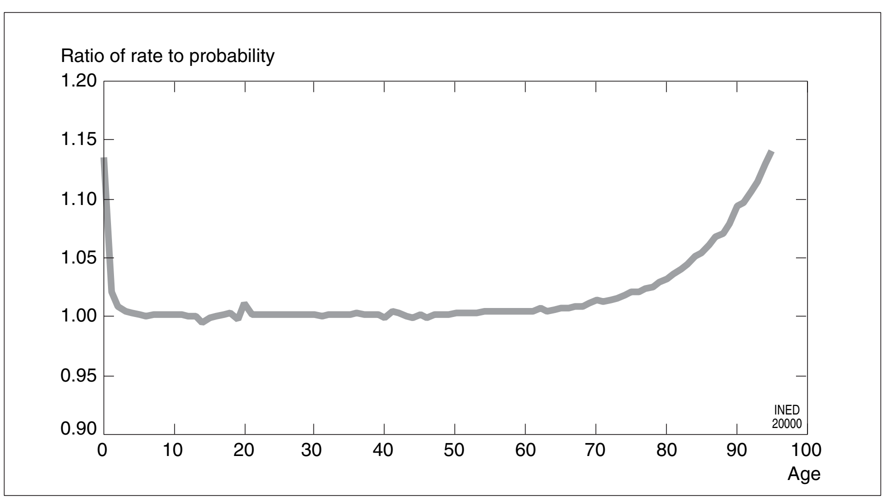


FIGURE 11–3 Ratio of type 3 rate to age-specific probability of dying in French male cohort born in 1899.

approximately age 80 (and, at most ages, is much smaller); but it increases very swiftly beyond age 80 and, toward age 90, returns to the level observed at under age 1. However, when actual data are used, the definitional difference is not the only reason why the rate differs from the probability.

Let us compare the result obtained above to the value that we would have obtained by applying to the age-74 rate the conversion formula given in Chapter 8 for determining the probability:

$$_{1}q_{74} = \frac{2 \cdot {}_{1}m_{74}}{2 + {}_{1}m_{74}} = \frac{2 \cdot 0.04135}{2 + 0.04135} = 0.04051$$

The difference is admittedly minimal, but it does exist. The reason is that the formula supposes a uniform distribution of deaths between two birthdays. In the present case, we are close to this hypothesis, which is not always the case. The best example of the problem is given by infant mortality.

In this same 1899 cohort, the probability of dying at age 0 is 185/1000 and the rate is 215/1000; applying the formula above, we would have found a probability of 194/1000—an overestimation of 5% (Table 11–4).

The discrepancy is obviously due to the fact that 75% of infant deaths are concentrated in the first triangle of the Lexis diagram, versus only 25% in the second triangle. Let us imagine that the total deaths at under age 1 were evenly distributed between the two Lexis triangles. The probability would have been 194/1000, with no change in the rate, and the application of the formula would have yielded the same result (last line of Table 11–4).

Again, we find a fairly large discrepancy between the probability and its estimate based on the rate at age 1, since the imbalance between the two Lexis triangles is still significant at that age. The error is 0.5%. At subsequent ages, however, this type of error becomes negligible (under 2/1000). In fact, it varies either way, no doubt because of the annual mortality fluctuations.

However, infant mortality is the exception that confirms the rule: At most other ages, the deaths are just about evenly distributed between the two Lexis triangles, and we can, without risk of serious error, estimate the probability from the rate using this classic formula if the rate is computed with the same data used to determine the probability.

b. *Other Approaches to Annual Probability*

The data needed to perform the calculations discussed here are not often available, however. Few countries have—as France does—long annual series of deaths by age and year of birth, and systematic estimates of the population by age at January 1. Other approaches to cohort-specific probabilities of dying are possible. The two most important are discussed in the following subsections.

Projective Probabilities

Some countries publish their current annual statistics of death by age reached in the calendar year. In Chapter 6 we defined a type 2 rate that corresponds to this type of classification. One can also calculate a probability of the same kind, generally called a *projective probability*, as it is perfectly suited to the needs of population projections. Let us take up our example of male French mortality at 74. The necessary data are illustrated in Figure 11–4 (parallelogram *ABCD*).

In this case, if there were no migrations, the population exposed to the risk would be, quite simply, the population aged 74 (in complete years) on January 1,

TABLE 11–4 Comparison between rates and probabilities at young ages in 1899 male cohort

Age	Death		Population on Jan. 1 of second year	Rate	Probability		Ratio
	1st year D1	2nd year D2	P	t	Computed q	Estimated from rate q'	q'/q
With the deaths observed in each of the two Lexis triangles							
0	59,716	19,433	369,000	0.2145	0.1846	0.1937	1.0493
1	6,760	3,003	340,200	0.0287	0.0281	0.0283	1.0054
2	1,880	2,919	334,900	0.0143	0.0142	0.0142	0.9985
3	1,821	1,782	332,700	0.0108	0.0108	0.0108	1.0001
Assuming deaths evenly distributed between the two Lexis triangles							
0	39574,5	39574,5	369,000	0.2145	0.1937	0.1937	1.0000

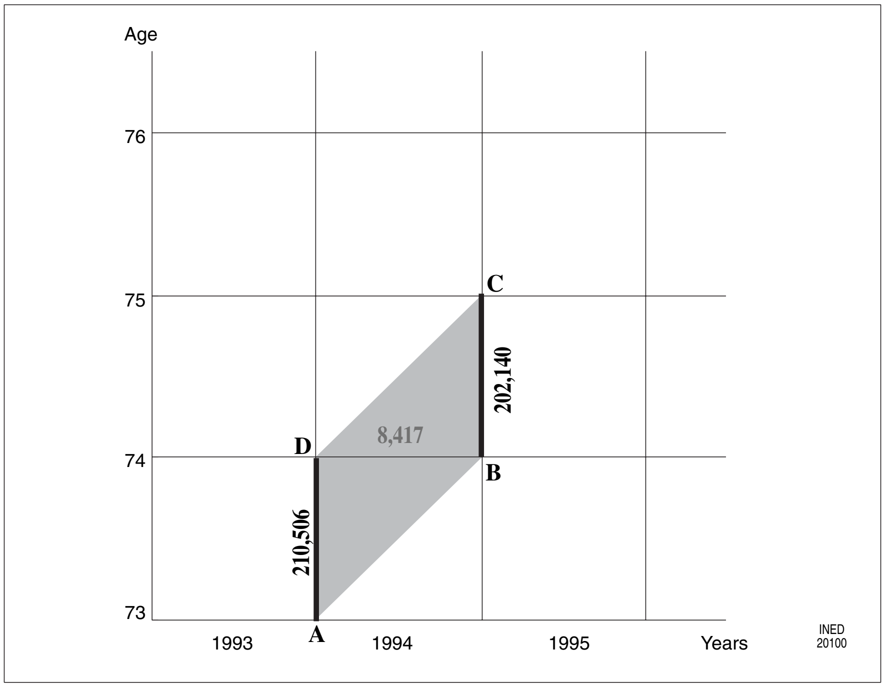


FIGURE 11–4 Data needed to determine the projective probability of dying at age 74 (France, males).

1994. But, as before, we need to factor in migrations. By analogy with the previous computation, it will be easily understood that this can be achieved by using as the denominator the cohort's average population on January 1, 1994, and January 1, 1995, to which we add half of the deaths in the parallelogram:

$$ {}_1q_{74} = \frac{D^{1920}_{1994,(73,75)}}{0.5\cdot\left(P^{1920}_{(73,74)}+P^{1920}_{(74,75)}\right)+0.5\cdot D^{1920}_{1994,(73,75)}} $$

$$ = \frac{8,417}{0.5\cdot(210,506+202,140)+0.5\cdot 8,417} = 0.03998 $$

If we compare this projective probability with the type 2 rate computed in Chapter 6, we find, unsurprisingly, a difference on the same order of magnitude as the one we found with the classic probability: the figures are 0.0400 for the projective probability versus 0.0408 for the type 2 rate.

Moreover, between the projective probability and the classic probability, we find a difference comparable to the one between the type 2 and type 3 rates: They are separated by roughly a half-year of age. The projective probability at 74 years (0.0400) is slightly below the classic probability at the same age (0.0405) because it is centered on the exact age of 74 whereas the classic probability is centered on 74.5 years.

It is fairly easy to shift from one probability to the other using interpolation (linear or, better yet, logarithmic) between two ages, as long as we remain in age brackets where mortality varies steadily with age. This is generally the case, except for mortality at very young ages.

Infant mortality poses a special problem here. The method for computing the projective probability disregards the first Lexis triangle. For this first triangle, we need to determine a probability that is neither classic nor projective, since it measures the risk of dying between birth and the following January 1. Chapter 14 shows that it may be useful to generalize the calculation of *partial probabilities* of this type. In the present case, however, the partial probability is something of an intruder in a series of probabilities by year of age—because, on average, it covers only a half-year of age. To estimate the classic probability at 0 year from the projective probabilities, we must use age-0 weights based on observations made in similar populations for which dual classifications of deaths by age and year of birth are available.

Indirect Calculation Based on Rates in the Squares of the Lexis Diagram

If the current annual statistics list deaths only by age in complete years, we cannot directly calculate any probability of dying. Thus, we must be content with estimating the deaths from the type 1 age-specific rates computed in the Lexis squares as described in

Chapter 6. However, we will never obtain a cohort-specific measure, since two contiguous cohorts cross each Lexis square. To find a measure that best approximates a classic probability for a given age, we must average the rates of two consecutive years in which the cohort studied reaches that age. For example, the probability of dying at age 74 for the 1920 cohort will be estimated from the type-1 rates at age 74 of the years 1994 and 1995, respectively, 0.0443 and 0.0418. As the average of the two rates is 0.0431, the rate-to-probability conversion formula gives us a probability of 0.0422.

This figure is much higher than the one obtained with the direct computation (0.0405): the probability is overestimated by more than 4%. This is the same problem already encountered in our comparison between the three types of age-specific rates: The disturbances produced by World War I in the timing of births of the 1919 cohort entail an overestimation of the type 1 rate at 74 years in 1994, and this impacts on the average of the two rates used here (Chapter 6)—a totally atypical situation. In most cases, the estimation of the classic probabilities is entirely acceptable, subject to the reservations voiced earlier concerning the uniformity of the age-specific distribution. However, we have deliberately returned to this example, to remind the reader that it has no universal value.

Likewise, we can estimate the projective probability at age x from the average of the type 1 rates of the two ages in complete years situated on either side of age x. Both rates, this time, belong to the same calendar year and, in our example of mortality at age 74 for the 1920 French male cohort, the result is not affected by the problem mentioned above. However, we find an overestimation of 1.5% by comparison with the direct method, which is due to the nonlinear variation of age-specific mortality.

Table 11–5 summarizes the results obtained using these different approaches for estimating the 1920 French male cohort's probabilities of dying at age 74.

II. CONSTRUCTION AND MAIN FUNCTIONS OF THE TABLE

Using age-specific probabilities of dying, we can (1) provide a complete description of the mortality process experienced by a cohort until its final extinction, (2) track the gradual reduction of its initial population from one age to the next, (3) determine the distribution by age at death of that initial population, and (4) calculate the average number of years lived by the individuals in the cohort (i.e., their *life expectancy*). For all these purposes, we need only construct the *life table*. Later, however, we will see that the life table can also be entered by means of a continuous computation, using mortality rates directly as estimators of the *force of mortality*.

1. Constructing the Table from the Probabilities of Dying

Let us begin with the series of age-specific probabilities of dying, ${}_1q_x$, for a given cohort. We set an arbitrary initial number of births (known as the life table *radix*): As a rule, the number is a power of 10 to facilitate the computations and the presentation of comparable results. We then construct the cohort's life table, using a series of iterations.

Let l_0 be the life table radix. We start by computing $d_{0.1}$, the number of deaths between ages 0 and 1:

$$d_{0,1} = l_0 \cdot {}_1q_0$$

Subtracting these deaths at under age 1 from the initial population, we obtain the number l_1 of survivors of the table at the first birthday:

$$l_1 = l_0 - d_{0,1}$$

Next, we apply the probability of dying at age 1 to these survivors to obtain the deaths in the table between ages 1 and 2:

TABLE 11–5 Classic and projective probabilities of dying at 74 for 1920 French male cohort, using different approaches

Calculation method	Classic probability		Projective probability	
	Result	Error, %	Result	Error, %
Direct calculation	0.04050		0.03998	
Estimation using type 3 rate	0.04051	0.02		
Estimation using type 2 rate			0.03998	0.00
Estimation using average of projective probabilities	0.04158	2.67		
Estimation using average of type 1 rates	0.04216	4.10	0.04058	1.50

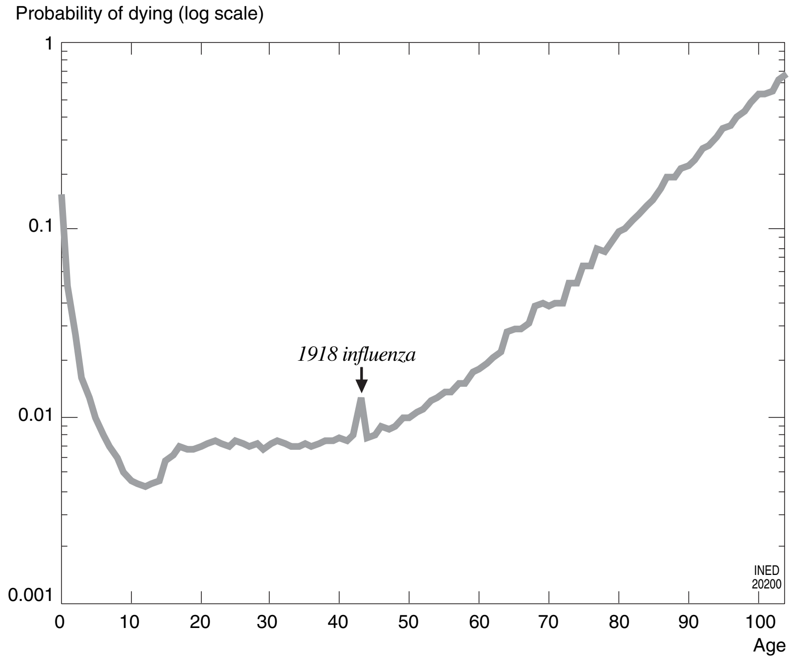


FIGURE 11–5 Age-specific probabilities of dying for the French female cohort born in 1875.

$$d_{1,2} = l_1 \cdot {}_1q_1$$

and so on, until the complete extinction of the initial population, under the effect of a final probability that is theoretically equal to unity:

$$d_{\omega-1,\omega} = l_{\omega-1} \cdot {}_1q_{\omega-1}$$

while:

$$l_\omega = l_{\omega-1} - d_{\omega-1,\omega} = 0$$

Figures 11–5, 11–6, and 11–7 illustrate the values of these three basic life-table functions for the French female cohort born in 1875.[7] The table is provided in the appendix.

In this cohort, the probability of dying starts from a very high level, with a value of about 150/1000 between birth and the first birthday. It bottoms out toward 10 to 15 years before surging at about 16 to 17 years to a level where it remains until approximately age 40. At all these ages, advances in health suffice to offset the loss of vitality due to aging. However, at age 50 and up, the age effect outweighs the health-progress effect, and the probability rises almost exponentially with age (Figure 11–5).

This general profile of the mortality curve results from a compromise between two deep-seated trends: the change in the risk of death with age and the change in health conditions over time. The profile also displays traces of specific events experienced by the cohort. Actually, the only significant accident experienced by this female cohort was due to the 1918 influenza pandemic. World War I, by contrast, caused no major disturbance. The cohort was 39 years old in 1914 and 42 years old in 1917. The mortality of women of those ages was barely affected, contrary to that of men, who—even at those ages—paid a heavy tribute to the war. On the other hand, the mortality curve of the 1875 female cohort is mildly disturbed by World War II. The second conflict, far more than the first, caused a general deterioration in health conditions, affecting both men and women, and the elderly as much as persons of combat age.

The survival curve (Figure 11–6) masks these dips and peaks in the probability of dying. However, it vividly illustrates the two major phenomena that presided over the cohort's extinction: (1) the very high infant mortality, which, right from the start, deprived the cohort of a large section of its initial population, and (2) the rising risk of death beyond age 50, which hastened the reduction in the cohort size.

The distribution of deaths by age is overwhelmingly dominated by the heavy infant mortality, which, of course, was experienced by the cohort at its largest size (Figure 11–7). All subsequent experiences seem marginal by contrast. When the risk of dying reaches new peaks after age 50, it applies to a very small group of survivors, and the second death distribution peak

[7] We choose this cohort because it is that of Jeanne Calment, who broke all known human longevity records when she died on August 4, 1997, at the age of 122 and a half (Allard *et al.*, 1998).

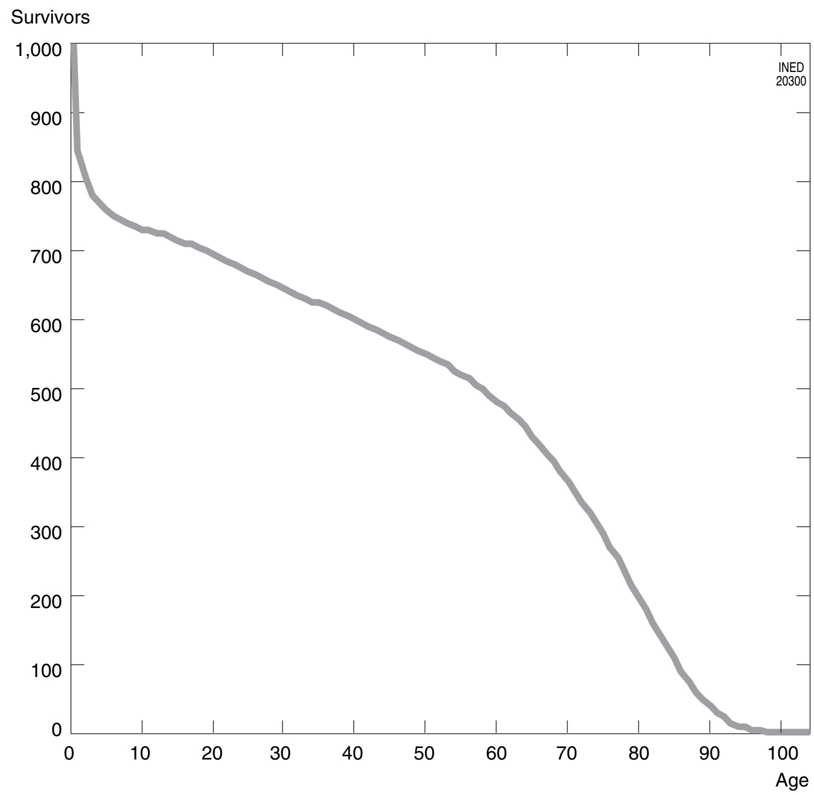


FIGURE 11–6 Survivors of the life table of the French female cohort born in 1875.

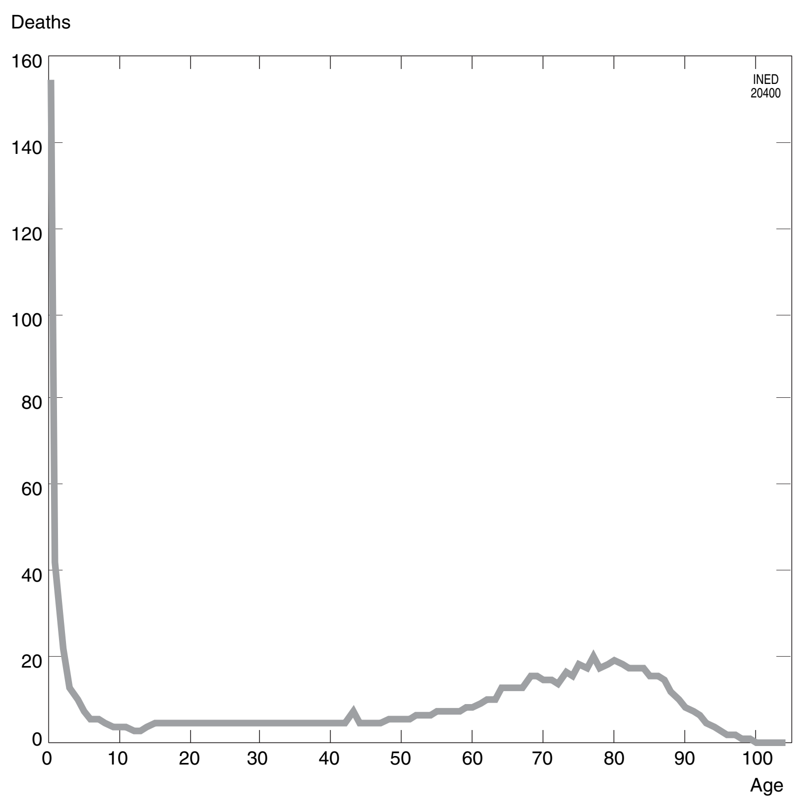


FIGURE 11–7 Life table deaths of the French female cohort born in 1875.

observed at 75 to 80 years or so is very modest in comparison. The impact of the Spanish influenza of 1918, while visible, is far more subdued here than on the probability curve.

When measured on an annual basis, these three main variables of the table can be combined to arrive at multiannual dimensions. The deaths are additive:

$$d_{x,x+a} = \sum_{x}^{x+a-1} d_{x,x+1}$$

The probabilities of survival are combined indirectly, via the probabilities of survival between two successive birthdays:

$$p_{x,x+a} = 1 - {}_aq_x = \prod_{x}^{x+a} {}_1p_x = \prod_{x}^{x+a} (1 - {}_1q_x) = 1_{x+a} / 1_x$$

and

$$l_{x+a} = l_x \cdot p_{x,x+a}$$

2. Entering the Life Table via Deaths, in a Closed Population

Constructing a cohort life table from probabilities is the only possible approach when using current annual data on a real population that experiences migration flows and changing health conditions. In the very infrequent case where we would use longitudinal observation data on a closed population, we could, of course, simply express the age distribution of actual deaths as a ratio of a power of 10 and equate it with the deaths in the table:

$$d_{(x,x+1)} = \frac{D_{(x,x+1)}}{B} \cdot l_0$$

For example, in the case of the 1992 Egyptian survey, if l_0 = 10000, the deaths in the table can be immediately deduced from the actual deaths (Table 11–6). From this, we can readily deduce the two other functions of the table, taking its radix as our starting point:

$$ {}_1q_0 = \frac{d_{0,1}}{l_0}$$

then, again:

$$l_1 = l_0 - d_{0,1}$$

and so on.

In this type of situation, as we lack the necessary perspective, we cannot expect to observe the complete distribution of deaths by age. We do know, however, that all the individuals of the cohort must die and we can deduce the expected deaths beyond the last age observed by subtracting the sum of actual deaths from the number of births. The table will remain incomplete, but we will at least be able to define its beginning.

TABLE 11–6 Actual deaths and deaths in table for children in 1992 Egyptian demographic and health survey born in 1973–1977

Exact age x	Actual deaths $D_{(x,x+1)}$	Deaths in table $d_{(x,x+1)}$
Births	**6,070**	**10,000**
0	796	1,311
1	207	341
2	96	158
3	52	86
4	20	33
5	11	18
6	16	26
7	8	13
8	7	12
9	3	5
10	3	5
11	3	5
12	4	7
13	1	2
14	4	7
From 15 to ω	4,839	7,972

In all other cases, the age distribution of deaths in the table must not be confused with that of actual deaths. More accurately, the deaths series in the table is the key result expected of any analysis of mortality. Indeed, it is the series that gives us the description of the only variation factor in mortality: its *tempo*.

3. Life Expectancy

The tempo of deaths in the table is, in particular, the factor that enables us to summarize a cohort's mortality by determining its *life expectancy at birth*. Whether deduced directly from a longitudinal observation or constructed from probabilities, the tables' death series gives us the distribution of the individuals composing the cohort by age at death, and we can use the data to obtain their average life span (their life expectancy at birth) by calculating the average of their ages at death. Typically, to compute this average, we assume a uniform distribution of deaths between two birthdays. We therefore consider that the deaths of age (x, x + 1) occur at age (x + 0.5). The life expectancy e_0 at birth is thus equal to:

$$e_0 = \frac{\sum_{0}^{\omega-1} (x+0.5) \cdot d_{x,x+1}}{\sum_{0}^{\omega-1} d_{x,x+1}}$$

But, as $d_{(x,x+1)} = l_x - l_{x-1}$, we can also write this in simpler form:

$$e_0 = 0.5 + \frac{\sum_{1}^{\omega} l_x}{l_0}$$

More generally, we can calculate life expectancy at any age *x:*

$$e_x = \frac{\sum_{j=x}^{\omega-1} (j - x + 0.5) \cdot d_{j,j+1}}{\sum_{j=x}^{\omega-1} d_{j,j+1}} = 0.5 + \frac{\sum_{j=x+1}^{\omega} l_j}{l_x}$$

If mortality were constant with age or rose continuously with age, life expectancy would obviously decrease in a very steady pattern with age. In fact, in the 1875 French female cohort, life expectancy starts by rising at young ages (Figure 11–8). Infant mortality is so high that, by the sheer fact of having survived, infant girls still alive at age 1 have 7.5 years more life expectancy than the total cohort at birth. Mortality remains high enough at ages 1 and 2 to keep life expectancy on an upward trend until age 3, when it is 10 years longer than life expectancy at birth. It is only beyond age 3 that life expectancy moves onto a steady downward path.

Without actually undermining what we have just said about the 1875 cohort, this life-expectancy computation calls for two comments.

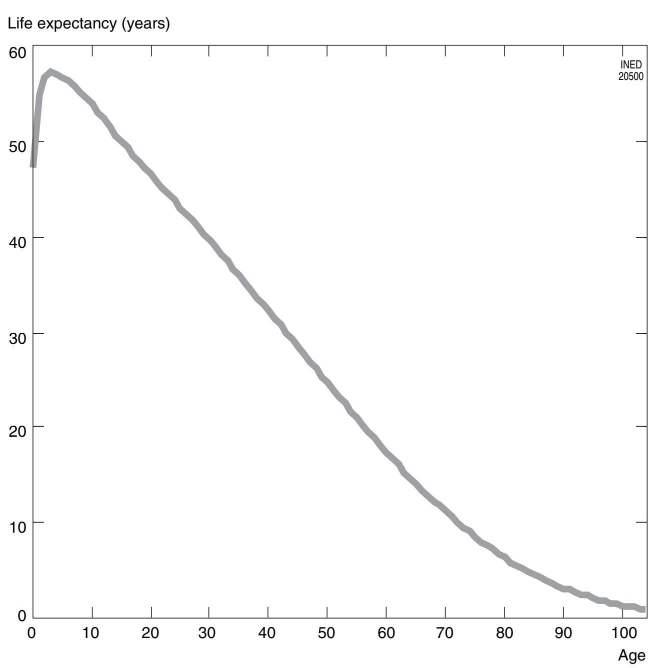


FIGURE 11–8 Age-specific life expectancy in French female cohort born in 1875.

First, if the actual distribution of deaths between two birthdays is indeed very close to the equidistribution hypothesis at nearly all ages, this profile—as we have already seen—does not apply to the start of life. The life expectancy calculation would be more accurate if we could determine probabilities by month of age for the earliest ages, especially when infant mortality is very high. Absent the necessary data, we can also apply a different weighting for the initial ages, based on empirical observations of populations with similar mortality levels. This refinement in the measure is, however, seldom used. The most common approach is to assume equidistribution at all ages of the table.

Moreover, the life expectancies computed in this manner are, in reality, no more than estimates (to a partial extent). The actual current statistical data used in the standard computation of probabilities of dying are no longer reliable after a certain age. This makes it impossible to pursue the calculation until the cohort's total extinction, especially when we are dealing precisely with the cohort whose last death, that of Jeanne Calment, occurred at age 122—long after the penultimate death. Concretely, the probabilities used here are reliable only until an age *k*—in this case, around 85 or 90 years (Vallin and Meslé, 2001). Pending the discussion in the section entitled "The Oldest-Ages Problem and Table Closure," which addresses the specific solutions for estimating with greater precision the probabilities of dying at the very oldest ages, we must make do with an incomplete table; to determine life expectancy, we must close off the table arbitrarily by estimating life expectancy at age *k* as realistically as possible. For any age *x* younger than *k*, we obtain:

$$e_x = \frac{\left(\sum_{j=x}^{k} (j - x + 0.5) \cdot d_{j,j+1}\right) + (k - x + e_k) \sum_{j=k}^{\omega-1} d_{j,j+1}}{\sum_{j=x}^{\omega-1} d_{j,j+1}}$$

$$= 0.5 + \frac{l_k(e_k - 0.5) + \sum_{j=x+1}^{k} l_j}{l_x}$$

In consequence, the life expectancy at birth of Jeanne Calment's cohort is 47.2 years, if we take 4.48 years as the life expectancy at 85.[8] The arbitrary closure of the table at age *k* has almost no effect on the result (as we shall see in greater detail in Chapter 14). By contrast, the closer *x* gets to *k*, the more life expectancy at age *x* depends on the cutoff point.

[8] This estimate of life expectancy at 85 years is actually derived here from an extrapolation of probabilities of dying at the oldest ages.

We can also use the table's arbitrary closure to obtain an estimate of life expectancy at birth in a retrospective longitudinal observation cut off at the survey date. For this purpose, we combine the start of the life table actually observed with a life expectancy taken from a *standard table* (see Josianne Duchêne's Chapter 128 on this topic in Volume 4) adjusted to the situation for the age at which the information gathered stops. In the case of the Egyptian survey given as an example at the start of this chapter, we could have taken into account the mortality of the cohort's first 15 years of life and closed the table with a life expectancy at age 15 of 46 years—which, according to the standard tables of Coale and Demeny (1983), is consistent with the actual level of infant and child mortality.[9] The estimated life expectancy of these Egyptian cohorts would be 48.9 years at birth and 58.3 years at age 1. This method is all the more approximate as the final age of observation is young and infant mortality is low. In the Egyptian case, we can clearly see that with nearly 80% of survivors at age 15, most of the mortality is experienced after age 15; the method's value is all the more limited as no one knows how the Egyptian health situation will evolve as the cohort studied grows older. The procedure may, however, be useful in contexts where infant mortality is very high—which are also, typically, the contexts where we know nothing about adult mortality.

III. FROM THE DISCRETE TO THE CONTINUOUS: ENTERING THE TABLE VIA THE FORCE OF MORTALITY

The life table functions described here in discrete mode can also be treated in a continuous mode. To enter the table in continuous mode, we must first specify the relationships between probability, rate, and force of mortality. For this, we need to introduce a new function into the table: the *number of years lived in an age interval*.

1. Number of Years Lived in an Age Interval or "Person-years"

Starting with the functions already defined in discrete mode, and in particular the l_x series, we can—by postulating a uniform distribution of deaths between two birthdays—compute the number ${}_1L_x$ of years lived in an age interval $(x, x + 1)$ as the half-sum of the survivors at the two birthdays on either side:

$${}_1L_x = \frac{l_x + l_{x+1}}{2}$$

The number is also (1) the number of *person-years* exposed to the risk of mortality at this age in the cohort and (2) the number of persons in the cohort who have reached the complete age $(x, x + 1)$ on January 1 of the xth year after their birth. The death rate at age x is thus equal, in the life table, to:

$${}_1\overline{m}_x = \frac{d_{(x,x+1)}}{{}_1L_x}$$

This ${}_1\overline{m}_x$ rate computed in the life table is slightly different from the ${}_1m_x$ rate calculated from the actual data insofar as the first is based, by definition, on the equidistribution of deaths in the age interval—a situation never fully attained in reality. In fact, as long as mortality decreases with age (until approximately 10 years), ${}_1m_x > {}_1\overline{m}_x$, and, later, ${}_1m_x < {}_1\overline{m}_x$ (Keyfitz, 1977).

To shift from the discrete to the continuous, we must look at what happens to the rate ${}_1\overline{m}_x$ in the table in an infinitesimal age interval. For this purpose, we need to define in continuous terms the sum of years lived in the age interval $(x, x + 1)$. If $l(a)da$ is the time lived by the cohort between ages a and $a + da$, ${}_1L_x$ can be written:

$${}_1L_x = \int_x^{x+1} l(a)da$$

This integral defining ${}_1L_x$ measures the area located between the l_x curve, the lines of equations $a = x$ and $a = x + 1$, and the x-axis (Figure 11–9).

2. The Force of Mortality and Its Relationship with the Death Rate

The *force of mortality* at age x is the limit value of the death rate of the table ${}_n\overline{m}_x$ when n tends toward 0:

$$\mu_x = \lim_{n \to 0} {}_n\overline{m}_x = \lim_{n \to 0} \frac{{}_nd_x}{{}_nL_x}$$

Taking the earlier expression of ${}_1L_x$, we find the equation already mentioned in Chapter 8:

$$\mu_x = \lim_{n \to 0} \frac{l_x - l_{x+n}}{\int_x^{x+n} l(a)da} = \frac{-1}{l_x} \cdot \frac{\partial l_x}{\partial x}$$

In other words, after solving this differential equation:

$$\mu_x = -\frac{\partial l_x / \partial x}{l_x}$$

[9] As it happens, it is the western model that best matches the structure of infant and child mortality measured by the 1992 Egyptian Demographic and Health Survey; the level observed is slightly below the male level 13 of the model, which gives a life expectancy of 45.5 years at age 15.

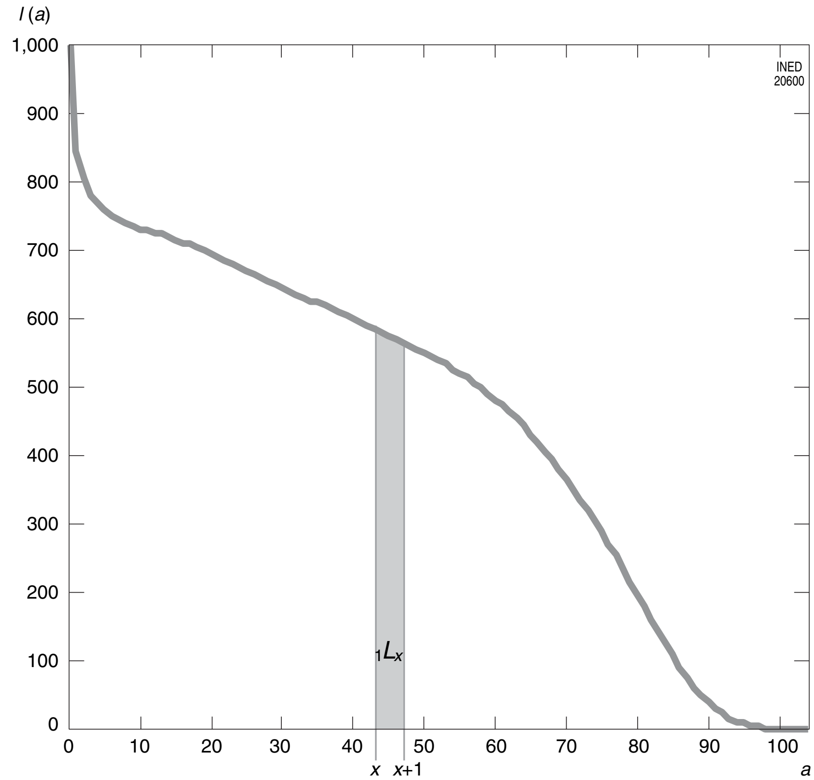


FIGURE 11–9 Area defined by ${}_1L_x$ in the life table of the French female cohort born in 1875.

or

$$\mu_x = \frac{-\partial(\ln l_x)}{\partial x}$$

Within an age interval, the force of mortality measures the mortality at each instant (i.e., the age-specific *instantaneous mortality* rate). For example, μ_{x+n} is the instantaneous rate for each age between x and $x + n$, where n varies between 0 and 1. In a way, this rate can be regarded as a sort of average of all the instantaneous rates of the interval. That is why the rate of the table, generally called the *central death rate*, is a good estimator of the force of mortality at mid interval:

$$\mu_{x+0.5} \cong {}_1\overline{m}_x$$

Concretely, as we have seen, we calculate the death rate, ${}_1m_x$, as the ratio of actual deaths in a given age interval to the number of person-years exposed to the risk, estimated by the half-sum of the actual cohort population at either end of the interval. Pollard (1973) has shown that, if one uses annual age groups, ${}_1m_x$, is equally a good estimator of the force of mortality $\mu_{x+0.5}$:

$$\mu_{x+0.5} \cong {}_1m_x$$

Thatcher *et al.* (1998) have shown that the error committed by equating the force of mortality with the actual rate depends on the relationship between the force of mortality and age. If the relationship were constant over the entire age interval, the error would be null. However, even when the force of mortality grows exponentially with age, the error remains modest as long as we use a one-year age interval: in the real universe studied by the authors, the error never exceeds 0.8%.

3. Force of Mortality and Probability of Dying

If we revisit the survival function, we can also study the relationship between the force of mortality and the probability of dying (i.e., the probability of dying defined in discrete terms).

Integrating $\mu_x = -[\partial l_x/\partial x]/l_x$, we obtain the analytical expression of S_x:

$$l_x = l_0 \exp\left[-\int_0^x \mu(a)da\right]$$

If $l_0 = 1$, l_x is equal to the probability ${}_xp_0$ of survival from birth to age x:

$${}_xp_0 = \exp\left[-\int_0^x \mu(a)da\right]$$

and the probability of dying in the same interval, the probability ${}_xq_0$, is equal to:

$$ {}_xq_0 = 1 - \exp\left[-\int_0^x \mu(a)da\right] $$

Last, the conditional probability ${}_1q_x$ of dying before age $x + 1$ when reaching age x is equal to:

$$ {}_1q_x = 1 - \exp\left[-\int_x^{x+1} \mu(a)da\right] $$

As an age interval comprises an infinity of discrete ages, the force of mortality at the discrete mid-interval age (x, $x + 1$) is regarded, for the sake of convenience, as equal to the probability of dying in the discrete computation mode:

$$ {}_1q_x \cong 1 - \exp[-\mu_{x+0.5}] $$

This relationship shows that, when the force of mortality $\mu_{x+0.5}$ tends toward infinity, ${}_1q_x$ tends toward 1. At the same time, the rate of the table, ${}_1\overline{m}_x$, tends toward 2, since ${}_1\overline{m}_x = 2{}_1q_x/(2-{}_1q_x)$.

Thus, the force of mortality can be proxied by:

$$ \mu_{x+0.5} \cong -\ln(1 - {}_1q_x) $$

When the age interval is 1 year, this approximation is even better than the one provided by the rate. If the force of mortality is constant, the error is obviously null. If it rises exponentially with age, Thatcher *et al.* (1998) have shown that the error can hardly exceed 0.04%.

4. Another Look at the Relationship Between Rate ${}_1m_x$ and probability ${}_1q_x$

Combining the relationships above, we obtain:

$$ \mu_{x+0.5} \cong {}_1m_x $$

$$ \mu_{x+0.5} \cong -\ln(1 - {}_1q_x) $$

$$ {}_1q_x \cong 1 - \exp[-\mu_{x+0.5}] $$

We can therefore write:

$$ {}_1m_x \cong -\ln(1 - {}_1q_x) $$

and

$$ {}_1q_x \cong 1 - \exp[-{}_1m_x] $$

Taken together, these relationships show that, if we define the year of age as our unit, we can generally use the rate, the probability, or the force of mortality interchangeably. But, whatever the length of the age interval, these equations also make it possible, when one of the indicators is known, to determine the other two.

5. Life Expectancy in Continuous Terms

Generalizing the definition given at the start of this section for the number of years lived in a 1-year age interval, ${}_1L_x$, we can integrate $l(a)da$ from age x to the extreme age ω to obtain the total number of years lived beyond age x:

$$ {}_{\omega-x}L_x = T_x = \int_x^{\omega} l(a)da $$

Dividing this total by the number of survivors at age x, we obtain life expectancy at age x:

$$ e_x = \frac{T_x}{l_x} = \frac{\int_x^{\omega} l(a)da}{l_x} $$

Life expectancy at birth is, of course, equal to:

$$ e_0 = \frac{T_0}{l_0} = \frac{\int_0^{\omega} l(a)da}{l_0} $$

IV. THE OLDEST-AGES PROBLEM AND TABLE CLOSURE

In determining the probabilities of dying at the oldest ages, we come up against two very different problems. In most cases, as explained in the second section of this chapter, this obliges demographers to interrupt the table construction at a given age and to arbitrarily assign it the most plausible life expectancy possible in order to close the table.

The first problem is that the empirical data on the oldest ages are often of poor quality. The ages of the very elderly are subject to reporting errors. One of the most classic error factors is the prestige surrounding the elderly, which sometimes encourages the reporting person to exaggerate. In countries where civil registration is recent or does not function well, this type of error can impair death statistics as well as population counts to a sometimes-phenomenal degree. For example, legends have emerged on the presence of supercentenarians in Georgia (in the Caucasus) or certain Andean valleys (Allard, 1991). In countries with a long tradition of comprehensive birth records, the risk is relatively low for death registrations, which are usually based on official documents. However, the risk does persist in censuses: These generally rely on respondents' declarations, which are not systematically checked. Such errors break the strict consistency between numerator and denominator needed to calculate probabilities from current statistics. The French data taken here as an example visibly suffer from this restriction. Above a certain age, the probabilities determined by empirical methods display an erratic profile that contrasts sharply with their smoother upward path at earlier ages (Figure 11–10).

As we can see, the empirical computation ceases to be reliable at approximately age 85 for most of the older cohorts shown here. However, we also note that the quality of the data improves over time and varies from one sex to the other. In the most recent female cohorts (1880, 1890, and 1900), the increase in probabilities with age remains steady until almost age 100 (up to the end of the observations available for the 1900 cohort). That is still the case for the 1890 and 1900 male cohorts but not for the 1880 cohort, whose probabilities are visibly incorrect beyond age 90. Moreover, the fluctuations at the oldest ages observed in the older cohorts are far wider among men than among women.

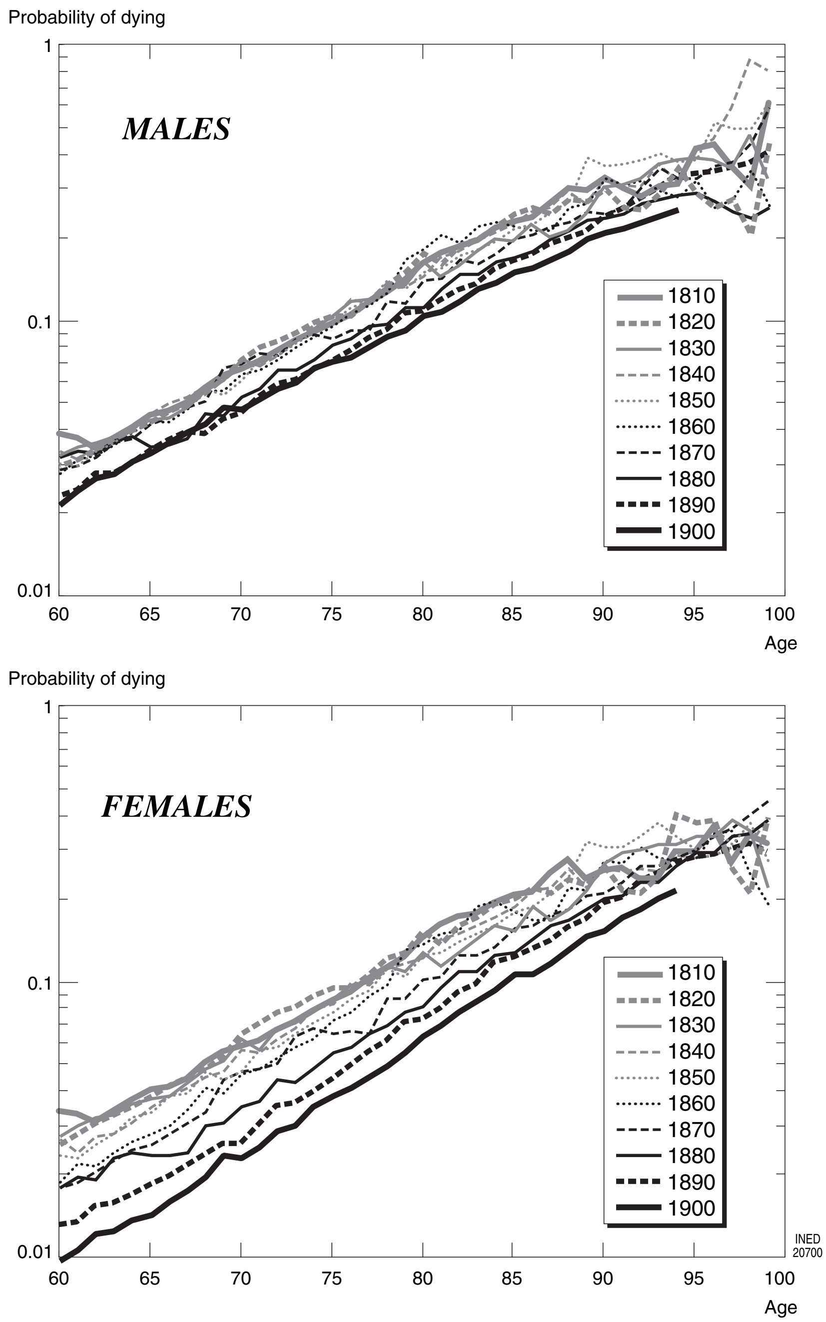


FIGURE 11–10 Probability of dying older than 60 years in French cohorts born between 1810 and 1900, in 10-year intervals, by sex.

This sex difference, however, is probably not due to a difference in data quality; it has more to do with the second problem posed by the measure of mortality among the very elderly.

This second problem is of an entirely different nature. The greater the decrease in cohort with age, the smaller the number of observations on which the computation rests. At the very oldest ages, numbers inevitably become so small that the probability calculation loses the value it derives from the law of large numbers at other ages. This phenomenon naturally occurs earlier among men than among women, who are far more numerous at the oldest ages, owing to male excess mortality.

When we estimate the age-related change in mortality in this manner, the randomness factor outweighs the theoretical risk that we are seeking to measure.

The problem becomes most acute when the cohort is reduced to a handful of survivors. Indeed, even at the very high mortality levels for these ages, chance may have it that no survivor dies in a full year of observation. The French female cohort born in 1875 offers the best illustration, since Jeanne Calment[10] died at age 122, in 1997, 5 years after the penultimate death in the cohort, which occurred in 1992, at age 117; the three previous deaths in the cohort occurred at the ages of 116 (in 1991), 114 (in 1989), and 109 (in 1984).[11]

The best way to definitively solve the first problem would be, of course, to improve data collection conditions to dispel any uncertainty over the measurement of the age of the elderly, in particular in censuses. The surge in the number of the very old, and the socioeconomic problems this entails, explains the growing interest in the extreme age category. France made a special effort to verify the ages of the over-90s in the 1990 census, and one can hope that such campaigns will intensify in most countries. Another solution is to perform specific surveys to measure the mortality of the very elderly, such as the survey on centenarians conducted in France by the IPSEN Foundation (Allard *et al.* 1996). This is also the goal of the program that has been under way for the several years at the University of Odense and the Max Planck Institute of Demographic Research, which aims to gather the largest possible set of reliable data on mortality at the oldest ages (Kannisto, 1994, 1996). However, these efforts clearly cannot solve the problem of the quality of old data—which are indispensable for constructing cohort life tables—and leave unsolved the issue of randomness due to small numbers.

On the first point, a special approach to old-age mortality proposed by Vincent (1951) provides a partial remedy to this difficulty: it is known as the *extinct cohorts* method. Concerning the second problem, in the wake of Gompertz (1825), several authors have suggested methods for extrapolating age-specific mortality that allow an estimation of mortality at ages when it is no longer observable.

1. The Extinct-Cohort Method

Since the age of the oldest individuals is less accurately identified by censuses when they are still alive than by the civil records at the time of their death, is there a way to measure their mortality using only the age information recorded when they die?

As seen earlier, in the absence of migration, we could construct a cohort life table simply by taking the number of births and deducting the deaths recorded at each successive age. There would be no need to refer to population data. However, the method is totally unsuitable for a population that experiences migration flows.

Vincent's (1951) ingenious method rests on the notion that a procedure inconceivable to perform from births is probably not impossible to carry out from deaths, beyond a certain age. Vincent argued that at the uppermost ages, at least at the national level, migration is so limited that it becomes entirely marginal relative to mortality, which is very high.

Accordingly, starting with the death of the cohort's last survivor, we can determine the number of survivors at any age x, by successively combining the deaths observed at the preceding ages:

$$l_x = \sum_{x}^{\omega-1} d_{x,x+1}$$

In this case, the survivors l_x and the deaths $d_{(x,\,x+1)}$ are exactly, by hypothesis, those of the sought-for life table. Constructed in this manner, however, the table has an indeterminate radix. To link this end-of-table to the table computed using the classic method for the younger ages, we must (1) reduce the total population in the death series thus calculated to the number of survivors at the age when the conventional calculation stopped or (2) go back to the probabilities and complete the series of classic probabilities by adding the probabilities determined by means of the extinct-cohort method for the oldest ages.

If we return to the example of the Jeanne Calment cohort, born in 1875, and apply the extinct-cohort

[10] Jeanne Calment was born in Arles (France) on February 21, 1875, at 7 AM. and died, also in Arles, on August 4, 1997, at the age of 122 and a half (Allard *et al.*, 1994, 1998).

[11] If, however, the ages at death reported to the civil registrar for these persons are correct. While Jeanne Calment's age at death has been thoroughly checked (Allard *et al.*, 1994, 1998), the same cannot be said for the earlier deaths in her cohort (Meslé *et al.*, 2000).

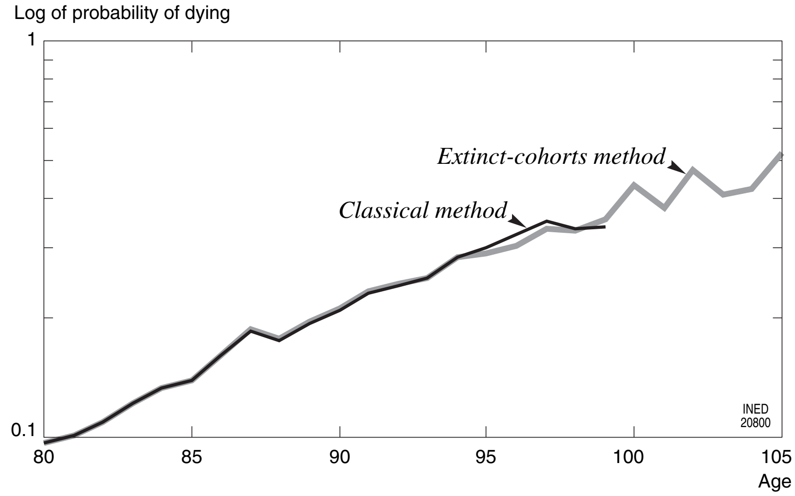


FIGURE 11–11 Comparison of classic probabilities of dying and probabilities given by the extinct-cohort method, beyond age 80 (France, 1875 cohort).

method to measure its mortality beyond age 85, its life expectancy at 85 is 4.7 years. Had we used the probabilities calculated with the classic method until age 100[12] (with a life expectancy of 1.5 years at that age), we would have found 4.5 years. The difference, therefore, is not very great at age 85, but the probability series beyond that age is very different and, the older the age, the more the life expectancy given by the classic method depends on the arbitrary value chosen to close the incomplete table (Figure 11–11).

2. Extrapolation Methods

Although the extinct-cohort method provides a partial solution to the problem of age-reporting errors in censuses, it is obviously powerless to deal with the problem of small numbers. It is difficult to accept the notion that, in the French female cohort born in 1875, the mortality reflecting the cohort's health condition was 0 from ages 117 to 122, then jumped to 100% at age 122 because Jeanne Calment remained the sole survivor for 5 years. What would we have observed if, at age 110, the cohort comprised not merely four women (Jeanne Calment and the three others who would live beyond 110) but 1000 or 100,000? The question is all the more sensitive as it raises issues about the ultimate age of life.

A number of authors have tried to model the observed part of the mortality curve to extrapolate it to the ages where it is no longer measurable, either because of poor data quality or owing to the random effect of small numbers. Although their priority has been to estimate mortality, these authors have also had to address the extreme-age issue, if only implicitly.

a. From DeMoivre to Gompertz

Back in 1725, Abraham DeMoivre sought to answer this question by formulating the hypothesis that the number of persons alive at a given age is proportional to the number of years between that age and the extreme age of life:

$$l_x = l_0\left(1 - \frac{x}{\omega}\right)$$

from which it appears that the force of mortality is equal to:

$$\mu_x = \frac{1}{\omega - x}$$

The hypothesis of a linear change in l_x with age was indeed fairly consistent with what was known about mortality at the time, particularly from Halley's table (1693). Today, of course, the assumption seems quite naive. Most of all, however, DeMoivre's formula implied that the issue of the limit age of life had been settled beforehand. The age of 86 suggested by DeMoivre is obviously well below what we may reasonably imagine today.

A century later, Benjamin Gompertz (1825)—relying on an admittedly far greater abundance of data—made the far more realistic proposal that the force of mortality μ_x should be deemed to follow an exponential function of age:

[12] Unfortunately, we cannot use the same method to determine probabilities of dying after age 100 because the population estimates by age at January 1 do not distinguish the number of individuals by age beyond 100.

$$\mu_x = Bc^x$$

where B and c are constants (curve [1] of Figure 11–12). Benjamin Gompertz saw in this mathematical model a description of "two generally coexisting causes; the one, chance, without previous disposition to death or deterioration; the other, a deterioration, or an increased inability to withstand destruction" (Gompertz, 1825; p. 517).

This formulation requires no preconceived notion on the limit age of life—hence the statement by Smith and Keyfitz (1977) that "Gompertz is as much remembered for having reasoned well: his Law is a better description of the upper ages than DeMoivre's hypothesis that populations collapse towards a specific age limit." In fact, however, Benjamin Gompertz was assuming that there is no age limit to life. Once the constants B and c have been set on the basis of the analysis of empirical data, the extrapolated values of the force of mortality can increase without limit; μ_x can tend toward infinity, at the same time as x, whereas the probability of dying ${}_1q_x$ will merely tend toward 1, without ever actually reaching it. Provided that the population is large enough, these assumptions theoretically allow for the survival of at least a few individuals beyond any limit. Contrary to what Vincent (1951) had conjectured when trying to determine the uppermost age from Gompertz's law, the refusal to set an arbitrary age limit is tantamount to denying the existence of a limit.

Nonetheless, Gompertz's formula enables us to extend the actual values of the life table with extrapolated values up to the age we want; typically, this is the age at which the real population studied contains few or no survivors. In these conditions, the method used to close the table matters little, as it will only affect the description of the mortality of those rare survivors.

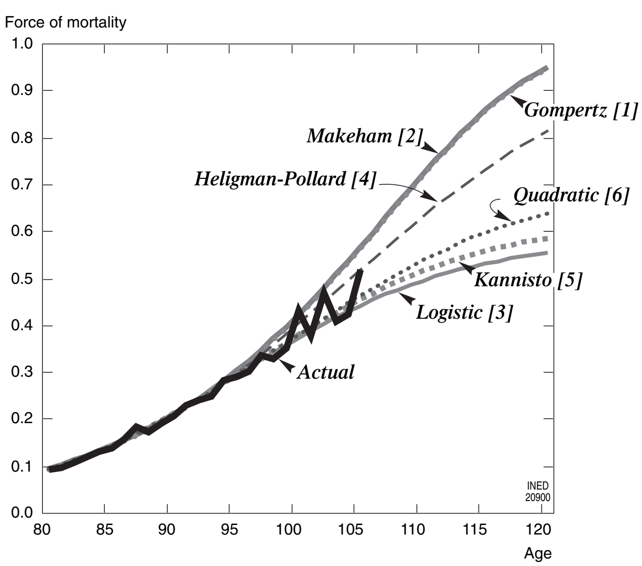


FIGURE 11–12 Comparison of actual probabilities of dying and probabilities obtained with various models, beyond age 80 (France, 1875 cohort).

b. From Exponential Functions to Logistic Functions

After Gompertz, a number of authors tried to improve the fit between the mortality-curve adjustment formula and the underlying philosophy. Although the exponential growth of mortality with age reflects the degeneration process, one must accommodate the fact that some causes of death are not directly linked to the process. Makeham (1860, 1867) held that some causes of death are independent of age and suggested extrapolating the force of mortality at the upper ages using a Gompertz law modified as follows (Figure 11–12):

$$\mu_x = A + Bc^x$$

where parameter A represents the risk of dying for age-unrelated causes (Thatcher *et al.*, 1998).

The hypothesis of age-unrelated causes of death is altogether questionable, but it is true that the adjustment of the mortality curve by means of the Gompertz formula has to contend with the fact that not all the causes of death vary identically with age, and that some most definitely do not follow an exponential path. That is why Jean Bourgeois-Pichat (1952), in constructing his "limit biological life table," confined his use of Gompertz's law to extrapolating the mortality of the causes that he deemed *endogenous*. Although Bourgeois-Pichat's approach was later taken up by Bernard Benjamin (1982), many authors have tried to refine the Gompertz model for all-causes mortality. In so doing, they have simply added parameters to the basic formula to take into account particular nonexponential factors in the observed-mortality curve (Duchêne, 1980). Other authors have investigated alternative adjustment models.

In the 1930s, Perks (1932) used empirical data to show that the μ_x values of a life table can be fitted very well to a logistic function of the form:

$$\mu_x = c + \frac{ae^{bx}}{1 + \alpha e^{bx}}$$

where a, b, c, and α are constants (Figure 11–12). In fact, when α is equal to 0, the form is identical to the Makeham function. That is why, at the ages when α is small (generally until 80–85 years or so), Makeham's law and Gompertz's law both provide wholly acceptable fits (Figure 11–12).

In the late 1950s, the path opened by Perks was followed again by Beard (1963, 1971) with the so-called Gamma-Makeham model. This was later reworked by James Vaupel *et al.* (1979) to establish a *frailty* model. Other authors, as well, drew inspiration from Perks, confirming the relevance of the Gamma-Makeham model in describing mortality at the oldest ages (Le Bras, 1976; Gavrilov and Gavrilova, 1991; Yashin and Iachine, 1995, 1997).

In their attempt to describe mortality at all ages, Heligman and Pollard (1980) put forward a new exponential model with eight parameters:

$$\frac{q_x}{p_x} = A^{(x+B)^c} + D \cdot \exp\left(-E(\ln x - \ln F)^2\right) + GH^x$$

where the last term is explicitly designed to describe the mortality at over 50 years (Figure 11–12). The authors assume that, beyond age 50, mortality obeys the following relationship:

$$\frac{q_x}{p_x} = GH^x$$

where, of course,

$$p_x = 1 - q_x$$

while G and H are constants: G is the basic, aging-related mortality; H is its growth rate with age.

The relationship can also be written:

$$q_x = \frac{ae^{bx}}{1+ae^{bx}}$$

This is equivalent to admitting that, from age 50 up, the probability of dying follows a logistic function, with values ranging between 0 and 1:

$$Logit(q_x) = \ln(a) + bx$$

In the 1990s, Väino Kannisto *et al.* (1994) showed that—in countries with reliable data—the age-related growth of the force of mortality slows from 75 to 80 years onward. At those ages, its shape becomes very similar to a logistic curve. These authors, as well, proposed a function of the type:

$$\mu_x = \frac{ae^{bx}}{1+ae^{bx}}$$

which can also be written:

$$Logit(\mu_x) = \ln(a) + bx$$

(Figure 11–12).

Another interesting proposal was made by Coale and Kisker (1990) to describe mortality between ages 85 and 110. Their model is of the form:

$$\ln(\mu_x) = a + bx + cx^2$$

where c is negative. This is known, in the literature, as the *quadratic model* (Figure 11–12).

To assess the capacity of these different models to fit the data actually observed, we have applied them to the mortality of the French female cohort of 1875 determined with the extinct-cohort method (Figure 11–11). The best fit is obtained with the logistic models.

For a long time, demographers have noted that the growth rate of the force of mortality decreases at the oldest ages. In addition, the slowing was attributed to the degradation of data quality at those ages. However, with the improvement in data and the rapid rise in the number of elderly due to greater life expectancy, it is increasingly being noticed that mortality ceases to follow the Gompertz model beyond a certain age (Horiuchi and Coale, 1990; Kannisto, 1994; Thatcher *et al.*, 1998; Wilmoth and Horiuchi, 1998). An initial explanation of this phenomenon may be found in the differences in frailty between individuals of the same cohort (Vaupel and Yashin, 1985; Vaupel and Carey, 1993).[13] The fall in the growth rate of mortality with age would, on this assumption, be due to a change in the cohort structure by frailty level. Each subset with the same frailty effectively follows a Gompertz distribution, whose level is specific to its degree of vulnerability. However, as the most vulnerable subsets are eliminated sooner, the rate of increase of the overall mortality risk slows. A second possible explanation is that the health improvements among the very elderly actually benefit the hardiest individuals rather than the weakest. Here, it is not the composition of the population by degree of frailty that changes, but rather the difference in mortality between frailty groups that widens.

If we held to the first explanation, this deviation from the Gompertz law at the upper ages might not be incompatible with an immutable theoretical extreme age at death. By contrast, with the second explanation, the hypothesis of the fixed limit age becomes much harder to defend.

An analysis of these dynamic relations between heterogeneity and mortality at the oldest ages in the French population (Barbi *et al.*, 2003) has shown that, in reality, both processes are at work.

Figure 11–13 shows that for French women, from the oldest to the most recent cohorts, the slowdown in age-related mortality growth observed for the total population (μ) is due to two concurrent processes: (1) a homogenization of the cohort group in terms of indi-

[13] See also Chapter 21, by James Vaupel and Anatoli Yashin.

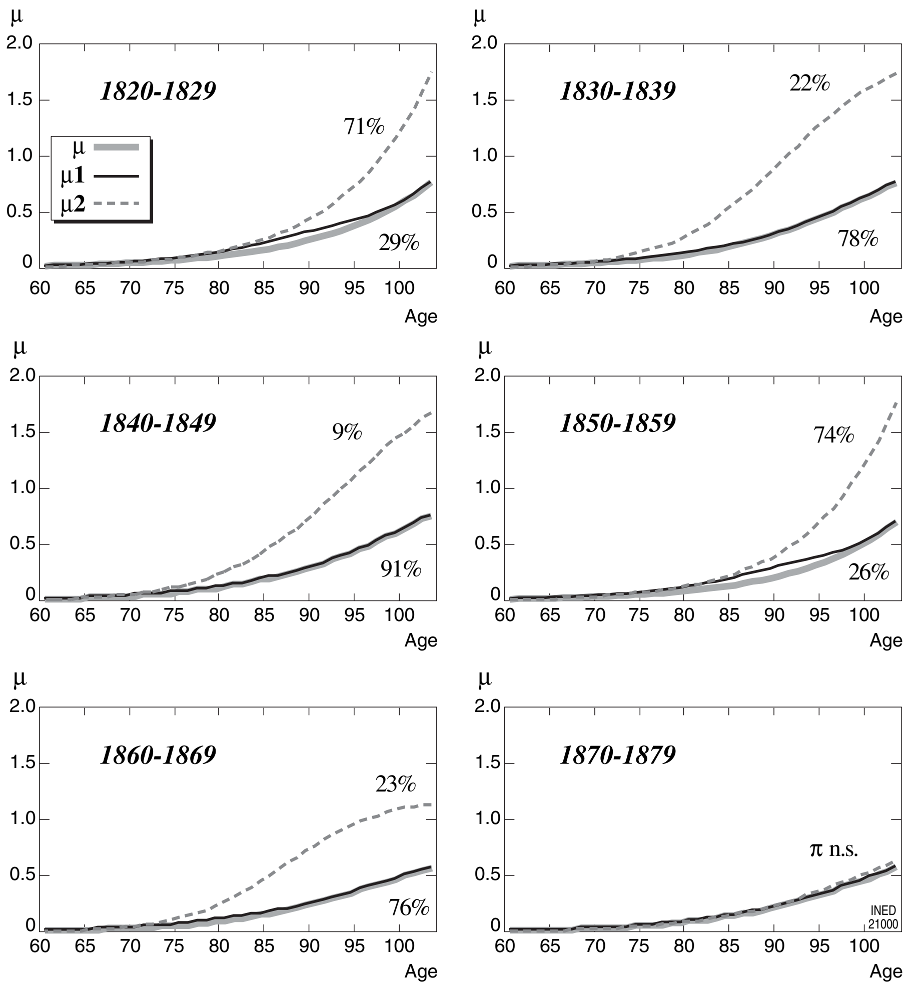


FIGURE 11–13 Estimation of the force of mortality for two subsets of individuals with different frailty (μ_1 for the frailest and μ_2 for the hardiest) and for the entire population (μ) in the six groups of French female cohorts. (Source: Barbi *et al.*, 2003.)

vidual frailty (the mortality of the frailest being much higher, their proportion in the cohort falls very rapidly with age) and (2) a slowdown in age-related mortality growth, among the frailest (μ_1) and the hardiest (μ_2) alike.

V. THE SPECIAL CASE OF INFANT MORTALITY

Although mortality at the oldest ages poses theoretical problems that are hard to solve, mortality at young ages also deserves special attention (but for historical and practical reasons rather than theoretical ones). The issue here is not so much to find a remedy to a still-unsolved problem but to avoid confusing the various indicators that have been enshrined in standard usage for practical reasons.

Owing to the preeminence of child mortality in the overall mortality of human populations, and to its enduring major importance in some populations even today, it was soon realized that child-mortality analysis should take precedence over all other investigations of mortality as a tool for conducting health policy. Numerous studies in historical demography have shown that until the mid-18th century in western Europe and even until the early 20th century in several countries of central/eastern Europe (Vallin, 1989, 1991), nearly half the members of a newborn cohort died before age 5 and almost 30% before their first birthday. Today, no population experiences such a massive loss in peacetime, but in some developing countries, particularly in rural areas, we can still observe losses in excess of 10% before age 1 and of almost 20% before age 5. In such situations, child mortality carries an extremely heavy weight in cohort life expectancy. Thus, we can understand why it has been tracked for so long, even in the absence of any resources for a sophisticated measurement of mortality.

In fact, the earliest current statistics on death distinguished deaths at under 1 year of age from total deaths. Even today, in many situations where only rough mortality statistics are available, the figures almost always show at least the total of deaths at all ages and the number of deaths of infants less than a year old.

The *infant mortality rate* is the simplest mortality indicator that one can calculate, after the crude death rate—and sometimes even ahead of it. However, the indicator designated by that standard term is neither a rate nor a probability according to any one of the several definitions discussed in this chapter. In reality, it is the ratio of deaths at under age 1 observed in the course of a calendar year to the number of births in that same year:

$$m_i = \frac{D_{0,1}}{B}$$

It is not a rate, as it does not express the ratio of deaths to an average population exposed to the risk. Nor is it a probability since (apart from the migration issue explored earlier) we can easily see in Figure 11–14, which plots the computation elements on a Lexis diagram, that only a portion of the deaths constituting the numerator occurred among children born during the year studied.

In the example of Figure 11–14, which concerns French male mortality in 1920, the infant mortality rate works out at:

$$m_i = \frac{35,433 + 13,692}{432,036} = 113.7/1000$$

whereas the age-0 mortality rate calculated in the Lexis square comes to:

$$m_0 = \frac{35,433 + 13,692}{(234,700 + 396,000)/2} = 155.8/1000$$

and the average of the two cohorts' probabilities of dying at age 0 contributing to the deaths used to compute the infant mortality rate is:

$$m_0 = \frac{26,417 + 35,433 + 13,692 + 14,925}{(260,455 + 432,036)} = 130.6/1000$$

The infant mortality rate is well below the death rate at age 0 (113.7/1000 versus 155.8/1000). This is quite normal insofar as the same number of deaths is divided in the first case by the number of births, and in the second by an average population of children aged under one year—who, all other things being equal, have already experienced part of the infant mortality. This initial differentiation factor is the same as the one that separates an age-specific rate from a probability.

Again, however, we can clearly see in this example that there is a great difference between the infant mortality rate and the average of the two classic probabilities at age 0 (113.7/1000 thousand instead of 130.6/1000). This is mainly due to the sharp fluctuation in the number of births linked to the end of World War II.

To obtain the exact measure of a true probability of dying, which would reflect the deaths composing the numerator of the infant mortality rate, we should (1)

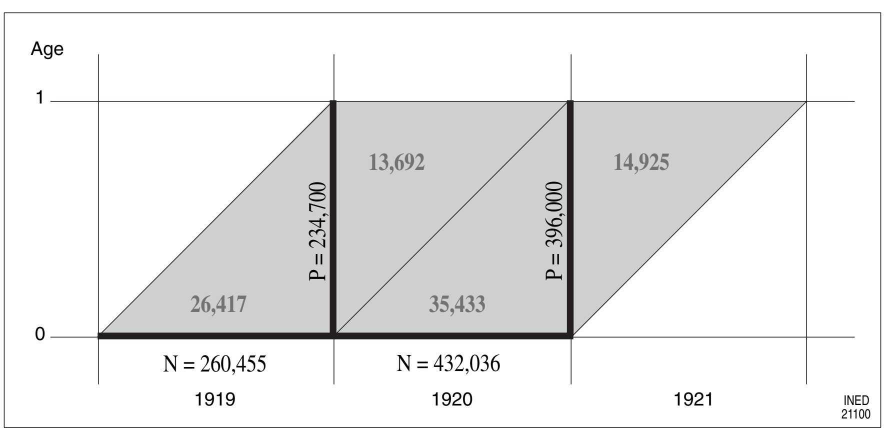


FIGURE 11–14 Lexis diagram locating the data that can be used for different measures of infant mortality (France, males, 1920). Note: To allow for the change in the territory covered by the statistic, the 209,185 births and 21,217 deaths at under age 1 of the 1919 cohort registered in 1919 have been weighted by the ratio 234,700 : 188,500 (population under age 1 on January 1, 1920, for the two territorial definitions).

compute the ratio of the deaths in the first Lexis triangle to 1920 births only and (2) determine the ratio of deaths in the second triangle to 1919 births:

$$m'_i = \frac{13,692}{260,455} + \frac{35,433}{432,036} = 134.6 / 1000$$

There remains a small difference between this adjusted infant mortality rate and the average of the two classic probabilities, owing to the fact that infant mortality varies from year to year. To obtain a measure from classic probabilities centered on 1920, we use three observation years. If mortality were constant, these two measures would be strictly identical.

However, to perform this type of correction—which transforms an infant mortality rate into a true probability of dying (if we disregard the migration issue)—we need the breakdown by cohort of deaths at under age 1 observed during the year. Such figures are seldom available. That is why demographers sometimes weight infant mortality rate by births; for this purpose, they refer to a standard distribution of deaths at under age 1 between the Lexis triangles corresponding to the mortality level of the population studied. For an infant mortality in the range of 100 to 150 per thousand, the ratio ranges between two thirds and one third. In the previous case, we would therefore have estimated a weighted infant mortality rate, which can be equated with a classic probability as follows:

$$m'_i = \frac{(35,433 + 13,692)}{432,036 \cdot 0.67 + 260,455 \cdot 0.33} = 130.9 / 1000.$$

This result is indeed close to the 134.6 computed with the true distribution.

Once again, however, we have taken a fairly uncommon example of a sharp fluctuation in the number of births to highlight the measurement differences tied to the differences in the definition of the various possible indicators of infant mortality.

In summary, the conventional infant mortality rate is never equivalent to a true death rate at age 0. On the other hand, the conventional rate is strictly comparable to the classic probability of dying at age 0 if the number of births is constant and if mortality does not vary in the course of the 3 years centered on the rate measurement. In other words, if the real fluctuations in births and deaths are small, the conventional rate can be equated with the probability; if the birth fluctuations are significant, we can equate the two after weighting the infant mortality rate by the number of births concerned.

Apart from these issues linked to the definition of commonly used indicators, let us note in conclusion that the measurement of infant mortality is sometimes impeded by problems of defining life and death. Registered deaths are registered precisely by virtue of the legal definition that makes it possible to distinguish between *live births* and *stillbirths*. This definition differs from country to country. Some statistical institutes try—with varying degrees of success—to compile additional information to identify, among the stillborn, those who actually *lived* without being able to attain the legal status of live births: these are known as *false stillbirths*. Such attempts are broadly consistent with the universal definition proposed by the World Health Organization (WHO).[14] To analyze infant mortality with precision, we need to investigate the definition that actually matches the statistical tables used. In France, for example, INSEE (the national statistical institute) has published a special statistical series on the stillborn that distinguished between the true, the false, and the undetermined stillborn. Since 1975, the Institute has simply incorporated the false stillbirths into the birth statistics as well as the death statistics—a decision that does not, however, ensure total consistency with the WHO definition.

References

ALLARD Michel, 1991. *À la recherche du secret des centenaires*. Paris, Le Cherche Midi, 273 p.

ALLARD Michel, LÈBRE Victor, and ROBINE Jean-Marie, 1998. *Jeanne Calment, from Van Gogh's time to ours, 122 extraordinary years*. New York, W. H. Freeman and Company, 136 p.

ALLARD Michel, LÈBRE Victor, and ROBINE, Jean-Marie, 1994. *Les 120 ans de Jeanne Calment, doyenne de l'humanité.* Paris, Le Cherche Midi, 162 p.

ALLARD Michel, VALLIN Jacques, ANDRIEUX Jean-Michel, and ROBINE Jean-Marie, 1996. In the search of the secret of centenarians: a French demographic and medical survey about centenarians, *in:* Graziella Caselli and Alan Lopez (eds.), *Health and mortality among elderly populations*, p. 61–86. Oxford, Clarendon Press, xvip. + 360 p. (International Studies in Demography).

BARBI Elisabetta, CASELLI Graziella, and VALLIN Jacques, 2003. *Trajectories of extreme survival in heterogeneous populations. Population* (English Edition), Vol. 58(1), p. 43–66.

BEARD Robert E., 1963. A theory of mortality based on actuarial, biological and medical considerations, *in:* IUSSP (ed.), *International Population Conference. Congrès international de la population, New York, 1961. Tome I*, p. 611–626. London, IUSSP and UNESCO, 798 p.

BEARD, Robert E., 1971. Some aspects of theories of mortality, cause of death analysis, forcasting and stochastic processes, *in:* William

[14] "Live childbirth is defined as the expulsion or complete extraction from the mother's body, irrespective of the gestation period, of a product of conception that, after this separation, breaths or displays another vital sign, such as heartbeat, pulsation of the umbilical cord, or actual contraction of a muscle subjected to an act of will, regardless of whether the umbilical cord has been severed or not, and of whether the placenta remains attached or not. Any product of such a birth is regarded as a live birth" (World Health Organization, 1977).

Brass (ed.), *Biological aspects of demography*, p. 57–68. London, Taylor & Francis, 176 p.

Benjamin Bernard, 1982. The span of life, *Journal of the Institute of the Actuaries*, vol. 109, p. 319–340.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Cleland John and Scott Chris, 1987. *The World Fertility Survey: an assessment.* Oxford, Oxford University Press, 1058 p.

Coale Ansley and Demeny Paul, 1983. *Regional model life tables and stable populations,* 2nd ed. New York and London, Academic Press, 496 p.

Coale Ansley J. and Kisker Ellen E., 1990. Defects in data on old-age mortality in the United States new procedures for calculating mortality schedules and life tables at the highest ages, *Asian and Pacific Population Forum*, vol. 4(1), 32 p.

Das Gupta Monica, Aaby Peter, Garenne Michel, and Pison Gilles, 1997. *Prospective Community Studies in Developing Countries*. Oxford, Clarendon Press, 350 p. (Coll. International Studies in Demography)

Duchene Josianne, 1980. *Un essai de modèlisation de la répartition des décès selon l'age et la cause dans les pays industrialisés*. Louvain la Neuve, Cabay, 182 p.

Dupâquier Jacques, 1996. *L'invention de la table de mortalité.* Paris, PUF, 178 p.

El zanaty Fatma H., Sayed Hussein, Zaky Hassan H., and Way Ann A., 1993. *Egypt Demographic and Health Survey 1992*. Cairo, National Population Council and Calverton, Macro International Inc., 317 p.

Gavrilov Leonid and Gavrilova Natalia, 1991. *The biology of life span: a quantitative approach*. New York, Harwood Academic, 385 p.

Gompertz Benjamin, 1825. On the nature of the function expressive of the law of human mortality. *Philosophical Transactions,* vol. XXVII, p. 513–585.

Halley Edmund, 1693. An estimate of the degrees of mortality of mankind . . . , *Philosophical Transactions,* vol. XVII, p. 596–610, 653–656.

Heligman Larry and Pollard John H., 1980. The age pattern of mortality, *Journal of the Institute of the Actuaries*, vol. 107, p. 49–80.

Horiuchi Shiro and Coale Ansley, 1990. Age patterns of mortality for older women: an analysis using the age specific rate of mortality change with age, *Mathematical Population Studies*, vol. 2(4), p. 245–267.

Horiuchi Shiro and Wilmoth John, 1998. Deceleration in the age pattern of mortality at older ages, *Demography,* vol. 4, p. 391–412.

ISI, 1986. *The World Fertility Survey. Final report.* Voorburg (The Netherlands), International Statistical Institute, 228 p.

Kannisto Väinö, 1994. *Development of oldest-old mortality, 1950–1990: evidence from 28 developed countries.* Odense (Denmark), Odense University Press, 108 p. (Odense Monographs on Population Aging no.1).

Kannisto Väinö, 1996. *The advancing frontier of survival.* Odense (Denmark), Odense University Press, 144 p. (Odense Monographs on Population Aging no.3).

Kannisto Väinö, Lauritsen, Jens; Thatcher, A. Roger, and Vaupel, James W., 1994. Reductions in mortality at advanced ages: several decades of evidence from 27 countries. *Population and Development Review*, vol. 20(4), December, p. 793–810.

Keyfitz Nathan, 1977. *Applied mathematical demography*. New York, John Wiley and Sons, 388 p. (Reedited in1985 by Springer-Verlag, New York, 442 p).

Le bras Hervé, 1976. Lois de mortalité et âge limite, *Population,* vol. 31(3), p. 655–692.

Makeham William M., 1860. On the law of mortality and the construction of annuity tables. *Assurance Magazine*, vol. 8, p. 301–310.

Makeham William M., 1867. On the law of mortality. *Journal of the Institute of the Actuaries*, vol. 13, p. 325–367.

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIXe siècle. *Population*, vol. 44(6), p. 1121–1158. (Also published in English in *Population: An English Selection* vol. 46(3), 1991, p. 33–62).

Meslé France, Vallin Jacques, and Robine Jean-Marie, 2000. *Survivre à plus de 110 ans en France. Gérontologie et société,* vol. 94, p. 101–120.

Moivre (de) Abraham, 1725. *Annuities on lives: or, the valuations of annuities upon any number of lives; as also, of reversion*. London, 125 p.

Perks Wilfred, 1932. On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries*, vol. 63, p. 12–57.

Pollard John, 1973. *Mathematical models for the growth of human populations*. Cambridge, Cambridge University Press, 186 p.

Smith David and Keyfitz Nathan, 1977. *Mathematical demography: selected papers*. Berlin, Heidelberg and New York, Springer, 514 p.

Thatcher Roger, Kannisto Väinö, and Vaupel James, 1998. *The force of mortality at ages 80 to 120*. Odense, Odense University Press, 104 p. + fig. and tabl. (Coll. Odense Monographs on Population Aging, vol. 5).

Vallin Jacques, 1973. *La mortalité par génération en France depuis 1899*. Paris, INED, PUF, 484 p. (Coll. Travaux et Documents, Cahier no. 63).

Vallin Jacques, 1989. La mortalité en Europe de 1720 à 1914: évolution à long terme et changements de structure par âge. *Annales de démographie historique*, p. 55–77.

Vallin Jacques, 1991. Mortality in Europe from 1720 to 1914: long term trends and changes in patterns by age and sex, *in:* Roger Schofield, David Reher, and Alain Bideau (eds.), *The decline of mortality in Europe,* p. 38–67. Oxford, Clarendon Press, 270 p.

Vallin Jacques, 1993. *L'avenir de l'espérance de vie*. Paris, INED//PUF, 106 p. (Congrès et colloques, no. 12).

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle*, INED/PUF, 102 p.

Vaupel James and Carey James R., 1993. Compositional interpretations of medfly mortality, *Science,* vol. 260, p. 1666–1667.

Vaupel James and Yashin Anatoli I., 1985. Heterogeneity's ruses: some surprising effects of selection on population dynamics, *American Statistician,* vol. 39, p. 176–185.

Vaupel James W., Manton Kenneth G., and Stallard Eric, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography,* vol. 16, p. 439–454.

Vincent Paul, 1951. La mortalité des vieillards. *Population*, vol. VI (2), p. 181–204.

Who, 1977. *International Classification of Diseases. 1975 Revision,* vol. 1. Geneva, World Health Organization, 782 p.

Yashin Anatoli I. and Iachine Ivan A., 1995. Survival of related individuals: an extension of some fundamental results of heterogeneity analysis, *Mathematical Population Studies*, vol. 5(4), p. 321–339.

Yashin Anatoli I. and Iachine Ivan A., 1997. How frailty models can be used in evaluating longevity limits, *Demography*, vol. 34(1), February, p. 31–48.

APPENDIX 11–1 Life Table of French Female Cohort Born in 1875

X	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x	x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x
0	0.154049	100,000	15,405	47.22	51	0.010634	54,478	579	23.95
1	0.050193	84,595	4,246	54.73	52	0.011059	53,899	596	23.20
2	0.027502	80,349	2,210	56.60	53	0.012259	53,303	653	22.45
3	0.016098	78,139	1,258	57.18	54	0.012496	52,649	658	21.72
4	0.012491	76,881	960	57.11	55	0.013554	51,992	705	20.99
5	0.009894	75,921	751	56.83	56	0.013668	51,287	701	20.27
6	0.007864	75,170	591	56.39	57	0.015185	50,586	768	19.55
7	0.006896	74,579	514	55.83	58	0.015310	49,818	763	18.84
8	0.005956	74,064	441	55.22	59	0.017641	49,055	865	18.13
9	0.005072	73,623	373	54.55	60	0.017932	48,190	864	17.44
10	0.004487	73,250	329	53.82	61	0.019304	47,326	914	16.75
11	0.004448	72,921	324	53.06	62	0.020905	46,412	970	16.07
12	0.004231	72,597	307	52.30	63	0.022409	45,442	1,018	15.41
13	0.004337	72,290	313	51.52	64	0.028426	44,423	1,263	14.75
14	0.004607	71,976	332	50.74	65	0.029473	43,161	1,272	14.16
15	0.005856	71,645	420	49.97	66	0.029950	41,889	1,255	13.58
16	0.006343	71,225	452	49.26	67	0.031490	40,634	1,280	12.98
17	0.006875	70,773	487	48.57	68	0.038462	39,354	1,514	12.39
18	0.006747	70,287	474	47.91	69	0.040659	37,841	1,539	11.86
19	0.006617	69,813	462	47.23	70	0.039111	36,302	1,420	11.35
20	0.006885	69,351	477	46.54	71	0.040694	34,882	1,420	10.79
21	0.007200	68,873	496	45.86	72	0.040508	33,463	1,356	10.22
22	0.007348	68,377	502	45.19	73	0.052291	32,107	1,679	9.63
23	0.007257	67,875	493	44.52	74	0.052152	30,428	1,587	9.14
24	0.006994	67,382	471	43.84	75	0.063097	28,842	1,820	8.61
25	0.007338	66,911	491	43.15	76	0.064150	27,022	1,733	8.16
26	0.007142	66,420	474	42.46	77	0.078671	25,288	1,989	7.69
27	0.007012	65,946	462	41.76	78	0.075801	23,299	1,766	7.30
28	0.007070	65,483	463	41.05	79	0.084156	21,533	1,812	6.86
29	0.006725	65,020	437	40.34	80	0.096753	19,721	1,908	6.44
30	0.007192	64,583	464	39.61	81	0.100663	17,813	1,793	6.08
31	0.007331	64,118	470	38.90	82	0.109283	16,020	1,751	5.70
32	0.007265	63,648	462	38.18	83	0.121472	14,269	1,733	5.34
33	0.006901	63,186	436	37.46	84	0.133219	12,536	1,670	5.01
34	0.006860	62,750	430	36.71	85	0.138747	10,866	1,508	4.70
35	0.007191	62,320	448	35.96	86	0.160105	9,358	1,498	4.38
36	0.007050	61,871	436	35.22	87	0.185436	7,860	1,457	4.12
37	0.007101	61,435	436	34.47	88	0.175424	6,402	1,123	3.94
38	0.007453	60,999	455	33.71	89	0.194571	5,279	1,027	3.67
39	0.007400	60,544	448	32.96	90	0.209926	4,252	893	3.44
40	0.007785	60,096	468	32.20	91	0.231730	3,359	778	3.22
41	0.007408	59,628	442	31.45	92	0.242579	2,581	626	3.04
42	0.008019	59,187	475	30.68	93	0.252294	1,955	493	2.86
43	0.012582	58,712	739	29.92	94	0.285658	1,462	418	2.65
44	0.007801	57,973	452	29.30	95	0.302464	1,044	316	2.51
45	0.008095	57,521	466	28.53	96	0.324409	728	236	2.39
46	0.008766	57,055	500	27.75	97	0.349563	492	172	2.29
47	0.008432	56,555	477	27.00	98	0.336889	320	108	2.26
48	0.009004	56,078	505	26.22	99	0.340687	212	72	2.15
49	0.009828	55,573	546	25.45	100		140	140	2.00
50	0.009978	55,027	549	24.70					

(Source: Vallin and Meslé, 2001)

C H A P T E R

12

Cohort Analysis of Fertility

GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

I. ANALYZING COHORT FERTILITY FROM RETROSPECTIVE DATA

Chapter 8 showed that measures of the quantum and tempo of fertility in a birth cohort[1] could be obtained either by a retrospective or a prospective observation using fertility frequencies (*age-specific fertility rates* [ASFR]), if the condition of *causal independence* (as defined in Chapter 9) between fertility and the censoring phenomena (mortality, migration) is satisfied. This condition is usually assumed, as previously discussed, when one raises the question of what would have been the quantum and tempo of fertility if there had been no mortality or migration affecting the cohort. This chapter will deal mainly with *retrospective* fertility data, although the *prospective* approach will also be considered.

1. Average Number of Children Born

As pointed out in Chapter 8, suppose that at a survey women are interviewed at the end of their childbearing period, say at age 55, on the number of live-born children they have had during their lifetime and on their age at childbearing (i.e., on their *birth history* or their *maternity history*[2]). Let F_{55} stand for the number of women interviewed and B_x for the number of births occurring at completed age x in this cohort of women. Then, the ratio $f_x = B_x/F_{55}$ represents the age-specific fertility *rate* or frequency at age x for those women interviewed at the end of their childbearing period.

The sum of these *rates* over all x, from puberty to menopause, will yield the average number of children born or average family size at the end of the childbearing period. This *cohort total fertility (TF) rate* (TFR), or *completed fertility rate* (CFR) as it is sometimes called, characterizes the intensity or quantum of fertility and can be written:

$$TF = \sum_{x=15}^{54} \frac{B_x}{F_{55}}$$

Two measures of the *tempo* of fertility can also be obtained from these data (i.e., the average age at childbearing $\bar{x}$ and the corresponding variance v). The mean age at childbearing is simply the weighted average of ages at birth, the weights being the ASFRs at each age:

$$\bar{x} = \frac{\sum x.B_x}{\sum B_x} \text{ or } \bar{x} = \frac{\sum x \frac{B_x}{F_{55}}}{\sum \frac{B_x}{F_{55}}} = \frac{\sum x f_x}{\sum f_x}$$

The variance of age is similarly obtained by the expression:

$$v = \frac{\sum (x-\bar{x})^2 .B_x}{\sum B_x} \text{ or } v = \frac{\sum (x-\bar{x})^2 \cdot \frac{B_x}{F_{55}}}{\sum \frac{B_x}{F_{55}}} = \frac{\sum (x-\bar{x})^2 f_x}{\sum f_x}$$

all sums covering the whole reproductive period.

[1] A cohort viewpoint is adopted here, although, as noted in Chapter 10, one must always remember the critique that has been addressed to cohort studies (Bhrolcháin, 1992; Bongaarts and Feeney, 1998) before embarking on a cohort approach instead of on a period approach.

[2] *Maternity* histories are broader than *birth* histories, as the former collect data on all pregnancies and not only on live births.

If one is only interested in the sole behavior of the subcohort interviewed at the end of the reproductive period, whatever the possible selection effects it has incurred by mortality and migration, it is not necessary to assume independence between fertility and mortality and migration, as this condition is then without purpose. To give other examples, one may examine whether cohabitation is followed by marriage or if change of residence occurs before or after marriage, without assuming independence between the various factors (Courgeau and Lelièvre, 1994). In both of these cases, individuals are not lost to follow-up as when censoring occurs; they only change their characteristics or states. On the contrary, when individuals are lost to follow-up, one should consider whether these individuals differ from those who remain under observation (i.e., if selection effects operate or not on the process one is studying).[3] This is the case when people are lost from observation through mortality or international migration or—in a prospective survey—if they are not followed anymore for various reasons, such as survey dropout. Any withdrawal from observation can bias the results of the study through selection effects (see e.g., Alderson, 1983, Chapter 6).

Consider for example the *Fertility and Family Surveys* (FFS) conducted in Europe under the auspices of the Population Activities Unit of the United Nations Economic Commission for Europe since the late 1980s (Nikander, 1998; Schoenmaeckers and Lodewijckx, 1999). The purpose of these sample surveys is to document various aspects of partnership and reproductive behavior among a sample of women in reproductive ages. The analysis is focused here on the survivors, and not on those who have died or migrated,[4] although *nonresponse* may severely bias the results (Frankfort-Nachmias and Nachmias, 1996, Chapter 8). Contrary to the vital registration system, surveys collect much more information on the characteristics of the females concerned; moreover, these facts are available individually and not as aggregates as in the case of vital statistics. If the latter are still useful for monitoring period trends in fertility, by age or marriage duration for example, they are not of much help for studying the factors that have led to fertility changes, as the amount of information vital statistics contain is minimum. Furthermore, as cohabitation has become common during the last decades in the developed countries, vital statistics are not suited for this situation either, as only the legal marital status of the individual is recorded. Thus, sample surveys have become the main source of data for fertility analysis, both in developed and in less-developed countries.

As stated in a report of the FFS (Macura and Klijzing, 1997; p. 3):

> thus far changes in partnership and reproductive behavior have mostly been documented using data from population census, vital registration and/or population registers. As the depth and breadth of these data vary a great deal across countries, the composite picture they can provide is at best patchy. Moreover, these data are often incapable of shedding light on the more subtle aspects of partnership behavior, such as the onset of sexual relationship, the use of modern contraception, and living apart together, to mention just a few. Also, these data often preclude an in-depth analysis of interactions, such as those between school attendance and work on the one hand, and union formation and childbearing, on the other.

The following sections will therefore deal mainly with *retrospective* fertility survey data, and more particularly with the single-round survey,[5] although the link with the approach using vital statistics will be identified when necessary. Both of these data sources are actually useful in practice. As noted above, and as Handcock *et al*. (2000) among others have pointed out, survey data contain more detail with respect to demographic events and the characteristics of the individuals who do or do not experience these events. On the other hand, registration data are collected for all events in the population. This eliminates sampling error and sample-selection bias. Registration data are also less affected by reporting biases than retrospectively collected data. Moreover, vital statistics enable year-to-year monitoring of fertility trends, for which surveys are less well equipped.

2. Fertility by Birth Order

Suppose a sample of women is interviewed at the end of their reproductive period. If the date of birth of each of their children born alive[6] is recorded, and most usually it is, the *birth history* (marital and extramarital) of each woman is recorded. Knowing the date of each birth and therefore the age of the mother at the birth of her children, one may easily compute the arithmetic average of ages, giving the average age at childbear-

[3] This problem cannot be solved by methods alone, including more sophisticated approaches such as survival analysis; one needs information on the causal relations between the variables concerned and between these and possible common causes.

[4] Although one probably assumes implicitly that the sample interviewed at age 50 represents the past experience of the whole population.

[5] Other types of surveys exist, but they are less common in demography (Tabutin, Chapter 121 in Volume 4; Cleland, 1996).

[6] In demography, fertility is usually restricted to *live births* only. Surveys also often collect data on late fetal deaths, although the quality of these data is probably poorer than that for live births. For the study of fetal mortality, see Chapter 32 by Gourbin in Section 2.

ing for this group of women. As seen above, the average number of births per woman[7] or average family size at the end of the reproductive period, also called total fertility rate (TFR) or completed fertility rate (CFR), is simply the total number of live births declared at a postreproductive age divided by the total number of women at that age, including those women who have never had a child (i.e., the infertile[8] women of ultimate parity 0). As one also knows the birth order of each child, age-specific fertility rates $f_x(k)$ may be computed at age x for each birth order k, yielding the average number $TF(k)$ of births of order k per woman:

$$TF(k) = \sum_x f_x(k)$$

One can also compute the average age at childbearing of birth order k, and the variance of ages around the mean as a measure of dispersion.

Note also that the average number of births of all orders is simply the sum of the average number of births of each order:

$$TF = TF(1) + TF(2) + \ldots + TF(k) + \ldots$$

or

$$TF = \sum_k \sum_x f_x(k)$$

As computers prefer crunching numbers rather than years, months, and days, one usually transforms dates expressed in months and years into *century-months* as follows. Take January 1900 as month 1. March 1950 is therefore equal to 50 years times 12 months plus 3 months (i.e., month 603 in the new system). With this recoding of dates, it is easy to compute the interval between two events. For example, if a woman was born during the month of February 1930 (code 362) and had her first child in July 1951 (code 619), her *age at first birth* is then equal to 619 − 362 = 257 months (i.e., 21 years and 5 months). One can compute the same interval for all women having had at least one birth. The arithmetic average gives the mean length of time between birth of the mother and birth of her first child[9] (i.e., the average age at first birth). A similar code can be developed for years, months, and days. In this case, one must however take into account the different number of days per month as well as the leap years.

One may do the same computation for intervals between births of order k and births of order $k + 1$, giving the average *birth interval*[10] between k and $k + 1$, another measure of the tempo of fertility. As birth intervals depend upon the total number of children born,[11] birth intervals should be compared only among women having the same ultimate family size. Changes in birth intervals are useful indicators of changes not only of the timing of births but also in the quantum of fertility; in high-fertility countries, if fertility decreases birth intervals are usually extended. In low-fertility countries, one may nevertheless observe significant changes in the timing of births, such as the postponement of the first birth, without major changes in the total number of children born (Granström, 1997).

Suppose now that these women aged 50 to 55 years, for example, are classified at the survey according to their achieved family size, w being the maximum number of children born:

Parity	0	1	2	. . .	k	. . .	w
No. women	F_0	F_1	F_2	. . .	F_k	. . .	F_w

Tables like these on family size are often available for censuses too, even those held a long time ago, where a question on the total number of children born has been asked. This distribution can also be combined with year of marriage, age at marriage, and duration of marriage (Anderson, 1998). As stated before, the average family size is then:

$$TF = \frac{\sum_0^{\omega} kF_k}{\sum_0^{\omega} F_k}.$$

Furthermore, the proportion p_k of women having borne exactly k children during their lifetime is:

$$p_k = \frac{F_k}{\sum_0^{\omega} F_k}.$$

The probability a_k of moving from parity k to parity $k + 1$, called the *parity progression ratio*,[12] can then be obtained as follows. Let $TF(1)$, $TF(2)$, . . . $TF(k)$, . . . represent as before the average number of births of order 1, 2, . . . k per woman at the end of

[7] *Descendance moyenne*, in French.

[8] Remember that the term *fertile* is translated by *féconde* in French, whereas *fecund* is translated by *fertile*! With confusions such as these, one understands why the British and the French have quarreled in the past!

[9] Called *intervalle proto-génésique* in French.

[10] *Intervalle intergénésique* in French. Because of the skewed distribution of interval lengths, caused by a small number of very long intervals, the *median* is a better measure than the *mean* (Newell, 1988).

[11] Intervals have to be short to have more than 12 children over one's lifetime!

[12] In French, *probabilitéé d'agrandissement des familles*.

their reproductive period, i.e. the proportion of women having had at least 1, 2, . . . k children. One then has:

$$a_0 = TF(1)$$

$$a_1 = TF(2)/TF(1)$$

$$a_2 = TF(3)/TF(2)$$

$$\ldots$$

$$a_k = TF(k+1)/TF(k)$$

$$\ldots$$

One obtains:

$p_0 = 1 - TF(1)$ and $TF(1) = 1 - p_0$

$p_1 = TF(1) - TF(2)$ and $TF(2) = TF(1) - p_1$ or $1 - p_0 - p_1$

$p_2 = TF(2) - TF(3)$ and $TF(3) = TF(2) - p_2$ or $1 - p_0 - p_1 - p_2$

$$\ldots$$

$p_k = TF(k) - TF(k+1)$ and $TF(k) = 1 - p_0 - p_1 - p_2 - \ldots - p_{k-1}$

$$\ldots$$

The *parity progression ratios* will therefore be equal to:

$a_0 = TF(1) = 1 - p_0$,
$a_1 = TF(2)/TF(1) = (1 - p_0 - p_1)/(1 - p_0)$,
$a_2 = TF(3)/TF(2) = (1 - p_0 - p_1 - p_2)/(1 - p_0 - p_1), \ldots$

$$a_k = TF(k+1)/TF(k) = (1 - p_0 - p_1 - p_2 - \ldots - p_k)/(1 - p_0 - p_1 - p_2 - \ldots - p_{k-1})$$

$$\ldots$$

These parity progression ratios state the *ex post facto* frequency of passing from parity k to parity $k+1$ for the birth cohort considered.

Alternatively, knowing the parity progression ratios, it is easy to derive the average number of births of order k per woman. From the above relations, one has:

$$TF(1) = a_0$$

$$TF(2) = a_1\, TF(1) = a_1 a_0$$

$$\ldots$$

$$TF(k+1) = a_k\, TF(k) = a_k a_{k-1} a_{k-2} \ldots a_1 a_0$$

The average number of children of all birth orders being $TF = \Sigma_k\, TF(k)$ one has:

$$TF = a_0 + a_0 a_1 + a_0 a_1 a_2 + \ldots + \prod_{j=0}^{k-1} a_j + \ldots$$

Birth history data obtained from sample surveys are not exempt from problems. First, as the condition of independence is probably not fulfilled, it is not recommended to interview women at very old ages to reconstruct the fertility history of their birth cohort. The subcohort of women surviving at the time of the survey might be a selected group with respect to fertility, if there is an association between fertility on one hand and mortality and international migration on the other hand. In this case, the results from the survey will not adequately stand for the experience of the whole cohort, including those women who have died or migrated before the interview. Another problem with fertility surveys, as with all surveys, is the sampling error. For example, for a sample of 5000 women, the sampling error for an age-specific fertility *rate* (1 year of age and 1 calendar year) is around 20% to 25% (Verma, 1982; Little, 1982). Survey data are therefore often aggregated into 5-year age groupings per calendar year or five calendar years per single year of age, or both.

If the sample is restricted to women aged 15 to 50, as is often the case, another problem arises when one derives period data from the sample; this is often done in countries where the vital registration system is defective. The further one goes backward in time from the survey date, the more data are missing at older ages. This is due to a censoring effect: as the cutting point of the sample has been fixed at, for example, 50 years of age, and as period analysis requires data from older and older cohorts as one goes back in time, one increasingly underevaluates period fertility as one moves away from the survey date.[13] In that case, fertility at higher ages has to be extrapolated, if one wishes to evaluate period fertility covering all reproductive ages. Another solution is to compare period fertility up to a certain age x only ($x < 50$). For example, one could compare period fertility until age 35 during the 15-year period preceding the time of the survey. In this case, no extrapolation is required but one loses the information for the cohorts aged 35 and over at the time of the survey.

Another censoring effect hampers fertility comparisons by cohort because the date of the survey serves as a cutting point for longitudinal observations. It would therefore be completely inappropriate to compare achieved family size or parity between two or more cohorts at the time of the survey, as their ages may significantly differ. One must thus compare

[13] Evaluating period data from a retrospective survey is useful in developing countries lacking a good vital registration system. To improve period estimates, data can be combined if one has conducted several comparable surveys over time.

fertility of the birth cohorts at the same age, even if this results in loss of valuable information for the older cohorts. One could also extrapolate the incomplete fertility of birth cohorts into the future for comparative purposes, but this solution cannot be recommended.

There are still other problems. If the survey is restricted to married women aged 15 to 50, in this case as Ryder (1982) has pointed out, marriage has always preceded the survey. Therefore, at young ages, say 15 to 20 years at the time of the survey, one selects women who have married young. As the latter usually have a higher achieved fertility than those who marry later, the parity for women aged 15 to 20 at survey time therefore overestimates the fertility of the birth cohort. The problem does not arise at older ages; for example, the fertility of women aged 40 to 45 takes account of all ages at first marriage. Note also that if the survey is restricted to married or ever-married women, it is not possible to derive age-specific fertility *rates* from the survey, as these require as a denominator the total population ever married and single. Finally, fertility surveys can be affected by several types of errors, particularly in the developing countries: errors due to poor reporting of date of birth or age of woman (see Chapters 4 and 10), errors relating to the total number of children born and to their date of birth, and so forth. Data must be verified and perhaps corrected before use (Brass, 1980; Srinivasan, 1980; Preston *et al*., 2001).

3. An Example Based on Moroccan Data

This example is based on original individual survey data obtained by the Demographic and Health Surveys (DHS) conducted in Morocco in 1987 and 1992.[14] The 1987 sample survey focuses on ever-married women aged 15 to 50. Marital status is known at the time of the survey only, although one also knows the number of marriages per woman and the date of her first marriage. Birth history data are collected for all ever-married women. The 1992 data are similar except that the population is composed of ever-married women and of spinsters. Birth history data are collected, however, only for the ever-married. Extra-marital fertility is very low in Morocco; thus, restricting the fertility data only to ever-married women is therefore justified.[15]

Figure 12–1 presents on a *Lexis diagram* (see Chapter 6) the retrospective fertility data for all birth orders by 5-year birth cohorts, 5-year age groups (y-axis), and 5-year periods before the survey (x-axis), taken from the 1987 Moroccan survey. As usual in Lexis diagrams, the time unit (in this case, 5 years) is identical on the x- and y-axes. The number of females interviewed are shown on the y-axis for each age group or cohort. The *average number of children born* per woman (or quantum of fertility) at the end of the reproductive period, taken here as the age group[16] 45 to 50, is simply the sum of all births in the cohort divided by the number of women interviewed (i.e., $TF = (346 + 964 + \ldots + 631 + 224)/759 = 7.4$ children per ever-married woman on average). The same result is obtained by summing the *age-specific fertility rates*:

$$TF = \frac{346}{759} + \frac{964}{759} + \ldots + \frac{631}{759} + \frac{224}{759} = 7.4$$

These rates are usually expressed on an annual basis, obtained by dividing the above ratios by 5 (see below this section). The quantum of fertility is then obtained by summing these average annual *rates* and multiplying the result by 5.

One similarly obtains the average family size, or incomplete average fertility, for those women of younger ages that have not yet reached the end of their reproductive period at the time of the survey (Table 12–1).

The data are censored *on the right* by the date of the survey[17]; these average numbers of children born may not be compared from one cohort to another because these women have not yet all reached the end of their childbearing period.

One can easily derive from the data the average age at childbearing, for those women aged 45 to 49 at survey time, assuming that in each age group births are delivered at mid period:

$$\bar{x} = \frac{15 \times 346 + 20 \times 964 + \ldots + 40 \times 631 + 45 \times 224}{346 + 964 + \ldots + 631 + 224}$$
$$= 28.88 \text{ years}$$

This measure may not be computed for the younger age groups, as these women are still in their childbearing ages.

For the oldest birth cohorts, aged 45–49 in completed years[18] at the time of the survey, who have

[14] "Enquête nationale sur la planification familiale, la fécondité, et la santéde la population du Maroc" (1987) and "Enquête nationale sur la population et la santé" (1992).

[15] This would not be the case, for example, in sub-Saharan Africa.

[16] Some women in this age group have probably not achieved their ultimate family size; as the cutoff age is 50, this underestimation is unavoidable. An ultimate age group of 50 to 55 would have been better in this respect.

[17] They are also censored *at the top* because the cutoff age is 50. *Period* fertility cannot therefore be reconstructed from this sort of data, as more and more data are missing as one goes backward in time.

[18] Forty-five to 50 in exact years.

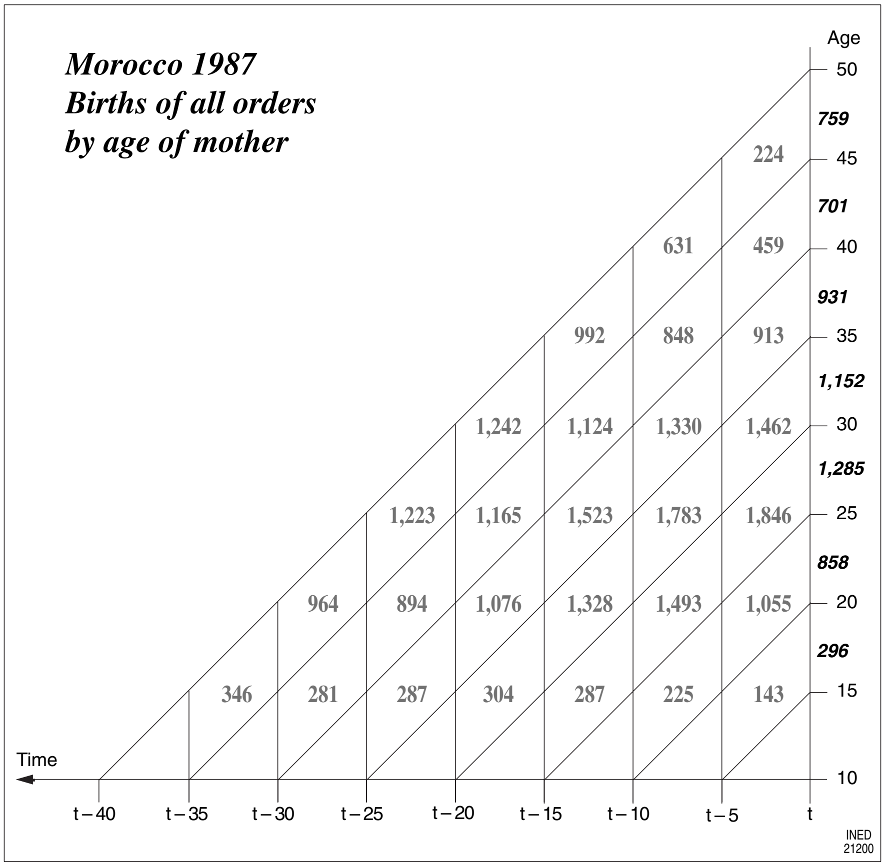


FIGURE 12–1 *Lexis diagram* representation of women and births by 5-year birth cohorts or 5-year age groups (vertical axis) and 5-year periods (horizontal axis), Moroccan Fertility Survey of 1987 (*t* indicates the time of interview, variable from one woman to another). (Source: DHS, Morocco, 1987)

TABLE 12–1 Incomplete average fertility (Morocco, 1987)

Age group (completed y)	Incomplete average fertility
40–44	6.81
35–39	5.51
30–34	4.23
25–29	2.82
20–24	1.49
15–19	0.48

essentially completed their childbearing period, one may compute an annual age-specific fertility rate for each 5-year age group by dividing by 5 the age-specific fertility rates obtained previously (Table 12–2). One knows that, in a country such as Morocco, there remain very few spinsters at that age; one may therefore equate the number of ever-married women to the total number of women whatever their marital status.[19]

[19] Age-specific fertility *rates* or frequencies are ratios of births by age of mother to the number of females of *all* marital statuses, not only those ever-married.

TABLE 12–2 Age-specific fertility rates by 5-year age groups for cohorts aged 45–49 years at time of survey (Morocco, 1987)

Average age of age group, y	Age-specific fertility rate
15	0.0912
20	0.2540
25	0.3223
30	0.3273
35	0.2614
40	0.1663
45	0.0590

On the contrary, one may not compute similar age-specific fertility rates for the younger cohorts, as the younger these are, the greater the difference between the number of ever-married women and the number of women of all marital statuses will be.

Adding up the age-specific fertility rates for the cohort aged 45 to 49 at survey, and multiplying the result by 5, yields the average family size for these women at the end of their reproductive period (i.e., 7.4

TABLE 12–3 Average number of children by birth order (Morocco, 1987)

Birth order	Average number
1	0.94
2	0.92
3	0.90
4	0.84
5	0.79
6 and +	3.03
Total	7.42

TABLE 12–4 Average age at childbearing by birth order (Morocco, 1987)

Order	Average age, y
1	20.75
2	22.84
3	25.19
4	27.08
5	29.10
6 and +	34.76

children per ever-married woman). One may once again obtain from these rates the average age at childbearing (i.e., 28.88 years).

Similar tables and Lexis diagrams (not shown for reasons of space) can be obtained for births by birth order. Considering an open category of six children or more, one obtains (Table 12–3) for the cohort aged 45 to 49 at the time of the survey, the average number of children born by birth order, at the end of the childbearing period.

If one adds up these order-specific averages (i.e., 0.94 + 0.92 + . . . + 0.79 + 3.03), one obtains once again 7.4, the average number of children born (all birth orders) per ever-married woman.

One can also obtain the order-specific average ages at childbearing[20] (Table 12–4) by weighting the mid-interval exact age by the number of births (or the fertility rate) at each age.

One sees that as the birth order increases, the average number of children per woman falls while the average age at childbearing increases, a rather obvious fact. These average ages are obtained assuming that over each interval of time, births are centered on the exact age at the mid of the age group. This is a rather large assumption: for example, the average age at birth of the first child, obtained from the individual data, is actually equal to 20.0 years instead of 20.75 as calculated from the aggregate data. Average durations should therefore always be obtained from the individual birth histories themselves, when they are available, rather than from the aggregated data.

Going back to the average number of children by birth order (Table 12–3), one sees that 94% of the women[21] have had at least one child, and 92% at least two children; thus 94% − 92% = 2% have had only one child during their lifetime. These women are of parity one (or achieved family size) at the end of their childbearing period. The probability of moving from one child (at least) to two children (at least), also called the *parity progression ratio*, is therefore:

$$a_1 = 0.92/0.94 = 0.979$$

Similarly, 92% − 90% = 2% have had only two children (parity two), yielding a parity progression ratio of 90%/92% = 0.978. One thus obtains the distribution of women by parity, and the parity progression ratios for each birth order (the open category 6+ not included). The average number of children of birth order one (i.e., 0.94), can also be considered as the parity progression ratio from 0 to 1 child (at least). The complement 1 − 0.94 = 0.06 is equal to the proportion of women of parity zero, which is the proportion of infertile ever-married women at the end of their childbearing period. In the case of Morocco at that time, this proportion should be close to the percentage of sterile women as voluntary infertility was low (see Chapter 33).

One may also compute the age-specific fertility rates by birth order for the oldest age group (Table 12–5). Multiplied by five, the sum of these rates yields once again the average number of children by birth order already shown in Table 12–3.

Figure 12–2 illustrates these results for the first five birth orders. One observes that the fertility schedule shifts to the right on the graph as birth order increases because a birth of order n can only occur after a birth of order $n - 1$, as already seen from the average ages at childbearing by order in Table 12–4. One also observes that fertility decreases with birth order but in the case of Morocco at that time, still remains high for order five. Indeed, as shown in Table 12–5, fertility above rank five (i.e., 6+) leads to an average of 3.03 children to complete the total family size of 7.41 children per woman.

[20] It is also possible to obtain from these order-specific average ages the average childbearing age for all birth orders (Wunsch and Termote, 1978).

[21] The average number of children of order k is equal to the proportion of women having had at least k children.

TABLE 12–5 Fertility rates by 5-year age groups and birth order for cohorts aged 45–49 years at time of survey (Morocco, 1987)

Average exact age	Birth order						Total all orders
	1	2	3	4	5	6+	
15	0.0546	0.0235	0.0082	0.0040	0.0011	0.0000	0.0914
20	0.0762	0.0754	0.0530	0.0271	0.0145	0.0079	0.2541
25	0.0372	0.0527	0.0648	0.0659	0.0498	0.0519	0.3223
30	0.0150	0.0245	0.0345	0.0445	0.0519	0.1568	0.3272
35	0.0032	0.0058	0.0148	0.0200	0.0290	0.1887	0.2615
40	0.0008	0.0016	0.0037	0.0058	0.0103	0.1441	0.1663
45	0.0005	0.0005	0.0003	0.0008	0.0013	0.0556	0.0590
Total fertility	0.9375	0.9200	0.8965	0.8405	0.7895	3.0250	7.4090

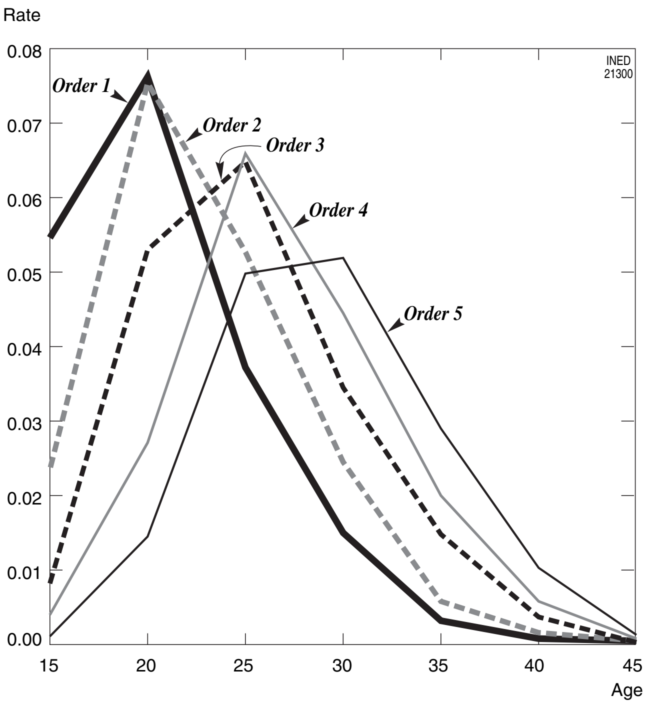


FIGURE 12–2 Fertility rates by age group and birth order. (Source: Morocco, 1987 DHS survey.)

To conclude this example, one can compare the 1987 survey data to those drawn from the 1992 survey. Comparisons are restricted to fertility of all birth orders, for lack of space. Figure 12–3 presents the data from 1992 similar to that shown in Figure 12–1. The number of ever- and never-married women (spinsters) are put on the *y*-axis (*ns* and *s*, respectively). One sees that as age increases, the number of spinsters (*s*) falls sharply; at higher ages, most women are ever-married.

Comparisons between 1987 and 1992 are restricted to ever-married women, for whom the data have been collected in both surveys. For the cohort aged 45 to 49 years in 1992, the ultimate family size (computed as before) is equal to 7.09 births per ever-married woman, while the average age at childbearing is 28.40 years. One observes a fall in family size from one survey to the other, from 7.41 to 7.09, and a slight reduction in the average age at childbearing (from 28.88 to 28.40), although statistical tests should be performed to determine if the differences are significant. Table 12–6 compares the complete or incomplete family sizes controlling for age at survey, between the 1987 and 1992 samples. Except for the first age group (15–19 years), one observes a decrease in the average number of children born in all the other age groups.

From the 1992 data, one may also compute the age-specific fertility rates by dividing in each age group the number of births by the total number of women, ever-married or not. One further divides these ratios by five to obtain rates per annum (Figure 12–4). Results are presented graphically in Figure 12–5. Although fertility is still incomplete in most cohorts, one observes a significant decline in fertility rates as one moves from the older cohorts to the more recent ones, indicating a fall in achieved family size and also most probably an increasing age at childbearing. One may also compare these results to the fertility rates previously obtained for the 45- to 49-year-old cohort in 1987.

If one had only compared the fertility schedules for those women aged 45 to 49 years at survey in 1992 and 1987, one would have noted only a very small and negligible decrease in fertility (Figure 12–6). Actually, from the 1992 data, this very recent decrease in fertility was only experienced by the younger birth cohorts.

Summing the fertility rates and multiplying by five, one obtains the average number of children born up to

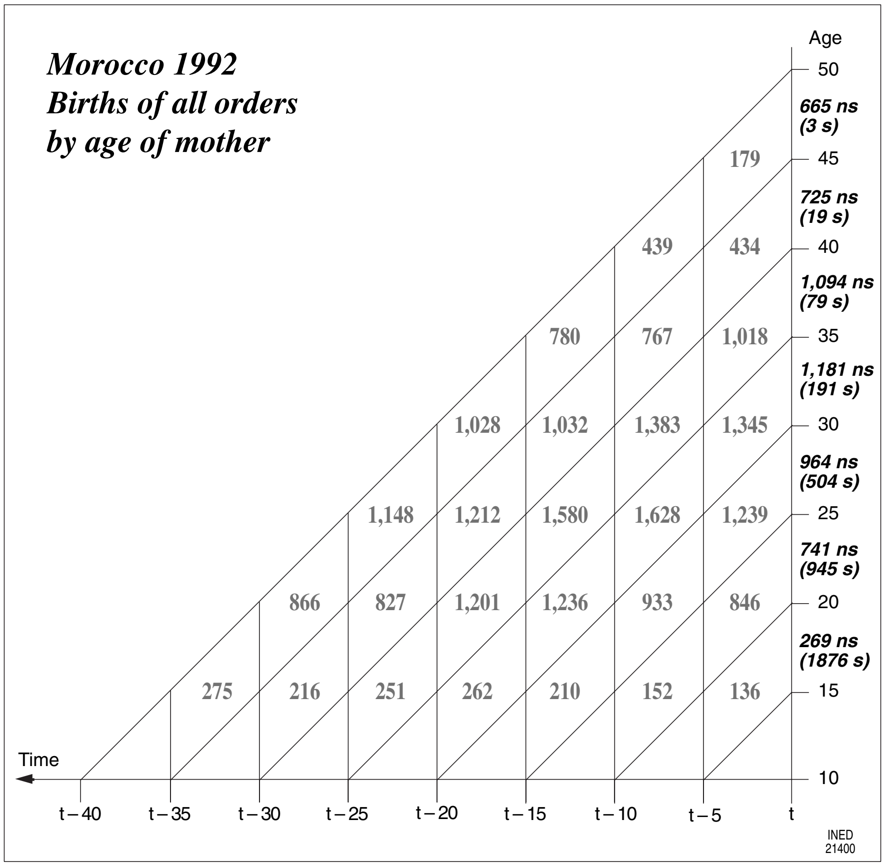


FIGURE 12–3 *Lexis diagram* representation of women and births by 5-year birth cohorts or 5-year age groups (vertical axis) and 5-year periods (horizontal axis), Moroccan Fertility Survey of 1992 (*t* indicates the time of interview, variable from one woman to another). (Source: DHS, Morocco, 1992)

a certain age. For example, one may compute this number at age 30 to 34 years for the various women interviewed in 1992 and older than 30 (Table 12–7). Once again, one sees the fall in fertility from the older to the younger cohorts.

Indicators computed from individual birth histories show an increase, from 20.00 to 20.21 years of age, in the average age at first birth for women aged 45 to 49 years in 1987 and 1992. This average age even reaches 21.08 for those women aged 30 to 34 years in 1992, considering that these women have more or less completed their fertility of birth order one at that age.

One may also compute from individual birth histories the proportion of women still infertile at ages 20 to 24, for those women aged respectively 20 to 24, 30 to 34, and 45 to 49 years at the time of the 1992 survey, 23.6%, 22.7%, and 20.8%, respectively. This confirms the fact, already noted, that the younger cohorts in Morocco are having their children later than the older ones.

II. COHORT FERTILITY FROM VITAL REGISTRATION DATA

Suppose now, for a given birth cohort, that births by age of mother are recorded by the vital (civil) registration system, and that the number of women by age in this birth cohort are estimated for each January 1.[22] As seen in Chapter 8, one may then compute age-specific fertility *rates* by dividing the number of births B_x borne by women of this given birth cohort at age x by the mid-period number of females F_x in this cohort still alive and present at that age. The fertility rates f_x at ages x, taken as the ratios B_x/F_x, will be equal to the rates obtained in retrospective analysis if women interviewed at the end of their reproductive period are not selected for their fertility.

Once again, the average family size will be obtained by summing the age-specific fertility rates over the

[22] These estimates are usually based on past census results.

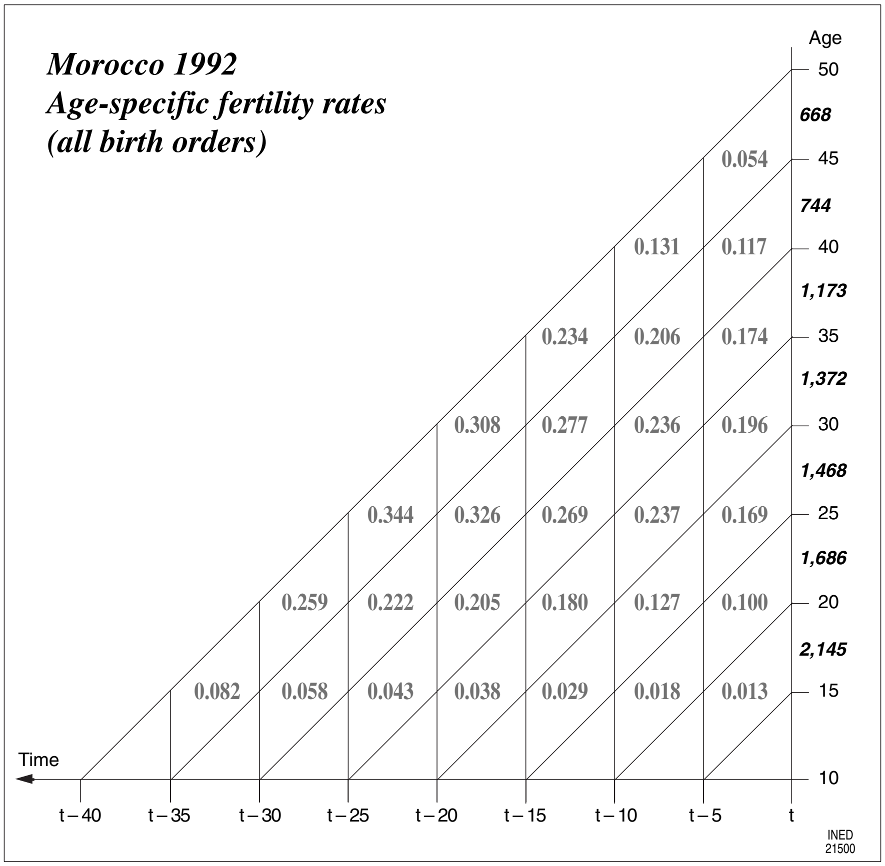


FIGURE 12–4 *Lexis diagram* representation of fertility rates by 5-year age groups and 5-year periods, Moroccan Fertility Survey of 1992 (*t* indicates the time of interview, variable from one woman to another).

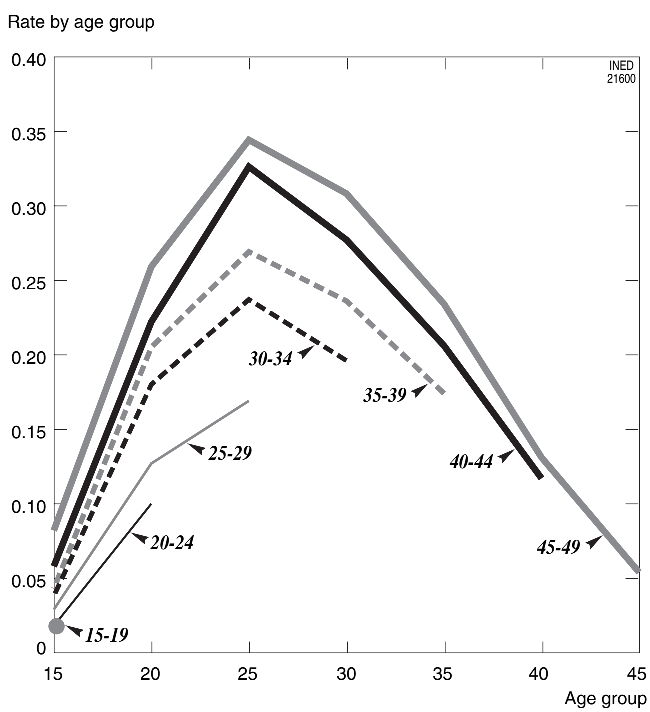


FIGURE 12–5 Fertility rates by age groups and birth cohorts, Moroccan fertility survey of 1992 (data from Figure 12–4).

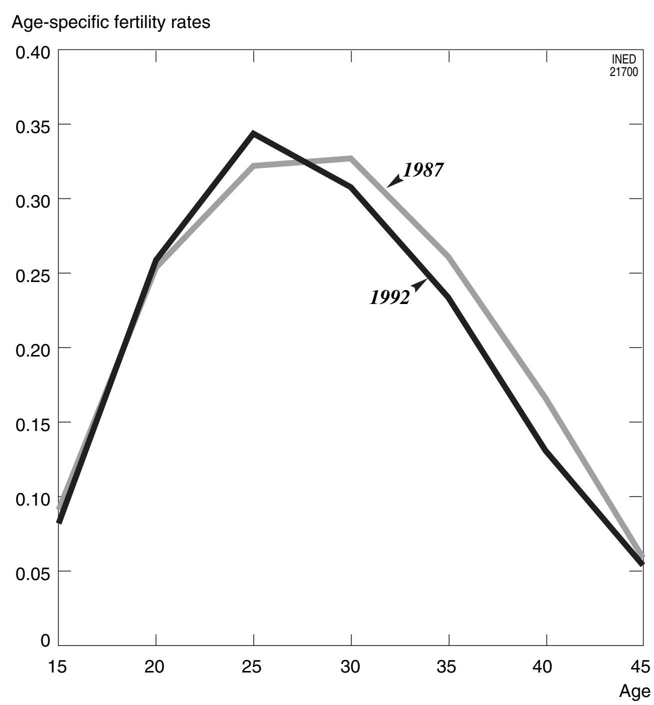


FIGURE 12–6 Fertility schedules of the cohorts aged 45 to 49 years at survey: Comparison of the Moroccan surveys 1987 and 1992.

TABLE 12–6 Average number of children by cohort (Morocco, 1987 and 1992)

Age at survey, y	Average no. children	
	1987	1992
15–19	0.48	0.51
20–24	1.49	1.35
25–29	2.82	2.47
30–34	4.23	3.79
35–39	5.51	4.97
40–44	6.81	6.19
45–49	7.41	7.09

TABLE 12–7 Incomplete family sizes at ages 30–34 years for different birth cohorts, Moroccan Fertility Survey 1992

Age groups at survey (cohorts), y	Family size at ages 30–34Y
45–49	4.97
40–44	4.42
35–39	3.77
30–34	3.26

TABLE 12–8 Age-specific fertility rates of French females born in 1931

Age[a] x	Fertility rate $f_{x,1931}$ (per 1000)	Year	Age[a] x	Fertility rate $f_{x,1931}$ (per 1000)	Year
15	0.6	1946	33	97.7	1964
16	0.4	1947	34	82.5	1965
17	13.0	1948	35	71.2	1966
18	34.3	1949	36	58.8	1967
19	65.9	1950	37	47.6	1968
20	101.1	1951	38	38.7	1969
21	135.5	1952	39	29.8	1970
22	164.6	1953	40	23.2	1971
23	183.6	1954	41	16.5	1972
24	192.1	1955	42	10.8	1973
25	190.8	1956	43	6.1	1974
26	186.2	1957	44	3.2	1975
27	175.3	1958	45	1.4	1976
28	164.2	1959	46	0.7	1977
29	147.3	1960	47	0.3	1978
30	136.2	1961	48	0.1	1979
31	121.2	1962	49	0.1	1980
32	110.2	1963	50	0.1	1981

[a] Age reached during the year. (Source: INSEE)

whole reproductive period, and the average age at childbearing can be computed by weighting each age at birth by the fertility rate at that age. One may also have recourse to births classified by age groups. Examples are given in demographic textbooks (see for example, Pressat, 1983; Wunsch and Termote, 1978).

For example, the Belgian birth cohort born in 1959 has given birth to 7494 children during the year 1990, 3841 births at completed age 30 and 3653 at completed age 31. The mid-period female population (on June 30, 1990) is estimated to be 78,663. The fertility rate at that age is then equal to 7494/78,663 = 0.09527 at exact age 31 on average. Similarly, if data on births and number of females are available each year (and this is not the case with the data published in Belgium), one may compute the fertility rates at each age for the whole reproductive period, for the cohort born in 1959. The sum of these rates over the reproductive life span will give the average number of births per woman in the absence of mortality and migration, if the latter have not selected women with higher or lower fertility than those remaining in the population.

Table 12–8 and Figure 12–7 present the fertility rates by age for the French birth cohort born in 1931, according to each year of observation from 1946 to 1981. The sum of these age-specific fertility rates, from ages 15 to 50, yields the completed family size for this birth cohort (i.e., 2611.3 births per 1000 females or 2.6 births per woman). The average age at childbearing is obtained as usual, by weighting the ages by the corresponding age-specific fertility rates:

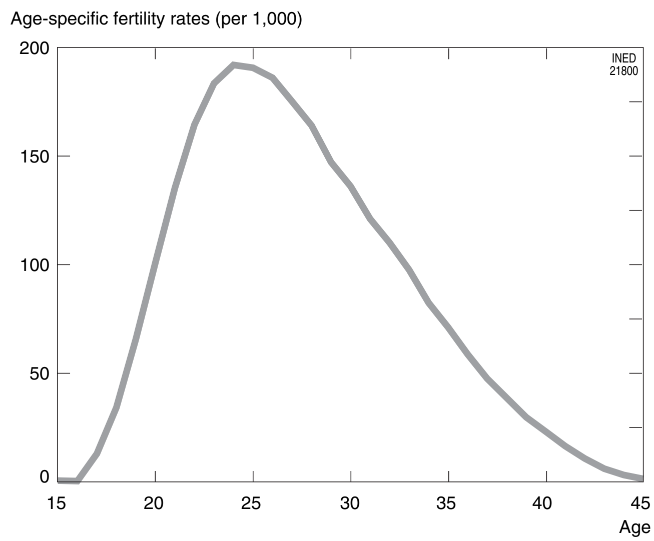


FIGURE 12–7 Age-specific fertility rates of French females born in 1931.

$$15 \times 0.6 + 16 \times 0.4 + \ldots + 49 \times 0.1 + 50 \times 0.1$$

and by dividing the result by the sum of the rates, yielding:

71,628.1/2,611.3 = 27.43 years

In this approach, the fertility rates can also be computed by birth order, say $f_x(k)$ for birth order k, data drawn from the birth certificates being taken as numerator and female population estimates as denominator. The sum over x of these fertility rates by order $\Sigma_x f_x(k)$ is equal to the average family size by order: $TF(1)$, $TF(2)$, . . . Once again one has

$$TF(\text{all orders}) = TF(1) + TF(2) + \ldots + TF(k) \ldots$$

and as shown previously the parity progression ratios can be obtained from these average numbers of children by order:

$$a_0 = TF(1)/1$$

$$a_1 = TF(2)/TF(1)$$

$$a_2 = TF(3)/TF(2)$$

$$\ldots$$

$$a_k = TF(k+1)/TF(k)$$

$$\ldots$$

The average ages at childbearing for births of all orders or by birth order are once again obtained by the weighted average of ages at childbearing, for example, for all orders:

$$\bar{x} = \frac{\sum_x x f_x}{\sum_x f_x}$$

and by order:

$$\bar{x}(1) = \frac{\sum_x x f_x(1)}{\sum_x f_x(1)}, \bar{x}(2) = \frac{\sum_x x f_x(2)}{\sum_x f_x(2)} \ldots . \bar{x}(k) = \frac{\sum_x x f_x(k)}{\sum_x f_x(k)}$$

If births by age are made available between *exact* ages 15 and 16, *exact* ages 16 and 17, . . . and so forth, the values of x will be the exact ages 15.5, 16.5, and so forth. If births are available between *completed* ages 15 and 16, *completed* ages 16 and 17, and so forth, the values of x will be 16, 17, . . . , and so forth in exact ages. Finally, other measures of the tempo of fertility can be computed, such as the variance of ages for all birth orders or by birth order.

III. FERTILITY BY MARRIAGE COHORTS

In many populations, fertility is mainly the fact of married couples, whatever the type of *marriage* (civil, religious, common law, etc.). This issue will be dealt with again in Section 2 in this volume, devoted to the determinants of fertility.[23] In addition to age, it is then useful to classify the retrospective survey data according to duration of marriage, as the latter is a proxy for the number of children born. Taking sufficiently high marriage duration, say 20 or 25 years of marriage, one can then compute the average number of children (all orders and by order[24]) per marriage as a measure of the quantum of marital fertility, and the average duration of marriage at childbearing as a measure of its tempo. One can also compute the average interval between marriage and first birth and the average intervals between successive births for this marriage cohort. If marriage-duration—specific fertility *rates* (frequencies) can be derived from vital statistics,[25] similar measures of average family size by marriage cohort and of average duration of marriage at childbearing can also be obtained in a prospective cohort approach. In both prospective and retrospective approaches, parity progression ratios by marriage cohort can also be computed. One may also compute the incomplete family sizes at lower durations of marriage. The approach is quite similar to that adopted in the previous section, marriage duration being taken into account instead of age.

The Moroccan survey data on births by ever-married women, considered previously, can be cross-classified according to *marriage duration* and *period* for those women in their first marriage at the time of the surveys. Considering that few women give birth to a child after 20 to 25 years of first marriage duration, one obtains an ultimate family size per first married woman of 7.05 in 1987 and 6.84 in 1992, and an average duration of first marriage at birth of the children of 10.02 and 9.48 years, following the same methods of calculation as those described for the birth cohorts. One can also compute the family size at shorter marriage durations for the cohorts interviewed in 1987 and 1992. The results are presented in Table 12–9. One observes once again, this time by marriage cohort, the fall in fertility already seen from the data by birth cohorts.

However, the elimination of age in favor of marriage duration alone does raise some problems. First, among the number of marriages observed after, for example, 25 years' duration, some were concluded at

[23] See in particular, Chapters 25 and 26 on family formation and dissolution, Chapter 28 on the determinants of marriage formation, and Chapter 29 on marital and extramarital fertility.

[24] Birth orders are defined here within marriage.

[25] This requires that births within marriage be registered according to the date of (last) marriage. Usually, the population of the marriage cohort surviving death and migration is not known, and the initial number of marriages is used instead as the denominator of the marriage-duration–specific fertility *rate* (frequency).

TABLE 12–9 Average number of children by marriage duration (Morocco, 1987 and 1992)

Marriage duration at survey	Survey	
	1987	1992
0–4	0.86	0.80
5–9	2.63	2.42
10–14	4.19	4.05
15–19	5.73	5.34
20–24	7.05	6.84

postreproductive ages, especially if remarriages are taken into account. As these marriages do not contribute to fertility, they should be excluded from the data set. Second, the average number of children at duration 25 is biased due to overrepresentation of women who have married at young ages, because these women have a greater chance of still being alive after 25 years of marriage than those who have married at older ages due to the decrease in the probability of survival with increasing age. A woman married at 15 years of age and having a marriage duration of 25 years is more likely to survive than a woman married at 33 and having the same duration of marriage, especially if mortality is high.

The bias is nevertheless easily corrected as follows. Having computed the average family sizes at 25 years' marriage duration TF_x for women marrying at various ages x, the corrected average number of children for the whole marriage cohort is obtained by the weighted average:

$$TF = \frac{\sum F_x TF_x}{\sum F_x}$$

where F_x represents the number of females marrying at age x derived from the vital statistics data at the time of marriage (i.e., 25 years ago), if these data are available. Finally, if the fertility survey restricts the interviews to married women of fertile ages only (from 15 to 50 for example), women of 25 years of marriage duration have all married before age 25, as the sample is confined to women currently aged under 50. Once again, women marrying young are therefore overrepresented due to the sampling procedure this time.

It is therefore necessary to take both age and duration of marriage into account to avoid some of these biases. One then computes the average number of children for females married at age x and having a marriage duration of y, interviewed at age $x + y$. One can also calculate for these subcohorts the average duration of marriage at childbearing and therefore their age at childbearing and the various parity progression ratios for these subcohorts. In temporal or spatial comparisons, this approach has the advantage of showing differences in the *marriage-mix* (i.e., the variations in the proportions of females marrying at specific ages). Another useful measure of the tempo of fertility is the age at *last confinement* in the subcohorts of marriages according to age at marriage; this age usually decreases when fertility starts falling in a high-fertility population. Once again, censoring hampers the analysis however, due to the cutoff survey date, as in the case of birth cohorts. These issues and others will be examined in Chapter 29 devoted to fertility in and outside marriage.

Cohabitation

As discussed in Section 1, nonmarital cohabitation has become more common in the developed countries, and data by marriage duration cannot take this factor into account. In addition, birth of children to cohabitating couples is becoming more common. It is useful, as an indicator of this trend, to compute the ratio of extramarital births as a percentage of all live births for the various birth cohorts considered in a survey. If questions on cohabitation and its duration have been asked in a survey or a census, one may tabulate births according to the partnership status of respondents at the time of the survey (e.g., marriage, consensual union, not in any partnership) and analyze fertility according to duration of cohabitation. It must be remembered, however, that consensual unions tend to break up easily. The full partnership history of the respondent must be obtained to ensure a good description of the fertility history of each person not limited to the partnership status at the time of the survey. For a thorough discussion of the definition and measure of cohabitation, see Toulemon (1996) and Chapter 29.

IV. FERTILITY BY PARITY COHORTS

Parity cohorts are defined by the fact that all women of the cohort have reached a specific parity during the same interval of time. For example, parity cohort 4 of 1960 represents the cohort of women who have borne their fourth child during the year 1960. In general, parity cohort k of 1960 relates to the number of women reaching parity k during the year. By definition, parity cohort 0 of 1960 is taken either as the number of women born or marrying during this year (i.e., the birth or marriage cohort of 1960). Birth histories can therefore be tabulated according to the time of occurrence of births of order k. Consider all those women who have had their first, second, third, . . . kth . . . birth in 1960 (i.e., the parity $k[k \geq 1]$) cohorts of 1960. If the survey is held in 1998, all these women are of postreproductive ages. Among those women who have had k births in 1960, some will have a birth of order $k + 1$ (or more) after that period, and others will have no more children. Let F_{k+1} and F_k stand for these two groups of women. Then the ratio $a_k = F_{k+1}/(F_k + F_{k+1})$ represents the quantum of fertility (in this case the *parity progression ratio*) for this parity cohort (i.e., the proportion of women having another child). The parity progression ratio a_0 for those born or married in 1960 (parity cohort 0) represents the proportion having at least one child. The complement to unity $1 - a_0$ is therefore equal to the proportion of childless females in this cohort (i.e., the percentage of infertile [although not necessarily infecund[26]] women).

For each of these parity cohorts, one can also compute the average time elapsed between births of order k and $k + 1$. For this sole purpose, one usually restricts the analysis to *fertile pregnancies* (i.e., pregnancies ending in one or more live births). This avoids having to deal with near-zero length intervals in the case of multiple births: The latter are coded one fertile pregnancy for the calculation of the interval. This measure of the tempo of fertility of the parity cohorts is equal to the birth interval between birth orders k and $k + 1$ for the parity cohort k of 1960. One can then compare the quantum and tempo of fertility of parity cohorts k for different years, say cohort 1960 compared to cohort 1955, meaning the differences in parity progression ratios a_k and in birth intervals k to $k + 1$.

Once again, birth intervals should be compared for women of the same ultimate fertility at the end of their reproductive period. As high-parity women usually have shorter birth intervals than lower parity females, birth intervals should be compared only for women of the same parity (e.g., the same ultimate family size). If one does not observe this rule, birth intervals seem to decrease with the order of birth, in high-fertility populations, which is not the case in fact. Indeed, the last birth intervals are usually longer than the others, for women of the same ultimate parity, due to decreasing fecundity with age. This erroneous conclusion stems from the fact that birth intervals of higher order are obviously computed for high-parity women only (i.e., families with short overall birth intervals). Another problem arises when computing the average interval between marriage and first birth. This interval is often influenced by the impact of premarital conception. In the absence of information on premarital conception, an approximate solution is to retain only those births born after 8 or 9 months of marriage.

The same sort of measures can also be obtained in a *prospective* cohort approach if births $B(k)$ of order k are registered each year by duration elapsed since the previous birth of order $k - 1$. Birth order may refer to all births borne by a woman during her lifetime or to births born in the current marriage only (first marriage or remarriage). The first option is better because it takes into account the full fertility history of the woman. In comparative studies, care must be taken to ensure that the same definition of birth order is used in all comparisons. For births of order $k = 1$, birth order 0 will be defined once again as the event-origin of the birth or marriage cohort of women, according to that observed. If this registration is available from vital statistics, and such is the case in several countries where the birth certificate contains the necessary questions, it is possible to study longitudinal cohort fertility trends by birth order, according to the interval elapsed since the preceding birth. Let $B(k)_i$ represent the number of live births of order k recorded during a year D at average duration i elapsed since birth $B(k - 1)$ borne by women during year A. The number of births $B(k - 1)$ born during year A corresponds to the number of women reaching parity $k - 1$ during year A, (i.e., to the parity cohort $k - 1$ of year A).

The *birth-interval–specific fertility rate* (frequency) computed for this parity cohort at average duration i is then obtained by dividing the number of births $B(k)_i$ observed at this duration by the number of survivors of the parity cohort. The number of survivors being usually unknown, except possibly for countries with a population register, one divides the number of births of order k by the *initial* population of the parity cohort, equal in this case to the number of births $B(k - 1)$ born during year A, due to the correspondence between order of birth and parity of mother. In general, one therefore computes the ratio:

$$f(k)_i = B(k)_i / B(k - 1)$$

This approximate frequency usually underestimates the true frequency, as the initial population number taken as denominator is generally greater than the actual but unknown population of survivors. Taking the sum of fertility frequencies (parity-specific fertility *rates*) $\Sigma_i f(k)_i$ leads to the average quantum of fertility in this parity cohort, that is to the average number of births of order k per woman of parity $k - 1$. This sum is therefore equal to the *parity progression ratio* for each particular parity cohort (i.e., the frequency of moving from parity $k - 1$ to parity k). This measure will be

[26] An *infertile* woman (i.e. having no children) is not necessarily *sterile*. See Chapters 30 and 33 on the biologic factors of fertility.

TABLE 12–10 Births B from birth cohort $c = 6$, according to period i, parity cohort j, and birth order[a] k

Parity cohort J	$B(k,j)$	Period i before survey 6	5	4	3	2	1	0	Total
		$B(k + 1 = 1, i, j)$, parity $k = 0$							
6	200	62	120	18					200
5	608		162	358	58				578
4	96			22	55	10			87
3	29				5	14	2		21
2	13					1	5	1	7
1	6						1	1	2
0	2								0
Total	954	62	282	398	118	25	8	2	895
		$B(k + 1 = 2, i, j)$, parity $k = 1$							
6	62	20	34	5					59
5	282		79	152	23				254
4	398			95	211	34			340
3	118				20	51	9		80
2	25					3	8	1	12
1	8						1	1	2
0	2								0
Total	895	20	113	252	254	88	18	2	747

[a] or parity, from the mother's viewpoint. (Source: Ryder, 1982)

denoted by the usual symbol a_{k-1}. As to the average duration:

$$\frac{\sum_i i \cdot f(k)_i}{\sum_i f(k)_i}$$

it represents the *average birth interval* in the parity cohort between births of orders $k - 1$ and k. Finally, as with all rates (frequencies), partial intensities of fertility $\sum_{i=0}^{i} f(k)_i$ may also be computed for segmented parity cohorts, indicating the cumulated fertility up to a certain time interval j elapsed since birth of order $k - 1$.

V. FERTILITY BY AGE, MARRIAGE DURATION, AND BIRTH INTERVAL

A thorough analysis of fertility should take account of age, time elapsed since last birth, and marriage duration. However, when using survey data on ever-married women aged 15 to 50 years at time of survey, Ryder (1982) has argued convincingly to drop marriage duration from the analysis, as marriage duration is mainly a proxy of achieved parity, and therefore code the birth cohort, the period of occurrence of marriage (*parity 0*), and the period of occurrence of each birth. Consider for example women aged 45 to 50 at the time of a survey. Number the 5-year age groups 15 to 20, 20 to 25, and so forth in this birth cohort from 0 to 6, and do the same for the calendar periods before the survey. With this recoding, one has age group 15 to 20 equals 0, age group 20 to 25 equals 1, and so forth. Moreover, the 5 years before the survey receive the period code 0, the previous 5 years the period code 1, and so forth. One can then produce a table giving the classification of births by order, for birth cohort number 6 (i.e., women aged 45 to 50), according to the duration in period difference elapsed since the previous birth, marriage[27] being taken as parity 0.

Table 12–10, derived from Ryder (1982), presents illustrative survey data for parity 0 (marriage) and parity 1; similar tables can be drawn for higher parities. One sees that birth cohort $c = 6$ is composed in this example of 954 married women. Of these, 200 married

[27] If one considers instead all women whatever their marital status, birth time of the mother can be taken as parity 0. The first birth interval between order 0 and order 1 is therefore equal to mother's age at first birth.

during period 6, 608 during period 5, and so forth. For the 200 women married during period 6, 62 have given birth to their first child during the same period, 120 during period 5, and 18 during period 4. Similarly, among the 608 women married during period 5, 162 gave birth to their first child during period 5, 358 during period 4, and 58 during period 3, and so on. Therefore, in this birth cohort, one observes 62 first births during period 6, 282 during period 5, and so forth. These figures are then taken as the number of women reaching parity 1 during period 6, and giving birth to their second child during period 6 (20 births), during period 5 (34 births), during period 4 (5 births), and so on.

From this table, one can compute the parity progression ratios. For a_0, one has for the whole cohort

$$a_0 = 895/954 = 0.938$$

However, a_0 varies according to age at marriage, as one can see by computing this parity progression ratio in the various subgroups classified according to the time of their marriage. The following parity progression ratios are obtained:

$$a_0(6) = 200/200 = 1.000$$

$$a_0(5) = 578/608 = 0.951$$

$$a_0(4) = 87/96 = 0.906$$

$$a_0(3) = 21/29 = 0.724$$

$$a_0(2) = 7/13 = 0.538$$

$$a_0(1) = 2/6 = 0.333$$

$$a_0(0) = 0/2 = 0.000$$

One notices that, as age at marriage increases, the parity progression ratios decrease. A similar calculation can be made for a_1. The overall parity progression ratio is equal to

$$a_1 = 747/895 = 0.835$$

and the values for the subgroups are equal to

$$a_1(6) = 59/62 = 0.952$$

$$a_1(5) = 254/282 = 0.901$$

$$\ldots$$

$$a_1(0) = 0/2 = 0.000$$

Once again, fertility decreases as the age at which parity 1 is attained increases. The small numbers involved in this case, influences results, and statistical tests should be made in comparative studies. Nevertheless, the trend in the parity progression ratios is quite obvious.

One may also calculate tempo measures from the data in Table 12–10. For parity 0, for example, the first child is born in all subgroups according to age at marriage at the modal duration one in period difference. In this case, the result is, however, not very interesting, as the 5-year period is so large that the measure is meaningless. One should compute instead the average birth interval from the individual data on the basis of the exact dates of marriage and first birth. The same can be said for the birth intervals 1 to 2, 2 to 3, and so forth.

Other measures of the quantum of fertility can also be computed, such as the *parity progression probabilities* by duration. Going back to Table 12–10 for parity $k = 0$, one sees that out of the 200 women married during period 6, 62 had their first birth during the same period. The probability of giving birth to a first child during duration 0, in period difference, is thus $b_0(1) = 62/200 = 0.310$. At duration 1 in period difference, the probability will be $b_1(1) = 120/(200 - 62) = 0.870$, and at duration 2 (i.e., 6 – 4) one has $b_2(1) = 18/(200 - 120 - 62) = 1.000$. One observes that in the present case

$$a_0 = 1 - (1 - b_0(1))(1 - b_1(1))(1 - b_2(1)) = 1.000$$

Parity progression *probabilities* can therefore be computed at each duration elapsed since the previous birth, with parity 0 being in this example the date of marriage, for all the values of k and j (i.e., for all parities k and periods j of occurrence of birth order k). The parity progression *ratio* is then equal to the complement to one of the products of the complements of the parity progression *probabilities.*

If the sample is small, one may group together all the women (whatever their birth cohort) having a child of order k during a specific period i into a parity cohort without reference to age, as was also the case with vital statistics data. One loses the information on age but one gains in having greater group sizes. Once again, care must be taken to avoid inadequate comparisons. As there is censoring due to the survey date, the fertility of the parity cohorts constituted during periods 3 and 2, for example, may only be compared until duration 2 in completed duration, at the time of the survey. As in the previous situation where age was also taken into account, one may compute parity progression probabilities $b_i(k + 1)$ of reaching parity $k + 1$ at duration i elapsed since the birth of order k. Because birth intervals greater than 10 years are relatively rare, Rodriguez and Hobcraft (1980; see also Brass, 1996) have proposed to use the *partial quantum* after 5 years (or 60 months) as a measure of the fertility of the parity cohort due to the high degree of association between this partial quantum and the complete parity progression ratio. If parity progression probabilities have been computed over a 10-year period, one has

$$a_k \cong 1 - \Pi_i(1 - b_i(k + 1))$$

for i varying from 0 to 9 in completed years. The partial quantum after 5 years (noted B_{60} by Rodriguez and Hobcraft, 1980) would then be equal to $1 - \Pi_i(1 - b_i(k + 1))$, with $i = 0$ to $i = 4$.

In this approach, one mixes age and marriage durations in each parity cohort. Fertility differences between two cohorts of parity k could therefore be due only to differences in the age or marital composition of the parity cohorts. It is probable, however, that this bias is relatively small, especially if births of order $k + 1$ are borne from women having the same age. Computation of parity-specific fertility may also be made for subgroups specific for various explanatory variables, such as education, occupation, and so forth. However, if multiple classifications are used, the size of the subgroups will become very small and the fertility indices will tend to behave erratically. Event-history analysis such as the Cox-type methods (see Chapter 23) applied to the individual data, should here be preferred.

Open and Closed Birth Intervals

When dates of birth are given, it is easy to compute birth intervals by birth order, commonly designated by the order of birth that terminates the interval (Hinde, 1998). For example, the first birth interval is the segment of time between marriage and first birth. Some demographers number the interval instead by the order of birth that opens the interval (Newell, 1988). The first birth interval would, in this case, be the time between the first and second births. It should be stressed that birth intervals may only be calculated for all *closed* intervals (i.e., those terminated by a live birth). In interview data, one may also take account at reproductive ages, of *open* intervals censored by the date of the interview. It can be shown that open intervals straddling the survey date, which will eventually be closed, are on the average longer than those that have been closed before this date (Henry, 1972). One cannot therefore estimate from these sole open intervals the average closed interval. Furthermore, the average of the sole closed intervals for women who have not yet achieved their total family size will underestimate the true overall length of the interval. One must also take into account the fact that some of the women with an open interval will never close this interval, as they will not proceed to having another birth. In this respect, when using open intervals, it is therefore useful to control for age and parity, as these influence the length of the interval. More generally, one may wish to determine the effect of different factors, such as the use of contraception or the level of education, on the probability of closing the interval (i.e., of moving to the next parity). Proportional hazard models, or other forms of event-history analysis, can be used for this purpose (Hinde, 1998).

Finally, the last closed interval and the open interval have sometimes been used as measures of period fertility, in view of examining the impact of birth control on recent fertility trends in less developed countries (Srinivasan, 1980). It is interesting to note that at a particular point in time, the distribution of women according to the length of the open birth interval is similar to a population age structure, except that in the latter case all *open* intervals (ages) will eventually be closed by death. This is not the case in a birth interval structure, as some of the females will never close the interval (i.e., move on to a higher parity). Nevertheless, the similarity between the age and the birth interval structures has been used by several demographers, including Henry, Pressat, Feeney, and Ross, to compare open and closed birth intervals. If fertility is stationary, it can be shown furthermore that the average of all intervals *closing during a specified period* before the survey is equal to the average overall interval. This result is similar to the fact that if mortality is stationary, the average age at death of those individuals dying during the past year is equal to the expectation of life at birth (Leridon and Toulemon, 1997).

References

Alderson Michael, 1983. *An introduction to epidemiology*. London, MacMillan, 335 p.

Anderson Michael, 1998. Fertility decline in Scotland, England and Wales, and Ireland: comparisons from the 1911 census of fertility. *Population Studies*, 52(1), 1–20.

Bongaarts John and Feeney Griffith, 1998. On the quantum and tempo of fertility. *Population and Development Review*, 24(2), 271–291.

Brass William, 1980. *Screening procedures for detecting errors in maternity history data*. London, World Fertility Survey, p. 30–49. (Occasional Papers, no. 22).

Brass William, 1996. Demographic data analysis in less developed countries, 1946–1996. *Population Studies*, 50(3), 451–467.

Cleland John, 1996. Demographic data collection in less developed countries, 1946–1996. *Population Studies*, 50(3), 433–450.

Courgeau Daniel and Lelièvre Èva, 1994. Risques compétitifs et indépendance: cadre théorique d’une réflexion. *Population*, 49(2), 481–498.

Frankfort-Nachmias Chava and Nachmias David, 1996. *Research methods in the social sciences*. London, Arnold, 600 p.

Granström Fredrik 1997. *Fertility and family surveys in countries of the ECE region: standard country report–Sweden*. New York, United Nations, 90 p.

Handcock Mark S., Huovilainen Sami M., and Rendall Michael S., 2000. Combining registration-system and survey data to estimate birth probabilities. *Demography*, 37(2), 187–192.

Henry Louis, 1972. *Demographie: analyse et modèles*. Paris, Larousse, 341 p.

Hinde Andrew, 1998. *Demographic methods*. London, Arnold, 305 p.

Leridon Henri and Toulemon Laurent 1997. *Démographie: approche statistique et dynamique des populations*. Paris, Economica, 440 p. (Coll. Économie et statistiques avancées).

Little Roderik J.A., 1982. *Sampling errors of fertility rates from the WFS*. London, World Fertility Survey, 55 p. (Technical Bulletins, no. 10).

Macura Miros and Klijzing Erik, 1997. Preface, *in:* Fredrik Granström (ed.), *Fertility and family surveys in countries of the ECE region: standard country report–Sweden*. New York, United Nations, 90 p.

Newell Colin, 1988. *Methods and models in demography*. Chichester, Wiley, 217 p.

Ni Bhrolcháin Maire, 1992. Period paramount? A critique of the cohort approach to fertility. *Population and Development Review*, 18(4), 599–629.

Nikander Timo, 1998. Fertility and family surveys in countries of the ECE region: standard country report–Finland. New York, United Nations, 87 p.

Preston Samuel H., Heuveline Patrick, and Guillot Michel, 2001. *Demography: measuring and modeling population processes*. Oxford, Blackwell Publishers, 291 p.

Rodriguez German and Hobcraft John, 1980. *Illustrative analysis: life table analysis of birth intervals in Colombia*. London, World Fertility Survey, 87 p. (Scientific Reports, no. 16).

Ryder Norman B., 1982. *Progressive fertility analysis*. London, World Fertility Survey, 66 p. (Technical Bulletins, no. 8).

Schoenmaeckers Ronald and Lodewijckx Edith, 1999. Demographic behaviour in Europe: some results from FFS country reports and suggestions for further research. *European Journal of Population*, 15(3), 207–240.

Srinivasan Krishnamurthy, 1980. *Birth interval analysis in fertility surveys*. London, World Fertility Survey, 48 p. (Scientific Reports, no. 7).

Toulemon Laurent, 1996. La cohabitation hors mariage s'installe dans la durée. *Population*, 51(3), 675–716 (also published in English as: Cohabitation is here to stay, *Population: An English Selection*, 9, 1997, 11–46.

Wunsch Guillaume J. and Termote Marc G., 1978. *Introduction to demographic analysis*. New York, Plenum Press, 274 p.

CHAPTER

13

Cohort Approach to External Migration Flows

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

For a birth cohort, emigration represents an outflow from the cohort. By contrast, immigration is an event of external origin that, by definition, cannot be included in a cohort approach in the strict sense, as it is determined by the behavior of a group of individuals who do not belong to the initial population (Santini, 1992). There is only one situation in which immigration can be regarded as an event that effectively concerns the entire cohort: this is when the study focuses on the internal migrations of a closed population (i.e., without external migrations), where emigration and immigration are simply each other's opposites (immigration to one region consisting of emigration from the other regions). But that is not our purpose here, because at this point we want to examine the role of migration in the overall dynamics of a population.

In classic analysis, however, immigration is regarded as a birth of sorts, with the individual entering the population studied at an age x different from zero; likewise, emigration is, in practice, equated with an early death, since the individual leaves the population studied. Instead of being strictly defined as a set of persons born in the same year in the same territory, the cohort becomes a set of persons born in the same year and living, at the age of observation, in the same territory. On these definitions, the construction of migration indicators becomes far simpler (admittedly, owing to the very special nature of immigration, this expedient is entirely debatable). Even more so than the previously formulated hypotheses on the independence of phenomena (for example, mortality and migration, or fertility, mortality, and migration), the hypothesis of an event that is totally exogenous but just as independent of the structure and dynamics of the initial population lacks credibility. How can one postulate the neutrality of the initial context or the population's unrestricted capacity to export migrants? We should also add that, in practice, because of the shortage of available data, we must often make do with an analysis of net migration, which ignores these issues altogether.

The gathering of information on external migration flows and their characteristics is indeed, almost always, the first and biggest problem. As already discussed in Chapter 4, few countries maintain population registers; even when a register does exist—as in Italy—it is not uncommon for large numbers of migrations to go unrecorded. The problem is particularly acute for immigration, as a large share of inflows occur clandestinely, and legal immigrants sometimes represent only a marginal share of the real total.

Just as a population's natural dynamics depend on the balance between births and deaths, so the migration dynamics are shaped by the difference between inflows and outflows. Obviously, we must be careful to record the direction in which the flows occur (i.e., to subtract departures and add arrivals). We will begin by adopting this method to measure the propensities to emigrate and the propensities to welcome immigrants that can influence the population's dynamics.

But this is not the most interesting approach for the study of mobility itself. What is actually more important is to measure the total intensity of flows in either direction in the course of each individual path. While this approach to migration is not directly applicable to the analysis of population dynamics, we believe it is essential to provide a broad outline here. This will enable us to refer to the approach in the chapters on mobility factors at the end of this volume (Chapters 22 and 23) and—more significantly—in Volume 2 on the determinants of migration.

As a first step, however, let us briefly describe the available data.

I. THE DATA CONSTRAINT

The data situations may be radically different and seldom ideal, and each may require a different approach. Let us rehearse the main possibilities before reviewing the existing methods.

1. Population Registers

As emphasized in the Introduction, the cohort approach can be fully developed when the quality of the population registers available is sufficient to provide a nearly complete record of individual movements and their characteristics. If the system has been operating for a long time, this type of registration enables us to follow each individual's complete sequence of inflows and outflows, from birth or from his or her first entry into the geographic area. If at a given moment in his or her life (for example, when looking for work), a person decides to emigrate, he or she will be deleted from his or her cohort population. If the person returns later (for example to work in the country again or to retire there), he or she will be added to the total again after having been enumerated as an immigrant. This roundtrip can obviously be repeated several times in a lifetime, giving rise to a sequence of multiple migrations. Obviously, other individuals from other geographic areas, but born in the same year as the cohort studied, may swell the ranks of that cohort.

Such information—alas seldom available—makes it possible to construct migration rates or probabilities of migrating for each birth cohort. The indicators may cover either all migrations observed in the cohort, or inflows and outflows separately. In any event, of course, we can only examine the frequency of movements observed relative to the size of the cohort studied. The indicator will be of debatable significance in the case of inflows, because the denominator will consist of a population that is not the source of the events (unless we are dealing with the return of an individual born in the geographic area). However, although immigration is an exogenous event for the population of arrival, it does constitute a fundamental element in the dynamics of any open population. Like emigration or mortality, it contributes to the change in the cohort's size as it ages.

The information gathered from the population registers not only lends itself easily to a classic analysis of flows. It can also provide a sometimes-abundant database for analyzing the interactions between demographic phenomena as well as the determinants of these phenomena, through the succession of the events that form individual histories such as marriage, divorce, widowhood, parenthood, and working career. This usefulness will be illustrated again in later chapters (particularly Chapters 23 and 24) discussing interactions between demographic phenomena and their determinants.

2. Prospective and Retrospective Surveys

We can try to offset the lack of comprehensive current statistics by performing selective sample surveys. There are two possibilities: *follow-up* (or *prospective*) surveys and *retrospective* surveys. Follow-up surveys are based on a sample of individuals taken at birth or at a given age, initially observed at a given date, then monitored by means of successive interviews at regular intervals. With such a sample, however, we can track departures only; by definition, arrivals do not form part of the initial cohort. If, instead of a sample of individuals, we have selected a sample of geographic areas, we can add new entrants to the initial cohort as they arrive, provided that they have experienced the cohort's constitutive event (e.g., birth, for a birth cohort) at the same date as the initial population. This approach, whose principle is fairly similar to that of a population register, has been used in several studies of countries without registers and subject to large departure flows.

As a rule, prospective surveys are conducted over a limited period of time only. In that case, the information gathered is confined to specific life segments between the time of sample selection and the final interview. Another drawback of follow-up surveys is that outflows followed by a return between two interviews may go unobserved.

The second alternative is a *retrospective* survey in which a sample of individuals is interviewed about its past migrations. By choosing a cohort that has reached a reasonably old age at the time of the survey, we can obtain a fairly long-term picture of the cohort's life. Again, however, we are faced with a series of problems that are not readily solved. The first is that retrospective surveys cannot tell us about departures, unless they have been followed by a return in the meantime (admittedly, arrivals are far easier to measure than in prospective surveys, as retrospective surveys can only cover individuals present at the end of the chosen time frame). Second, because of the length of the retrospective observation period, neglecting hypotheses on independence between phenomena and on population homogeneity (Chapter 9) may create serious difficulties. We assume that the persons who could not be interviewed because they have died or emigrated would have had the same migration behavior as those present at the time of the survey (independence hypothesis). We also assume that the outflows from the observation field do not alter the average status of the remaining population (homogeneity hypothesis). Although tolerable over short periods, such hypotheses may well become unacceptable over so long a period. Moreover, the information gathered in these surveys depends largely on respondents' cooperation and memory, and is often vulnerable to major observation errors.

Yet, despite all these shortcomings, the retrospective survey is generally preferred to a prospective survey because it is much cheaper, far quicker to perform, and capable of producing results immediately—as against the many years required by the prospective method.

Many retrospective surveys have been carried out in recent decades, and they have served as a basis for the main methodologic developments in the cohort analysis of migrations (Courgeau 1973, 1980, 1984, 1985a, 1985b, and 1988; Piché *et al*., 1984; Piché, 1990; Riandey, 1988; Bottai, 1990). Among the most innovative studies are the developments in the analysis of individual biographies (Courgeau and Lelièvre, 1989), discussed in Chapter 23.

3. Contribution of Censuses

Many censuses include a few questions on migrations. The three most typical items of information collected are place of birth, place of residence at a given date prior to the census, and date of last migration. With these items, we can reconstruct at least a portion of a cohort's migration history. Naturally, at the overall national level, the census questions only allow a tracking of arrivals. However, if the population studied is that of a particular region, we will also capture emigrants through the results obtained in all the other regions. The place-of-birth question gives us a record of only one migration per individual, with no guarantee of identifying all those who have migrated at least once in their lives, since some may have returned to their birthplaces after a stay elsewhere. If the question is supplemented by others on place of residence at one or more precensus dates, the information will be somewhat fuller, but will inevitably remain very incomplete. In any event, we are still faced with the above-mentioned risks inherent in retrospective surveys. The risk of observation error, in particular, is great because the census is a comprehensive operation carried out quickly and cheaply.

4. Intercensal Net Migration

Censuses, combined with civil-registration data, offer another possibility: the reconstruction of intercensal net migration. If we know the population by year of birth (or, failing that, by age) in two successive censuses, and if we have annual statistics of births and deaths by year of birth (or, failing that, by age), we can deduce the overall net migration for each cohort in the intercensal period. This approach, of course, will give us a very limited view of migrations, since it is only a balance: inflows and outflows cannot be measured separately. Yet, for countries that lack population registers, this is practically the only way to estimate net external migration—so much so that it features prominently in most demographic-analysis manuals.

If we do not have an annual breakdown of deaths by year of birth (or age), we can obtain an approximate result by means of a fairly cumbersome procedure based on life tables showing the presumed mortality of the cohorts studied. This procedure is emphasized in many manuals of demographic analysis (Wunsch and Termote, 1978; Livi Bacci, 1981).

II. AGE-SPECIFIC INFLOW AND OUTFLOW RATES

If we have a systematic record of migrations at our disposal and if the system is sufficiently reliable, we can monitor a cohort using the same approach as in the study of mortality and fertility: if we have figures for the cohort's size and the migration inflows and outflows each year, we can compute immigration and emigration rates.

Like maternity, migration is a renewable event. The same person can migrate several times. However, when analyzing migration dynamics, we must count inflows and outflows separately. Now a single individual cannot leave a second time unless he or she has returned (i.e., unless he or she has made an initial entry), and the ranking of movements in one direction depends entirely on the ranking of movements in the other. As we cannot amalgamate inflows and outflows here, we will save the opportunities for analysis offered by ranking for the overall study of intracohort mobility (Chapter 22). For the moment, we confine our analysis to migration regardless of rank.

As with fertility and mortality, the usable rates are the type 3 rates (Lexis parallelograms between birthdays) if we have a breakdown of migrations by age and year of birth or the type 2 rates (prospective parallelograms) if the statistics are given by year of birth and year of observation (see Chapter 6 and subsequent

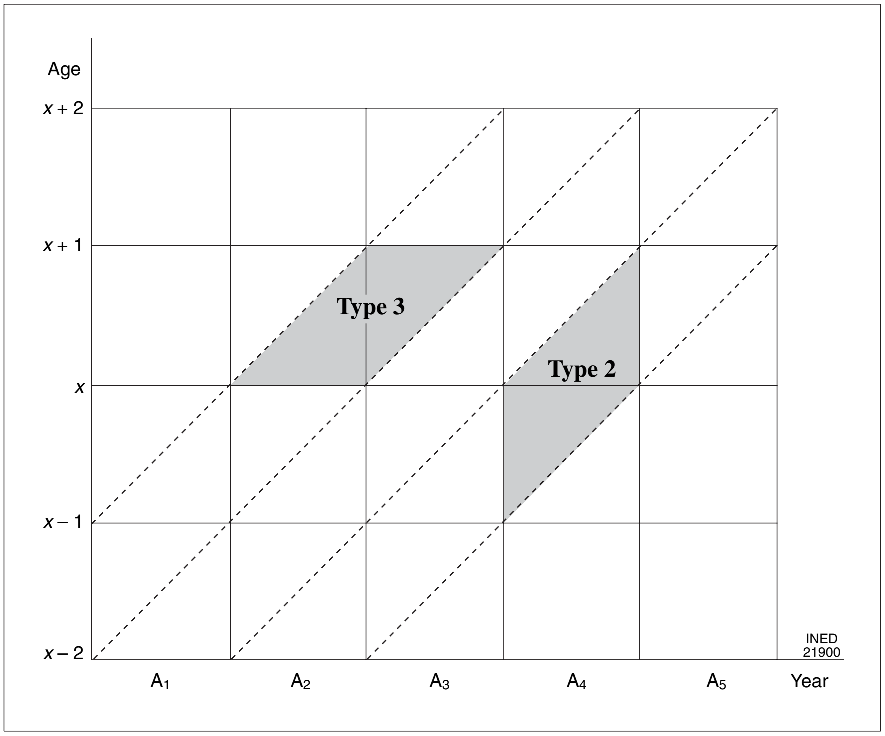


FIGURE 13–1 Representation on a Lexis diagram of the elements needed to compute an age-specific emigration or immigration rate in a cohort.

chapters for different types of rate). But, as in the study of mortality, if we use the type 2 rates, we will have to solve the special problem posed by the first year of age, which straddles an initial Lexis triangle and the first prospective parallelogram (Figure 13–1).

Let D_x^i, I_x^i, and E_x^i be the deaths, inflows, and outflows for cohort i at age x, and P_x^i the number of cohort individuals aged x. The type 3 rate of emigration e_x^i between the exact ages x and $x + 1$ is equal to:

$$e_x^i = \frac{E_x^i}{P_x^i + \frac{1}{2}(D_x^i - I_x^i + E_x^i)} = \frac{E_x^i}{0.5(P_x^i + P_{x+1}^i)}$$

With broader age groups, of n years, the rate is equal to:

$${}_ne_x^i = \frac{E_{x,x+n}^i}{0.5(P_{x,x+n}^i + P_{(x+1),(x+n+1)}^i)}$$

In most cases, however, we calculate type 2 rates. Let $E_{x,x+1}^i$ be the outflows from cohort i at age $(x,x + 1)$ in year t and $P_{x,t}^i$ the number of cohort individuals of age x on January 1 of year t. The emigration rate e_x^i is equal to:

$$e_x^i = \frac{E_{x,x+1}^i}{0.5(P_{(x-1),t}^i + P_{x,t}^i)}$$

Subject to all the reservations expressed earlier on such rates, we can also compute age-specific immigration rates i_x^i in the same manner.

We will accordingly obtain complete series of age-specific emigration or immigration rates for a cohort. We can sum these rates to obtain a synthetic indicator of propensities to depart (EF) and to arrive (IF). ("EF" and "IF" stand for emigration flow and immigration flow, respectively.)

$$EF^i = \sum_{x=0}^{\omega} e_x^i \text{ and } IF^i = \sum_{x=0}^{\omega} i_x^i$$

Again, using an age-group approach, we obtain:

$$EF^i = n \cdot \sum_{x=0}^{\omega} {}_ne_x^i \text{ and } IF^i = n \cdot \sum_{x=0}^{\omega} {}_ni_x^i$$

Of course the difference between the immigration rate and emigration rate at each age gives a net migration rate, and we can also sum these rates to obtain a summary measure of net migration.

Lastly, we can determine a weighted mean age at emigration and mean age at immigration:[1]

[1] Reminder: this formula applies to type 3 rates. If we are dealing with type 2 rates, we will have:

$$\bar{x}^i(e) = \frac{0.25 \cdot e_0^i \sum_{x=0}^{\omega} x \cdot e_x^i}{\sum_{x=0}^{\omega} e_x^i}$$

$$\overline{x}^i(e) = \frac{\sum_{x=0}^{\omega}(x+0.5)e_x^i}{\sum_{x=0}^{\omega} e_x^i} \text{ and } \overline{x}^i(i) = \frac{\sum_{x=0}^{\omega}(x+0.5)i_x^i}{\sum_{x=0}^{\omega} i_x^i}$$

III. ESTIMATING NET MIGRATION: A ROBUST APPROACH DESPITE ITS LIMITS

Absent fuller data on migration flows, we can attempt to estimate the *cohort-specific net migration between two successive censuses*. Two procedures are possible, depending on whether we are working directly with statistics on deaths by age and year of birth or using life tables.

1. Estimation of Overall Net Migration

a. Using Deaths by Year of Birth

If we have the current statistics on death by year of observation and year of birth for the entire intercensal period, we can obtain the cohort-specific net external migration by subtracting cohort deaths from the intercensal net migration (or from the difference between births and the sizes of the corresponding cohorts in the following census), for each cohort.

Obviously the cohort(s) studied must be fully identified. The ideal solution would be for censuses to take place on January 1, which would allow a perfect congruence between death statistics and populations by age, but this is hardly ever the case. If the two censuses have not been carried out at the start of the year but have taken place at the same time of year, we can at least be sure of exactly identifying the same cohort from one census to the other.

In the latter case, however, the 12-month period of births that defines a cohort is not bounded by two January 1sts, but by the two census dates.

Let us consider two censuses performed at an interval of exactly n years, the first in $t1$ and the second in $t2$ (Figure 13–2). Let $P^i_{x,t1}$ be the size of cohort i aged x in the first census, $P^i_{x+n,t2}$ the size of the same cohort at age $x + n$ in the second census, and $D^i_{(x,x+n),(t1,t2)}$ the sum of deaths of cohort i registered during the intercensal period. The cohort's net migration is equal to:

$$NM^i_{x,x+n} = P^i_{x,t1} - P^i_{x+n,t2} + D^i_{(x,x+n),(t1,t2)}$$

If we are studying a cohort born after the first census, we must proceed identically starting with births. Let B^j_t be the births of a cohort j born in one of the intercensal years t, $P^j_{x,t2}$ the size of the same cohort aged x at the date of the second census, and $D^j_{(0,x),(t,t2)}$ the cohort deaths between birth and the second census. The net migration for each of these cohorts is equal to:

$$NM^{j}_{0,x} = B^j_t - P^j_{x,t2} + D^j_{(0,x),(t,t2)}$$

We can also perform the computation for conventional cohorts (i.e., individuals born in the same calendar year). In that case, however, the complete ages at the time of census are not the ages between exact birthdays but the ages between exact ages not expressed as round numbers.

For example, the size of the Italian male cohort born in 1941 was estimated at 354,892 in the October 24, 1971, census[2] (at an age ranging between 29.8 and 30.8 exact years), then at 351,357 in the October 25, 1981, census. Between the two dates, the civil statistics recorded 5391 male deaths for the same 1941 cohort.[3] The intercensal net external migration for the 1941 cohort is therefore equal to:

$$NM^{1941}_{(C1971,C1981)} = 351{,}357 - 354{,}892 + 5{,}391 = 1{,}856$$

Similarly, the 1976 cohort, born after the first census, comprised 390,149 male births, whereas its population enumerated in 1981 totaled 397,603 males, and 10,410 deaths were recorded for the cohort between birth and the 1981 census. The cohort's net external migration between birth and the 1981 census is therefore equal to:

$$NM^{1976}_{(1976,C1981)} = 390{,}149 - 397{,}603 + 10{,}410 = 2{,}956$$

In reality, this net-migration estimate is by default, as the deaths registered in the intercensal period include the deaths of immigrants occurring after they entered Italian territory. By subtracting all these deaths, we underestimate the number of entries. However, because we do not know the number of immigrants, we cannot perform the necessary adjustment with this method.

b. Using Two Life Tables

Even in countries with full statistics, the classification of deaths by year of birth and calendar year is not always available, and the procedure for estimating net migration described in the previous section is rarely followed. A more commonly used alternative is the survival ratio method, which consists in estimating the deaths between two censuses by applying to each

[2] Value adjusted for census error (Caselli *et al.*, 1989).

[3] This result is obviously not published as such in the current statistics. It was kindly provided by ISTAT (the Italian statistical institute), which obtained it by comparing dates of birth and dates of death.

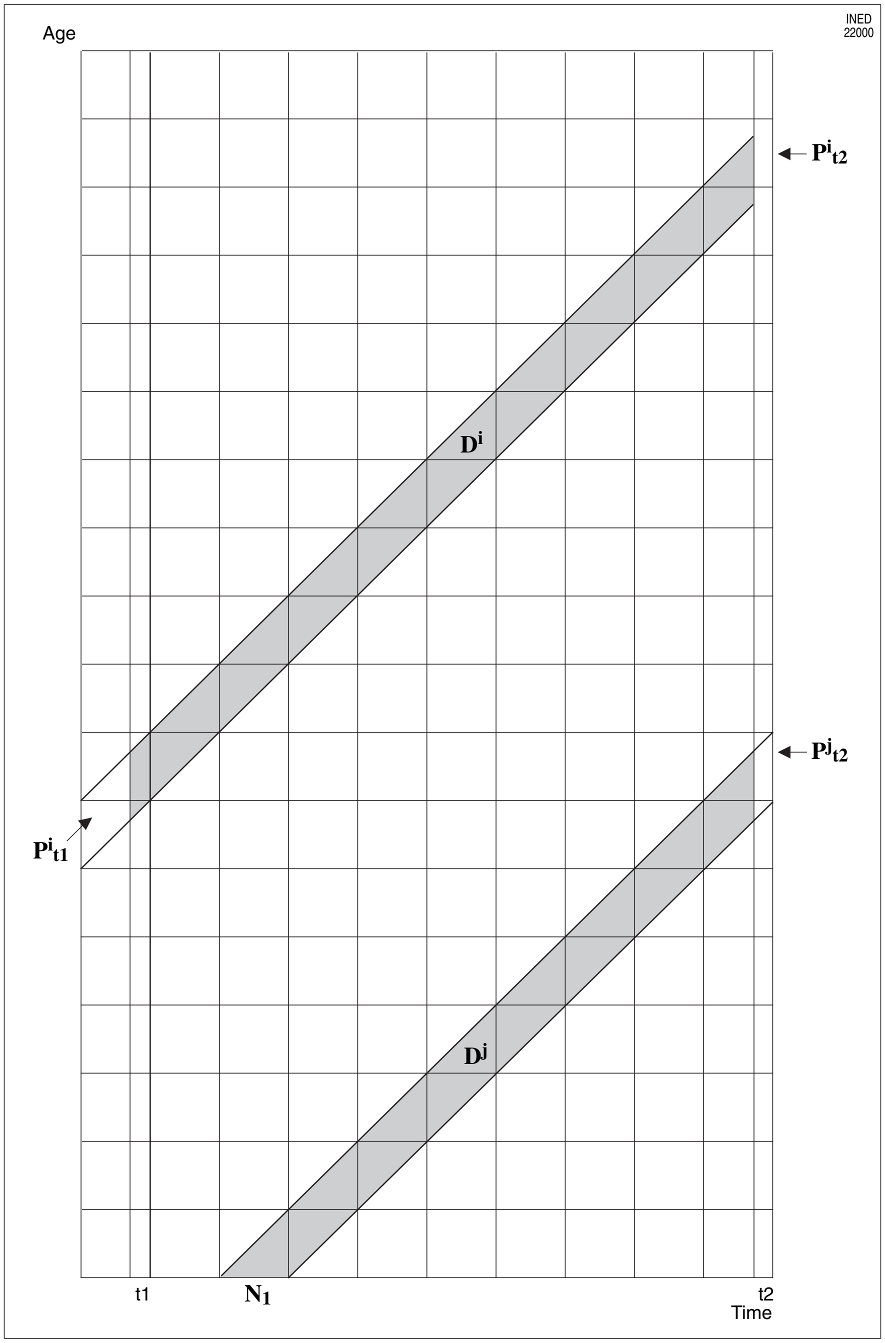


FIGURE 13–2 Position on Lexis diagram of elements needed to estimate a cohort's net migration between two censuses.

cohort the probabilities of dying derived from a suitably chosen series of life tables:

$$NM^i_{x,x+n} = P^i_{(x,x+n),t2} - [P^i_{x,t1} - P^i_{x,t1} \cdot q^i_{x,x+n}]$$

where $q^i_{x,x+n}$ is cohort i's probability of dying between age x in $t1$ and age $x + n$ in $t2$. However, because the aim is to determine a population expected in the second census with no migration to compare it with the enumerated population, we can rewrite the formula more simply (method I) as:

$$NM^i_{x,x+n} = P^i_{(x,x+n),t2} - P^i_{x,t1} \cdot p^i_{x,x+n}$$

where the product included in the second member of the equation represents the expected population, $p^i_{x,x+n}$ being cohort i's probability of survival from one census to the next. The latter value is, of course, a prospective probability of the $L_{x+n,t2}/L_{x,t1}$ type, generally calculated from two life tables constructed for the periods surrounding either census. Naturally, for each cohort *j* born between the two censuses, the period of exposure to the risk of dying is shorter than *n* and the probability of survival is of the $L_{x,t2}/l_0$ type.

However, this method not only suffers from all the drawbacks of the previous one but raises additional problems as well. First, the relevance of the period life tables used to estimate the mortality of the cohorts studied is open to question. Second, if we apply the probability of survival to the initial population, we are applying it *ipso facto* to emigrants beyond the duration of their stay in the geographic area. Conversely, it is true that we ignore the mortality of immigrants since their entry, but, even if the inflows and outflows were broadly of equal magnitude, the problem would be only partially overcome, since the entry and outflow dates are not necessarily equivalent. In reality, the hypothesis underlying this method is that all migration flows occur just before the second census.

According to this method, the intercensal net migration of the Italian male 1941 cohort is estimated at:

$$NM^{1941}_{1971,1981} = 351{,}357 - 354{,}892 \cdot 0.98711 = 1{,}040$$

and that of the cohort born in 1976 is:

$$NM^{1976}_{1976,1981} = 390{,}149 - 397{,}603 \cdot 0.97437 = 3{,}179$$

Following a less commonly used approach, we can work backward from the second census to estimate a cohort size in the preceding census on the assumption of no migration (method II):

$$NM^{i}_{x,x+n} = \left(P^{i}_{(x,x+n),t2} \cdot \frac{1}{P^{i}_{x,x+n}} \right) - P^{i}_{x,t1}$$

However, we will then encounter the same problem in reverse, because we have implicitly assumed that all the migrations occurred just after the first census.

The end result, however, is barely different. For the male 1941 cohort, we obtain:

$$NM^{1941}_{1971,1981} = \left(351,357 \cdot \frac{1}{0.98711} \right) - 354,892 = 1,053$$

and for the male 1976 cohort:

$$NM^{1976}_{1976,1981} = \left(390,149 \cdot \frac{1}{0.97437} \right) - 397,603 = 2,809$$

As Courgeau (1988) has shown, either approach allows us to estimate the net number of migrants, but not the number of net migrations. To estimate the latter, we need to make additional assumptions on the distribution of migration flows (inflows and outflows) within the intercensal interval. The estimation of deaths is hindered by lack of information on the time distribution of migration flows. Several authors (Wunsch, 1969; Dionne, 1970; Courgeau, 1988) propose an identical solution to reduce the estimation error by introducing an estimation of migrants' deaths into the net-migration calculation. Because the distribution of migration flows between the two censuses is unknown, these authors assume an equidistribution of net migration during the period. On the further assumption that mortality (at least at adult ages) rises gradually with age, they attach more weight to the mortality of the second half of the intercensal interval than to the first. Let q_{I} and q_{II} be the probabilities of dying for these two half-intervals. We apply to the migrants of the first half-interval the probability of dying $0.5q_{\text{I}}$, and to the survivors among them the probability q_{II}; to the migrants of the second half-interval, we apply the probability $0.5q_{\text{II}}$. The total number of migrants' deaths accordingly comes to:

$$0.5\left(\frac{NM^{i}_{x,x+n}}{2} \right) q_I + \left[\frac{NM^{i}_{x,x+n}}{2} - 0.5 \frac{NM^{i}_{x,x+n}}{2} q_I \right] q_{II} + 0.5\left[\frac{NM^{i}_{x,x+n}}{2} \right] q_{II}$$

or

$$\frac{NM^{i}_{x,x+n}}{2} [0.5q_I + 0.5q_{II} + (1 - 0.5q_I) q_{II}]$$

In this manner, at the price of some further hypotheses (Wunsch and Termote, 1978; Courgeau, 1988), we can obtain a correction factor $\dfrac{1}{1 - 0.5q^{i}_{x,x+n}}$ to estimate net migrations as follows (method III):

$$NM^{i}_{x,x+n} = [P^{i}_{(x,x+n),t2} - P^{i}_{x,t1} + q^{i}_{x,x+n} P^{i}_{x,t1}] \cdot \left[\frac{1}{1 - 0.5q^{i}_{x,x+n}} \right]$$

For the 1941 cohort, this gives:

$$NM^{1941}_{1971,1981} = [351,357 - 354,892 + (0.01289 \cdot 354,892)] \cdot \left[\frac{1}{1 - 0.5 \cdot 0.01289} \right] = 1,046$$

and for the 1976 cohort:

$$NM^{1976}_{1976,1981} = [390,149 - 397,603 + (0.02563 \cdot 397,603)] \cdot \left[\frac{1}{1 - 0.5 \cdot 0.02563} \right] = 2,772$$

As we can see in Table 13–1, whatever the procedure, the methods based on life tables yield very similar results for the 1941 cohort. In particular, the difference between *net migrants* and *net migrations* is minimal in this example. By contrast, these approximations and the results obtained with the method

TABLE 13–1 Comparison of intercensal net migration figures for Italian male cohorts born in 1941 and 1976, obtained with different procedures

Method	1941 Cohort		1976 Cohort	
	Net figure	Index value[a]	Net figure	Index value[a]
Deaths method	1856	100	2956	100
Method I (net migrants)	1040	56	3179	107
Method II (net migrants)	1053	57	2809	95
Method III (net migrations)	1046	56	2772	94

[a] Deaths method = 100.
From censuses of October 24, 1971, and October 25, 1981.

involving direct use of death statistics—for that same 1941 cohort—differ by a factor of two. No such difference is found with the 1976 cohort. The main reason is that the number of officially registered births is far more reliable than any cohort population enumerated at a given age. The imprecision of an evaluation based on the difference between observations subject to error is much more important here than any methodologic refinements. Not only are the censuses far from perfect, but also the enumeration errors from which they suffer vary from one census to another—and, sometimes, can even change direction. In any event, of course, these deficiencies must inspire caution in accepting the results for intercensal net migration, whatever the method used.

2. From Overall Net Migration to Age-specific Rates

Whatever the procedure, the methods discussed in the previous sections give only an overall net figure for the intercensal period (10 years in the case of Italy) or for a shorter period of variable length when we study cohorts born between censuses. The next step consists of decomposing this net figure by age, taking the ages reached by cohort studied in the course of the period (i.e., the relevant parallelograms [or triangles] on the Lexis diagram [*ABCD* or *EFG*, Figure 13–3]). Almost all demography manuals recommend a straightforward uniform distribution of the overall net migration between the years of age concerned. This means assigning the same net migration to each age in each cohort, whereas we know that the age distribution of migrations is—on the contrary—far from uniform (Figure 4–19 of Chapter 4) and also displays a fairly similar profile in many cases (Castro and Rogers, 1979; Livi Bacci, 1981; Capocaccia and Caselli, 1990). International migration is fairly significant at the youngest ages, usually falls sharply toward ages 10 to 15 years, increases steeply up to age 30 or so, and declines gradually thereafter, becoming negligible after age 70. In other words, for an accurate distribution of an overall net migration observed in the course of a rather long age interval, we must allow for the age-related variations in the propensity to migrate. We must also remember that migration is sensitive to the economic cycle. Last, we need consider annual fluctuations, even when breaking down an overall net migration figure for a single cohort.

To take simultaneous account of age effects and period effects on the distribution of the net intercensal migration of each cohort, we can use an age-period (AP) model developed for a reconstruction of the Italian population by province at each January 1 (Capocaccia and Caselli, 1990). This method rests on the following two hypotheses:

1. Migration flows follow the same distribution by age and calendar year as that of the movements recorded in the population register (*registro anagrafico*[4]);
2. The age structure of these flows is constant from one calendar year to the next.

The first hypothesis has not been tested empirically in Italy, but it seems fairly logical and consistent with observations made in other countries. The second tallies with the results of a comparative analysis of deletions and new entries, by age, in the population

[4] Italy requires place-of-residence changes to be reported. In theory, a citizen who changes residence must report his or her arrival to the Ufficio Anagrafico (registry) of the new locality and report his or her departure to the registry of the old locality. This registration, however, covers only a portion of the real population flows, as many moves go undeclared. But it is safe to assume that the under-recording is broadly independent of migrants' ages and that it varies little over time.

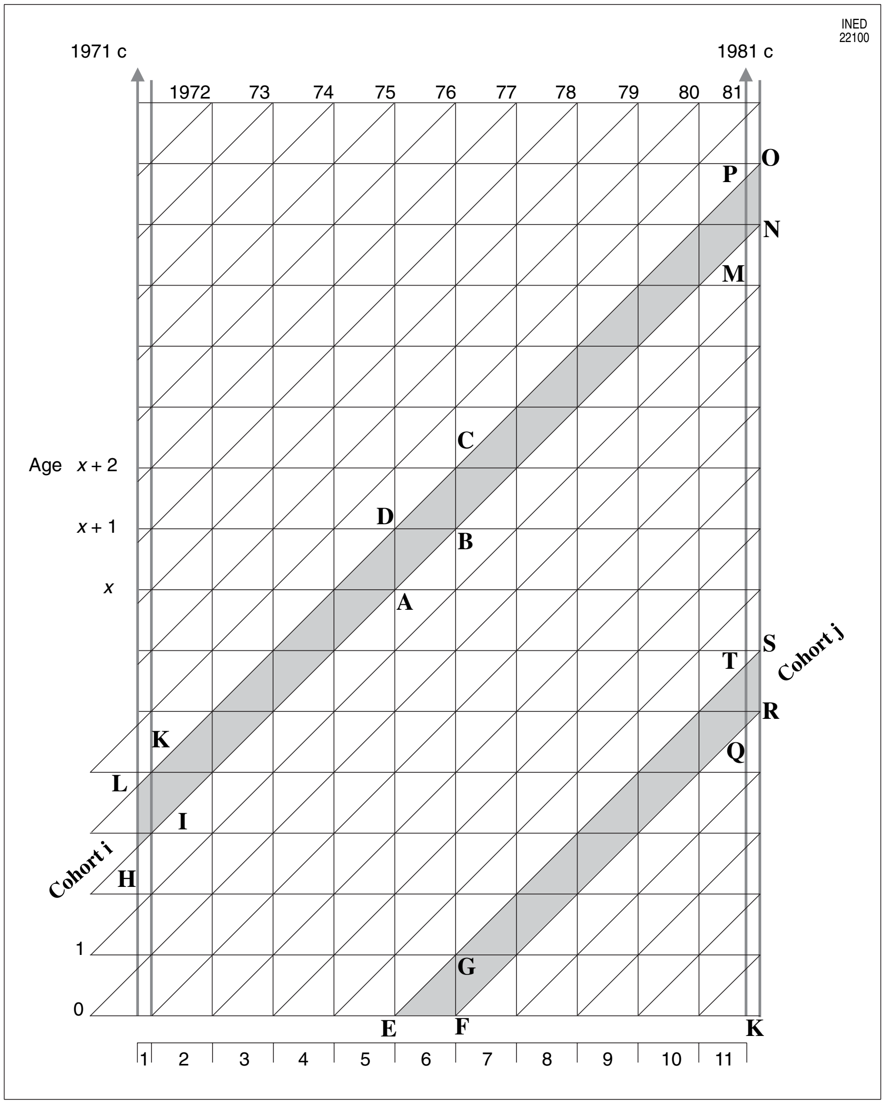


FIGURE 13–3 Visual representation on Lexis diagram of an age-specific net migration within an overall intercensal net migration estimated for a given cohort (Source: Capocaccia and Caselli, 1990).

registers of the 92 Italian provinces during the 1971 to 1976 and 1977 and 1981 periods.[5]

Intercensal net migration by cohort for 1971 to 1981 was estimated using the procedure described above, using death statistics. To estimate the age-specific probabilities of migrating,[6] we use a logistic function whose independent variables are age and calendar year: this is the AP model:

$$pm_{i,t} = (1 + \exp(a_i + b_t + c))^{-1}$$

where $pm_{i,t}$ is the probability for an individual aged i of migrating in t; the parameters a and b denote age effect and period effect respectively, c being a constant that describes the average propensity to migrate for the total population. These parameters are estimated by the least-squares method, applied to the statistics of deletions from the population register: This is preferred to new entries in the register, since Italy was an emigration country during the period. The results for a (age effect) and b (period effect) are reported in Figure 13–4. The age-effect profile is entirely consistent with the descriptions in the specialized literature. The period-effect profile shows that the propensity to migrate was greater at the start of the period and gradually declined thereafter.

[5] A later study has found the same age structure in the 1981 to 1986 and 1986 to 1991 periods.

[6] We can determine probabilities of migrating if we assume that each migration flow involves a different individual: this means treating migration as a nonrenewable event.

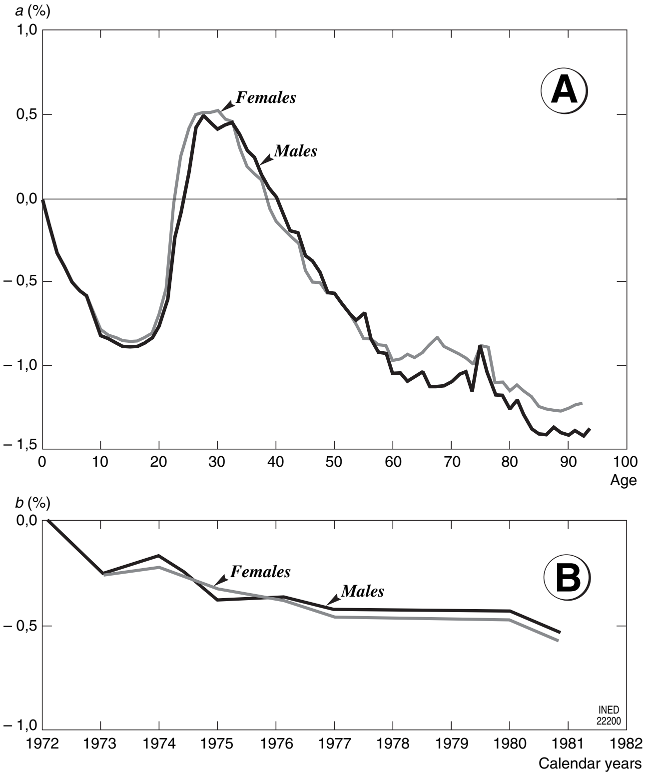


FIGURE 13–4 Profiles by age (parameter *a*) and calendar year (parameter *b*) of Italian net migration between 1971 and 1981, obtained with an age-period model applied to deletions from population registers (Source: Capocaccia and Caselli 1990).

For the cohorts concerned, parameter *a* of the model provides an estimate of probabilities of migrating between two birthdays. However, our goal is to determine prospective probabilities that take into account each cohort's propensity to migrate in the course of a calendar year: we want to fit the equations precisely to the data derived from the comparison between two censuses that have not taken place on January 1 (Figure 13–3). For details of the calculation, see Capocaccia and Caselli (1990).

Once we have calculated the age-specific probabilities of migrating for each cohort, their total intercensal net migration is distributed as follows:

If the cohort is born before the first census:

$$NM^i_{x,t} = NM^i_{(x,x+n),(t1,t2)} \cdot \frac{pm^i_{x,t}}{\sum_{t=2}^{n+1} pm^j_{x,t}}$$

If the cohort is born between the two censuses:

$$NM^i_{x,t} = NM^j_{(0,x+n),(t,t2)} \cdot \frac{pm^j_{x,t}}{\sum_{t=2}^{n+1} pm^j_{x,t}}$$

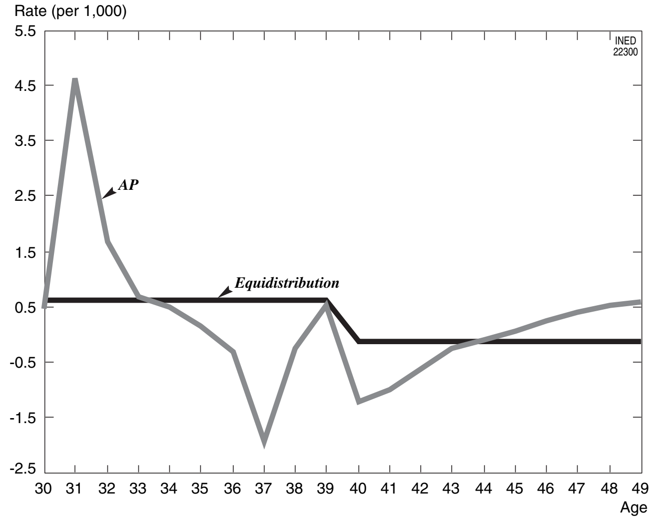


FIGURE 13–5 Age-specific net migration for Italian male 1941 cohort, using two methods for distributing the overall intercensal net figure.

The ratio in the second member of these equations gives the proportion of total net migration to be assigned to each age *x* reached in the calendar year for the cohort studied.

We can then derive indicators for migration *intensity* and *tempo*:

a rate (type 2):

$$nm^i_{(x,t)} = \frac{NM^i_{x,t}}{0.5(P^i_{x,t} + P^i_{(x+1),(t+1)})}$$

an average number of net migrations (total net migration) reached by each cohort in the intercensal interval:

$$NMF = \sum_{x}^{x+n} nm^i_{x,t}$$

and a mean age of these cohorts at net migration in the intercensal interval:

$$\bar{x}(nm) = \frac{\sum_{x=0}^{\omega} x \cdot nm_{x,t}}{\sum_{x=0}^{\omega} nm_{x,t}}$$

Table 13–2 and Figure 13–5 illustrate the results obtained for the Italian male 1941 cohort between the 1971 and 1981 censuses, and between the 1981 and 1991 censuses. This gives us a picture of 20 years of the cohort's life. The table and figure compare the application of the generally recommended method of net-migration equidistribution and the AP model distribution method. In the first case, we obviously obtain

TABLE 13–2 Age-specific net migration and age-specific net migration rate between 1972 and 1991 for Italian male 1941 cohort, using two methods for distributing the overall net figure

Age	Year	Uniform distribution			Distribution based on AP model		
		Net migration by age	Population at Jan. 1	Rate (per 1000)	Net migration by age	Population at Jan. 1	Rate (per 1000)
30	1972	215	354,300	0.61	166	354,300	0.47
31	1973	215	354,000	0.61	1640	354,000	4.62
32	1974	215	355,200	0.61	596	355,200	1.68
33	1975	215	355,300	0.61	240	355,300	0.68
34	1976	215	355,100	0.61	173	355,100	0.49
35	1977	215	354,800	0.61	48	354,800	0.14
36	1978	215	354,300	0.61	–109	354,300	–0.31
37	1979	215	353,600	0.61	–688	353,600	–1.95
38	1980	215	352,300	0.61	–95	352,300	–0.27
39	1981	215	351,500	0.61	178	351,500	0.51
40	1982	–50	350,974	–0.14	–425	350,974	–1.21
41	1983	–50	349,786	–0.14	–352	349,786	–1.01
42	1984	–50	348,575	–0.14	–224	348,575	–0.64
43	1985	–50	347,513	–0.14	–86	347,513	–0.25
44	1986	–50	346,501	–0.14	–35	346,501	–0.10
45	1987	–50	345,424	–0.14	22	345,424	0.06
46	1988	–50	344,400	–0.15	84	344,400	0.24
47	1989	–50	343,265	–0.15	135	343,265	0.39
48	1990	–50	342,167	–0.15	177	342,167	0.52
49	1991	–50	340,972	–0.15	203	340,972	0.60
50			339,672			339,672	
Net migration reached between ages 30 and 49		1,650		4.66	1,648		4.66
Mean age at net migration between ages 30 and 49				37.0			36.9

constant rates with identical signs in each intercensal period, whereas the AP method gives a more realistic picture of age-related changes in migration.

In both methods, the total net migration after 20 years is naturally the same, and the difference in mean age is actually minimal (37.0 versus 36.9 years). Only the distribution is radically different. Of course, the problems relating to the quality of census observations remain unsolved!

CONCLUSION

At the start of the chapter, we saw the scope of the problems raised by the analysis of migrations in terms of their contribution to the overall population dynamics. For the natural dynamics (i.e., the balance between births and deaths), the cohort analysis of events occurring in the life of individuals (maternity, death) is used to construct unambiguous indicators of the progressive extinction or cumulative fertility of a cohort. On the other hand, the dynamics of migration (i.e., the balance between inflows and outflows) are a source of multiple difficulties. The first are, of course, observational, as the administrative sources in this field are most often either non-existent or severely deficient. But there is also—and above all—a conceptual difficulty: while emigration can be understood as an endogenous phenomenon, specific to the population studied (although there needs to be an *elsewhere* to which people can migrate), immigration depends on the behavior of external populations to which we should be able to refer in order to construct a satisfactory representation model. In most cases, though, we know nothing about the external areas from which the immigrants come. Traditionally, demographers get around this problem by calculating immigration rates as if inflows could be measured against the number of residents—but this does very little to reduce the sense of dissatisfaction. We can easily see that this is at best an expedient to describe the various components of a population's capacity to grow.

By contrast, migration is a demographic phenomenon in itself that—beyond the overall population dynamics—can, through different channels, affect the population's structure and flows. First, as we shall see, fertility and mortality are not unrelated to mobility, and this would alone be sufficient reason to go back to the analysis of migration processes as a factor in demographic behavior. But domestic migration plays a crucial role in the distribution of a population on its territory, and hence in shaping its structure. This, in turn, can have major indirect effects on the population dynamics. Last, migration is a phenomenon whose economic, social, cultural, political, and other consequences are immense, which is why we will return to migration in Chapter 22, using an entirely different approach to mobility. This approach is capable of taking into account all the categories of movements that occur in the life of an individual—from simple changes of residence to international migrations; it focuses on the act of migration *per se*, as distinct from its direction (inflow or outflow).

References

Baccaïni Brigitte and Courgeau Daniel, 1995. *Approche individuelle et approche agrégée: utilisation du registre de population norvégien pour l'étude des migrations, Seminar on l'analyse spatiale de données biodémographiques*, Paris, 22 p.

Bottai Marco, 1992. Storie familiari e storie migratorie: un'indagine in Italia, *in:* Alberto Bonaguidi (ed.), *Prospettive metodologiche nello studio della mobilità della popolazione*, p. 145–174. Pisa, Pacini Editore, 202 p.

Capocaccia Riccardo and Caselli Graziella, 1990. *Popolazione residente per età e sesso nelle province italiane. Anni 1972–1981*, Fonti e Strumenti, n. 2, Rome, Dipartimento di Scienze Demografiche dell'Università di Roma "La Sapienza", 251 p.

Caselli Graziella, 1998. Migrationi internazionali, *in: Enciclopedia del Novecento*, p. 252–264. Rome, Istituto dell'Enciclopedia Italiana Treccani, vol. XI, 855 p.

Caselli Graziella, Golini Antonio and Capocaccia Riccardo, 1989. Una correzione alla struttura per età della popolazione italiana al censimento del 1971, *Genus*, vol. 45(3–4), p. 1–20.

Courgeau Daniel and Lelièvre Èva, 1989. *Analyse démographique des biographies*. Paris, INED, 268 p. (Also in English: *Event history analysis in demography*, Oxford, Clarendon Press, 1992, 226 p.).

Courgeau Daniel, 1973. Migrants et migrations, *Population*, vol. 28(1), p. 95–129.

Courgeau Daniel, 1980. *Analyse quantitative des migrations humaines*. Paris, Masson, 225 p. (Anthropologie physique, no. XII).

Courgeau Daniel, 1984. Relations entre cycle de vie et migrations, *Population*, vol. 39(3), p. 483–514.

Courgeau Daniel, 1985a. Interaction between spatial mobility, family and career life-cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel, 1985b. Changement de logement, changements de département et cycle de vie, *L'Espace Géographique*, no. 4, p. 289–306.

Courgeau Daniel, 1988. *Méthodes de mesure de la mobilité spatiale: Migrations internes mobilité temporaire, navettes*, Paris, INED, 301 p.

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, no. 1, p. 7–26. (Also in English: From the group to the individual: what can be learned from migratory behaviour, *Population. An English Selection*, 1995, vol. 7(1), p. 145–162).

Dionne Claude, 1970. Estimation des soldes migratoires par la comparaison de deux recensements, *Recherches Économiques de Louvain*, vol. 36(4), p. 309–340.

Ginsberg R. B., 1978. The relationship between timing of moves and choice of destination in stochastic models of migration, *Environment and Planning*, A(10), p. 667–679.

Haenszel William, 1967. Concept, measurement and data in migration analysis, *Demography*, vol. 4(1), p. 253–261.

Livi Bacci Massimo, 1981. *Introduzione alla demografia, Firenze*. Florence, Loescher, 439 p.

Morrison Peter A., 1967. Duration of residence and prospective migration: the evolution of a stochastic model, *Demography*, no. 4, p. 553–561.

Myers George C., Mcginnis Robert, and Masnick George, 1967. The duration of residence approach to a dynamic stochastic model of internal migration: a test of the axiom of cumulative inertia, *Eugenics Quaterly*, vol. XIV(2), p. 121–126.

Piché Victor, 1990. Potentialités et originalité d'une enquête rétrospective pour l'étude de la migration, *in: Conference on the role of migration in African development: issues and policies for the 90s*, p. 283–313. Dakar, Union pour l'étude de la population africaine (UEPA), 313 p.

Piché Victor, Gregory Joel, and Lavoie Jean-Pierre, 1984. L'analyse historique des migrations: la pertinence de l'analyse longitudinale à partir des histoires rétrospectives, *Genus*, vol. XL(3–4), p. 25–45.

Riandey Benoit, 1988. La qualité du suivi des échantillons dans les enquêtes démographiques, un bilan, *Population*, vol. 43(4–5), p. 829–854.

Rogers Andrei, 1980. Introduction to multistate mathematical demography, *Environment and Planning*, vol. 12(5), p. 489–498.

Rogers Andrei, 1995. *Multiregional demography: Principles, methods and extensions*. New York, Chichester, xi + 236 p.

Santini Antonio, 1992. *Analisi demografica: fondamenti e netodi*. Scandicci (Florence), La Nuova Italia, 422 p.

Wendel Bertil, 1953. A migration schema: theories and observations, *Lund Studies in Geography*, no. 9, p. 34–37.

Willekens Frans, 1992. Synthetic biographies: extension of a multistate life table, *in:* Alberto Bonaguidi (ed.), *Prospettive metodologiche nello studio della mobilità della popolazione*, p. 83–98. Pisa, Pacini Editore, 202p.

Wunsch Guillaume, 1969. Le calcul des soldes migratoires par la méthode de la "population attendue." Caractéristiques et évolution des biais, *Population et Famille*, no. 18, p. 49–61.

Wunsch Guillaume J. and Termote Marc G., 1978. *Introduction to demographic analysis*. New York, Plenum Press, 274 p.

P A R T

III

PERIOD ANALYSIS REVISITED

The Hypothetical Cohort and its Relations with Actual Birth Cohorts

GRAZIELLA CASELLI AND JACQUES VALLIN

The factors of natural population change affect nothing if not the life and death of individuals. To account for them in the most practical way possible, therefore, we have analyzed at length the relevant processes within birth cohorts whose reproduction can be tracked through their recorded fertility, and their attrition through their death rate. We attempted to address the factors of migration flows (immigration and emigration) in the same way, but with less success due to the difficulties of observation. As a result, we moved from observing concrete events which punctuate the lives of the individuals to calculating rates or probabilities which characterize fertility behavior, propensity to migrate, and the process of attrition within specific sets of individuals—cohorts—sharing the attribute of having been born in the same year and so being the same ages at the same times.

As a result, we somewhat disregarded the population in the sense of the continuously subsisting total population of each birth cohort comprising it. The fertility, mortality or propensity to migrate of a population in a given period is in fact predicated by the period-specific combination of life experiences of the different birth cohorts concerned. Each cohort adds to the population dynamics through its behavior at the age attained in the period of interest.

The combined behaviors of different birth cohorts at different ages are what determines natural and migratory population change. And changes in these combinations are the main focus of population policies, be they public health, migration or family policy.

Longitudinal or cohort analysis is a purely secondary means of tracking both short-term and longer-term trends. Not only because the statistical perspective required means that it can be done only long after the event, but short- or even long-term changes generally affect most of the age groups constituting the population at the time they occur, while new events or new changes occur as the cohorts age. So, to assess the influence of different social, economic, cultural factors on population development, and to give direction to the policies to deal with them, we must have time-specific indicators for fertility, mortality and migration and track their development over time. Whence the need to reconstruct by period analysis what was previously done by cohort analysis.

This is what will be done here, starting with Chapter 14, which applies the theoretical construct of the hypothetical cohort to the three main components of population change in turn: mortality, fertility, and migration.

At this juncture, there is a good case to be made for having simple indicators of these phenomena. Bearing in mind the problems arising from confounding variables, *standardization methods* can be used to define comparative measurements (Chapter 15). Chapter 16 analyzes the relationships between age-specific rates and total rates, and describes various *methods of differential analysis* of two life expectancies or two total fertility rates.

Comparing period and cohort indicators shows how far, in some cases, both can differ from each other, the former taking account of age-specific timing factors pertaining to different birth cohorts with different life histories at a single point in time within the same population, through period-specific recombination within a hypothetical cohort. Chapter 17 will clarify the relations between these two approaches through *translation models*.

Finally, this third part concludes with a discussion of *APC* (age-period-cohort) *models* that seek to link period and cohort analyses to pinpoint what is age-, period-, or cohort-specific in the development of a demographic phenomenon.

C H A P T E R

14

The Hypothetical Cohort as a Tool for Demographic Analysis

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

To construct period indicators for tracking the changes over time in the basic phenomena of population dynamics, we will resort to artificial constructs known as *hypothetical cohorts* (also known as *fictitious* or *synthetic* cohorts). These imaginary groups are assumed to display throughout their lives the fertility, mortality, and migration behaviors measured by the age-specific rates observed in a given period. The rates obviously reflect the behaviors of different real cohorts during the period. The relationships between these behaviors and the behaviors of the same cohorts at earlier ages and at older ages are presumably quite different from the links between these behaviors and those of other cohorts at other ages in the period considered. But this association of age-specific behaviors in a given period describes the state of the phenomenon studied in the period examined. To track the change in the phenomenon, we will pretend that the age-specific rates observed in each time fraction are those of a single cohort, the hypothetical cohort, for which we can measure life expectancy, mean number of children per woman, and other variables.

We examine mortality, fertility, and migrations.

I. THE PERIOD LIFE TABLE

The *period life table* assigns to a hypothetical cohort the age-specific risks of dying observed in a given period. It can be constructed only by measuring the risks. In this respect, it differs from the life table of a real cohort, which can—in theory, and under certain observation conditions—be constructed directly from actual deaths (see Chapter 11). There is only one set of circumstances in which we could define the observed age-specific deaths as the deaths in the *period life table*. Three conditions would be required: (a) an identical age-specific mortality for all the cohorts contributing to the period measure, (b) an identical initial number of births for all the cohorts, and (c) no change in cohort sizes due to migration flows at any time in their lives. No real population, of course, ever meets these criteria, which are those that define what we will later refer to as a *stationary population* (Chapter 20). Consequently, we can hardly ever regard the age-specific deaths observed in a given period as a possible starting point for a period life table.

1. From Age-specific Risks to the Period Table

We have already seen that, depending on the data available, different approaches can be used to estimate the age-specific risks of dying (Chapter 11). How should we go about describing the reality of a given observation period as faithfully as the available data allow? Let us start with the simplest case: a calendar year for which we want to measure the year-of-age-specific risk of deaths. Figure 14–1 illustrates the position on a Lexis diagram of the various measures of the risk of dying that may be used to construct a life table by year of age for an individual calendar year.

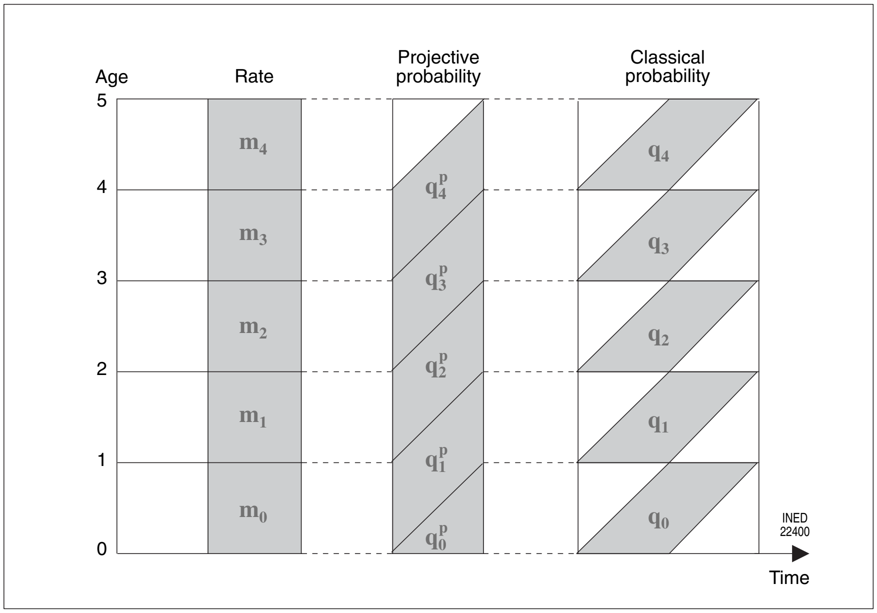


FIGURE 14–1 Positions on Lexis diagram of different measures of risk of dying suitable for constructing a calendar-year life table by year of age.

a. Using Rates as the Starting Point

Obviously we can start with the year-of-age-specific mortality rates (Figure 14–1, first column) and estimate the corresponding probabilities of dying. There are three advantages to this approach: the data are often accessible in the proper form, they concern precisely one calendar year, and they identify each year of age just as exactly. However, the probabilities are only approximations that assume an equal distribution of deaths in the two Lexis triangles. As seen earlier, the conversion formula based on this hypothesis is usually acceptable. But there are two cases where it may not be acceptable: (a) when the two cohorts crossing the same Lexis square were born at a time of abrupt swings in the birth rate and (b) above all, at the ages when the age-related swings in mortality within a given 1-year age interval are too wide (Chapter 11). Although the first exception is rare, the second is always a problem when measuring infant mortality and mortality at the oldest ages. In any event, the oldest ages pose other problems (as seen in Chapter 11), which we shall also discuss later in this chapter.

For infant mortality, two alternatives are possible. First, we can approximate the probability at age 0 by the classic infant mortality rate or, better, by an infant mortality rate obtained after weighting the deaths by the initial number of births in the two cohorts concerned (Chapter 11). We can also use empirical rate-to-probability conversion tables, such as the one published by Reed and Merrell (1939), which we shall discuss in connection with abridged tables.

b. Using Projective Probabilities as the Starting Point

Without stepping outside the specific framework of an observation year, we can calculate the projective probabilities of dying (Chapter 11). This requires a classification of the year's deaths by year of birth of the deceased. The drawback, as we know, is that the probabilities measure mortality not between two birthdays, but between two ages in complete years (Figure 14–1, second column). We can obviously deduct interbirthday probabilities by interpolation between two consecutive values. As a rule, a simple linear interpolation suffices, except at very young or very old ages, when the mortality function diverges sharply from the linear pattern even between two birthdays. However, this offers no solution to the specific problem of infant mortality, which lies somewhere in between a standard projective probability at 1 year and an initial probability at age 0 that concerns only one triangle of the Lexis diagram. Again, we can make up for this by using the classical infant mortality rate or, better, a rate weighted by the population at birth of the two cohorts involved.

c. Using Classical Probabilities as the Starting Point

Finally, if we have the double classification of deaths by age and year of birth, we can construct a period table from the classical probabilities of dying between two birthdays (Figure 14–1, third column). We will, however, encounter the same problem already discussed in connection with the calculation of mortality rates (Chapter 6): Each of these probabilities covers two contiguous calendar years; by combining the set of these age-specific probabilities, the mortality obtained lies somewhere between those of each of these 2 years, without taking either one fully into account. To incorporate the entire mortality of the year studied into this measure, we actually need to determine the classical probabilities over a 3-year period, centered on the year studied (Figure 14–2).

The drawback of this approach is that it involves a partial compensation between the mortality levels of the 3 years in question. This attenuates any possible fluctuations, which might have been worth studying.

d. Using Partial Probabilities as the Starting Point

With aggregated data, the only way to construct an annual life table by year of age from real probabilities is by computing the *partial probabilities* defined in each Lexis triangle (Figure 14–3). We then combine the probabilities two by two to obtain the classical probabilities between two birthdays determined on the sole basis of the data on the year studied (Vallin, 1973):

$$ {}_1q_x = 1 - (1 - q_x^1) \cdot (1 - q_x^2) $$

This assumes that we have the double classification of deaths by year of age and year of birth.

Whatever the method used to establish the probability of dying between two birthdays (estimate or direct calculation), we construct the life table of the year in exactly the same way as a cohort life table, starting from a radix l_0 and proceeding by successive iterations until the gradual extinction of the population. This is done by applying the age-specific probabilities of death, which give us the series of survivors l_x and deaths $d_{(x,x+1)}$ for the table. As we would do with cohort tables, we also calculate the table's other functions, in particular the stationary population of the table (person-years exposed to the risk) ${}_1L_x$ and the life expectancy at birth, e_0 or, at age x, e_x.

2. Closing the Table

Naturally, as with the life table of a real-life cohort, we need to adopt a solution for *closing* the table to minimize the problem of measuring mortality at the oldest ages. For this purpose, we can use the same procedures as those already mentioned in connection with the cohort life table (arbitrary choice of life expectancy at age k, estimation of probabilities at the oldest ages by means of the extinct-cohort method, or extrapolation using a function to provide the best possible fit for risks of dying at earlier ages). If we use the extinct-cohort method, once we have estimated the probabilities longitudinally, we will need to recombine them in cross-sectional form by selecting only those concerning the year studied. The distance from the year studied needs to be all the greater as we must estimate probabilities for ages far removed from the limit age of life.

In addition to these various methods for estimating mortality at the oldest ages, already discussed in connection with cohort tables (Chapter 11), there are two methods more immediately suited to the requirements of period-table computation.

Later on, we will see that—for a *stationary population*—life expectancy at birth is equal to the inverse of the crude death rate (Chapter 20) and, more generally, life expectancy at age x is equal to the inverse of the death rate of the age group x, $x + \omega$:

$$ e_x = \frac{1}{m_{(x,x+\omega)}} $$

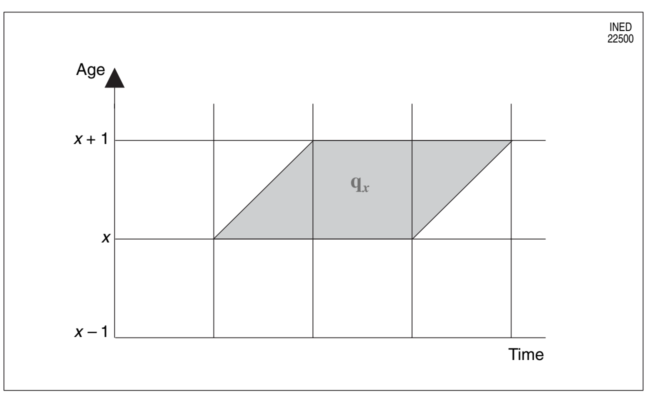


FIGURE 14–2 Probability between two birthdays taking full account of the central year (i.e., the focus of our study).

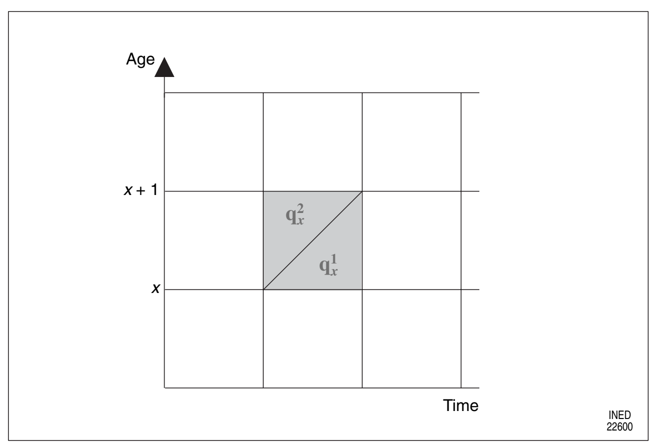


FIGURE 14–3 Partial probabilities between two birthdays allowing an exact coverage of the year studied.

As the stationarity conditions are rarely satisfied, it would of course be totally unrealistic to estimate life expectancy at birth from the crude death rate. On the other hand, given the relatively minor importance of the precision with which we need to estimate life expectancy at an old age k, we can—as Greville (1943) suggests—use the above formula to close a life table. This makes more sense for a period table, because we need only determine the final mortality rate at age x, $x + \omega$, in the open interval of age k, used in the death and population statistics of the year studied.

Coale and Caselli (1990) outlined another method. The authors point out that, thanks to civil registration, reasonably reliable statistics on age-specific deaths are usually available, at least for recent years: As deaths are officially reported, often with certified documentary evidence, the age of the deceased is known with a fairly high degree of precision. By contrast, the population's age distribution is determined far less accurately, as it is based on the census, which collects voluntary declarations without significant checks on the information supplied.

The method consists in calculating the age distribution at the oldest ages from the observation of deaths, a procedure that somewhat resembles the extinct-cohort method. If, as in the latter method, we admit that migrations beyond a certain age are negligible, the ratio of the population size at an age x to the size at the following age, $x + 1$, will depend both on mortality and on the population increase at that age. With ${}_1r_x$ as the population's growth rate at age x, the Coale–Preston equation (1982) gives us:

$$P(x) = P(x+1) \cdot \exp({}_1r_x) + {}_1D_x \cdot \exp\left(\frac{{}_1r_x}{2}\right)$$

where $P(x)$ is the mean population of age x in the observation period and ${}_1D_x$ the deaths at age x, $x + 1$ in the period. If we know the number of deaths, we need only estimate ${}_1r_x$ to determine the population at different ages through successive iterations. If we assume that the population at the oldest ages is closed (i.e., without migration), we can estimate the growth rate of the population aged x from the rate of increase in deaths at age x. To avoid the disturbances due to annual fluctuations in mortality, we use a mean annual growth rate over several years. Once we have calculated the age-specific populations, we can measure the risks of dying at older ages as before.

3. Abridged Tables

The available death statistics are not always detailed enough to calculate the probability of dying by year of age. The figures, for example, may consist in a classification of deaths by 5- or 10-year age groups, or even by age groups of variable size. In such cases, we can construct *abridged life tables*. The direct computation of period probabilities is usually impossible.[1] Our starting point, therefore, will almost always be the death rates by age groups: the problem is to determine to what extent we can use the classical rate-to-probability conversion formula, and what are the possible alternatives.

We know that if, in a given age interval a, deaths are uniformly distributed by age, the probability of dying between ages x and $x + a$ can be deduced from the mortality rates at the same ages as follows (Chapter 8):

$${}_aq_x = \frac{2 \cdot a \cdot {}_am_x}{2 + a \cdot {}_am_x}$$

As we saw in the preceding sections, the hypothesis of a uniform distribution of deaths in an age interval is usually acceptable when dealing with annual age groups. This is generally not the case when we are working with multiyear age groups. All other things being equal, the problem increases as the age interval widens. As we know, the survival function is far from linear and, over a fairly wide age interval, the mean age at death diverges significantly from the interval's midpoint. As a rule, at young ages, the mean age lies below the interval midpoint as long as mortality declines with age; on the contrary, at the older ages, when mortality increases with age, the mean age lies above the midpoint.

Reworking Long's demonstration (1968), Santini (1992) suggests a way to offset this drawback. He proposes that in a given age interval x, $x + a$, the total of years lived is equal to:

$${}_aL_x = a \cdot (l_x - d_{(x,x+a)} + {}_a\delta_x \cdot d_{(x,x+a)})$$

where ${}_a\delta_x$ represents the mean fraction of the interval lived by the deceased. The rate is accordingly equal to:

$${}_am_x = \frac{d_{(x,x+a)}}{a \cdot (l_x - d_{(x,x+a)} + {}_a\delta_x \cdot d_{(x,x+a)})}$$

and the probability to:

$${}_aq_x = \frac{a \cdot {}_am_x}{1 + (1 - {}_a\delta_x) a \cdot {}_am_x}$$

If we postulate the uniform distribution of deaths, this equation obviously yields the classical formula given earlier.

[1] Unless we are working on data from a retrospective survey processed in period form—as is the case, for example, in the Demographic and Health Surveys (DHS).

The problem is to estimate ${}_a\delta_x$ taking reality into account.

Several authors have proposed models to convert rates into probabilities as realistically as possible. Reed and Merrell (1939), analyzing 24 U.S. life tables built between 1900 and 1930, showed that the expression

$$ {}_a q_x = 1 - \exp\left[-a \cdot {}_a m_x - 0.008 \cdot a^3 \cdot ({}_a m_x)^2\right] $$

gives a good estimate of the probability in the universe of tables studied, except for the probabilities ${}_1q_0$ and ${}_4q_1$, for which the authors offer specific coefficients to allow for the under-recording of deaths.

Greville (1943) proposed the following equation:

$$ {}_a q_x = 1 - \exp\left[-a \cdot {}_a m_x - \frac{a^3}{12} \cdot ({}_a m_x)^2 \cdot (\ln {}_a m_x)'\right] $$

but Nathan Keyfitz (1977, p. 43) noted that, when $(\ln {}_a m_x)' = 0.096$, we end up exactly with Reed and Merrell's formula.

Keyfitz and Frauenthal (1975) offered another expression:

$$ {}_a q_x = 1 - \exp(-a \cdot ({}_a m_x + C)) $$

or

$$ C = \frac{({}_a P_{(x-a)} - {}_a P_{(x+a)}) \cdot ({}_a m_{(x+a)} - {}_a m_{(x-a)})}{48 \cdot {}_a P_x} $$

where ${}_aP_x$ is the population aged $(x, x + a)$.

In fact, if we develop the Keyfitz–Frauenthal equation, we realize it is a special case of the Reed–Merrell formula (Santini 1992).

4. Selected Results

Let us return to the French example of the cohort of Jeanne Calment, born in 1875, which we have discussed several times in connection with the cohort life table. It is interesting to compare the cohort table with the period tables compiled for 1875, the year of her birth; 1997, the year of her death; and 1936, the year in which she reached 61 years, the midpoint of her life.

Figure 14–4 compares the age-specific probabilities. The cohort's mortality is obviously equal to that of the year 1875 at age 0 and to that of the year 1997 at the oldest ages. However, it is between ages 0 and 15 that the cohort and the year 1875 are indistinguishable,

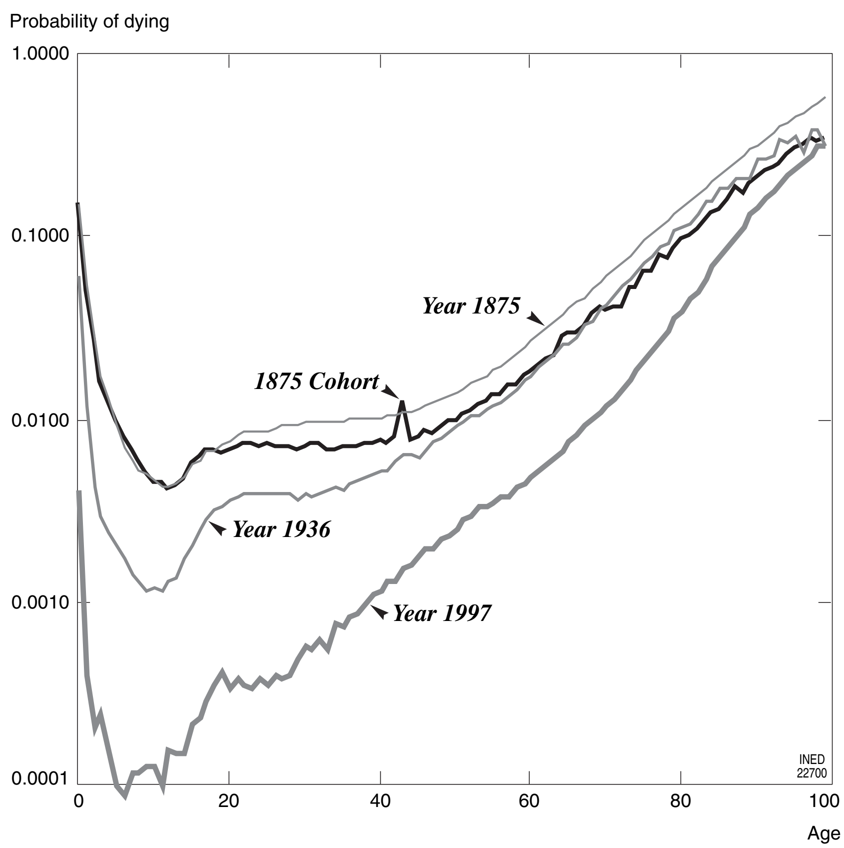


FIGURE 14–4 French female probabilities of dying for Calment's cohort compared with probabilities of period tables for year of her birth (1875), year of her death (1997), and year when she reached her midpoint in life (1936).

given the insignificant gains at the youngest ages between 1875 and 1890. Conversely, the mortality of the 1875 cohort lags behind that of the year 1997, since the gains at the oldest ages have only recently become substantial. The cohort naturally reaches the mortality level of the year 1936 toward age 60. In fact, the most characteristic aspect of this comparison between the cohort and the 3 years selected relates to the steady rise from the 1875 level toward the 1936 level, accomplished between ages 15 and 60. In these 45 years, the cohort reaped the full benefits of health progress, which enabled it to grow older throughout the period without a greater risk of dying (except for the peak due to the Spanish influenza). By contrast, after age 60, the cohort's mortality lingered very close to the 1936 figure, well above the 1997 level. The explanation is that the over-60s hardly benefited from healthcare advances before the 1970s.

This comparison clearly shows the very different possible uses of the cohort tables and period tables. The cohort table resembles none of the period tables. In this sense, it is clear that the device of the hypothetical cohort totally disproves the notions about the life experiences of actual populations that we sometimes derive from a reading of period indicators. On the other hand, only period tables can give us a precise view of the improvement in health standards over time.

Figures 14–5, 14–6, and 14–7 compare the four identical tables on three criteria: survivors at each age, age distribution of deaths in the table, and age-specific life expectancy. The survivor curve shows the extraordinary weight of infant and child mortality. Because it shares the high infant mortality then prevailing with the 1875 table, the cohort's survival rate at nearly all ages lingers far below the values for 1936 and 1997; in fact, it stays very close to the 1875 figure. The divergence does not become significant until after age 40, but this is too late to radically reverse the situation, because more than half the initial population has already died.

The main differentiating factor in the age distribution of deaths is infant mortality (Figure 14–6). But the most significant distinction between the period tables and the cohort table is the greater concentration of deaths around the second peak. From one period table to the next, the peak is increasingly concentrated, but at the same time it shifts rightward. As a result, the deaths are more scattered in the cohort than in any of the three period tables, even including the 1875 table.

For life expectancy, the third criterion, the curves once again display, of course, the weight of infant

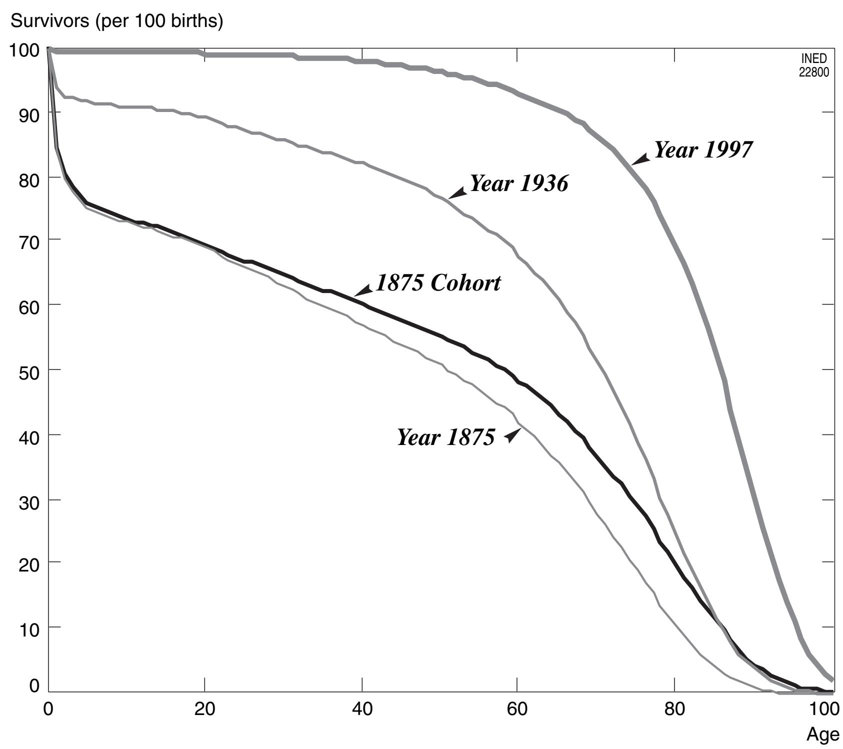


FIGURE 14–5 Survivors of French female life table of Calment's cohort compared with those of period tables for year of her birth (1875), year of her death (1997), and year when she reached her midpoint in life (1936).

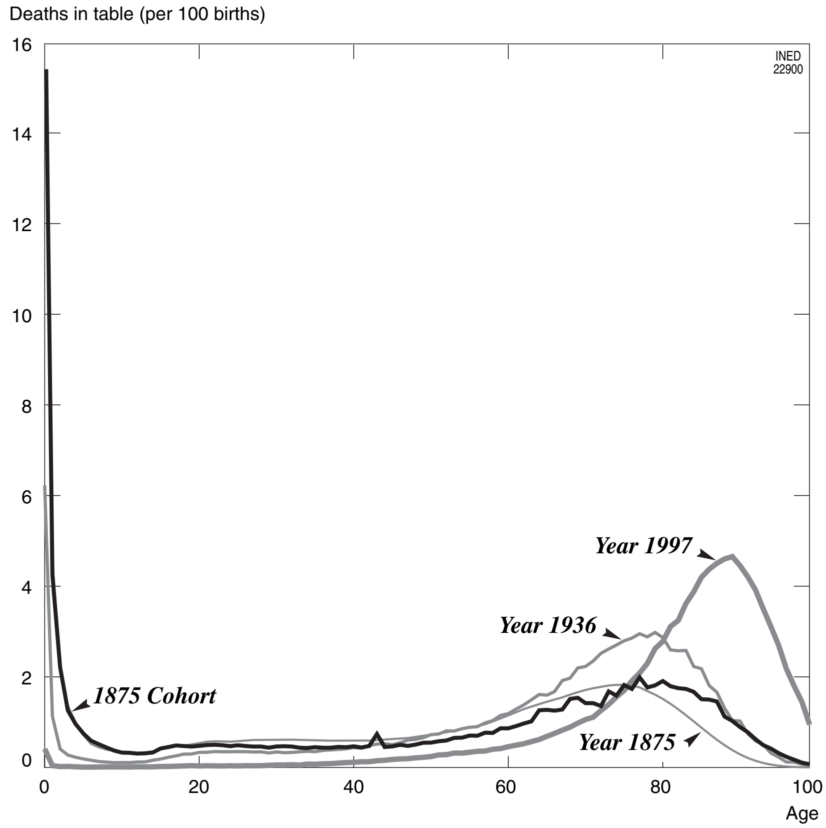


FIGURE 14–6 Deaths of French female life table of Calment's cohort compared with those of period tables for year of her birth (1875), year of her death (1997), and year when she reached her midpoint in life (1936).

mortality. Here, it is also responsible for a close resemblance between the cohort table and the 1875 table: Both are characterized by a sizable increase in life expectancy in the first years of life (Figure 14–7).

Unlike the survival curves, however, the life expectancy curves of the 1875 cohort and the year 1875 are entirely distinct from the outset. Whereas the two infant and child mortalities are very similar, the cohort life expectancy at the youngest ages incorporates the future gains at the adult ages. Indeed, the cohort life expectancy is quick to catch up with that of the year 1936, nearly 20 years ahead of schedule! Yet, the life expectancy at birth for the year 1936 was almost 15 years longer than that of the 1875 cohort. The reason is that unlike the survival curve, which depends only on earlier mortality, life expectancy depends solely on future mortality. The life expectancy curve of the year 1997 lies far above the other three. It embodies all the progress achieved since 1875. Infant mortality has fallen so low that the leap in the earliest ages still visible in 1936 has totally disappeared, and life expectancy decreases almost linearly with age from 0 to 80 years.

The life tables are given in the appendix.

Figure 14–8 compares the change in life expectancy at birth for France since the start of the 19th century given by the period tables with the values shown in the cohort tables. The representation of such a comparison raises several problems. First, we need to choose which observation year (period table) to compare with which birth year (cohort table). Should we compare the life expectancy of a cohort with (1) that given by the period table of the year of its birth, (2) that of the table of the year in which the cohort dies out, or (3) that of the year in which the cohort reaches its mean age at death? None of these solutions is truly satisfactory. In this chapter, we have decided to compare the mortality of each cohort with that shown in the period life table of its year of birth. We can thus measure the gap between (a) the life expectancy that could have been displayed by the cohort if mortality had remained constant and (b) the life expectancy determined from the cohort's actual history. This choice, however, raises a second problem: since we cannot know a cohort's actual history until after its total extinction, we cannot make the comparison on actual data except for cohorts born over a century ago. This would exclude from the comparison the period life expectancies observed in the

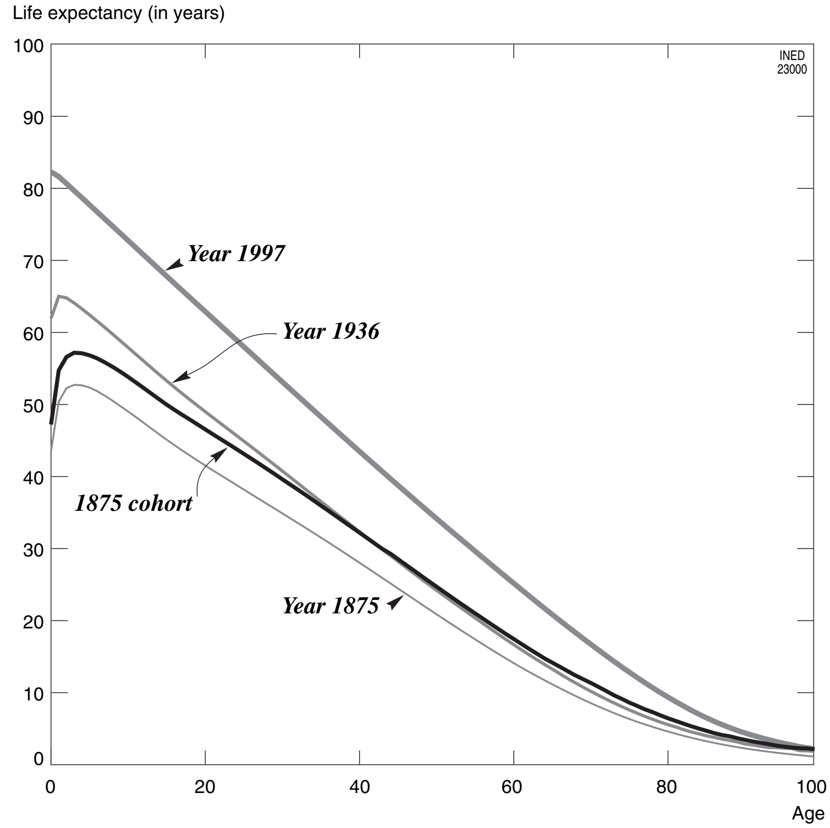


FIGURE 14–7 Life expectancies of French female life table of Calment's cohort compared with those of period tables for year of her birth (1875), year of her death (1997), and year when she reached her midpoint in life (1936).

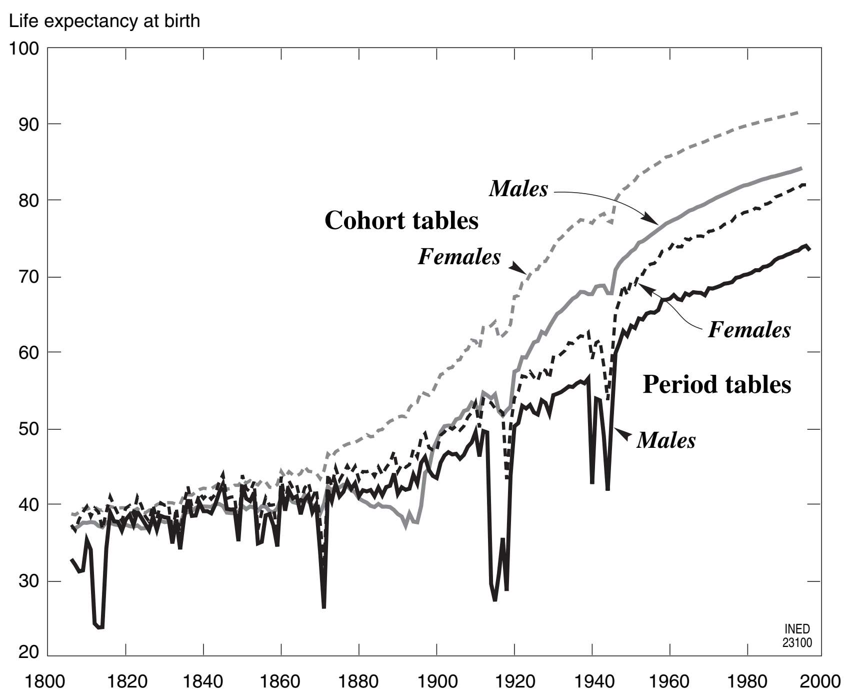


FIGURE 14–8 French female life expectancy at birth since 1806, by calendar year and cohort (cohort values are assigned to year of birth). (Source: Vallin and Meslé, 2001.)

20th century. To overcome this obstacle, we have used the results of an approximately 100-year projection of the period age-specific risks of dying, calculated from a straightforward extrapolation of recent trends (Vallin and Meslé, 2001).

As a rule, of course, the health progress achieved after each cohort's birth enables it, on average, to outlive the life expectancy of its birth-year table. This phenomenon is highly visible here for women. The difference, however, remains very modest for the cohorts born in the early 19th century. The gap does not begin to widen in earnest until the cohorts born after 1850. Initially, it stood at 1 or 2 years, with wide swings due to the persistence of major disturbances at the time. By 1880, the gap had reached 4 years, rising to 9 years at the start of the 20th century. The subsequent pattern naturally depends on the projections but, if these were realized, the gap would continue to widen for some time, reaching about 15 years on the eve of World War II. Beyond, however, it would start to narrow because of the slower gains in period life expectancy, which increasingly depend on the fall in mortality at older ages. In consequence, the gap would be reduced to 13 years for female cohorts born in 1950 and to 10 years for those born 1990 (but, at this point, cohort life expectancy is almost entirely invented by the projection).

The male pattern, which is very different here, shows that significant exceptions to this general phenomenon can result from major historical accidents. In France, male cohort survival was violently disrupted by the two world wars, particularly the first. Despite all the health improvements they enjoyed after their birth, males born between 1880 and 1897 had a far shorter life expectancy than if the mortality conditions of their year of birth had remained constant. All male cohorts born between 1875 and 1902 had a lower life expectancy at birth than the value given by the table for their birth year. In particular, the life expectancy of the 1894 cohort was more than 5 years shorter than that of the year 1894.

II. HYPOTHETICAL INTENSITY AND TEMPO OF FERTILITY

As we have already emphasized (Chapter 8), the subject of the study of fertility (childbearing) differs from the subject of the study of mortality (deaths) on two counts: First, it is not fatal; second, it is renewable. The analysis of fertility must therefore aim to measure two different variables: tempo and intensity. The intensity will seldom equal unity. Its value is just as likely to exceed unity (because fertility is a renewable event) as to be below unity (because we are dealing with a nonfatal event). The probabilistic approach, therefore, cannot be applied to the treatment of fertility as it can to the study of mortality. However, as we saw in the cohort analyses of Chapter 12, if we distinguish a given mother's childbirths by order, the fertility of each order is determined by nonrenewable events and is open to probabilistic analysis. In this case, for the reasons set out earlier, the phenomenon's intensity within the same cohort obviously cannot exceed unity and is usually below it. The conversion of cohort fertility rates into period fertility rates theoretically obeys the same logic, but—and this is where the artificiality of the cross-sectional computation is most clearly demonstrated—the conversion may sometimes lead to parity-specific intensities greater than unity, whereas no woman can obviously have more than one child of the same order.

1. From Age-specific Fertility Rates to the Total Fertility Rate and Mean Age at Childbirth

As discussed earlier, there are different ways to calculate year-of-age–specific annual fertility rates (Chapter 6). In practice, two methods prevail (Figure 14–9): the year-of-age- and observation-year-specific rate, $f_{x,t}$ (type-1 rate, in a Lexis square), and the year-of-birth– and calendar-year–specific rate $f^{p}_{x,t}$ (type 2 rate, in a *projective* Lexis parallelogram). The determination of the rate is most commonly restricted to the interval between ages 15 and 50, since fertility before 15 years and after 50 years is regarded as negligible. However, in some populations where early or late fertility is more frequent, it may be useful to define a wider interval.

As with mortality, if we assume a uniform distribution of births in the age interval, we can shift from one type of rate to the other simply by taking the half-sum of two consecutive rates. Although theoretically false, this hypothesis generally suffices for a valid comparison between series of different fertility rates. Moreover, unlike with mortality, we do not face any problem at the extreme ages. In particular, whatever the computation mode, the first rate is always comparable with the following ones: the first projective parallelogram is not truncated by age 0, since the calculation is only useless (and not impossible) below the minimum fertility age.

Depending on the data available, we may also need to use age-group–specific fertility rates (Chapter 6). One of the most common situations is having to work with five-year age groups.

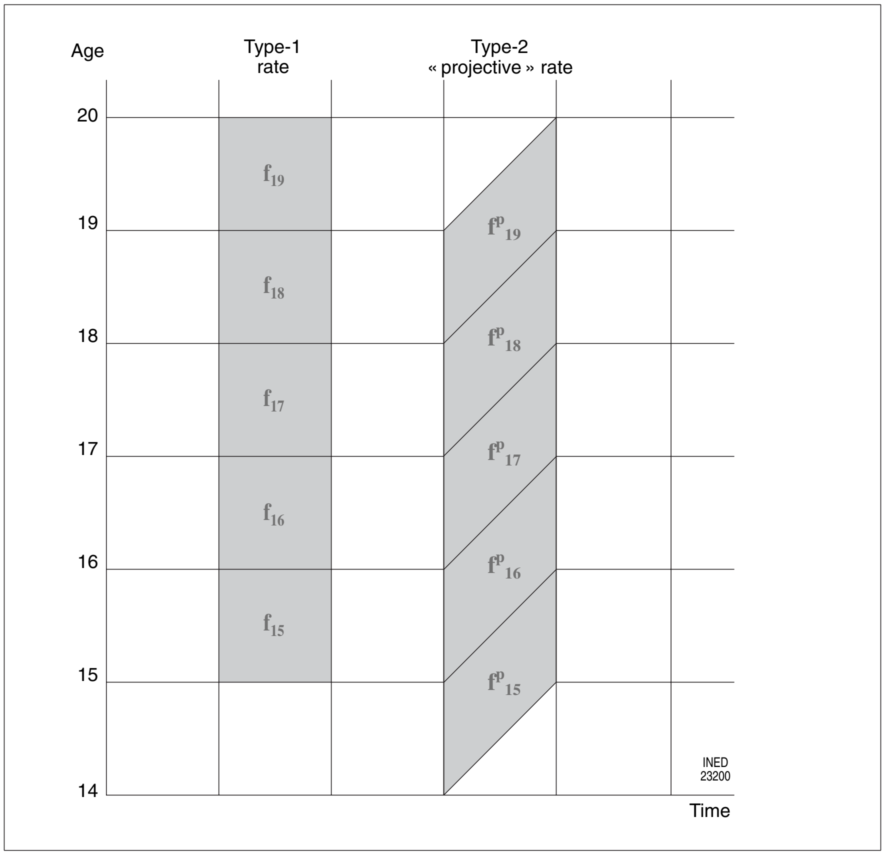


FIGURE 14–9 Position on Lexis diagram of the two most common measures of the age-specific fertility rate allowing the determination of the total fertility rate.

a. Total Fertility Rate

As we did with mortality, we will construct summary or synthetic period fertility indicators that resemble those already compiled on a cohort basis. For this, we combine the age-specific rates observed in a given period within a hypothetical cohort. At the total fertility level, the sum of the age-specific rates gives us the mean number of children per woman of this fictitious cohort. In mortality studies, the convention is to use the term "life expectancy" interchangeably for the hypothetical cohort exhibiting the period rate and for the value measured in real cohorts. In fertility studies, by contrast, the terminology is richer—perhaps too rich—since the expression *completed fertility* is strictly confined to the cohort measure; the period measure is designed differently according to the authors. French-speaking demographers refer in particular to the *sum of reduced births* (*somme des naissances réduites*) (Henry, 1972), *synthetic fertility index* (*indice synthétique de fécondité*) or even "period fertility indicator" (*indicateur conjoncturel de fécondité*) (Calot, 1984). More simply, in the rest of the world and notably among English-speakers, the standard term is "total fertility rate" (TFR) (*tasso di fecondità totale* in Italian, etc.). In our opinion, the *total fertility rate* or *TFR* is the term that best expresses the concept of summing the rates in period form, and we will therefore use it in the remainder of our discussion. However, we will reserve it strictly for period computations, whereas English-language authors often use it interchangeably for real as well as hypothetical cohorts.

If $f_{x,t}$ is the age-x-specific fertility rate in year t, the total fertility rate for the year t is:

$$TFR_t = \sum_{x=\alpha}^{\omega} f_{x,t}$$

where α and ω are the extreme ages of the female child-bearing period.

Naturally, if we are working on age groups, we must multiply each rate[2] by the dimension of the age interval a:

[2] The computation of period fertility rates in a hypothetical cohort effectively consists in applying annual rates to a female population of each age, assumed to be constant. If we are working with rates for 5-year age groups, we must apply each rate five times in a row to this constant female population. Conversely, if we were using month-of-age-specific rates, we would have to divide the sum by 12.

$$TFR_t = \sum_{x=\alpha}^{\omega} af_{(x,x+a),t}$$

If a is the same for all age groups, we obtain:

$$TFR_t = a\sum_{x=\alpha}^{\omega} f_{(x,x+a),t}$$

Most often, we will be dealing with 5-year age groups, and the TFR will be equal to:

$$TFR_t = 5 \cdot \sum_{x=\alpha}^{\omega} f_{(x,x+5),t}$$

The TFR has no standard equivalent in mortality, because its main purpose is to express the total intensity of the phenomenon (necessarily equal to 1 in mortality).[3] On the other hand, we must supplement this intensity indicator by a measure of the tempo of fertility, which does resemble life expectancy: the mean age at childbirth.

b. Mean Age at Childbirth

Unlike with the TFR, but as with life expectancy, the standard terminology does not distinguish at this level between the measure of the tempo of fertility of a cohort and that of a hypothetical cohort reflecting current conditions. In both cases, demographers speak of *mean age at childbirth*. By contrast, we can avoid confusion in regard to the age distribution of the births themselves: the expression *tempo of births* should be confined to the cohort analysis (where it is usually noted α_x), while the term *age-specific proportions of births* should be reserved for period measures. In a hypothetical cohort of women who, at each age, displayed the age-specific fertility measured in year t, the proportion b_x of births to women aged x in this cohort would be equal to:

$$b_x = \frac{f_{x,t}}{\sum_{x=\alpha}^{\omega} f_{x,t}}$$

Continuing the period analysis, we will, of course, find a mean age at childbirth of:

$$\overline{x} = \frac{\sum_{x=\alpha}^{\omega} (x+0.5) f_{x,t}}{\sum_{x=\alpha}^{\omega} f_{x,t}}$$

[3] We should note, however, that Sardon (1993) suggested using this method to construct a summary mortality indicator. In fact, this involves calculating a comparative mortality rate (see Chapter 15) on the basis of a population with equal numbers of individuals at all ages.

if type 1 rates are used, or

$$\overline{x} = \frac{\sum_{x=\alpha}^{\omega} x \cdot f_{x,t}}{\sum_{x=\alpha}^{\omega} f_{x,t}}$$

if type 2 rates are used.

If we work with 5-year age groups, we obtain

$$\overline{x} = \frac{\sum_{x=\alpha}^{\omega} (x+2.5) f_{(x,x+5),t}}{\sum_{x=\alpha}^{\omega} f_{(x,x+5),t}}$$

2. Comparison Between Cohort Indicators and Period Indicators

We will take as an example here the change in the fertility of Italian women, using the abundant database compiled by Santini (1995), which he kindly made available to us. Figure 14–10 compares the age-specific fertility rates of the 1952 cohort with those of (1) its birth year; (2) the year 1967, in which it reached age 15, the presumed start of the childbearing period; (3) the year 1981, in which it reached its mean age at childbirth (29 years); and (4) a year close to that of the end of its childbearing period (here, 1994, the most recent year for which data are available).

This comparison illustrates—as we already did for mortality—how the information supplied each year by period measures diverges over time from actual cohort fertility.

The 1952 cohort obviously has a much lower fertility (1.85 children per woman) than the period rates for its birth year (2.34), and we can see here that the discrepancy is entirely due to a tempo difference. Both curves peak at exactly the same level. In contrast, the cohort completed its fertility earlier than we would have expected from the pace indicated by the age-specific rates of its birth year: its fertility rates rose faster at the earliest ages; more important, they fell much sooner and more quickly after the age of peak fertility. This results from the combination of opposing period trends.

Between the year of the cohort's birth and the year in which it entered its childbearing period, Italy experienced the baby boom tied to the acceleration of its economic development in the 1960s. The TFR rose to 2.53 in 1967, well above the completed cohort fertility (CFR) of 1.85, but also above the *TFR* for 1952. We can see here that the difference between the years 1952 and 1967 is essentially due to the rise in fertility at early

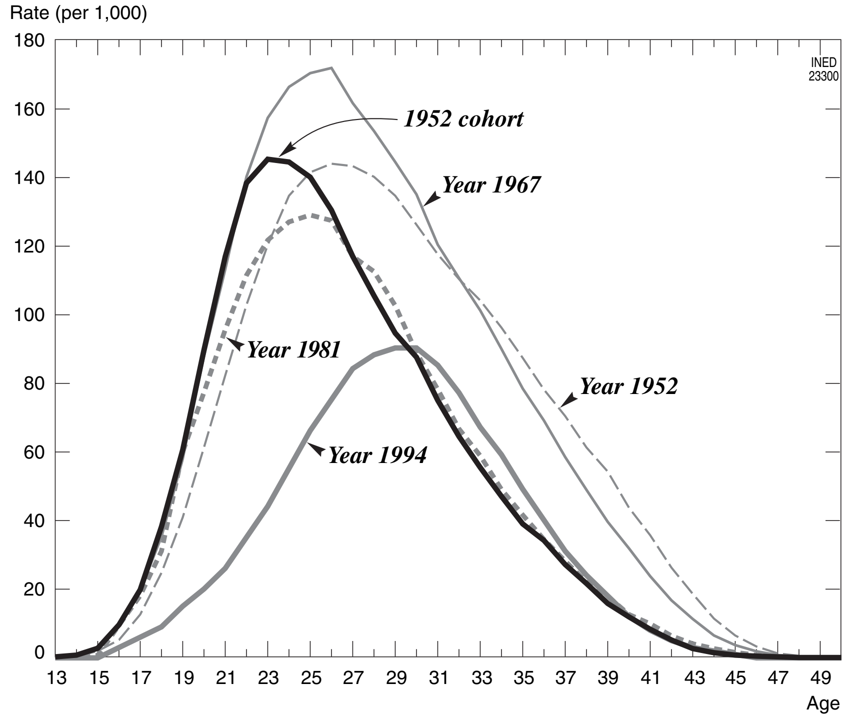


FIGURE 14–10 Age-specific fertility rates of the cohort of Italian women born in 1952 compared with those of its birth year (1952), of the year in which it reached age 15 (1967), of the year in which it reached its mean age at childbirth (1981), and the most recent year for which data are available (1994).

ages and to the increase in the peak. By contrast, the fertility rates over age 30 are the same. It is because the 1952 cohort entered its childbearing period in this very special phase of rising fertility that it began with high fertility rates at the young ages. The phenomenon was short-lived, however: The cohort's fertility leveled off very quickly, well below the peak observed in 1967.

Indeed, soon after the baby boom, Italy began experiencing the drop in period rates that drove its fertility to the world's lowest levels. By 1981, the year in which the cohort reached its mean age at childbirth, the TFR had fallen to 1.76, a value very close to the completed cohort fertility. This is understandable, since in both cases the age-specific fertility is heavily concentrated around its mode. We do note, however, that the 1981 TFR is slightly below the completed cohort fertility (1.78 versus 1.85), because of the cohort's higher early fertility due to the baby boom.

By the time the 1952 cohort reached the end of its childbearing period, Italy's period fertility had literally collapsed: the TFR was only 1.19 in 1994. However, this had virtually no effect on the cohort, as the gap between the age-specific fertility rates of 1981 and 1994 was actually confined to the rates at ages below 30 years—rates that the cohort had exceeded well before.

Figure 14–11 compares the change in Italian women's TFR with the pattern of their completed fertility over nearly three fourths of a century. As with mortality, the comparison between these two different kinds of indicators is problematic. For life expectancy, as we have seen, the age-specific mortality function is highly scattered, with two modes located at the two extremes: There is clearly no satisfactory solution, and we must choose one or the other depending on what we want to examine. For fertility, the situation is somewhat different. The phenomenon is heavily concentrated around a fairly central mode, which tends to be close to the mean age at childbirth. The solution that comes naturally to mind, therefore, is to emphasize the comparison of the completed fertility of a given cohort with the TFR of the year in which the cohort reaches its mean age at childbirth.

Two problems, however, immediately arise. First, the mean age at childbirth is not constant, and we must address the following question: Should we shift forward in time all the cohorts of the series studied by an identical number of years, roughly equivalent to the mean value of the mean age at childbirth in all the cohorts, or should we shift each cohort by its exact mean age? The first approach will cause a slight bias in the sought-for comparison, but the second method will create irregular intervals between two successive

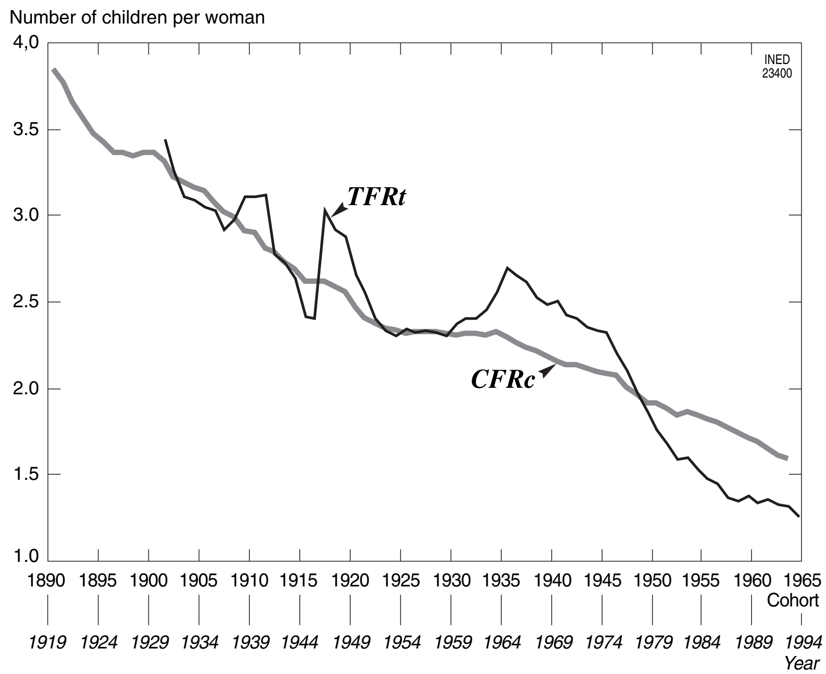


FIGURE 14–11 Change in Italian total fertility rate since 1930 compared with that of completed cohort fertility since 1890.

cohorts, which may make the chart harder to read. In fact this problem is usually minor, as the mean age at childbirth changes slowly; in most cases, demographers choose the first solution. Here, we have shifted all the cohorts by 29 years (Figure 14–12). The second problem is more fundamental: However rational it may seem, this exercise remains a comparison between two indicators of different kinds, which should therefore never be compared without losing sight of their different significance. It is because he forgot (or pretended to forget) this difference that one author, in making the comparison, felt entitled to conclude that one indicator—completed fertility—was objective and hence more trustworthy, while the other—the TFR—was ruled to be deceitful (Le Bras, 1991). In reality, the comparison must, above all, help the analyst and the reader to clearly identify the intrinsic difference between the two measures.

The example of Italy is particularly interesting, as it concerns the tumultuous recent history of the country's fertility (Figure 14–11). In the three fourths of a century observed, the downtrend is unquestionably steep regardless of which indicator is used. From this standpoint, there is no basic difference in the result. However, there is a striking contrast between the sharp swings in the TFR and the steady curve of completed fertility. This highly classic contrast perfectly reveals the intrinsic difference between the two indicators. Although the TFR faithfully registers all the bumps in the cycle, completed fertility may be regarded as a summary of changes in behavior within each cohort. By comparison with the TFR, these changes may seem fairly chaotic. Often, however, they simply represent intracohort phenomena driven by expectations or adjustments aimed at achieving a desired family size.

Indeed, neither the surge in fertility triggered by the profamily incentives of the Fascist regime in the 1930s (also observed in Germany in the same period), nor the birth deficit of World War II prevented Italian women from gradually but steadfastly reducing their completed fertility. The TFR upswing in the Fascist era actually made up for a lag that developed in earlier years, and the official measures enacted at the time presumably anticipated the population's desire to readjust family size. Even more clearly, the birth deficit due to World War II was compensated by a very characteristic recovery phase just after the Liberation; at that time, however, Italy did not experience a baby boom comparable to that observed in many European countries.

However, Italy registered a true baby boom in the 1960s, linked to the acceleration of its economic and social development, as we can readily see from the TFR curve. This phenomenon reflects a deep-seated change in behavior, in that it persisted long enough to produce a distinct upturn in completed fertility. For about 15 years—from the cohorts born in the early 1920s to the cohorts born in about 1935—completed fertility

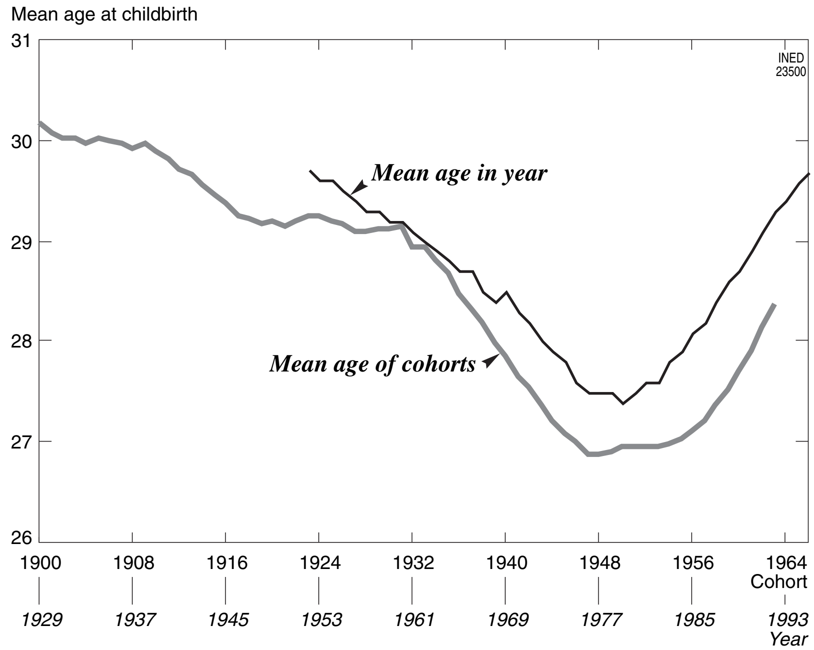


FIGURE 14–12 Period mean age at childbirth since 1952 and mean age at childbirth of cohorts since 1900, in Italy.

stopped falling and stabilized at around 2.3 children per woman.

In the end, though, the long-term decline in fertility resumed: the abrupt collapse of the TFR in the 1970s presumably signals a new, radical change of behavior whose impact on completed fertility cannot yet be fully measured for lack of sufficient historical perspective.

The changes in behavior affect not only the intensity of fertility, but also its tempo. Even if the fluctuations in the mean age at childbirth are confined, as we have said, to a relatively narrow range compared with that of the TFR or even of completed fertility, the mean age has changed significantly during the century in Italy (Figure 14–12). We can see here, however, that unlike the pattern for the intensity indicators, the two indicators of mean age at childbirth move on parallel lines. In both cases, the long-term reduction in the mean age suddenly quickened in midperiod (1960s or cohorts born after 1930); after bottoming out in the mid 1970s and in the cohorts born in the early 1950s, the mean age has since climbed back abruptly.

This similar pattern does come with a major difference, however: Throughout the period, the mean age at childbirth of each cohort was consistently below that of the year in which the cohort reached the mean age. Later, we shall see that what is true of total fertility is not necessarily true of parity-specific fertility.

However, the mean age at childbirth is only a summary index of fertility tempo. Changes in behavior can also alter the age distribution of births, and such shifts are not necessarily measured by the mean age.

To illustrate this point, Figure 14–13 takes up the comparison in Figure 14–10 between the cohort born in 1952 and four selected years: The year of its birth (1952), the year in which it reached age 15 (1967), the year in which it reached its mean age at childbirth (1981), and the most recent year for which data are available (1994). We simply replaced the age-specific fertility rates by the tempo for the cohort and the proportions of age-specific births for the calendar years.

The youngest of all distributions is that of the 1952 cohort. As described earlier, it reached the years of the Italian baby boom in the very first years of its childbearing period and later experienced the return to the downtrend in Italian fertility. These two events combined their effects to give this cohort what ultimately turned out to be a very early fertility.

However, we are more interested here in the changes in the period distribution of births. The contrast is striking between (1) the very steady movement, from 1952 to 1967 and then to 1981, toward a birth distribution increasingly concentrated around an ever-higher mode, and (2) the abrupt switch, from 1981 to 1994, to an entirely different distribution in which the mode, without changing level, suddenly shifted 5 years. Until 1981, the fall in the age at childbirth had been achieved simply through an increased concentration around a near-constant mode. By contrast, the

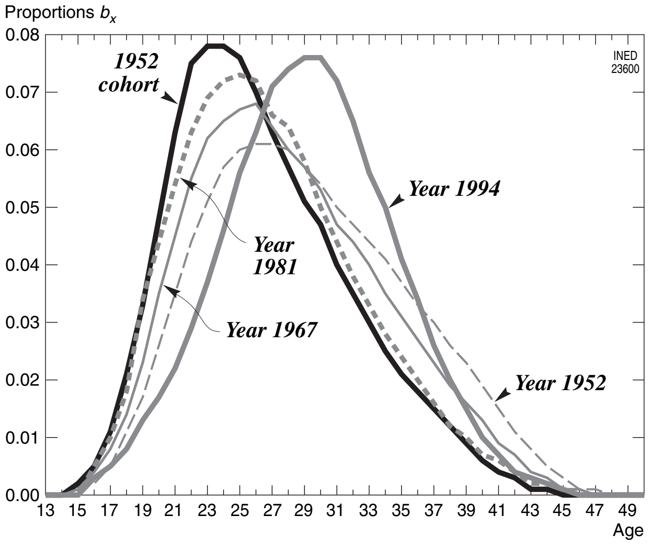


FIGURE 14–13 Fertility tempo of Italian female 1952 cohort compared with proportions of age-specific births observed in 1952, 1967, 1981, and 1994.

recent aging is being accomplished almost entirely through a shift in the mode, without a reduction of the concentration around it.

3. Translation Models

Through several comparisons between period indicators and cohort indicators, we have seen the extent to which circumstances can cause period indicators to diverge from cohort indicators, simply because of the recombination at a given moment, within a hypothetical cohort, of age-specific timing factors among different cohorts that have experienced different histories. The divergence is greater in fertility than in mortality owing to the fact that the change concerns both final intensity and tempo. It is therefore especially interesting to examine how the measure of period fertility may be affected by the changes in fertility tempo and intensity for the cohorts concerned.

Let CFR^i be the sum of reduced events of cohort i and TFR_t the sum of reduced events of year t. From a cohort perspective (but the situation would be symmetrical in a period approach), each rate $f^i_{x,t}$ can be viewed as a proportion of the completed fertility CFR^i achieved by cohort i in t:

$$f^i_{x,t} = \alpha_x CFR^i$$

The quantity α_x denotes the share of final intensity achieved by the cohort in t at age x. The period sum of reduced events of year t can be written as:

$$TFR_t = \sum_{x=15}^{49} \alpha^i_x CFR^i$$

This equation clearly expresses the fact that the period sum of reduced events is a function of the intensity and tempo of cohort fertility. The variations in tempo from one cohort to the next cause changes in the intensity of period fertility; ultimately, the observed cohort intensity is determined by the tempo and intensity that characterize the cohorts living through the period studied. *Translation models* rely on this relationship to measure the influence of changes in successive cohorts' fertility behavior on the variation in period fertility.

The resolution of the problem in its most general form is, however, very complex. We need to take into account all the possible types of change, from one cohort to the next, in both components of cohort fertility: intensity and tempo. Norman Ryder (1963, 1964) was the first to formulate a model for doing so. Keilman comprehensively addresses this issue in Chapter 17.

4. Parity-specific Fertility

As we saw in our discussion of cohorts (Chapter 12), we can study fertility as a series of nonrenewable events by distinguishing births by order. We can thus also consider a translation of parity-specific fertility indicators into period indicators. However, thanks to retrospective surveys (presented in Chapter 12), the current conditions of period observation are very different from those we can use in a cohort approach. Such surveys, provided they cover sufficiently large samples, could no doubt allow period reconstructions, at least for the few years immediately preceding the survey.[4] The Family Survey conducted in France with each population census may, for example, provide the necessary information. But this kind of operation does not allow a computation of parity-specific period fertility indicators except for a few periods scattered in time. Consequently, they cannot provide a steady tracking of phenomena, as we would expect of period analysis. The latter depends essentially on current statistics.

Now, while current birth statistics are fairly often compiled by birth order (or mother's parity), the same is not true of January 1 population estimates. These never give the distribution of women by parity. The standard practice, therefore, is to calculate parity-specific fertility rates, expressed as the ratio of births of each order, not exclusively to women of the parity immediately below (i.e., the mothers who have not yet had a child of the order examined), but to all women

[4] These surveys typically cover samples of women of childbearing age at the time of the survey, and we cannot calculate period rates at all ages except in the most recent period.

regardless of their parity.[5] In other words, instead of computing a specific occurrence-exposure rate (known in French as a *first-category rate* [*taux de première catégorie*]) as follows (see Chapter 8):

$$_1f_x^k = \frac{B_x^k}{\bar{F}_x^{(k-1)}}$$

where B_x^k stands for births of order k of a mother aged x and $\bar{F}_x^{(k-1)}$ stands for the mean number of women aged x and of parity k-1, we calculate a fertility frequency (known in French as a *second-category rate* [*taux de deuxième catégorie*]):

$$_2f_x^k = \frac{B_x^k}{\bar{F}_x}$$

typically written in simpler form as f_x^k, where $\bar{F}_x$ represents the mean number of all women aged x, regardless of their parity. This measure of parity-specific fertility has a drawback and an advantage. The drawback is its bias: By including in the denominator women who are not concerned by the phenomenon, we underestimate the fertility of each order. On the other hand, the values obtained for each order are additive, so that by summation we can obtain the TFR for all orders in the aggregate.

For instance, for order k and observation year t, we find the k-specific TFR:

$$TFR_t^k = \sum_{x=\alpha}^{\omega} f_{x,t}^k$$

and we obtain the all-orders TFR by summing the order-specific TFRs:

$$TFR_t = \sum_{k=1}^{n} TFR_t^k = \sum_{k=1}^{n} \sum_{x=\alpha}^{\omega} f_{x,t}^k$$

We can also determine the proportions of order-k births:

$$b_{x,t}^k = \frac{f_{x,t}^k}{\sum_{x=\alpha}^{\omega} f_{x,t}^x}$$

and an order-k mean age at childbirth equal to:

$$\bar{x}_t^k = \frac{\sum_{x=\alpha}^{\omega} (x+0.5) f_{x,t}^k}{\sum_{x=\alpha}^{\omega} f_{x,t}^k}$$

Let us return to the Italian example. Figure 14–14 illustrates the change in the order-specific TFR for orders 1, 2, 3, 4, and higher. This distinction basically enables us to grasp two phenomena. First, only the order-1 and -2 fertility rates are highly exposed to purely short-term fluctuations (jump in the 1930s, World War II deficit, and postwar catch-up). By contrast, orders 3 and higher are also subject—rather conspicuously—to the more fundamental changes that followed: the baby boom of the 1960s and the sharp drop of the 1970s to 1980s, although these events have a sizable effect on the first two orders as well.

Second, although the order-1 and -2 fertility rates are barely sensitive to the century-long downtrend, the entire burden of the decline has been shouldered, until recently, by orders 3 and higher. We also note, however, that the recent slump also affects order-2 fertility and even, to a lesser extent, order-1 fertility. This clearly denotes a qualitative change in the century-long decline of Italian fertility. Until the 1970s drop, the downtrend not only failed to undermine first-birth fertility, but even coincided with an upturn in the latter, partly thanks to the decline in sterility. Now, the behavior of a growing number of Italian couples is actually compromising the entry into the childbearing period.

This breakdown of fertility by birth order applied to Italy also gives us the full measure of the artificiality of the hypothetical cohort: in a real cohort, the parity-specific TFR can never, by definition, exceed unity, because no woman can have more than one child of the same order; with the hypothetical cohort used here, the order-1 TFR exceeds unity in some years.[6] This phenomenon occurs in the years of unusually high fertility, during the postwar catch-up (1947 and 1948) and in the Italian baby boom (1964 and 1965), despite the fact that the fertility frequencies underestimate real parity-specific fertility.

Table 14–1, illustrated by Figure 14–15, presents the Italian age- and parity-specific rates for the year 1964. Figure 14–16 plots the mean age of mothers at the birth of their first child in Italy, again by comparing the period and cohort data. As already noted with all-parities fertility, the two indicators move in a fairly similar pattern. However, the two curves virtually overlap here; in the all-parities measure, the cohort indicator ran consistently above the period indicator.

[5] The term *fertility rate* is ambiguous, as pointed out in Chapter 8. As in the latter, we distinguish here *occurrence-exposure fertility rates* from *fertility frequencies*.

[6] This is due to the type of indicator used: the anomaly is very visible with the fertility frequencies (*taux de seconde catégorie*) used here, but is eliminated when occurrence-exposure rates (*taux de première catégorie*) are used (see Chapter 8 for definitions).

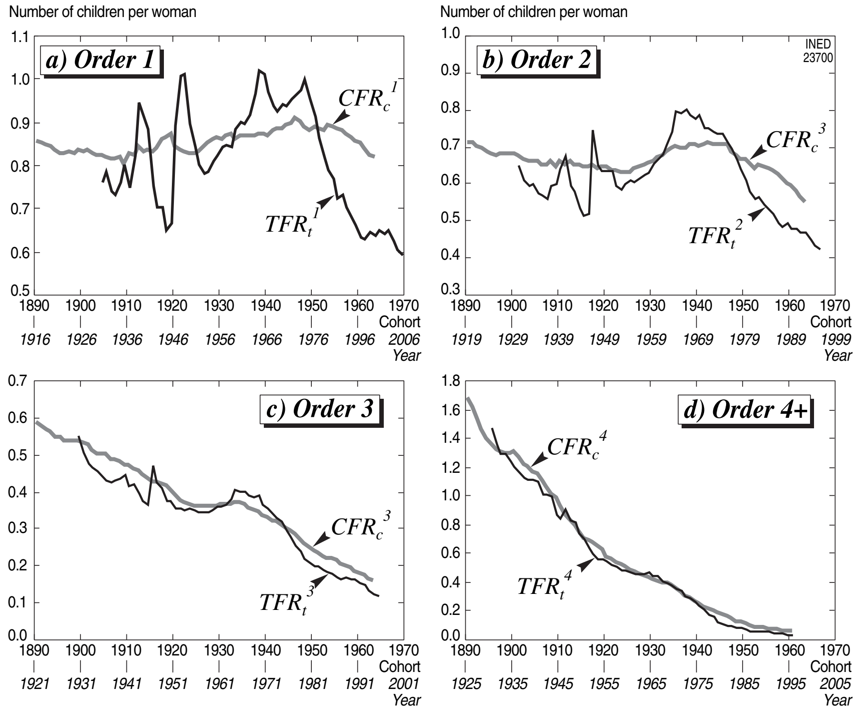


FIGURE 14–14 Italian parity-specific total fertility rate (1950–1994) compared with parity-specific completed fertility of cohorts from 1890 to 1963.

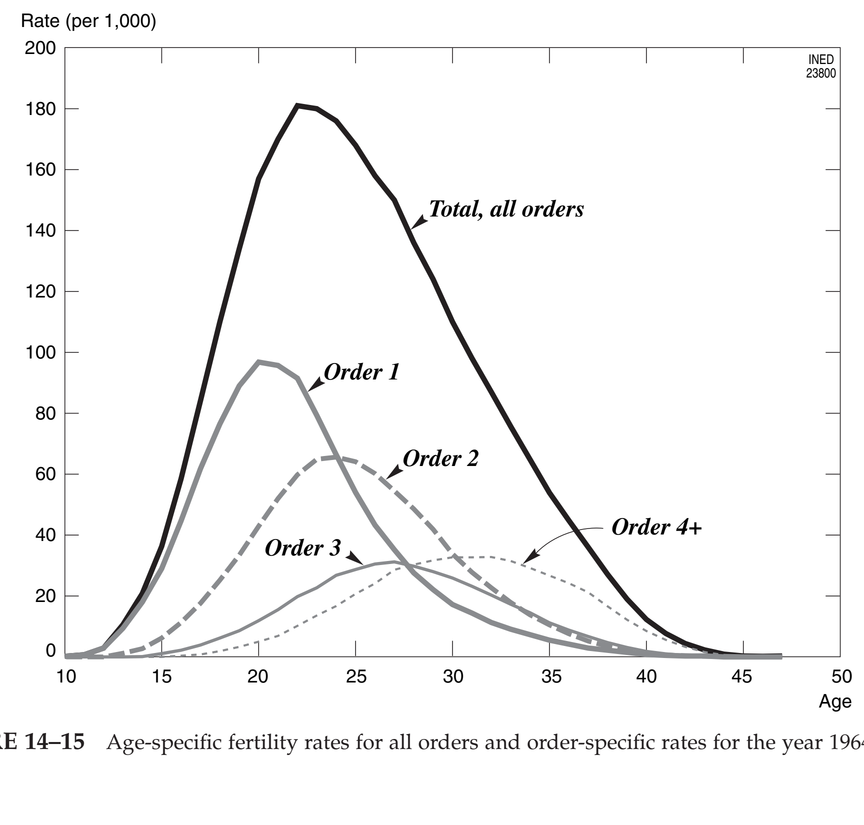


FIGURE 14–15 Age-specific fertility rates for all orders and order-specific rates for the year 1964 in Italy.

TABLE 14–1 Age-specific Fertility Rate and Order-specific *TFR* in Italy, in 1964

Age	Order				All orders
	1	2	3	4+	
13	0.2	0.0	0.0	0.0	0.2
14	0.7	0.0	0.0	0.0	0.7
15	2.9	0.0	0.0	0.0	2.9
16	9.4	1.2	0.0	0.0	10.6
17	18.0	2.7	0.1	0.0	20.8
18	29.0	6.2	1.1	0.0	36.3
19	44.7	11.3	2.2	0.4	58.5
20	61.8	17.6	3.9	0.8	84.1
21	76.5	25.2	6.2	1.9	110.0
22	89.0	33.4	8.6	3.3	134.0
23	96.8	42.9	11.9	5.0	157.0
24	95.7	52.0	15.5	6.9	170.0
25	91.5	59.8	19.8	10.2	181.0
26	79.3	64.9	22.7	13.6	180.0
27	66.4	65.6	26.8	16.7	176.0
28	54.0	64.1	28.7	20.7	168.0
29	43.3	60.2	30.5	24.1	158.0
30	35.2	54.5	31.2	28.6	150.0
31	27.6	48.6	29.8	30.3	136.0
32	22.1	41.9	28.0	31.7	124.0
33	17.2	33.8	25.9	32.7	110.0
34	14.4	27.8	23.2	32.7	98.1
35	11.3	22.7	20.3	32.8	87.0
36	9.1	17.9	17.3	31.5	75.7
37	7.3	13.8	14.5	29.2	64.8
38	5.5	10.5	11.2	26.7	53.9
39	4.2	7.6	8.8	24.2	44.8
40	2.9	5.2	6.7	21.2	36.0
41	2.2	3.5	4.7	16.6	27.1
42	1.5	2.2	3.1	12.2	19.0
43	0.8	1.2	1.8	8.5	12.3
44	0.5	0.6	1.0	5.6	7.7
45	0.3	0.4	0.5	3.3	4.4
46	0.3	0.2	0.2	1.6	2.3
47	0.0	0.2	0.1	0.6	0.9
48	0.0	0.0	0.1	0.2	0.3
49	0.0	0.0	0.0	0.1	0.2
50	0.0	0.0	0.1	0.2	0.3
TFR	1021.6	799.7	406.5	474.1	2702.9

TFR, total fertility rate.
From Antonio Santini database

5. Age- and Order-specific Birth Probabilities

The paradox of greater-than-unity order-specific TFRs is eliminated by construction when we switch to probabilistic calculus for constructing order-specific fertility tables: This involves a recombination by period not of the fertility rates but of the probabilities of giving birth to a order k child at age x. Since we are dealing here with nonrenewable events, we can indeed (as with mortality) compute order-specific birth probabilities by determining the ratios of order k births occurring between two birthdays to the initial number of women having borne k-1 children. Whelpton (1946, 1954) was the first to use summary fertility indices compiled from fertility tables.

The problem we face again here, however, is data availability. Only surveys provide the identification of female parity needed to correctly determine the number of women in the denominator. We therefore cannot construct order-specific fertility tables except on an occasional basis, when a survey or census offers a favorable opportunity.

In such circumstances, the probabilities are seldom calculated directly, but—as we already saw with mortality—they can be estimated from the rates without risk of serious error. The only requirement is that the rates should be true occurrence-exposure rates as described in Chapter 8 (see also Federici, 1965; Pressat, 1969). We can then construct a fertility table for order k as follows: we take an agreed initial population of women aged α having had k–1 children; through successive iterations—as with the life table—we apply to that population the order-k birth probabilities to obtain the number of parity-(k–1) women remaining at each birthday and the order-k births of the table between two birthdays. The total number of births in the table can never exceed the initial number of parity-(k–1) women, and the final intensity of order-k fertility obtained by dividing the two values will not exceed unity. In fact, the figure will usually be below unity, given that maternity is not a fatal event and that all women having had $k - 1$ children will not necessarily go on to have a kth child.

The problem is that we seldom have the data available on the parity distribution of the female population. This is why the computations are often confined to fertility frequencies. Shryock and Siegel (1973) proposed an indirect estimation of occurrence-exposure rates from frequencies;[7] Giorgi (1993) showed that this procedure yields a result that is very close to that of the direct calculation.[8] The estimate must, however, be determined within each cohort. We can accordingly write, for all orders $k > 1$:

$$_1f_x^k = \frac{f_x^k}{\sum_{15}^{x-1} f_x^{k-1} - \sum_{15}^{x-1} f_x^k}$$

and, of course, for order 1:

[7] See also Santini (1992).
[8] See also the applications by Das Gupta and Long (1985).

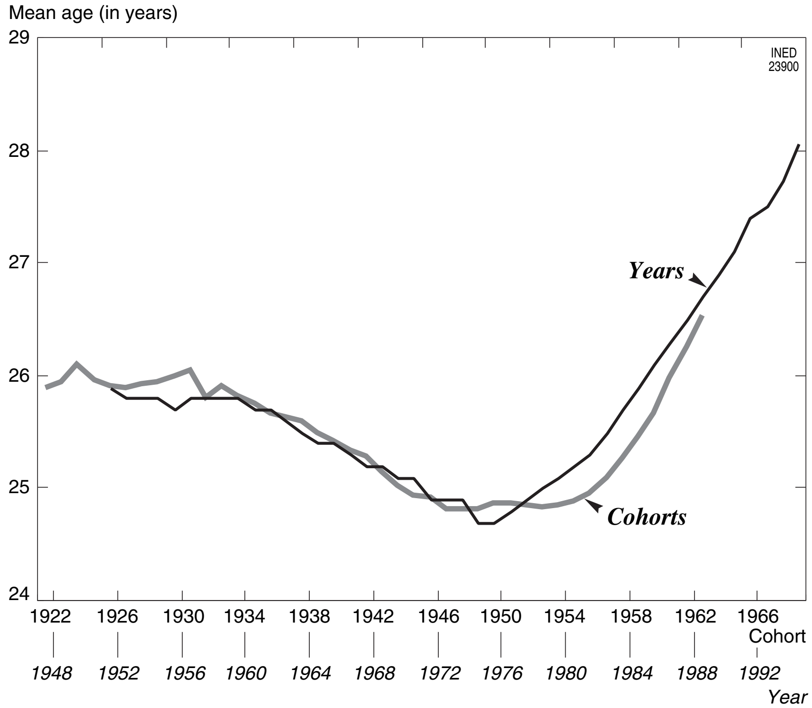


FIGURE 14–16 Mean age of mother at birth of first child, based on period data (1952–1994) compared with mean age observed for cohorts from 1922 to 1963, in Italy.

$$ {}_1f_x^1 = \frac{f_x^1}{1 - \sum_{15}^{x-1} f_x^1} $$

From these occurrence-exposure rates (which, for narrow age intervals, can be equated with birth probabilities), Giorgi (1993) constructed, on a probabilistic basis, *parity-specific period fertility tables* for Italy since 1950. The sum of the table's order-*k births* gives the hypothetical cohort's order-*k* completed fertility. Thanks to these data and the cohort figures compiled by Santini (1995),[9] we can compare over a long period (Table 14–2 and Figure 14–17) the order-1 completed fertilities for each period (CFR_p^1) and each cohort (CFR_c^1) with the order-1 total fertility rate calculated from the fertility frequencies (TFR_{t2}^1). Two facts are immediately obvious: (1) The paradox of greater-than-unity TFRs is eliminated by using occurrence-exposure rates or birth probabilities for calculating the sum of births in the table; (2) the births of the period table closely follow the changes in completed fertility of cohorts reaching their mean age at order-1 childbirth. The inefficiency of the fertility frequencies is evident, as is the imperative need to avoid equating their sum with the completed fertility of a hypothetical cohort.

[9] We thank the two authors for giving us access to the computer files of their findings.

The main goal here, however, is not to compare cohort measures with period measures but to describe the sequence of events with as little bias as possible.

Table 14–2 also gives mean ages at childbirth calculated from fertility frequencies and from the period-table births. Again, we can see the wide gap between the two measures and the need to avoid the measure derived from fertility frequencies (Figure 14–18).

Ultimately, the combination of final intensities of the parity-specific fertility tables and the combination of the final intensities of TFRs expressed as occurrence-exposure rates offer an identical measure of the final intensity of total period fertility—which, however, differs considerably from the *TFR based on fertility frequencies*. Figure 14–19 plots the curves for the two types of rate that measure total fertility regardless of parity distribution. For the sake of comparison, we have added the completed-fertility curve.

The period covered by the chart includes both the baby boom and the recent collapse of Italian fertility. Whatever the measurement method used, the TFR swings are much sharper than those of completed fertility. But the frequency-based TFR greatly exaggerates the real difference indicated by the occurrence-exposure TFR. All the same, the abrupt plunge in the occurrence-exposure TFR after the baby boom and even more so in the 1970s has dragged down completed fertility: Although slower, this decline does seem inevitable.

TABLE 14–2 Order 1 TFR^1_{t2} and Mean Age at Order 1 Childbirth Calculated from Fertility Frequencies, Compared with Sum of Births (Final Intensity) and Mean Age at Childbirth Given by Parity-1 Fertility Tables in Italy, 1952–1990

Cohort		Period				
Cohort	Completed fertility	Year	Intensity		Mean age at childbirth	
			TFR^1_{t2}	Sum of births in table	Calculated from fertility frequencies	Calculated from births in table
1926	846.0	1952	798.6	846.3	26.0	26.0
1927	853.2	1953	795.4	842.6	26.0	26.0
1928	860.4	1954	827.4	853.8	26.0	25.9
1929	864.6	1955	842.8	857.5	25.9	25.8
1930	863.2	1956	859.7	861.7	25.9	25.7
1931	869.6	1957	866.3	865.6	25.9	25.7
1932	873.1	1958	860.6	865.1	26.0	25.8
1933	861.3	1959	906.3	880.7	25.9	25.6
1934	870.2	1960	919.5	886.0	25.9	25.5
1935	871.5	1961	938.7	891.9	25.8	25.4
1936	869.2	1962	948.0	898.4	25.8	25.3
1937	870.9	1963	981.1	908.1	25.7	25.1
1938	872.5	1964	1036.0	921.7	25.6	24.8
1939	873.2	1965	1016.8	921.2	25.5	24.8
1940	871.0	1966	975.1	915.8	25.4	24.9
1941	882.9	1967	935.9	907.8	25.4	24.9
1942	885.9	1968	924.4	905.9	25.3	24.9
1943	887.2	1969	940.8	913.4	25.3	24.8
1944	894.9	1970	938.7	912.2	25.1	24.7
1945	898.7	1971	969.5	919.5	25.0	24.6
1946	910.5	1972	987.2	922.2	24.9	24.4
1947	902.1	1973	998.6	926.0	24.8	24.3
1948	892.4	1974	1024.1	936.3	24.8	24.2
1949	880.7	1975	982.9	929.1	24.7	24.2
1950	889.7	1976	928.4	922.1	24.8	24.4
1951	885.5	1977	882.8	917.6	24.9	24.6
1952	881.9	1978	843.7	910.6	25.0	24.8
1953	893.9	1979	811.8	904.9	25.0	25.0
1954	890.7	1980	805.4	903.0	25.0	25.1
1955	886.9	1981	779.9	895.2	25.1	25.3
1956	881.8	1982	709.2	874.4	25.3	25.8
1957	873.0	1983	682.2	864.4	25.5	26.1
1958	864.5	1984	653.7	852.2	25.7	26.5
1959	859.3	1985	633.7	841.6	25.9	26.9
1960	853.8	1986	606.1	823.5	26.1	27.2
1961	841.6	1987	599.3	815.0	26.3	27.6
1962	825.6	1988	613.4	816.6	26.5	27.7
1963	822.5	1989	605.6	803.4	26.7	28.0
1964		1990	617.2	801.5	26.9	28.1

Sources: Databases compiled by Antonio Santini (columns 2, 4, and 6) and Piero Giorgi, 1993. Una rilettura della fecondità del momento per ordine di nascita in Italia nel periodo 1950–1990 considerando la struttura per parità, *Genus*, vol. 49(3–4), p. 177–204. (columns 5 and 7).

6. Taking into Account Duration Since Last Birth and Constructing New Summary Indicators of Period Fertility

In an interesting article, Rallu and Toulemon (1993) showed that the TFR was only one of several summary indicators of period fertility. The basic approach in period analysis is to measure the fertility of a hypothetical cohort that would be subject throughout its life to the conditions of the year studied. If we want to take this notion to its ultimate conclusion, we should not simply standardize the indicators on the basis of age by summing age-specific rates. We should also incorporate other variables that can influence fertility, notably birth order, but also duration since last birth, marital status, educational attainment, socio-

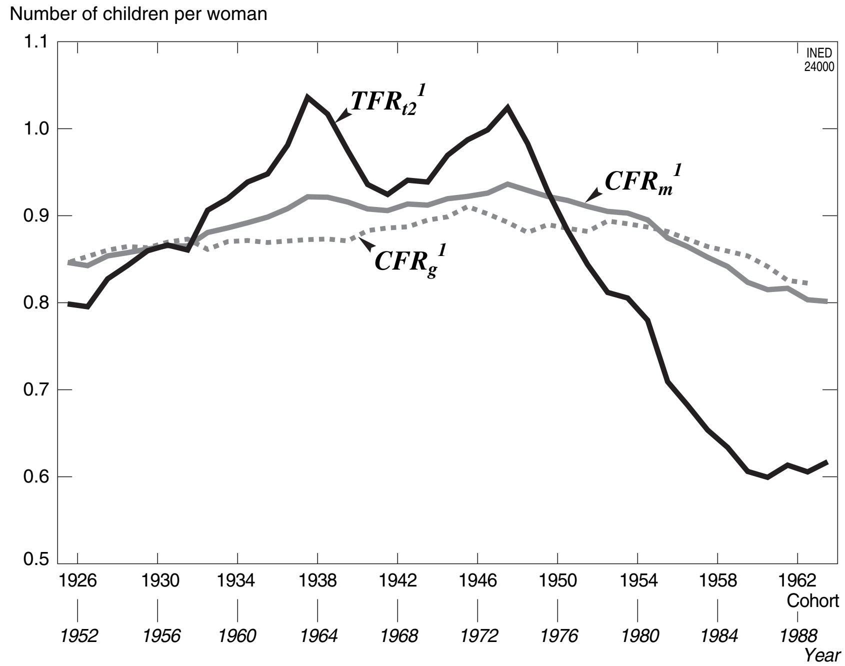


FIGURE 14–17 Order 1 period (CFR^1_m) and cohort (CFR^1_g) completed fertilities compared with order 1 total fertility rate calculated from fertility frequencies (TFR^1_{t2}).

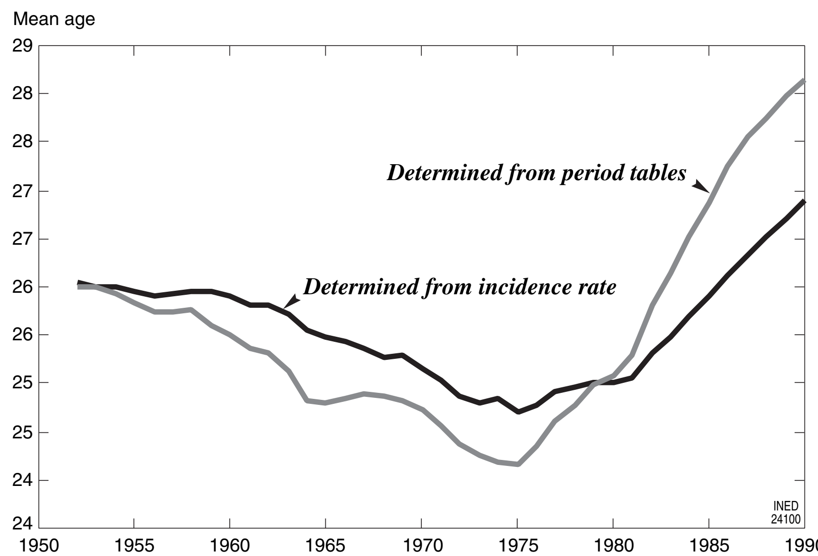


FIGURE 14–18 Mean age at childbirth calculated from period fertility tables compared with value obtained from fertility frequencies (Italy, 1950–1990).

occupational category, and so on. The authors show that, depending on the type and degree of detail chosen for these standardizations, we can obtain summary indicators of period fertility whose values may differ rather significantly.

In particular, Jean-Louis Rallu and Laurent Toulemon compare the classic TFR as defined earlier with two other indicators, each containing an additional refinement. The first is a summary index of parity- and age-specific fertility (*PATFR*); the second is a summary index of parity-, age-, and duration-specific fertility (*PADTFR*).[10]

[10] In French, the authors designate the standard *TFR* as *ISF* (*indice synthétique de fécondité*), and their two new indicators as *ISFRA* (*indice synthétique de fécondité d'après les quotients par rang et âge*: here *PATFR*) and *ISFRAD* (*indice synthétique de fécondité d'après les quotients par rang, âge et durée*: here *PADTFR*).

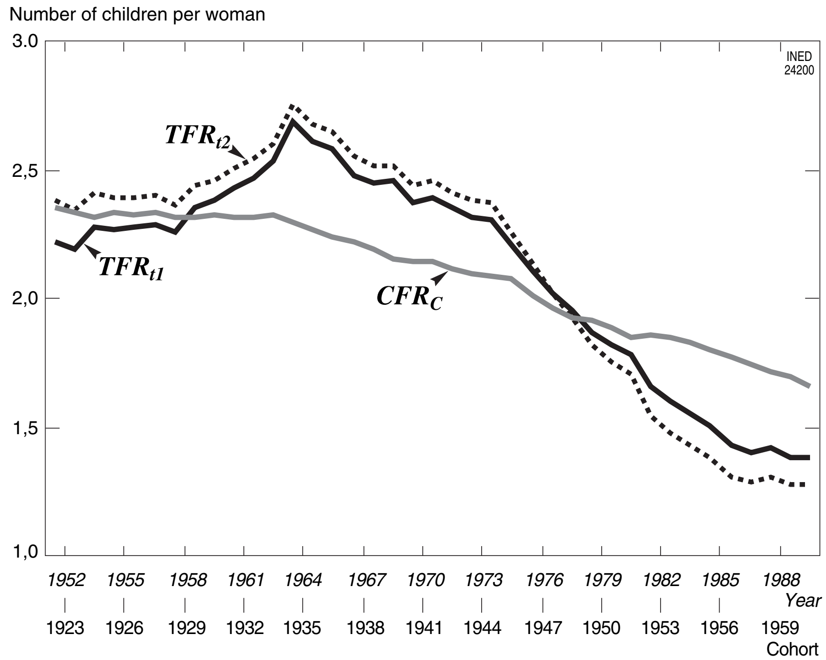


FIGURE 14–19 Occurrence-exposure and frequency total fertility rates 1952 to 1988, compared with completed fertility of cohorts from 1923 to 1959 (Italy).

To measure the impact of parity, the authors take up Guy Desplanques' (1985, 1986) suggestion that fertility behavior in a given year is defined by the probability of childbearing as a function of the mother's age and the number of children already born. Let us assume we can calculate $q^k_{x,t}$, the order-k birth probability at age x in year t (i.e., the probability that a woman aged x who has already had k-1 children at the start of the year will give birth to a child in year t). If so, we can construct a fertility table combining age and parity. As the table radix, we take a number F^0_{15} of childless women aged 15. We can then estimate the number F^k_{50} of women who would have k children at the end of their childbearing period, at age 50, if they experienced at each successive age the age- and order-specific fertility conditions prevailing in year t. The *PATFR* of year t is ultimately expressed as:

$$PATFR_t = \frac{\sum_{k=1}^{n} k \cdot F^k_{50}}{F^0_{15}}$$

Likewise, if we can calculate for year t the birth probabilities $q^{k,d}_{x,t}$ for age x, order k, and duration d since the previous birth, we can construct a fertility table combining age, parity, and duration. Taking a radix of $F^{0,0}_{15}$ childless women aged 15, and making d vary within limits enabling us to compute order-specific probabilities, we obtain—as in the previous calculation—the number of women having given birth to k children by age 50. This assumes that the hypothetical cohort experienced at each successive age the age-, order-, and duration-specific fertility conditions prevailing in the observation year t. Their sum divided by the initial population yields the PADTFR for year t.

Combining, the information drawn from the French 1982 Family Survey (Desplanques, 1985) and the current statistics obtained from civil registration and censuses, Rallu and Toulemon (1993a and b) demonstrated that, from 1975 to 1989, these various summary indicators of period fertility gave significantly different measures of changes in fertility in certain periods (Figure 14–20).

For 1975, the two summary indicators gave roughly the same value for France as the conventional TFR (1.94 for the PATFR and 1.92 for the PADTFR versus 1.93 for the TFR). In 1989, on the other hand, the divergence was substantial: The TFR stood at 1.81, compared with a PATFR of 1.87 and a PADTFR of 1.94.

One can, however, question the significance of such refinements in period measurements. "How far should one go in including additional covariates, provided that we have sufficient data?" asks Nico Keilman (1993; p. 426), adding,

> Statistical tests exist to determine whether an extra covariate improves the model's fit. But more important is the question whether it is easier to understand the time pattern exhibited by a certain summary indicator than the development shown

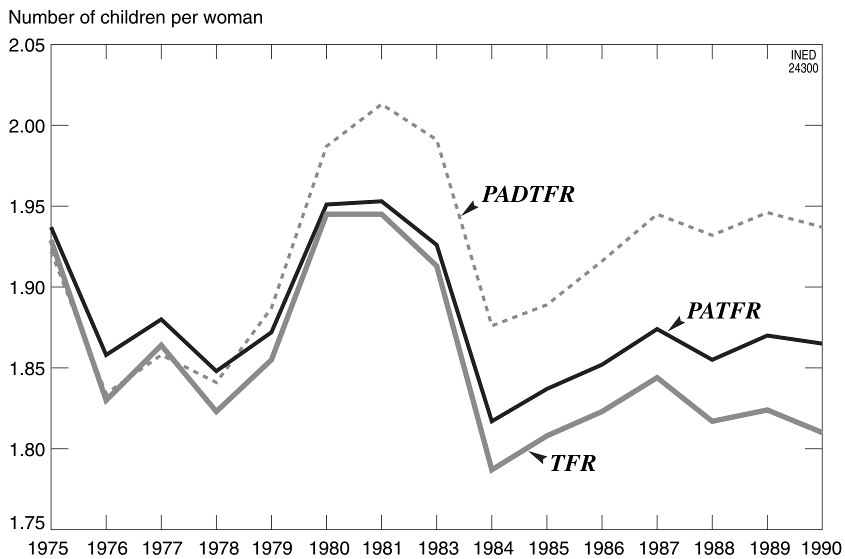


FIGURE 14–20 Period fertility in France as measured by Rallu and Toulemon's parity- and age-specific fertility rate (*PATFR*) and parity-, age-, and duration-specific fertility rate (*PADTFR*) compared with standard total fertility rate (*TFR*) (1975–1989).

by a less detailed one. For a forecaster it would be relevant to know which of these two is easier to extrapolate.[11]

7. Parity Progression Ratios

The mean number of children per woman in a order k period fertility table offers a cross-sectional view of the probability a_k, for a woman having had $k - 1$ children, that she will have at least one more child, in a hypothetical cohort that would exhibit during its entire childbearing period the same behavior as the one displayed in the observation year by the various cohorts involved. This is a period expression of the concept of *parity progression ratio*, defined on a cohort basis (Chapter 12).

For order 1, the complement to unity of the intensity gives a measure of the proportion of infertile women in the hypothetical cohort. We can also take the order k births of the table to compute the order-k-specific mean age at childbirth. We can continue the period reconstruction by combining the final intensities of order-specific fertility to obtain a summary indicator of total fertility by transposing the CFR to the hypothetical cohort:

[11] *Jusqu'où peut-on aller dans cette décomposition si les données sont disponibles? . . . Des tests statistiques permettent de dire si l'inclusion d'un facteur supplémentaire améliore l'ajustement par le modèle. Mais il est plus important de se demander si l'évolution chronologique décrite par un indicateur est plus facile à comprendre que celle décrite par un indicateur plus simple. Pour un prévisionniste, il s'agirait de se demander quelle est la série la plus facile àextrapoler* (Keilman, 1993).

$$a = a_0 + a_0 \cdot a_1 + a_0 \cdot a_1 \cdot a_2 + \ldots + \prod_{k=0}^{n} a_k$$

where n is the final birth order observed in the population.

This second measure of final intensity of all-parities fertility computed for a hypothetical cohort from a period table differs significantly from the TFR calculated from age-specific rates if these are frequency rates, but are instead very similar if we use occurrence-exposure rates.

We can also determine parity progression ratios from frequency parity-specific TFRs. At parity 0, the proportion of women who, in the hypothetical cohort, will have had at least one child is given by the order-1 TFR, which we can interpret as the probability for childless women of having a first child. For the observation year t:

$$a_{0,t} = TFR^1_{t2}$$

Subsequently, assuming the order 2 TFR gives us the proportion of women having had at least two children, we can obtain the probability for a woman having had one child of having a second by dividing this proportion by the preceding one:

$$a_{1,t} = \frac{TFR^2_{t2}}{TFR^1_{t2}}$$

and so on:

$$a_{k,t} = \frac{TFR^{(k+1)}_{t2}}{TFR^k_{t2}}$$

Again, we can recombine these parity progression ratios to determine the final intensity of the hypothetical cohort's total fertility:

$$a = a_0 + a_0 \cdot a_1 + a_0 \cdot a_1 \cdot a_2 + \ldots + \prod_{k=0}^{n} a_k$$

III. EXTERNAL MIGRATIONS

Continuous tracking of external migrations using period indicators is generally hampered by the lack of a system for gathering current statistics on the country's inbound and outbound migration flows. Few countries maintain records of migration flows comparable to the civil registration of births and deaths. Of the countries that do operate such a system, even fewer can claim that it operates efficiently.[12]

1. Countries with Comprehensive Migration Registers

If such a system exists and is sufficiently reliable, we can—as with mortality and fertility—rearrange the emigration and immigration rates in period form to construct summary indicators for hypothetical cohorts that display, at each age, the migration behaviors adopted in an observation year by the cohorts reaching those ages in that year. As with fertility, we can use two types of rates: type 1 rates (in the Lexis squares) if the migrations are classified by age and observation year, or type 2 rates (projective parallelograms), if the statistics are kept by birth year and observation year. But, as with mortality, if we use type 2 rates, we will need to solve the special problem posed by the first year of age, which straddles the first Lexis triangle and the first projective parallelogram. Migration, like fertility, consists of renewable events, and we need to distinguish between two situations. In the first, we do not identify the migration order: as a result, we can calculate only *migration frequencies*. In the second case, we can classify migrations by order, which theoretically paves the way for a computation of occurrence-exposure rates. Obviously, the first situation is more common.

For emigration, we can accordingly sum the *age-specific emigration rates* (or emigration frequencies), $e_{x,t}$, of year t (whether they are type 1 or type 2 is of little importance here). We thus obtain a *total emigration rate* (*TER*) for year t:

$$TER_t = \sum_{x=0}^{\omega} e_{x,t}$$

By contrast, to compute the mean age, we will need to take into account the type of rate used. For a type 1 rate, the expression is very simply:

$$\overline{x}_t = \frac{\sum_{x=0}^{\omega} (x+0.5) e_{x,t}}{\sum_{x=0}^{\omega} e_{x,t}}$$

but, for a type 2 rate, the expression becomes:

$$\overline{x}_t = \frac{0,25 \cdot e_{0,t} \sum_{x=1}^{\omega} x \cdot e_{x,t}}{\sum_{x=0}^{\omega} e_{x,t}}$$

Of course, as with fertility, we can also work on age groups. The formulas are the same (see previous section).

On the immigration side, we face the same problem here as the one already encountered in cohort analysis: As a rule, we do not know the immigrants' population (or, more accurately, populations) of origin. For lack of a better method, demographers usually divide the age-specific inflows by the receiving population of the same age to obtain the immigration rate at age x in year t, $i_{x,t}$. This rate, of course, differs radically from the emigration rate, because none of the individuals included in the numerator belongs to the population forming the denominator. The inconsistency of such rates is vividly illustrated by some extreme situations. In the United States, for example, at the time of the westward Gold Rush, it was not uncommon for the number of immigrants in a given area to exceed that of the native inhabitants. Here is an even more striking example: If we had had to calculate immigration rates to Reunion Island (in the Indian Ocean) when the island was discovered totally devoid of human inhabitants, we would have had to divide the number of immigrants by a zero denominator. In most cases, however, the calculation is of some value, as it gives a measure of the scope of the immigration phenomenon relative to the host population, as well as its age- and sex-specific distribution.

Once the basic principle is accepted, the computations are exactly identical to those for emigration: we merely need to replace $e_{x,t}$ by $i_{x,t}$ in the previous equations to obtain the *total immigration rate* (*TIR*) and the mean age at immigration.

One important consequence should be noted, however. If, instead of a country's total population, we

[12] In Europe, for example, a system of this kind operates fairly efficiently in the Scandinavian countries but very poorly in Italy; most countries have no migration register at all.

TABLE 14–3 Age-specific Net Migration Rate (per 1000) Between ages 30 and 50 of the Italian Male Cohort Born in 1941 Compared with Rates for 1972 and 1980

Age	Cohort		Year					
	1941		1972		1980		1990	
	Net value	Rate	Net value	Rate	Net value	Rate	Net value	Rate
30	166	0.47	166	0.47	21	0.05	–4	–0.01
31	1640	4.62	–301	–0.76	–916	–2.19	91	0.23
32	596	1.68	–1714	–4.29	–706	–1.71	89	0.23
33	240	0.68	–346	–0.89	–595	–1.42	115	0.29
34	173	0.49	302	0.83	–27	–0.08	177	0.46
35	48	0.14	2	0.01	–227	–0.67	268	0.70
36	–109	–0.31	100	0.27	–76	–0.22	202	0.55
37	–688	–1.95	–177	–0.49	130	0.38	293	0.80
38	–95	–0.27	45	0.12	–95	–0.27	153	0.42
39	178	0.51	143	0.40	–617	–1.58	200	0.52
40	–425	–1.21	292	0.80	–146	–0.37	112	0.29
41	–352	–1.01	–477	–1.24	57	0.15	149	0.36
42	–224	–0.64	–319	–0.89	108	0.30	176	0.43
43	–86	–0.25	–10	–0.03	–181	–0.52	134	0.33
44	–35	–0.10	–707	–1.94	–92	–0.26	196	0.61
45	22	0.06	325	0.91	–478	–1.35	183	0.55
46	84	0.24	361	1.02	–384	–1.09	78	0.23
47	135	0.39	–6	–0.02	360	1.04	250	0.74
48	177	0.52	–95	–0.27	–67	–0.19	177	0.52
Total	1445	4.1	–2416	–6.0	–3931	–10.0	3039	8.3
Total absolute values		15.5		15.4		3.6		8.3
Mean age		36.9		38.7		39.1		41.0

examine the population of one of its territorial units by using the registration data, we can avoid the difficulty inherent in measuring immigration. For this purpose, we treat immigration as the emigration from all the territorial units toward the unit studied (but this does assume that international migrations are negligible).

2. Estimating Intercensal Annual Age-specific Net Migration

In the absence of a continuous recording of inflows and outflows, we can—as in cohort analysis (Chapter 13)—confine our study to net migration by subtracting the natural surpluses provided by vital statistics from the intercensal excesses. The $nm_{x,t}$ rates of age- and cohort-specific net migration estimated by means of the method described in Chapter 13 can be used here for a period computation. They will provide summary intensity and tempo indicators for hypothetical cohorts: the TMR of year t (TMR_t) and the *mean age at migration* of year t ($\bar{x}_t$).

We must naturally be careful not to confuse such rates, which measure relative balances, with the rates reflecting propensities to die, have children, emigrate, and so on. These rates can be negative or positive, depending on the direction of the net flow, and their sum within a hypothetical cohort, as within a real cohort, can be equal to zero without signifying the total lack of migration in the cohort. Accordingly, as we saw in cohort analysis, it makes more sense to sum the absolute values of the net rates. Table 14–3 and Figure 14–21 compare the results obtained for Italian males in the years 1972, 1980, and 1990 with the values for the 1941 cohort.

For practical reasons, we have included in the table only the adult ages from 30 to 50.[13] This is because the data used to construct the table are the ones already described in Chapter 13: They were obtained from an age breakdown of the intercensal 10-year net migra-

[13] The age reported in the table is the complete age on the initial January 1. The age listed as 48 therefore covers the exact ages from 48 to 50 years.

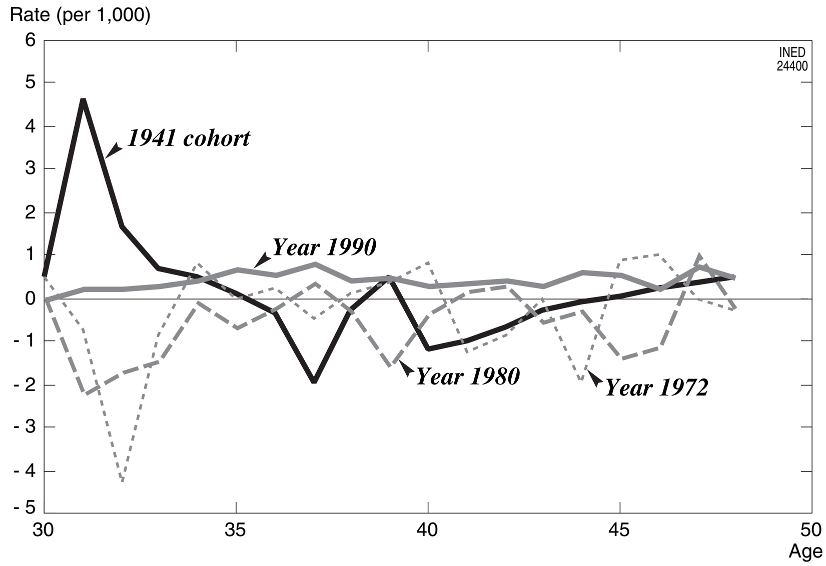


FIGURE 14–21 Age-specific net migration rate between ages 30 and 50 of the Italian male cohort born in 1941 compared with rates for 1972 and 1980. (Author's calculation based on Italian census data.)

TABLE 14–4 Age-specific External Immigration Rate, 1986–1991, for Individuals Older than 5 Years, as given by 1991 Italian Census

Age	Population enumerated in 1981	Population residing abroad in 1986	Immigration rate 1986–1991 (per 1000)	Annual average rate (per 1000)
5–14	6,260,054	53,556	8.59	1.72
15–24	8,832,952	76,451	8.69	1.74
25–34	8,753,829	133,957	15.42	3.08
35–44	7,736,280	71,141	9.24	1.85
45–54	7,086,236	41,929	5.93	1.19
55–64	6,659,574	38,252	5.76	1.15
65–74	4,907,618	17,818	3.64	0.73
75+	3,792,567	5,863	1.55	0.31
Total	54,029,110	438,967	8.16	1.63

tion values for each cohort. For this purpose, we used a model of additions to and removals from the municipal population registers by age and year, which requires substantial reconstruction work.

The cells of the table framed in heavy black lines are those in which we find an exact correspondence between the cohort rate and the rate of one of the selected years (this is because the rates are type 2 rates, in projective parallelograms of the Lexis diagram).

The table effectively summarizes Italy's recent external migration history, which has witnessed a shift from a long period of net emigration to a new period of net immigration. The total net migration observed here between ages 30 and 50 is negative in 1972 and 1980, but positive in 1990. In the same interval, we see an increase in the mean age at net migration of the 30 to 50 age group. Unfortunately, the detailed reading of the age-specific results is complicated here by the inconsistencies of the databases used. For example, the peak observed at age 31 in the 1941 cohort seems to be directly linked to the pre- and postcensal deletion and reregistration procedures discussed in Chapter 13, since the cohort reached age 31—marked by heavy migration—just after the 1971 census.

3. Estimating Migration Flows from the Retrospective Questions in the Census Schedules

Many census schedules include retrospective questions on migrations covering such variables as place of birth, date of arrival in current place of residence, and place of residence at an earlier date. The question on

TABLE 14–5 Lazio: Age-specific rates of emigration (for individuals aged over 5 years) to other regions and rate of immigration from other regions or abroad, 1986–91, as given by 1991 Italian census

Age	Population enumerated in region in 1991	Population residing in other regions or abroad in 1986	Immigration rate (per 1000)		Population enumerated in other regions in 1991 residing in the region in 1986	Emigration rate (per 1000)		Population of rest of Italy enumerated in 1991	Rate of emigration of rest of Italy to Lazio (per 1000)	
			1986–1991	Annual average		1986–1991	Annual average		1986–1991	Annual average
5–14	550,354	12,827	23.58	4.72	7,986	14.68	2.94	5,714,271	1.90	0.38
15–24	800,125	24,728	31.39	6.28	15,507	19.68	3.94	8,042,047	2.88	0.58
25–34	830,524	41,817	51.65	10.33	32,133	39.69	7.94	7,932,989	5.32	1.06
35–44	723,048	18,092	25.34	5.07	12,532	17.55	3.51	7,018,791	2.50	0.50
45–54	662,721	10,366	15.76	3.15	7,258	11.04	2.21	6,174,542	1.55	0.31
55–64	606,684	7,084	11.75	2.35	7,284	12.08	2.42	6,052,694	1.11	0.22
65–74	430,054	4,422	10.34	2.07	5,738	13.41	2.68	4,476,249	1.00	0.20
75+	296,700	3,164	10.72	2.14	4,066	13.78	2.76	3,792,567	0.91	0.18
Total	4,900,210	122,500	25.32	5.06	92,504	19.12	3.82	49,204,150	2.38	0.48

TABLE 14–6 Matrix of Italian Interregional Migration Flows, 1986–1991 by 1991 Italian Census

Region of residence in 1991 census	Piedmont	Val d'Aosta	Lombardy	Trentino Alto Adige	Veneto	Friuli V. Giulia	Liguria	Emilia-Romagna	Tuscany	Umbria	Marche
Piedmont	4,032,931	814	14,845	653	2,401	716	9,985	1,909	2,234	346	518
Val d'Aosta	1,928	105,344	449	39	116	45	231	68	80	17	27
Lombardy	14,144	306	8,256,833	2,369	9,866	3,007	6,883	9,738	5,966	987	1,948
Trentino A. Adige	700	34	3,110	822,247	2,994	444	340	909	529	97	171
Veneto	4,619	111	11,461	2,915	4,103,724	6,129	1,372	5,016	2,037	347	685
Friuli V. Giulia	1,439	46	3,527	584	6,370	1,118,192	618	980	696	140	271
Liguria	8,895	143	7,844	192	851	367	1,574,327	1,488	3,401	143	241
Emilia-Romagna	3,866	98	15,190	902	5,775	1,224	2,713	3,662,054	4,706	712	3,237
Tuscany	3,549	121	7,424	549	1,838	876	4,882	3,414	3,314,914	1,687	873
Umbria	735	16	1,410	133	387	182	338	752	2,378	754,613	867
Marche	1,373	34	3,346	235	933	417	576	3,309	1,254	1,061	1,332,817
Lazio	4,435	101	7,934	665	3,070	1,583	2,168	2,918	5,758	2,940	2,061
Abruzzo	1,473	50	3,199	160	704	379	546	1,662	1,095	403	1,941
Molise	475	9	720	26	161	86	87	434	336	98	164
Campania	5,097	58	9,378	331	1,593	1,086	1,428	2,696	3,300	384	472
Puglia	5,462	78	10,642	304	2,097	826	1,312	3,598	2,968	435	953
Basilicata	1,198	30	1,704	48	244	106	170	578	836	110	116
Calabria	5,734	244	8,774	214	1,203	404	1,329	2,198	3,903	784	393
Sicily	6,759	133	12,056	391	1,905	842	2,003	2,354	2,852	253	353
Sardinia	3,807	89	4,670	170	691	317	2,261	1,283	2,091	209	223
Total	4,108,619	107,859	8,384,516	833,127	4,146,923	1,137,228	1,613,569	3,707,358	3,361,334	765,766	1,348,331

Region of residence in 1991 census	Lazio	Abruzzo	Molise	Campania	Puglia	Basilicata	Calabria	Sicily	Sardinia	Outside of Italy	Total
Piedmont	3,895	737	371	7,940	8,047	1,572	9,321	10,727	3,123	24,612	4,137,697
Val d'Aosta	205	61	14	195	172	25	762	194	241	850	111,063
Lombardy	9,414	2,165	736	19,935	19,230	2,702	18,796	24,500	5,295	66,783	8,481,603
Trentino A. Adige	1,148	180	46	901	624	61	535	701	242	7,664	843,677
Veneto	4,574	687	194	5,140	4,231	354	2,286	4,092	1,161	31,289	4,192,424
Friuli V. Giulia	2,112	279	76	2,139	1,607	109	597	1,351	419	13,458	1,155,010
Liguria	2,071	275	77	2,274	1,714	268	2,568	3,027	1,576	9,469	1,621,211
Emilia-Romagna	5,239	1,555	653	12,419	9,417	1,641	6,408	8,628	2,158	26,307	3,774,902
Tuscany	9,234	800	370	9,512	3,753	1,045	3,670	6,933	2,566	21,542	3,399,552
Umbria	7,199	380	107	1,902	734	118	658	748	360	5,200	779,217
Marche	5,077	1,773	221	2,069	2,462	222	510	1,093	354	9,783	1,368,919
Lazio	4,777,710	4,311	1,125	18,993	7,668	1,255	6,225	7,023	3,283	38,984	4,900,210
Abruzzo	7,867	1,146,634	1,211	2,399	2,256	217	471	747	339	14,209	1,187,962
Molise	1,669	836	299,659	1,845	1,004	57	117	156	64	5,892	313,895
Campania	9,587	941	871	5,163,246	4,051	1,698	2,632	2,800	679	36,761	5,249,089
Puglia	5,805	1,310	546	4,498	3,709,864	1,869	1,782	2,670	550	29,498	3,787,067
Basilicata	1,206	90	52	2,193	2,687	554,373	759	310	91	7,752	574,653
Calabria	6,643	265	127	4,627	2,681	935	1,863,237	5,037	224	27,963	1,936,919
Sicily	4,762	448	130	3,271	2,984	332	3,165	4,554,895	956	50,958	4,651,802
Sardinia	4,794	220	70	1,500	712	72	321	1,235	1,527,510	9,993	1,562,238
Total	4,870,211	1,163,947	306,656	5,266,998	3,785,898	568,925	1,924,820	4,636,867	1,551,191	438,967	54,029,110

the place of residence at an earlier date is the one that lends itself most easily to the calculation of age-specific migration rates. However, in terms of external flows for a given country, the only information yielded by the question concerns immigration: this can be seen in Table 14–4 for precensal Italian migrations in the years 1986 to 1991.

The only emigration that can be measured on the basis of this information consists in internal migrations. We can identify the emigrants from any administrative division thanks to the questions put to the inhabitants of all the others. If we are studying the population of a specific administrative division, we can use the census data to measure both immigration and emigration, as already mentioned in our discussion of population registers.

For example, from a matrix of migration flows between Italian regions (Table 14–5 gives an example for the region of Lazio) for the period 1986 to 1991 (of which Table 14–6 reports only the all-ages totals) we can calculate, for any region, the age-specific emigration rates as well as the immigration rates.

References

Calot Gérard, 1984. *La mesure des taux en démographie Âge en années révolues ou âge atteint dans l'année, incidence du choix de la définition; application à la fécondité générale (France, 1946–1981).* Paris, INED et PUF, 322 p. (Travaux et documents, Cahier no. 104).

Calot Gérard and Caselli Graziella, 1991. Détermination d'une table de mortalité: la conversion des taux en quotients, *Population*, vol. 46(6), p. 1441–1490.

Chiang Chin Long, 1968. *Introduction to Stochastic Processes in Biostatistics*. New York, Wiley, 313 p.

Coale Ansley J. and Caselli Graziella, 1990. Estimation of the number of persons at advanced âges from the number of death at each âge in the given year and adjacent years, *Genus*, vol. 46(1–2), p. 1–23.

Coale Ansley J. and Preston Samuel H., 1982. Age structure, growth, attrition, and accession: a new synthesis, *Population Index*, vol. 48(2), p. 217–259.

Das Gupta Prithwis and Long John F., 1985. Ultimate parity distribution and mean age at childbearing by birth order for American women: 1917–1980, *in : General Population Conference, Florence 1985*, vol. 4, p. 219–230. Liège, IUSSP, 556 p.

De Simoni Alessandro, 1990. L'évolution récente de la fécondité italienne. Une analyse des tables de fécondité par rang de naissance. *Cahiers Québécois de Démographie*, vol. 19(1), p. 123–136.

Deplanques Guy, 1985. *Fécondité générale. Résultats de l'Enquête Famille*. Paris, INSEE, 248 p. (Archives et documents, no. 143).

Federici Nora, 1965. *Lezioni di demografia*. Rome, De Santis, 859 p.

Giorgi Piero, 1993. Una rilettura della fecondità del momento per ordine di nascita in Italia nel periodo 1950–1990 considerando la struttura per parità, *Genus*, vol. 49(3–4), p. 177–204.

Greville Thomas, 1943. Some methods of constructing abridged life tables, *Records of the American Institute of Actuaries*, vol. 32, Part 1, p. 29–43.

Henry Louis, 1972. *Demographie: analyse et modèles.* Paris, Larousse, 341 p.

Keilmann Nico, 1993. Du fondement et de l'usage des taux, *in:* Les mesures de la fécondité transversale. Réflexions autour d'un article, *Population*, vol. 48(2), p. 423–427.

Keyfitz Nathan and Frauenthal James, 1975. An improved life table method, *Biometrics*, vol. 31, Part 4, p. 889–899.

Keyfitz Nathan, 1977. *Applied mathematical demography*. New York, John Wiley and Sons, 388 p. (Republished in 1985 by Springer-Verlag, New York, 442 p).

Le Bras Hervé, 1991. *Marianne et les lapins*. Paris, Olivier Orban, 284 p.

Pressat Roland, 1969. *L'analyse démographique*, deuxième édition. Paris, PUF, 321 p.

Rallu Jean-Louis and Toulemon Laurent, 1993a. La mesure de la fécondité transversale. I. Construction des différents indices, *Population*, vol. 48(1), p. 7–26.

Rallu Jean-Louis and Toulemon Laurent, 1993b. La mesure de la fécondité transversale. II. Application à la France de 1946 à 1949, *Population*, vol. 48(2), p. 369–404.

Reed Lowell J. and Merrell Margaret, 1939. A short method for constructing an abridged life table, *American Journal of Hygiene*, vol. 30, p. 33–62.

Ryder Norman B., 1963. The translation model of demographic change, *in: Emerging techniques in population research*. New York, Milbank Memorial Fund, 306 p.

Ryder Norman B., 1964. The process of demographic translation, *Demography*, vol. 1(1), p. 74–82.

Santini Antonio, 1992. *Analisi demografica: fondamenti e netodi*. Scandicci (Florence), La Nuova Italia, 422 p.

Santini Antonio, 1995. *Continuità e discontinuità nel comportamento riproduttivo delle donne italiane nel dopo guerra: tendenze generale della fecondità delle coorti nelle ripartizioni tra il 1952 e 1991*. Florence, Dipartimento Statistico, 422 p. (Coll. Working Papers, no. 53).

Sardon Jean-Paul, 1993. Un indicateur conjoncturel de mortalité: l'exemple de la France. *Population*, vol. 43(2), p. 347–368.

Shryock Henri S. and Siegel Jacob S., 1973. Natality: measures based on vital statistics, *in:* Henri S. Shryock and Jacob S. Siegel *The methods and materials of demography*, vol. 2, p. 462–499. Washington, Bureau of the Census, 2 vol., 888 p. + annexes.

Vallin Jacques, 1973. *La mortalité par génération en France depuis 1899*. Paris, INED, PUF, 484 p. (Travaux et Documents, Cahier no. 63).

Vallin Jacques and Meslé France, 2001. *Tableau de mortalité française 1806–1997 et projections jusqu'en 2102.* Paris, INED, 44 p. + Cdrom.

Whelpton Pascal, 1946. Reproduction rates adjusted for age, parity, fecundity and marriage. *Journal of the American Statistical Association*, (41), p. 501–516.

Whelpton Pascal, 1954. *Cohort fertility: native white women in the United States*. Princeton, Princeton University Press, 492 p.

Wunsch Guillaume and Termote Marc, 1978. *Introduction to demographic analysis*. New York, Plenum Press, 274 p.

APPENDIX 14–1 Life Table of French Female Cohort Born in 1875

X	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x	x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x
0	0.154049	100,000	15,405	47.22	51	0.010634	54,478	579	23.95
1	0.050193	84,595	4,246	54.73	52	0.011059	53,899	596	23.20
2	0.027502	80,349	2,210	56.60	53	0.012259	53,303	653	22.45
3	0.016098	78,139	1,258	57.18	54	0.012496	52,649	658	21.72
4	0.012491	76,881	960	57.11	55	0.013554	51,992	705	20.99
5	0.009894	75,921	751	56.83	56	0.013668	51,287	701	20.27
6	0.007864	75,170	591	56.39	57	0.015185	50,586	768	19.55
7	0.006896	74,579	514	55.83	58	0.015310	49,818	763	18.84
8	0.005956	74,064	441	55.22	59	0.017641	49,055	865	18.13
9	0.005072	73,623	373	54.55	60	0.017932	48,190	864	17.44
10	0.004487	73,250	329	53.82	61	0.019304	47,326	914	16.75
11	0.004448	72,921	324	53.06	62	0.020905	46,412	970	16.07
12	0.004231	72,597	307	52.30	63	0.022409	45,442	1,018	15.41
13	0.004337	72,290	313	51.52	64	0.028426	44,423	1,263	14.75
14	0.004607	71,976	332	50.74	65	0.029473	43,161	1,272	14.16
15	0.005856	71,645	420	49.97	66	0.029950	41,889	1,255	13.58
16	0.006343	71,225	452	49.26	67	0.031490	40,634	1,280	12.98
17	0.006875	70,773	487	48.57	68	0.038462	39,354	1,514	12.39
18	0.006747	70,287	474	47.91	69	0.040659	37,841	1,539	11.86
19	0.006617	69,813	462	47.23	70	0.039111	36,302	1,420	11.35
20	0.006885	69,351	477	46.54	71	0.040694	34,882	1,420	10.79
21	0.007200	68,873	496	45.86	72	0.040508	33,463	1,356	10.22
22	0.007348	68,377	502	45.19	73	0.052291	32,107	1,679	9.63
23	0.007257	67,875	493	44.52	74	0.052152	30,428	1,587	9.14
24	0.006994	67,382	471	43.84	75	0.063097	28,842	1,820	8.61
25	0.007338	66,911	491	43.15	76	0.064150	27,022	1,733	8.16
26	0.007142	66,420	474	42.46	77	0.078671	25,288	1,989	7.69
27	0.007012	65,946	462	41.76	78	0.075801	23,299	1,766	7.30
28	0.007070	65,483	463	41.05	79	0.084156	21,533	1,812	6.86
29	0.006725	65,020	437	40.34	80	0.096753	19,721	1,908	6.44
30	0.007192	64,583	464	39.61	81	0.100663	17,813	1,793	6.08
31	0.007331	64,118	470	38.90	82	0.109283	16,020	1,751	5.70
32	0.007265	63,648	462	38.18	83	0.121472	14,269	1,733	5.34
33	0.006901	63,186	436	37.46	84	0.133219	12,536	1,670	5.01
34	0.006860	62,750	430	36.71	85	0.138747	10,866	1,508	4.70
35	0.007191	62,320	448	35.96	86	0.160105	9,358	1,498	4.38
36	0.007050	61,871	436	35.22	87	0.185436	7,860	1,457	4.12
37	0.007101	61,435	436	34.47	88	0.175424	6,402	1,123	3.94
38	0.007453	60,999	455	33.71	89	0.194571	5,279	1,027	3.67
39	0.007400	60,544	448	32.96	90	0.209926	4,252	893	3.44
40	0.007785	60,096	468	32.20	91	0.231730	3,359	778	3.22
41	0.007408	59,628	442	31.45	92	0.242579	2,581	626	3.04
42	0.008019	59,187	475	30.68	93	0.252294	1,955	493	2.86
43	0.012582	58,712	739	29.92	94	0.285658	1,462	418	2.65
44	0.007801	57,973	452	29.30	95	0.302464	1,044	316	2.51
45	0.008095	57,521	466	28.53	96	0.324409	728	236	2.39
46	0.008766	57,055	500	27.75	97	0.349563	492	172	2.29
47	0.008432	56,555	477	27.00	98	0.336889	320	108	2.26
48	0.009004	56,078	505	26.22	99	0.340687	212	72	2.15
49	0.009828	55,573	546	25.45	100		140	140	2.00
50	0.009978	55,027	549	24.70					

APPENDIX 14–2 French female life table for year 1875

x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x	x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x
0	0.154049	100,000	15,405	43.55	51	0.014976	50,147	751	20.16
1	0.054720	84,595	4,629	50.39	52	0.015734	49,396	777	19.46
2	0.027288	79,966	2,182	52.28	53	0.016686	48,619	811	18.77
3	0.017505	77,784	1,362	52.73	54	0.017694	47,808	846	18.08
4	0.013015	76,422	995	52.66	55	0.018764	46,962	881	17.39
5	0.009489	75,428	716	52.35	56	0.019897	46,081	917	16.71
6	0.006939	74,712	518	51.84	57	0.021098	45,164	953	16.04
7	0.005991	74,194	444	51.20	58	0.022993	44,211	1,017	15.38
8	0.005355	73,749	395	50.51	59	0.025055	43,195	1,082	14.73
9	0.004991	73,354	366	49.78	60	0.027301	42,112	1,150	14.09
10	0.004671	72,988	341	49.02	61	0.029744	40,963	1,218	13.48
11	0.004396	72,647	319	48.25	62	0.032403	39,744	1,288	12.87
12	0.004358	72,328	315	47.46	63	0.034945	38,456	1,344	12.29
13	0.004582	72,013	330	46.67	64	0.037682	37,113	1,398	11.72
14	0.004934	71,683	354	45.88	65	0.040630	35,714	1,451	11.15
15	0.005687	71,329	406	45.11	66	0.043803	34,263	1,501	10.61
16	0.006054	70,923	429	44.36	67	0.047219	32,762	1,547	10.07
17	0.006705	70,494	473	43.63	68	0.051498	31,215	1,608	9.54
18	0.006894	70,021	483	42.92	69	0.056153	29,608	1,663	9.03
19	0.007323	69,539	509	42.21	70	0.061215	27,945	1,711	8.54
20	0.007790	69,029	538	41.52	71	0.066718	26,234	1,750	8.07
21	0.008294	68,492	568	40.84	72	0.072698	24,484	1,780	7.61
22	0.008507	67,923	578	40.18	73	0.079668	22,704	1,809	7.16
23	0.008566	67,346	577	39.52	74	0.087277	20,895	1,824	6.74
24	0.008522	66,769	569	38.86	75	0.095577	19,072	1,823	6.34
25	0.008785	66,200	582	38.19	76	0.104622	17,249	1,805	5.95
26	0.009057	65,618	594	37.52	77	0.114471	15,444	1,768	5.59
27	0.009340	65,024	607	36.86	78	0.123221	13,676	1,685	5.25
28	0.009450	64,417	609	36.20	79	0.134073	11,991	1,608	4.92
29	0.009561	63,808	610	35.54	80	0.145798	10,383	1,514	4.60
30	0.009673	63,198	611	34.88	81	0.158451	8,870	1,405	4.30
31	0.009787	62,586	613	34.22	82	0.172088	7,464	1,284	4.02
32	0.009902	61,974	614	33.55	83	0.186764	6,180	1,154	3.75
33	0.009929	61,360	609	32.88	84	0.202531	5,026	1,018	3.49
34	0.009957	60,751	605	32.21	85	0.219441	4,008	879	3.26
35	0.009984	60,146	600	31.52	86	0.237542	3,128	743	3.03
36	0.010012	59,546	596	30.84	87	0.256876	2,385	613	2.82
37	0.010039	58,950	592	30.14	88	0.277479	1,772	492	2.62
38	0.010155	58,358	593	29.44	89	0.299379	1,281	383	2.43
39	0.010271	57,765	593	28.74	90	0.322593	897	289	2.26
40	0.010390	57,172	594	28.03	91	0.347124	608	211	2.10
41	0.010509	56,578	595	27.32	92	0.372962	397	148	1.94
42	0.010631	55,983	595	26.61	93	0.400075	249	100	1.80
43	0.010944	55,388	606	25.89	94	0.428414	149	64	1.67
44	0.011267	54,782	617	25.17	95	0.457904	85	39	1.55
45	0.011598	54,165	628	24.45	96	0.488445	46	23	1.44
46	0.011939	53,536	639	23.73	97	0.519909	24	12	1.33
47	0.012291	52,897	650	23.01	98	0.552139	11	6	1.23
48	0.012914	52,247	675	22.29	99	0.584945	5	3	1.12
49	0.013568	51,572	700	21.58	100		2	140	1.00
50	0.014255	50,873	725	20.87					

APPENDIX 14–3 French Female Life Table for 1936

x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x	x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x
0	0.062253	100,000	6,225	61.94	51	0.009723	76,041	739	23.42
1	0.012014	93,775	1,127	65.02	52	0.010740	75,302	809	22.65
2	0.004399	92,648	408	64.80	53	0.010799	74,493	804	21.89
3	0.002973	92,241	274	64.09	54	0.011604	73,689	855	21.12
4	0.002461	91,966	226	63.28	55	0.012186	72,834	888	20.36
5	0.002070	91,740	190	62.43	56	0.012523	71,946	901	19.61
6	0.001752	91,550	160	61.56	57	0.013844	71,045	984	18.85
7	0.001452	91,390	133	60.67	58	0.014794	70,062	1,037	18.11
8	0.001309	91,257	119	59.75	59	0.016657	69,025	1,150	17.37
9	0.001180	91,137	108	58.83	60	0.017518	67,875	1,189	16.66
10	0.001218	91,030	111	57.90	61	0.019304	66,686	1,287	15.95
11	0.001185	90,919	108	56.97	62	0.021106	65,399	1,380	15.25
12	0.001340	90,811	122	56.04	63	0.023341	64,019	1,494	14.57
13	0.001376	90,690	125	55.11	64	0.025773	62,524	1,611	13.91
14	0.001767	90,565	160	54.19	65	0.026199	60,913	1,596	13.26
15	0.002063	90,405	186	53.28	66	0.028263	59,317	1,676	12.60
16	0.002507	90,218	226	52.39	67	0.033301	57,641	1,919	11.96
17	0.002861	89,992	257	51.52	68	0.035399	55,721	1,972	11.35
18	0.003292	89,735	295	50.67	69	0.040873	53,749	2,197	10.75
19	0.003318	89,439	297	49.83	70	0.043240	51,552	2,229	10.19
20	0.003744	89,143	334	49.00	71	0.047674	49,323	2,351	9.62
21	0.003783	88,809	336	48.18	72	0.053669	46,971	2,521	9.08
22	0.003973	88,473	352	47.36	73	0.058806	44,450	2,614	8.57
23	0.003949	88,121	348	46.55	74	0.064596	41,836	2,702	8.07
24	0.003961	87,773	348	45.73	75	0.071418	39,134	2,795	7.59
25	0.004011	87,426	351	44.91	76	0.078565	36,339	2,855	7.14
26	0.004008	87,075	349	44.09	77	0.088161	33,484	2,952	6.71
27	0.003946	86,726	342	43.27	78	0.094374	30,532	2,881	6.31
28	0.003993	86,384	345	42.43	79	0.107783	27,651	2,980	5.91
29	0.003725	86,039	320	41.60	80	0.116257	24,670	2,868	5.56
30	0.003996	85,718	343	40.76	81	0.119019	21,802	2,595	5.23
31	0.003839	85,376	328	39.92	82	0.134049	19,207	2,575	4.87
32	0.003917	85,048	333	39.07	83	0.155297	16,633	2,583	4.55
33	0.004078	84,715	345	38.22	84	0.158569	14,050	2,228	4.29
34	0.004249	84,369	358	37.38	85	0.184338	11,822	2,179	4.00
35	0.004144	84,011	348	36.53	86	0.187683	9,643	1,810	3.80
36	0.004505	83,663	377	35.68	87	0.211974	7,833	1,660	3.56
37	0.004664	83,286	388	34.84	88	0.212112	6,173	1,309	3.38
38	0.004800	82,897	398	34.00	89	0.213555	4,863	1,039	3.15
39	0.004991	82,500	412	33.17	90	0.268058	3,825	1,025	2.87
40	0.005341	82,088	438	32.33	91	0.265619	2,799	744	2.74
41	0.005217	81,649	426	31.50	92	0.285718	2,056	587	2.56
42	0.005935	81,223	482	30.66	93	0.344707	1,468	506	2.38
43	0.006407	80,741	517	29.84	94	0.328984	962	317	2.37
44	0.006534	80,224	524	29.03	95	0.353856	646	228	2.28
45	0.006345	79,700	506	28.22	96	0.288114	417	120	2.25
46	0.006892	79,194	546	27.40	97	0.395036	297	117	1.96
47	0.007678	78,648	604	26.58	98	0.398045	180	72	1.92
48	0.007952	78,045	621	25.78	99	0.319981	108	35	1.86
49	0.008501	77,424	658	24.99			74	140	1.50
50	0.009435	76,766	724	24.20					

APPENDIX 14–4 French Female Life Table for 1997

x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x	x	$_1q_x$	l_x	$d_{(x,x+1)}$	e_x
0	0.00413	100,000	413	82.30	51	0.00289	96,155	278	33.24
1	0.00041	99,587	41	81.63	52	0.00296	95,877	284	32.33
2	0.00021	99,546	21	80.67	53	0.00334	95,593	319	31.43
3	0.00025	99,525	25	79.69	54	0.00333	95,274	317	30.53
4	0.00016	99,500	16	78.70	55	0.00354	94,956	336	29.63
5	0.00010	99,484	10	77.72	56	0.00377	94,620	357	28.74
6	0.00009	99,475	9	76.73	57	0.00386	94,264	364	27.84
7	0.00012	99,466	12	75.73	58	0.00429	93,900	403	26.95
8	0.00012	99,454	12	74.74	59	0.00447	93,497	418	26.06
9	0.00013	99,442	13	73.75	60	0.00493	93,079	459	25.18
10	0.00013	99,429	13	72.76	61	0.00532	92,620	493	24.30
11	0.00010	99,416	10	71.77	62	0.00569	92,127	524	23.43
12	0.00016	99,406	16	70.78	63	0.00624	91,603	572	22.56
13	0.00015	99,390	15	69.79	64	0.00679	91,032	618	21.70
14	0.00015	99,375	15	68.80	65	0.00759	90,413	686	20.84
15	0.00022	99,360	22	67.81	66	0.00839	89,727	753	20.00
16	0.00024	99,338	24	66.82	67	0.00926	88,974	824	19.16
17	0.00029	99,314	29	65.84	68	0.01011	88,151	891	18.34
18	0.00035	99,286	35	64.86	69	0.01122	87,259	979	17.52
19	0.00042	99,251	42	63.88	70	0.01226	86,280	1,058	16.71
20	0.00034	99,209	34	62.91	71	0.01304	85,222	1,111	15.91
21	0.00039	99,176	39	61.93	72	0.01496	84,111	1,258	15.12
22	0.00036	99,137	36	60.95	73	0.01680	82,853	1,392	14.34
23	0.00034	99,101	34	59.97	74	0.01910	81,461	1,556	13.57
24	0.00039	99,067	39	58.99	75	0.02128	79,905	1,700	12.83
25	0.00035	99,029	35	58.02	76	0.02444	78,205	1,911	12.10
26	0.00041	98,994	41	57.04	77	0.02732	76,293	2,084	11.39
27	0.00039	98,954	39	56.06	78	0.03094	74,209	2,296	10.69
28	0.00041	98,915	41	55.08	79	0.03648	71,913	2,623	10.02
29	0.00049	98,874	48	54.10	80	0.04015	69,290	2,782	9.38
30	0.00057	98,826	56	53.13	81	0.04677	66,508	3,111	8.75
31	0.00056	98,770	55	52.16	82	0.05117	63,397	3,244	8.16
32	0.00063	98,714	62	51.19	83	0.06016	60,153	3,619	7.57
33	0.00056	98,652	55	50.22	84	0.06856	56,534	3,876	7.02
34	0.00077	98,597	76	49.25	85	0.07969	52,658	4,196	6.50
35	0.00074	98,521	73	48.29	86	0.09025	48,462	4,374	6.02
36	0.00083	98,448	82	47.32	87	0.10217	44,088	4,504	5.57
37	0.00087	98,366	86	46.36	88	0.11626	39,584	4,602	5.14
38	0.00099	98,281	97	45.40	89	0.13298	34,982	4,652	4.76
39	0.00111	98,183	109	44.45	90	0.14684	30,330	4,454	4.41
40	0.00116	98,074	114	43.49	91	0.16263	25,876	4,208	4.08
41	0.00130	97,961	127	42.54	92	0.18071	21,668	3,916	3.78
42	0.00131	97,833	128	41.60	93	0.19729	17,752	3,502	3.50
43	0.00153	97,705	149	40.65	94	0.21794	14,250	3,106	3.24
44	0.00163	97,556	159	39.71	95	0.24171	11,144	2,694	3.00
45	0.00183	97,397	178	38.78	96	0.25738	8,451	2,175	2.79
46	0.00195	97,218	190	37.85	97	0.28652	6,276	1,798	2.59
47	0.00200	97,029	194	36.92	98	0.32428	4,478	1,452	2.43
48	0.00220	96,835	213	35.99	99	0.31323	3,026	948	2.35
49	0.00233	96,622	225	35.07	100		2,078	140	2.20
50	0.00251	96,397	242	34.15					

C H A P T E R

15

Confounding Variables, Standardization, and the Problem of Summary Indices

GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

In several chapters (Chapters 3 and 11, among others), it was pointed out that crude death rates were not the best indicators of mortality for temporal or spatial comparisons, as the crude death rate is highly dependent upon the age structure of the population. It was stated that a better measure could be obtained by *controlling* for age, the latter being a *confounding variable* in this case. Death rates specific for age and sex were therefore computed for this purpose and served for constructing the life table. The present chapter will deal first with the concept of confounding variable, then move on to the problem of controlling for the confounding variable(s), and finally to the topic of deriving suitable summary measures or indicators controlling for the effect of possible confounders. As a byproduct, this chapter will also tackle the problem of *interaction* between variables, in a causal framework. The illustrative example discussed in the following section is drawn from the field of morbidity; it has the advantage over others of dealing with variables that are well documented and understood, and for which the causal structure is well established.

I. CONFOUNDING VARIABLES

Suppose that one is interested in the determinants of cardiovascular diseases (CVD), one of the major medical causes of death in the low-mortality countries. It is known that high blood pressure has an adverse effect on CVD; therefore, in a primary prevention approach one would wish to reduce the causes of excess blood pressure. One such cause could be coffee drinking[1]; one should therefore test if there is an association between coffee drinking and blood pressure in a random sample of individuals. To simplify the presentation, we will use simple dichotomies such as drinkers versus nondrinkers; in an actual study, the measurement would be more precise. Suppose 600 people have been included in the study, among whom 400 are drinkers and 200 are nondrinkers. The average diastolic[2] blood pressure in the two groups is 89 mm Hg and 75 mm Hg, respectively, the absolute difference being a significant[3] 14 mm Hg. The possible conclusion is that coffee drinking has a considerable impact on (or at the least has a high association with) blood pressure, represented by the causal graph

$$C \rightarrow B$$

where *C* stands for the variable *coffee drinking* and *B* for *blood pressure*. The arrow means in this case that *C* is a cause of (or influences) blood pressure. The obvious conclusion is that people should stop drinking coffee to reduce blood pressure to avoid CVDs.

[1] The example is taken from Anderson *et al*. (1980), Chapter 2.
[2] The pressure between two heart contractions.
[3] In practice, significance tests should be performed to determine if the difference is not obtained by chance. We will not consider this particular issue here.

TABLE 15–1 Population Numbers and Diastolic Blood Pressure (in mm Hg) According to Coffee Drinking and Weight

Drinking	Weight		
	Overweight	Not overweight	Total
Drinkers	300 (94.0)	100 (74.0)	400 (89.0)
Nondrinkers	50 (90.0)	150 (70.0)	200 (75.0)
Total	350 (93.4)	250 (71.6)	600 (84.3)

Source: Anderson *et al.*, 1980.

TABLE 15–2 Distributions According to Weight

Drinking	Weight		
	Overweight	Not overweight	Total
Drinkers	75	25	100
Nondrinkers	25	75	100
Total[a]	58	42	100

[a] Weighted average.

Coffee drinking is however not the only possible cause of higher blood pressure. Obesity for example is also a well-known cause of this risk factor. What happens then if one takes body weight also into account? Once again, this variable is dichotomized for simplification into overweight and not overweight. Let *W* represent the variable *weight*. The assumed causal graph is then

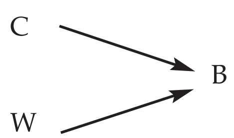


Table 15–1 gives the cross-classification of the number of individuals in the sample according to their coffee drinking behavior and their weight. Blood pressure is given for each cell of the table between brackets.

One sees that being overweight leads to a much higher blood pressure, on average, than not being overweight, both for drinkers and nondrinkers. Indeed, the difference is equal to a high 20 mm Hg in both groups (94–74 and 90–70). On the other hand, if one *controls* for weight (i.e., if one considers separately those overweight and those not overweight), the impact of coffee drinking is very slight: only 4 mm Hg in both groups (94–90 and 74–70). Should one then discourage people from drinking coffee,[4] as the difference between both groups is slight?

Consider now the two extremes. The highest blood pressure can be found on average among those who both drink and are overweight (94.0 mm Hg), while the lowest blood pressure is to be found in the nondrinkers and nonoverweight group (70.0 mm Hg). The difference between the two groups is 24 mm (i.e., the effect of weight [20.0 mm Hg] plus the effect of coffee [4.0 mm Hg]). The two effects are *additive* in this example; in other words, as will be discussed later, there is no *interaction* between weight and coffee drinking (the two causes or risk factors) on blood pressure (the effect or outcome variable).

[4] The answer might depend upon the coffee one is drinking, Italian or American!

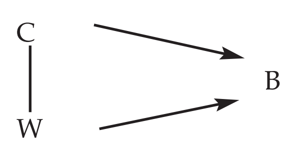
How did one reach the wrong conclusion in the first place, i.e. that coffee is bad for one's health? Computing the distribution of individuals according to the two variables, one obtains the results shown in Table 15–2. The distributions according to body weight are quite different between the drinkers and the nondrinkers, as overweight individuals also tend to be strong coffee drinkers. It is indeed well known in epidemiology that there is often a clustering of risk factors for a variety of health conditions. One observes in this case a positive relationship between weight and coffee drinking, leading to an overall difference of 14 mm Hg biased in favor of those who both drink and are obese. The converse is true for the nondrinkers.

Body weight is therefore a *confounding factor* in this example; it masks the true impact of coffee drinking, the possible risk factor one is studying.

Two conditions have to be both satisfied to have a confounding variable:

1. The risk groups must differ on this factor.
2. The factor itself must influence the outcome.

These conditions are both satisfied in this example, as the structure according to body weight differs among the drinkers and the nondrinkers (condition 1). There is indeed an *association* between weight and coffee drinking, the overweight individuals drinking more coffee than the others. Furthermore, excess weight is itself a cause of high blood pressure (condition 2). The causal graph is therefore as follows, the nonoriented line between body weight, *W*, and drinking, *C*, indicating an association between both variables.



If one or both of these conditions is not fulfilled, the variable is not a confounder and does not have to be

TABLE 15–3 Distributions According to Weight of Drinkers and Nondrinkers

Drinking	Weight		
	Overweight	Not overweight	Total
Drinkers	25	75	100
Nondrinkers	25	75	100

TABLE 15–4 Average Diastolic Blood Pressure (in mm Hg)

Drinking	Weight		
	Overweight	Not overweight	Total
Drinkers	84	84	100
Nondrinkers	80	80	100

controlled for. Take the case where both drinkers and nondrinkers have the same distribution according to body weight,[5] waiving the first condition, such as in Table 15–3.

In this case, the blood pressure of drinkers is equal to $(94 \times 25 + 74 \times 75)/100 = 79$ mm Hg, and for the nondrinkers $(90 \times 25 + 70 \times 75)/100 = 75$ mm Hg. The difference $79 - 75 = 4$ mm Hg gives the right answer. Consider now the case (Table 15–4) where blood pressure does not differ according to body weight (condition 2 not satisfied), the distribution of patients being the same as in the original situation.

The average blood pressure for the drinkers will now be $(84 \times 300 + 84 \times 100)/400 = 84$ mm and for the nondrinkers $(80 \times 50 + 80 \times 150)/200 = 80$ mm, which is a simple application of the rule that the average of a constant is the constant itself!

Returning to the impact of age structure on the crude death rate, the example given at the beginning of this chapter, one can now conclude that in a comparative analysis one must control for age if the populations being compared have a different age structure (condition 1), which is not always the case, and if age has an impact on mortality (condition 2), which is indeed the case. Therefore, crude death rates are perfectly adequate summary measures of mortality if the populations being compared have more or less the same age structures.[6] However, one does not have to

[5] There is therefore no association in this case between body weight and coffee drinking.

[6] This also explains why crude birth rates are usually better summary measures than crude death rates, as the age structure of the population *at fertile ages* is rather similar whatever the global age structure of the population, young or old.

TABLE 15–5 Population Numbers and Blood Pressure (between brackets) According to Drinking and Bodyweight

Drinking	Weight		
	Overweight	Not overweight	Total
Drinkers	300 (94.0)	100 (74.0)	400 (89.0)
Nondrinkers	50 (90.0)	150 (60.0)	200 (67.5)
Total	350 (93.4)	250 (65.6)	600 (81.8)

control the structure of the population according to the length of one's hair, for example, as this variable probably has no impact on mortality.[7]

In the presence of confounders, one's conclusion may differ dramatically according to the fact that one controls or one does not control for the confounding variables. Rouanet (1985) provides a good example, called the Bombach paradox. In this example, the initial conclusion is *reversed* when one controls for the confounding variables![8] Note that if body weight is a confounding factor when one examines the impact of coffee drinking on blood pressure, conversely coffee drinking is itself a confounding variable in the study of the relationship between weight and blood pressure. All depends on the assumption one is testing (i.e., the topic of the research project). Finally, one is never sure in actual research that one is controlling for all the confounders without exception; there may be hidden (i.e., unknown) variables that may wreak havoc with our results.[9] Science is an endless game in pursuit of knowledge.

II. INTERACTION BETWEEN CAUSES

Consider the following situation (Table 15–5), very similar to the original example given in Table 15–1, except that the average blood pressures are slightly different.

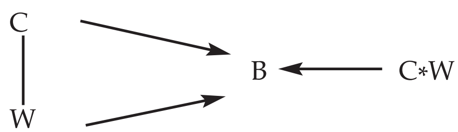
One sees that the impact of coffee drinking on blood pressure remains slight for those overweight (4 mm Hg), but for those not overweight, coffee drinking leads this time to a significant increase in blood pres-

[7] No bias is introduced if one does control for this variable, but it is a waste of time. In addition, decomposing the population into categories leads to smaller numbers in the cells and thus to greater random fluctuations.

[8] To take another example, a ranking of countries by mortality level may significantly differ according to the indicator one takes, the crude death rate or the expectation of life at birth. The latter controls for the age structure of the population, whereas the former does not.

[9] The randomized experimental approach should therefore always be preferred if feasible; see however Wunsch (1994).

sure (14 mm Hg). The difference between the extremes (94.0 − 60.0 = 34.0) is no longer the sum of the drinking effect and the weight effect, as these effects differ among categories. In other words, the impact of drinking depends on the modality of the confounding variable (overweight or not) and one speaks in this case of an *interaction effect* between causes on the outcome factor (blood pressure). The causal graph corresponding to this situation is the following, where the symbol $C{*}W$ represents the interaction effect.



In this example, there is a *weak interaction* between causes, in the sense that for overweight and nonoverweight individuals, coffee drinking increases the blood pressure; the impact has the same direction for both categories. In some situations one can have *strong interaction* effects, with crossovers between the series of rates, as shown in Table 15–6.

In this case (which is not a true situation), coffee drinking would increase blood pressure among those not overweight but decrease the pressure among overweight individuals. It has been shown, to give a true example, that smoking has a negative impact on Alzheimer disease except for carriers of the APOEε4 allele, for whom it has on the contrary a protective effect!

Interaction effects can sometimes be removed by a change in *scale*; for example, absolute differences may not be the same between the modalities of the confounding variable (as in the example of Table 15–5) but in some cases the relative differences or the odds are the same. One can therefore observe interaction effects on a linear scale but they disappear on a log-linear or logit-linear scale (Anderson, 1980; Wunsch and Thiltgès, 1995). Moreover, as Wilcox and Russell (1983) have shown, apparent strong interaction effects may disappear if one *norms* the distributions of the confounding variables, that is if one transforms them into distributions of zero average and a standard deviation of one.

TABLE 15–6 Population Numbers and Blood Pressure (between brackets) According to Drinking and Weight

Drinking	Weight	
	Overweight	Not overweight
Drinkers	300 (90.0)	100 (74.0)
Nondrinkers	50 (95.0)	150 (60.0)

III. STANDARDIZATION AND THE PROBLEM OF SUMMARY INDICES

One has just seen in section II that the variable body weight can be controlled for by measuring the blood pressure of drinkers and nondrinkers according to the fact that one is overweight or not. More generally, in mortality studies, one can control for the age structure by computing death rates specific for age. In regression analysis, a variable can be controlled for by including it in the regression equation, if a suitable functional form of the regression equation is chosen. Suppose now that one has computed death rates by single ages for the 52 countries one wishes to compare. Assuming approximately 100 age groups, this gives a table of $100 \times 52 = 5200$ cells. No ordinary human being can visually process such a large table unless he or she has exceptional cognitive abilities. One can also put the 52 graphs of death rates by age on the same figure, but once again, the result will probably not be visually interpretable as many graphs will overlap. Demographers therefore usually summarize their data for comparative purposes, even if this means losing valuable information. If one is comparing only two or three countries, it is therefore much preferable to put the data on a graph and discuss the results.

In the field of mortality, a summary measure of the death rates m_x often used is the *expectation of life at birth*,[10] given by the expression:

$$e_0 = \frac{\sum x \cdot d_x}{\sum d_x} = \frac{\sum x \cdot L_x \cdot m_x}{\sum L_x \cdot m_x}$$

where d_x and L_x are the life-table deaths and the stationary population at completed age x, respectively. As the number of deaths depends on the size of the stationary population at age x, one could also compute an average age at death based not on the absolute numbers of life-table deaths d_x, but on the relative numbers d_x/L_x. As the latter ratio is equal to the death rate at age x, the average age at death would simply be equal to $\Sigma x m_x / \Sigma m_x$. The difference between the two measures depends on the form of the weights applied to the ages x. Life-table deaths vary according to an overturned S, while death rates follow a U shape that continuously increases with age after adolescence.

[10] One can transform rates into probabilities, draw up a life table, compute the life expectancy at birth, and use this measure as a summary indicator of mortality. Note that the life expectancy is *not* a standardized index, even if it is independent of the actual population age structure, as the weights L_x (the stationary population by age) applied to the death rates vary from one life table to the other, contrary to a standard population.

1. Direct Standardization

A variable such as the death rate by age can also easily be summarized by one of the ordinary techniques in elementary descriptive statistics. For example, in comparative studies, one can simply take for each country the *arithmetic average* of death rates as a summary measure of mortality for this country. This would give equal weight to each age group. Alternatively, one can compute a *weighted average* of rates, where the various age groups would be given different weights according to the average age structure of the various populations one is comparing. Let m_{xj} and p_x represent respectively the death rate at age x (for a country j) and the relative average population number at that age (i.e., relative to the total average population of all the countries being compared). The death rate for country j would then be:

$$\sum m_{xj} p_x \Big/ \sum p_x = \sum m_{xj} p_x$$

as Σp_x equals 1.

In demography and in epidemiology, this method is known as the *direct standardization* of rates and the series p_x according to age as the *standard population.* The method is a simple *linear combination* of death rates, the standard population acting as weights. The basic principle is to use the *same standard population* for all the cases being compared. As seen above, one of the two conditions for having a confounding variable (age, in the present case) is that the risk groups differ on this factor. If one uses the same standard population structure p_x for all the series of rates being compared, this condition is obviously not satisfied and the age structure is therefore not a confounder. For example, the World Health Organization (WHO) is using a standard world population for direct standardization purposes. Unfortunately, the WHO decided in 1992 to alter this standard (Davies, 1995). Changing the standard has made comparisons over time and between sexes impossible without recalculation.

There is nothing about the technique that restricts its applicability to age. One can, for example, standardize the effects of differences in birth-order distributions between populations whose infant probabilities of dying we are comparing (Preston *et al*., 2001).

Standardized indices are used for comparative purposes only; the value of the index is of no interest per se. The choice of weights is actually completely arbitrary, and therefore different choices of weights will yield different results. One can for example set the weights equal to one, as in the numerator of the simple arithmetic average of rates. This is done in fertility analysis: the *total fertility rate* (a summary measure of fertility in period analysis) is indeed taken as the sum of age-specific fertility *rates* (frequencies) (i.e., as a linear combination of fertility rates with equal weights at each age). The *total fertility rate* is therefore a directly standardized measure of fertility, with a *rectangular* standard population of one person per age group.

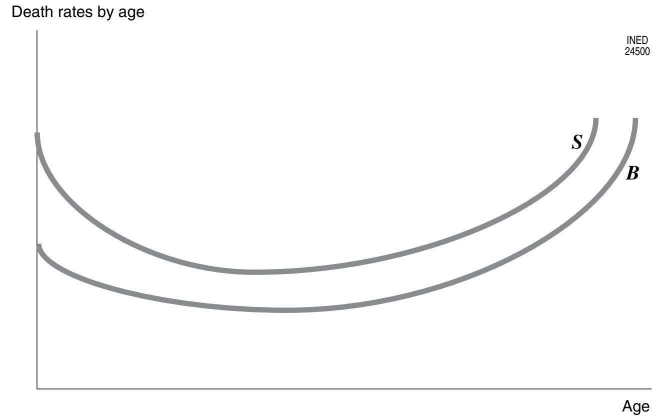


FIGURE 15–1 Age-specific death rates for countries Borduria (*B*) and Syldavia (*S*) in the absence of interaction.

Do standardized measures adequately represent the actual differences between the rates? This problem is similar to computing the averages of two distributions and asking if the difference between the means correctly represents the difference between the distributions themselves. Consider two sets of death rates by age for two imaginary countries,[11] Syldavia (*S*) and Borduria (*B*), as in Figure 15–1. In this example, the difference between rates $m_x(S)$ and $m_x(B)$ is the same whatever the age. In this case, it is easy to show (Wunsch and Thiltgès, 1995) that the absolute difference between standardized measures correctly represents the actual difference between rates, whatever the standard population. This is due to the absence of interaction effects between country and age in this example. Consider now data shown in Figure 15–2. In this case, Syldavia still has a higher morality at all ages than Borduria, but the difference is more marked at young ages than at older ages. There is *weak interaction* (as defined in section II) between country and age on mortality in this situation. Standardization will always yield a higher measure for Syldavia than for Borduria, whatever the standard population, but the difference between the two summary measures will depend on the choice of the standard population. Standardization correctly gives the order (higher versus lower mortality) but not the difference between rates by age. As these differences vary according to age, no

[11] My thanks to Hergé and his comic strips.

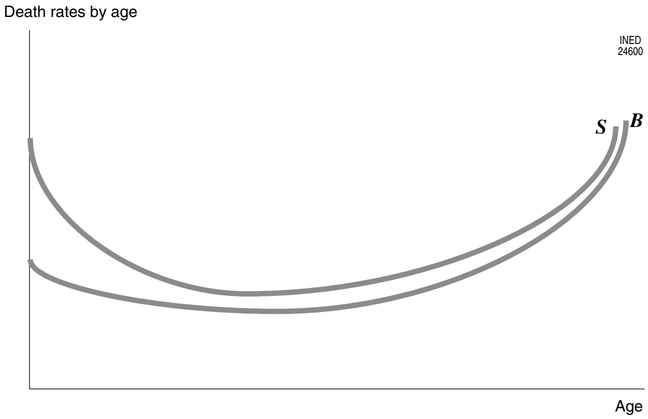


FIGURE 15–2 Age-specific death rates for countries Borduria (*B*) and Syldavia (*S*) in the presence of weak interaction.

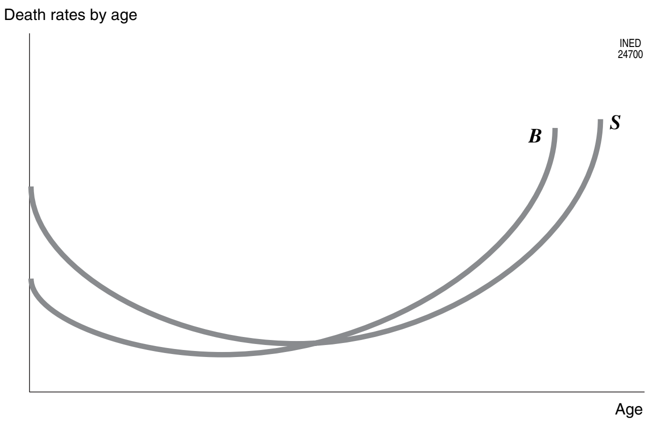


FIGURE 15–3 Age-specific death rates for countries Borduria (*B*) and Syldavia (*S*) in the presence of strong interaction.

single summary measure can correctly represent this fact.

Figure 15–3 presents a case of *strong interaction* between country and age: Syldavia has a higher mortality at young ages but a lower mortality at higher ages, i.e. a crossover of age-specific mortality curves. No single summary measure can adequately represent this situation. According to the choice of the standard population, one will conclude that Syldavia has a higher, an equal, or a lower mortality than Borduria, which is not a very helpful result! Having recourse to standardization, one must therefore test for interaction effects. If there is no interaction or if it is weak, standardized measures can be used with confidence to rank the countries one is comparing according to the variable of interest (e.g., mortality), whatever the standard population. If there is strong interaction, standardization breaks down and should not be used, as no single summary

TABLE 15–7 Deaths, Populations, and Death Rates by Region and Age

Variables	Age		
	1	2	Total
Syldavia			
Deaths	60	80	140
Persons	600	400	1,000
Rate	0.100	0.200	0.140
Borduria			
Deaths	60	393	453
Persons	200	800	1,000
Rate	0.300	0.491	0.453
Total			
Deaths	120	473	593
Persons	800	1,200	2,000
Rate	0.150	0.394	0.297

measure can adequately represent the actual situation in this case.[12]

2. The Mantel-Haenszel Statistic

The *Mantel-Haenszel (M-H) statistic* is useful when working with odds and *odds ratios* instead of absolute or relative rates or risks, the *odds* being the ratio of a rate to its complement to unity. Consider Table 15–7, which shows the number of deaths, the populations exposed to risk, and the death rates for the two countries Syldavia (S) and Borduria (B) and for two age groups 1 and 2.[13] In this example, one observes that there is no interaction at the odds level between country and age. Indeed, one derives from the data:

$$\frac{\text{Odds } 0.300}{\text{Odds } 0.100} = \frac{\text{Odds } 0.491}{\text{Odds } 0.200} = 3.86$$

Applying the definition of the odds, this is equivalent to saying that

$$\frac{k'60(600-60)}{k'60(200-60)} = \frac{k''393(400-80)}{k''80(800-393)} = 3.86$$

where k' and k'' are two arbitrary constants ($\neq 0$). Using some elementary mathematics, one derives

[12] This would also be the case with the expectation of life at birth, as it is also a sole indicator of the level of mortality.

[13] The example is drawn from Wunsch and Thiltgès (1995).

$$\frac{k'60(600-60)+k''393(400-80)}{k'60(200-60)+k''80(800-393)}=3.86$$

The (M-H) statistic (Kahn and Sempos, 1989) takes k' and k'' equal to the reciprocal of the total populations in the two age groups, respectively 800 and 1,200. This leads to:

$$M-H=\frac{60(600-60)/800+393(400-80)/1200}{60(200-60)/800+80(800-393)/1200}=3.86$$

The result is 145.3/37.6 = 3.86: the relative difference in odds or odds ratios obtained above. The M-H statistic gives the true difference between odds in the case of a logit-linear model with no interaction, controlling for age. If there is a common odds ratio (no interaction), the M-H statistic provides it. If interaction is present, the M-H statistic provides a weighted average of the odds ratios. Significance tests for the overall odds ratio can be found in the epidemiologic literature (Kahn and Sempos, 1989). Note that the M-H statistic is not a standardized index, as the cross-products involve the actual population numbers.

3. Indirect Standardization

Suppose now that the death rates by age are unknown because one has no breakdown of the number of deaths by age. One only knows the total number of deaths and the age-structure of the population. In this case, *indirect standardization* can be used. Suppose one compares once again Syldavia (S) and Borduria (B). Let $P_x(S)$ and $P_x(B)$ represent the population of the two countries decomposed by age x, and let $D(S)$ and $D(B)$ stand for the total number of deaths in the two countries. Take this time an arbitrary standard set of death rates by age m_x and compute the two ratios

$$\frac{D(S)}{\sum P_x(S)m_x} \quad \text{and} \quad \frac{D(B)}{\sum P_x(B)m_x}$$

These *standardized mortality ratios* (or more generally, the *standardized incidence ratios*) are simply ratios of the observed deaths to the expected deaths for the two countries, for any particular arbitrary set of standard death rates by age. The same standard set is applied to the two situations. One can also compute *indirectly standardized rates* by multiplying the standard mortality ratios by the crude death rate for the standard population. Indirectly standardized *relative risks* are then taken as the ratios of the indirectly standardized rates or of the standard mortality ratios. Once again, as with direct standardization, the standard is perfectly arbitrary, but Hoem (1991) has devised an iteration method to derive a unique standard series of rates from the data set itself.

How adequate is indirect standardization for controlling for confounders? Let $m_x(S)$ and $m_x(B)$ represent the unknown death rates by age for the two countries. As the number of deaths is equal to the sum of the products of death rates by the actual population numbers at each age, one can write the two ratios above as

$$\frac{\sum P_x(S)\cdot m_x(S)}{\sum P_x(S)\cdot m_x} \quad \text{and} \quad \frac{\sum P_x(B)\cdot m_x(B)}{\sum P_x(B)\cdot m_x}$$

The standardized mortality ratios are therefore respectively equal to the weighted average of $m_x(S)$ divided by the weighted average of the standard set of rates m_x and to the weighted average of $m_x(B)$ divided once again by the weighted average of m_x. Indirect standardization therefore compares the averages of the unknown rates in terms of the averages of the standard set of rates. Note however that the latter are not identical, as the weights applied to the set m_x are not the same. In the first case, the weights are the population numbers by age of Syldavia $P_x(S)$ whereas in the second case, they are equal to $P_x(B)$, the population numbers of Borduria.

Indirectly standardized ratios are therefore not independent of the actual population structures, contrary to the case of direct standardization. However, suppose that one knows one of the series of rates, say $m_x(B)$, and that one takes this series as standard. The standardized ratios become

$$\frac{\sum P_x(S)\cdot m_x(S)}{\sum P_x(S)\cdot m_x(B)} \quad \text{and} \quad \frac{\sum P_x(B)\cdot m_x(B)}{\sum P_x(B)\cdot m_x(B)}$$

The last ratio is equal to one while the first is equal to the ratio of the weighted average of $m_x(S)$ to the weighted average of $m_x(B)$, using the same weights $P_x(S)$. In this case, indirect standardization does give a meaningful result: the comparison between the averages of the two series of rates, using the same set of weights. Indirect standardization can thus be recommended here, as the result is similar to the ratio of two *directly* standardized indices using $P_x(S)$ as the standard population.

Indirectly standardized rates are biased, in the sense that their difference does not reflect the true difference between death rates by age even in the absence of interaction (Wunsch and Thiltgès, 1995). This was not the case with direct standardization, which is an unbiased procedure in the absence of (linear) interaction. On the other hand, it has been shown (Anderson *et al*. 1980) that indirect standardization is more precise (lesser variance) than direct standardization. It could then be recommended before direct standardization when population numbers are small. Procedures for computing confidence intervals for directly standardized or indirectly standardized rates can be found in Estève *et al*. (1993) and Kahn and Sempos (1989).

4. Double Standardization

The purpose of the methods presented in the preceding sections was to eliminate the influence of population structures (i.e., the composition of the population according to the confounding variable or variables) in the study of population change, by holding these structures constant over time or from one area to the other. One may also adopt the opposite point of view: Frequencies on the outcome factor can be held constant in spatial or temporal comparisons to determine the sole impact of changes in population structures on population movement.

Let E and E' represent the total number of events observed during periods (e.g., years) t and t' for two populations, say K and K'. Let P_x and P'_x be the corresponding midperiod population structures according to the confounding variable (e.g., age) and f_x and f'_x the corresponding frequencies. The sole influence of differential frequencies will be given by the standardized indices $\Sigma P_x f_x$ and $\Sigma P_x f'_x$ for example, taking P_x as standard. As to the sole influence of differential population structures, it can be expressed by the comparison between $\Sigma P_x f'_x$ and $\Sigma P'_x f'_x$, taking this time f'_x as standard. Multiply both relative differences between the standardized indices:

$$\frac{E'}{E} = \frac{\sum P_x f'_x}{\sum P_x f_x} * \frac{\sum P'_x f'_x}{\sum P_x f'_x}$$

As $\Sigma P_x f_x = E$ and $\Sigma P'_x f'_x = E'$, the above product yields the relative difference between the total number of events observed. The product can therefore be considered as a *double standardization* for both structures and frequencies. The relative difference between the total number of events at time t and t' can thus be decomposed into two parts, the first indicating the impact of changes of frequencies and the second the impact of changes in population structures.

Instead of applying double standardization to the *relative* difference between the total number of events observed, one can also have resort to the double standardization of the *absolute* difference between the number of events as follows:

$$E' - E = \sum P'_x f'_x - P_x f_x$$

or

$$E' - E = \sum (P'_x - P_x) f'_x + \sum (f'_x - f_x) P_x$$

The first sum represents the component due to differential population structures and the second sum the component due to differential frequencies.

Letting $f'_x = f_x + a_x$ and $P'_x = P_x + b_x$, one can also write

$$E' - E = \sum (P_x - b_x)(f_x + a_x) - \sum P_x f_x$$

i.e.,

$$E' - E = \sum (P'_x - P_x) f_x + \sum (f'_x - f_x) P_x + \sum a_x b_x \quad \text{(Eq. 1)}$$

If one neglects the last term, one breaks down the difference $E' - E$ into the two components $\Sigma(P'_x - P_x)f_x$ and $\Sigma(f'_x - f_x)P_x$. Note that as:

$$\begin{aligned}\sum f'_x (P'_x - P_x) &= \sum f_x (P'_x - P_x) + \sum a_x (P'_x - P_x) \\ &= \sum f_x (P'_x - P_x) + \sum a_x b_x,\end{aligned}$$

the *interaction* term $\Sigma a_x b_x$ is in fact included in the first term of Equation 1. The two relations are therefore identical.

The method described above can easily be modified to measure the components of a difference between two *crude rates*. If $P = \Sigma P_x$ and $P' = \Sigma P'_x$, then

$$\frac{E'}{P'} = \frac{E}{P} = \sum f'_x \left(\frac{P'_x}{P'} - \frac{P_x}{P} \right) + \sum (f'_x - f_x) \frac{P_x}{P}$$

Variants of this formula can be found in the literature (see for example the various papers on standardization included in Clogg, 1993).

5. An Example

Table 15–8 presents the age-specific fertility rates and the proportions of females at reproductive ages for two countries,[14] the United States and Ireland. The *crude female birth rates*[15] in these populations are, respectively, 34.78/1000 and 42.96/1000, a relative difference of +24%. As the age structures differ between the two countries, one seeks to compare their fertility on the basis of two summary indexes that, contrary to the crude rates, are independent of the population structures.

[14] The example is based on data from the end of the 1960s for the United States and Ireland, and is taken from Wunsch and Termote (1978).

[15] Number of births divided by the total *female* population.

TABLE 15–8 Age-specific Fertility Rates (ASFRs) and Proportions of Females by Age–Groups

Age groups	*ASFRs per* 1000		*Proportions of females per* 10,000	
	United States	Ireland	United States	Ireland
15–19	68.9	13.4	868	878
20–24	174.0	127.1	744	631
25–29	142.6	230.6	601	517
30–34	79.3	211.3	547	507
35–39	38.5	143.2	582	542
40–44	10.6	54.5	626	574
45–49	0.7	4.2	602	577
All ages 0 to ω			10,000	10,000

To compare the average fertility between the two countries, the following summary indexes have been computed:

a. the sum of age-specific fertility rates (ASFRs) (per 1000):

United States: 514.6
Ireland: 784.3
Ratio Ireland/United States: 784.3/514.6 = 1.5241

b. The arithmetic average of ASFRs (per 1,000):

United States: 73.51
Ireland: 112.04
Ratio: 112.04/73.51 = 1.5241

c. The geometric average of ASFRs (per 1,000):

United States: 32.55
Ireland: 59.75
Ratio: 59.75/32.55 = 1.8356

d. Direct standardization, American structure as standard (per 1,000):

United States: 34.78
Ireland: 48.04
Ratio: 48.04/34.78 = 1.3813

e. Direct standardization, Irish structure as standard (per 1,000):

United States: 31.16
Ireland: 42.96
Ratio: 42.96/31.16 = 1.3787

f. Indirect standardization, American ASFRs as standard (per 1,000):

United States: 34.78/34.78 = 1
Ireland: 42.96/31.16 = 1.3787
Ratio: 1.3787/1 = 1.3787

g. Indirect standardization, Irish ASFRs as standard (per 1,000):

United States: 34.78/48.04 = 0.7240
Ireland: 42.96/42.96 = 1
Ratio: 1/0.7240 = 1.3813

In all these cases, the average fertility of Irish women is higher than the average fertility of American women. Nevertheless, the proportional difference between the fertility schedules varies from 38% to 83% according to the index chosen. There is however a crossover between the two fertility schedules, American fertility rates being higher at lower ages and lower at higher ages. This strong interaction effect between age and country could lead to opposite conclusions depending on the standard age structure chosen. For example, taking a very young age structure would bias the index in favor of American fertility whereas taking an older structure would lead to the opposite conclusion. The use of a sole indicator by country can never translate the fact that American age-specific fertility is higher than the Irish at young ages and lower afterward. It is therefore always necessary to check for possible interaction effects before comparing standardized indexes.

6. Statistical Methods

When possible confounding variables are numerous, standardization is no longer practical and has to give way to the more convenient and powerful techniques of statistical analysis. Due to lack of space, only a brief presentation of some statistical adjustment approaches can be given here. The reader is therefore referred to the works given as reference for further details.

Firstly, consider the data themselves. If they are derived from a sample, the number of events or persons per cell can be small or even nil. Before proceeding to standardizing the data, it can be useful as a first step to fit a *log-linear model* in view of smoothing the observed data and obtaining nonzero counts in the cells. It also enables testing for interactions. As Bishop *et al*. (1975; p. 131) stress, fitting log-linear models before standardizing has actually two purposes:

> 1. Knowledge of the structure of the data enables us to determine whether the whole array can be described by a simple statistic or whether partitioning the array into smaller segments is appropriate;
> 2. Computing the statistics from the fitted values rather than from the observed counts removes some of the variability due to random fluctuations.

As to improving direct standardization by statistical techniques, one can for example apply *principle* *components analysis* and compute the *first* principle component from the set of data on age-specific death rates by country, which is equal to deriving the first eigenvector of the variance-covariance matrix, then use the factor score coefficients as weights in the linear combination of rates (Duchêne and Wunsch, 1980). The factor score on this first principle component would then be a summary measure of mortality independent from the actual age structure of the population. There are several advantages to this procedure. First, the choice of the standard population is no longer arbitrary, as the vector of coefficients is derived from the data themselves. Second, the first principle component is the *best* linear combination of rates, because it is the vector representing the maximum spread of the scatterplot. Third, the method yields a measure of the loss of information due to the combination of rates (i.e., the ratio of the first eigenvalue to the sum of all eigenvalues of the data matrix).

Other demographers (Hoem, 1987) have used *multiplicative models* of the type

$$m_{xk} = \Theta_k \alpha_x$$

where Θ_k is a group- or risk-specific scaling factor and α_x is an age- or confounder-specific process factor. This type of specification is similar to a two-component age-period-cohort model. It can also be considered, as Hoem has shown, as a special kind of proportional hazard model.

More simply, one can apply a *linear regression model* without interaction (Leridon and Toulemon, 1997), such as:

$$m_{xi} = m + a(x) + b(i) + \varepsilon$$

where x refers to age (the confounding variable) and i to the variable of interest (country, for example). Variable ε is an error term, normally distributed and of zero mean. Once the parameters are estimated, rates adjusted for the confounder can be obtained from the estimates derived from the regression equation, using the direct standardization approach. Instead of a normally distributed variable model, as in the case of ordinary linear regression, one may prefer in some situations a Poisson model, such as when dealing with small-sample data, leading to a *log-linear regression* equation estimated by maximum likelihood (Estève *et al.*, 1993). A convenient approach for adjusting data with multivariate models is using *multiple regression for categorical variables* (Kahn and Sempos, 1989), such as logit or logistic regression models. Both the outcome variable and the predictors are attributes in this case (i.e., take the values 0 or 1) and are usually computed from individual sample data. Once again, this approach is analogous to direct standardization but includes the possibility to control for many confounders at one time. Using the regression coefficients, we can easily modify the crude rate into various adjusted rates assuming that everyone has a same attribute (e.g., the same sex) with all the other variables including the confounder(s) unchanged. We can then compare the differences on the adjusted rates of the outcome variable (e.g., survival or death) between males and females in this case, holding all covariates constant.

Toulemon (1992) has compared the results of some of these approaches, using data on abortions for females by occupational categories and controlling for the confounding impact of age and marital status. Three types of models are fitted: linear, log-linear (i.e., multiplicative), and logistic, taking account of direct and indirect standardization and having recourse both to the classical methods of standardization and to regression analysis. Due to the presence of interaction effects, there are significant differences between the results. In particular, the common method of direct standardization cannot be recommended in this example, as it overestimates significantly the abortion rates in some categories. The other methods lead to more coherent conclusions in this case; they are also less sensitive to small cell counts than direct standardization.

Consider now a comparison of fertility or mortality rates by geographical areas (such as states, provinces, municipalities, and so forth), a common situation in demography. In this case, it is not only necessary to control for the age structure of the populations in the various areas, but also to check if the data are structured geographically. For example, is there a spatial clustering of death rates possibly indicating common risk factors among areas? Standardized mortality rates by areas can be computed as usual and the resulting map inspected for geographical structure, but if the areas are small and/or the outcome variable rare (such as for example, mortality due to lip cancer by districts), the map showing the standardized rates will often emphasize the areas with small population numbers leading to extremely high or low rates due to chance fluctuations. Moreover, the method is not well suited for taking account of multiple confounders and other covariates, as pointed out previously, and of hierarchical data structures. In this case, use a *multilevel spatial model* incorporating some form of local clustering taking account of possible area effects.

A commonly used solution is a conditional autoregressive spatial model where the effect of an area depends upon the effects in the neighboring areas (Leyland, 2001). A clustering model of this type can also include covariates representing possible risk factors. For example, in an analysis of lip cancer by districts in Scotland (Leyland, 2001), one may also include in the model the percentage of the population employed in agriculture, fishing, and forestry as a measure of exposure to sunlight, a potential risk factor. Further levels can be included in the spatial model, such as considering the effect of the country in addition to that of the district, in an intercountry and interdistrict comparison. Country effects may reflect cultural or behavioral differences among the countries having an impact on the outcome. Spatiotemporal models can furthermore be fitted if the data are available, taking account not only of differences in space but also in time.

If differences are observed in standardized rates among areas for example, one should try to find the causes or determinants of these differences. It is sometimes possible to postulate the relations between presumed causes and effects; one may then try to test this causal structure by statistical techniques, if suitable data are available. To take an example from Lopez-Rios *et al*. (1992), we examined the impact of health care use on regional adult mortality differences in Spain, assuming that use depends on the supply of health care and on the level of social development of the regions, these two variables being influenced themselves by the level of regional economic development. Each of these *latent variables*, latent in the sense that they cannot be measured directly, are represented by a set of measurable manifest variables or multiple indicators. The confounding variable is once again the age structure of the populations of the regions. A *covariance structure model* was fitted to the data with the LISREL software, outcome and predictor latent variables being linked by a series of linear equations.

Let

$$\eta' = (\eta_1\ \eta_2 \ldots \eta_m)$$

and

$$\chi' = (\chi_1\ \chi_2 \ldots \chi_n)$$

stand, respectively, for the vectors of latent outcome variables and predictor variables. The structural equation model can then be written, in matrix notation:

$$\eta = B\eta + \Gamma\chi + \sigma$$

where B and Γ are matrices of parameters between latent variables and

$$\sigma' = (\sigma_1\ \sigma_2 \ldots \sigma_m)$$

is a vector of residuals (errors in equations). The elements of B represent the direct effect of variables η on the other variables η, while the elements of Γ represent the direct effects of the variables χ on the variables η. The vector variables η and χ are latent or unobservable, but they can be represented by measurable indicators:

$$y' = (y_1\ y_2 \ldots y_p) \textit{ and } x' = (x_1\ x_2 \ldots x_q)$$

For example, the indicators for the latent variable *social development* in this example are the proportion of women per region with secondary education, the proportion of females occupied in the tertiary sector, the proportion of accommodations with no running water, and so forth. Measurable and latent variables are related by the equations:

$$y = \Lambda y\eta + \varepsilon \textit{ and } x = \Lambda x\chi + \delta$$

where ε and δ are vectors of error terms (measurement errors). These equations stand for the multivariate regression of y on η and of x on χ.

Results can be expressed in terms of *path coefficients* (or standardized regression coefficients) between causes and effects, as shown by the causal graph in Figure 15–4, all coefficients being in this case statistically significant.

We have dealt in more detail with this approach because a structural equation model implies clarifying the relational structure between variables, according to one's conceptual framework dealing with the problem at hand. It thus forces the researcher to define his or her concepts, to justify the indicators chosen, and to specify the explanatory variables, covariates, confounders, and the postulated relations between them, based on the theoretical assumptions he or she has. For example, one sees that in addition to the age structure of the regions, one needs in this case to control also for the impact of the level of social development, as this variable is significantly related to both mortality and health care use. It is therefore a confounder, as defined in section I, in the relationship between these two latter

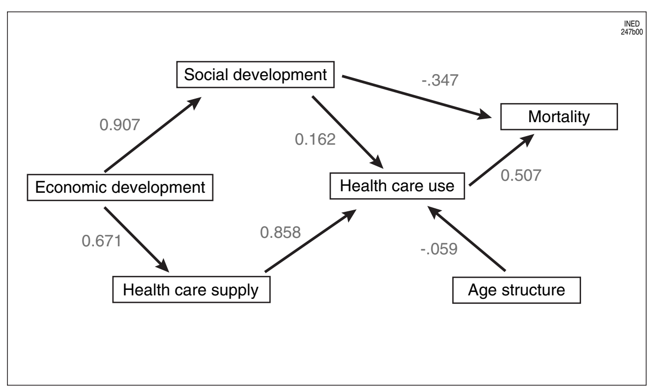


FIGURE 15–4 The health care system and regional mortality in Spain (1985). Path coefficients between latent variables.

variables. Actually, the impact of differential age structures between regions on health care use is quite slight in this example, and induces no important confounding effect.

To conclude, there are no roundabout statistical models: Each problem must be analyzed and the best model selected according to the problem, the data, and the underlying theory. Models are simplifications of reality; one should therefore always test the fit of the model (i.e., to see to what extent the model reproduces the data). Linear models, for example, are not adequate if the relationship between the outcome and explanatory variables is not linear, although a change in scale can sometimes improve the linear fit. To give a simple example, an exponential relationship between two variables can be made linear by taking logarithms. Moreover, except being God (or one's equivalent according to one's faith), one is never sure that all confounding variables have been taken care of. If there are hidden (i.e., unknown confounders), the statistical exercise can lead to false conclusions. As stated previously (see also Wunsch, 1988), science is therefore a game of (endless?) pursuit in search of better theories, data, and models.

References

Anderson Sharon, Auquier Ariane, Hauck Walter W., Oakes David, Vandaele Walter, and Weisberg Herbert I., 1980. *Statistical methods for comparative studies.* New York, John Wiley & Sons, 289 p.

Bishop Yvonne M. M., Fienberg Stephen E., Holland Paul W., 1975. *Discrete multivariate analysis*, Cambridge, MIT Press, 557 p.

Clogg Clifford C., (ed.), 1993. Standardization and decomposition of group differences, *in:* Donald J. Bogue, Eduardo E. Arriaga and Douglas L. Anderton (eds.), *Readings in population research methodology*, vol. 1, p. 1–69. Chicago, Social Development Center, n.p.

Davies Tom, 1995. World standard populations, *International Journal of Epidemiology*, 24(6), 1261.

Duchêne Josianne and Wunsch Guillaume, 1980. Population-type optimale et composante principale, *Population et Famille*, 49, p. 23–30.

Estève Jacques, Benhamou Ellen and Raymond Luc, 1993. *Méthodes statistiques en épidémiologie descriptive.* Paris, Les éditions de l'INSERM, 307 p.

Hoem Jan M., 1987. Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review, *International Statistical Review*, 55(2), p. 119–152.

Hoem Jan M., 1991. La standardisation indirecte améliorée et son application à la divortialité en Suède (1971–1989), *Population*, 46(6), p. 1551–1568.

Kahn Harold A. and Sempos Christopher T., 1989. *Statistical methods in epidemiology.* New York, Oxford University Press, 292 p.

Leridon Henri and Toulemon Laurent, 1997. *Démographie: approche statistique et dynamique des populations.* Paris, Economica, 440 p. (Coll. économie et statistiques avancées).

Leyland A. H., 2001. Spatial analysis, *in:* A. H. Leyland and H. Goldstein H. (eds.) *Multilevel modelling of health statistics*, Chichester, Wiley, 217 p.

Lopez Rios Olga, Mompart Anna, and Wunsch Guillaume, 1992. Système de soins et mortalité régionale: une analyse causale, *Revue européenne de démographie*, 8(4), p. 363–379.

Preston Samuel H., Heuveline Patrick, and Guillot Michel, 2001. *Demography: measuring and modeling population processes.* Oxford, Blackwell Publishers, 291 p.

Rouanet Harry, 1985. Barouf à Bombach, *Bulletin de Méthodologie Sociologique*, vol. 6, p. 3–27.

Toulemon Laurent, 1992. Population-type et autres methodes de standardisation: application à la mesure du recours à l'avortement selon la PCS, *Population*, 47(1), p. 192–204.

Wilcox Allen J. and Russell Ian T., 1983. Perinatal mortality: standardizing for birthweight is biased, *American Journal of Epidemiology*, 118(6), p. 857–864.

Wunsch Guillaume J. and Termote Marc G., 1978. *Introduction to demographic analysis*. New York, Plenum Press, 274 p.

Wunsch Guillaume J., 1988. *Causal theory and causal modeling.* Leuven, Leuven University Press, 200 p.

Wunsch Guillaume J., 1994. L'analyse causale en démographie. *in:* Robert Franck (ed.), *Faut-il chercher aux causes une raison?* p. 24–40. Paris, Librairie Philosophique J. Vrin and Institut interdisciplinaire d'études épistémologiques, 448 p. (Coll. Science, histoire, philosophie).

Wunsch Guillaume J. and Thiltgès Évelyne, 1995. Une confusion standardisée: variables confondantes et standardisation, *Genus*, 50(3–4), p. 27–59.

CHAPTER

16

Relationships Between Age-Specific Rates and Synthetic Indicators: Decomposition of a Difference

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italty

Synthetic indicators are a convenient means to track changes in demographic phenomena. However, it is just as important to assess the influence of individual components on a synthetic indicator to (1) understand those changes, (2) evaluate the differences observed between countries or regions, and (3) identify the main issues that can be addressed by demographic or social policies. For this purpose, we shall examine how the difference between two values of a synthetic indicator can be decomposed into fractions attributable to each of those components.

Concerning mortality, our approach at this stage merely involves assessing the impact of each age-specific or age-group-specific difference in mortality on a difference between two life expectancies.[1] For fertility, our emphasis will not be so much on decomposing a difference in average age at childbearing into the specific effects of each difference in age-specific fertility; rather, we will try to identify the distinctive roles of intensity (quantum), rank, and tempo of births in the expression of a difference in the total fertility rate.

[1] In Chapter 42 on causes of death (Volume II of this treatise), France Meslé shows that it is also interesting—if not even more so—to apply the same approach to mortality by causes of death.

I. HOW DIFFERENCES IN AGE-SPECIFIC MORTALITY AFFECT A DIFFERENCE IN LIFE EXPECTANCY

To what extent do each of the differences observed at different ages affect the difference between two life expectancies—for example, those of males and females, those of France and Russia, or those of France in 1950 and in 1995? There is no simple answer. In the life table, the change in a death probability for a given age has a dual effect: (1) a main effect, namely, a change in the number of deaths in the table at that age; and (2) an interaction effect (i.e., a change in the number of survivors at the following age, which will, in turn, alter the number of deaths at all subsequent ages).

In a continuous approach, the problem can be solved by means of a minor approximation (Pollard, 1982); for discrete data, the solution is harder to apply. Many authors have proposed solutions. Korchek-Tchepurkovski (1968) was the first to tackle the issue, followed by Andreev (1982), Pollard (1982, 1988), Arriaga (1984), Pressat (1985), and Valkovics (1991).

Using a somewhat different method, Andreev (1982) and Arriaga (1984) actually come up with the same solution, where the contribution $\varepsilon_{(x,x+a)}$ of the change in mortality at a given age to the life-expectancy difference ($e_0^2 - e_0^1$) equals

$$\varepsilon_{(x,x+a)} = l_x^1(e_x^2 - e_x^1) - l_{x+a}^1(e_{x+a}^2 - e_{x+a}^1)$$

The drawback of this formula is that it yields a different result when we measure the contributions to the same life-expectancy difference expressed in reverse ($e_0^1 - e_0^2$). One remedy is to perform the calculation in both directions and take the average of the two values obtained as the final estimate.

The method proposed by Pollard avoids this drawback. The author shows that the problem can be solved analytically using either of the two formulas:

$$e_0^2 - e_0^1 = \int_0^{\omega} (\mu_x^1 - \mu_x^2)\,{}_xp_0^2 e_x^1 dx \quad \text{or}$$

$$e_0^2 - e_0^1 = \int_0^{\omega} (\mu_x^1 - \mu_x^2)\,{}_xp_0^1 e_x^2 dx$$

where ${}_xp_0^1$ is the probability of survival from birth to age x in the first table and ${}_xp_0^2$ that of the second table, μ_x^1 and μ_x^2 being the forces of mortality at age x.

Arguing that there is no reason to prefer one method to the other, Pollard suggests using a third by averaging the two weightings that differentiate them:

$$w_x = 0.5 \cdot ({}_xp_0^2 e_x^1 + {}_xp_0^1 e_x^2)$$

which gives:

$$e_0^2 - e_0^1 = \int_0^{\omega} (\mu_x^1 - \mu_x^2)\, w_x dx$$

To return to the discrete calculation, Pollard estimates the integral of the instantaneous rate from the table's survivors:

$$\int_0^a \mu_x dx \cong -\ln \frac{l_{x+a}}{l_x} = {}_aQ_x$$

from which he obtains:

$$e_0^2 - e_0^1 \cong \sum_{x=0}^{\omega} ({}_aQ_x^1 - {}_aQ_x^2)\, w_{(x+0.5a)}$$

The approximation is good if we are using a complete life table (by year of age) and if it is not arbitrarily cut off at too low an age. The wider the age intervals, the less precise the result; typically, it is the final age interval that is the most problematic. If, for example, the table stops at 80 or 85 years of age, as is often the case, there can be a significant disturbance in the result, in particular when the rise in life expectancy depends heavily on the change in mortality at the most advanced ages.

Table 16–1 reports an initial decomposition of the difference between French male life expectancy in 1996 and 1989, based on complete life tables. We can see, for example, that of the total 1.46-year gain in life expectancy at birth recorded during this period, 0.23 years is due to the fall in infant mortality alone. The changes in mortality in all the other age groups produced far more modest gains. Another peak in gains

TABLE 16–1 Decomposition of the Difference Between French Male Life Expectancy in 1989 and 1996, Using John Pollard's Method Applied to Complete Life Tables

Age x	Survivors l_x		Life expectancy e_x		Weight w_x	$Q_x = -\ln \frac{l_{x+1}}{l_x}$		Contribution of age x
	1989	1996	1989	1996		1989	1996	
0	100,000	100,000	72.49	73.94	73.22	0.0087	0.0054	0.23
1	99,137	99,461	72.12	73.34	72.22	0.0007	0.0005	0.01
2	99,068	99,409	71.17	72.38	71.23	0.0004	0.0003	0.01
3	99,026	99,379	70.20	71.40	70.23	0.0004	0.0003	0.01
4	98,986	99,354	69.22	70.42	69.24	0.0003	0.0002	0.00
5	98,960	99,332	68.24	69.44	68.25	0.0003	0.0002	0.00
...	...	...	...	...	...	...	...	...
55	87,102	88,179	22.54	23.58	20.21	0.011	0.009	0.04
...	...	...	...	...	...	...	...	...
75	53,771	57,265	9.26	9.87	5.31	0.051	0.047	0.02
...	...	...	...	...	...	...	...	...
95	2,359	3,339	2.45	2.58	0.07	0.3629	0.3246	0.00
96	1,641	2,414	2.30	2.38	0.05	0.3987	0.3822	0.00
97	1,101	1,647	2.18	2.25	0.03	0.3801	0.3805	0.00
98	753	1,126	1.96	2.07	0.02	0.4206	0.4262	0.00
99	495	735	1.72	1.90	0.01	0.4959	0.3575	0.00
Total								1.46

TABLE 16–2 Contributions of Changes in Mortality for Different Age Groups to Life-Expectancy Gains Registered in France for 10-year Periods since 1899

Age	Period									
	1899–1909	1909–1920	1920–1929	1929–1939	1939–1949	1949–1959	1959–1969	1969–1979	1979–1989	1989–1996
					Male					
0–1	2.74	0.31	1.04	1.93	0.66	2.23	0.76	0.71	0.19	0.23
1–14	1.59	0.22	0.91	1.19	0.82	0.57	0.15	0.14	0.12	0.07
15–29	0.36	–0.12	0.64	0.75	1.03	0.32	–0.07	–0.02	0.15	0.13
30–59	–0.07	1.08	–0.38	0.44	2.41	0.78	0.01	0.51	0.61	0.29
60–74	–0.18	0.46	–0.34	0.18	0.58	0.47	–0.30	0.96	0.78	0.38
75+	–0.03	0.16	–0.18	0.11	0.13	0.35	0.04	0.34	0.54	0.36
Total	4.40	2.12	1.68	4.60	5.64	4.71	0.60	2.64	2.39	1.46
					Female					
0–1	2.54	0.20	1.10	1.77	0.50	1.91	0.63	0.60	0.18	0.11
1–14	1.70	0.18	1.19	1.39	0.83	0.57	0.17	0.13	0.07	0.06
15–29	0.42	0.02	0.36	1.02	1.18	0.58	0.02	0.07	0.09	0.03
30–59	0.34	0.66	0.36	1.03	1.53	1.14	0.30	0.55	0.42	0.11
60–74	–0.02	0.57	–0.22	0.55	0.65	0.96	0.27	1.03	0.67	0.23
75+	–0.04	0.27	–0.23	0.29	0.25	0.70	0.36	0.80	0.94	0.67
Total	4.94	1.89	2.56	6.06	4.94	5.85	1.75	3.19	2.38	1.21

is observed at around age 55, but its level is well below the gain registered in the first year of life. By summing all these gains we obviously find the total difference between the two life expectancies.[2]

One of the major advantages of Pollard's formula is that it allows the contributions of each age to be estimated from the mortality rates (i.e., we can estimate ${}_aQ_x$ using $a \cdot {}_am_x$). This makes it very easy to extend the formula's application to measuring the contribution of causes of death, for which mortality rates rather than life tables are generally available. In so doing, however, we increase the estimation error, especially if we are working on age intervals of more than 1 year. It is better, therefore, to work with the smallest possible age intervals and combine the results, which are always additive.

Using Pollard's approach as his starting-point, Pressat proposes a different weighting formula,

$$e_0^2 - e_0^1 \cong \sum_{x=0}^{\omega}\left[\frac{l_x^1 + l_x^2}{2}(e_x^2 - e_x^1) - \frac{l_{x+a}^1 + l_{x+a}^2}{2}(e_{x+a}^2 + e_{x+a}^1)\right],$$

which, however, will not work with mortality rates.

In fact, at the practical level, all these methods yield very similar results. The differences in the degree of precision of the initial data (fineness of age groups, method for computing the life tables used, consistency between these tables and age-specific rates, etc.) are much more important than the ones due to the theoretical distinctions between the methods employed (El Makrini, 1998).

We use Pollard's method here to illustrate the contributions of differences in age-specific mortality to a difference between two life expectancies. Our example concerns the life-expectancy gains registered in France for 10-year periods, since 1899 (Table 16–2 and Figure 16–1).

In the early 20th century, most of the increase in life expectancy was due to the fall in infant mortality and juvenile mortality. Between 1899 and 1909, the fall in male mortality among those younger than 15 years was single-handedly responsible for a 4.3-year gain in life expectancy, or 98% of the total 4.4-year gain achieved by males. The situation was very different in the subsequent 10-year period (1909–20) owing to the disturbance caused by World War I, but the fall in infant mortality soon became the predominant factor again—until World War II upset the pattern once more. The decline in infant and juvenile mortality regained its pre-eminence in the postwar years. After the 1950s, however, mortality at early ages fell to very low levels, and the potential gains from its further reduction are dwindling. Indeed, life expectancy would eventually have leveled off if an entirely novel phenomenon had

[2] Or at least an approximation of that total, insofar as concrete computations always give rise to small discrepancies due to data flaws.

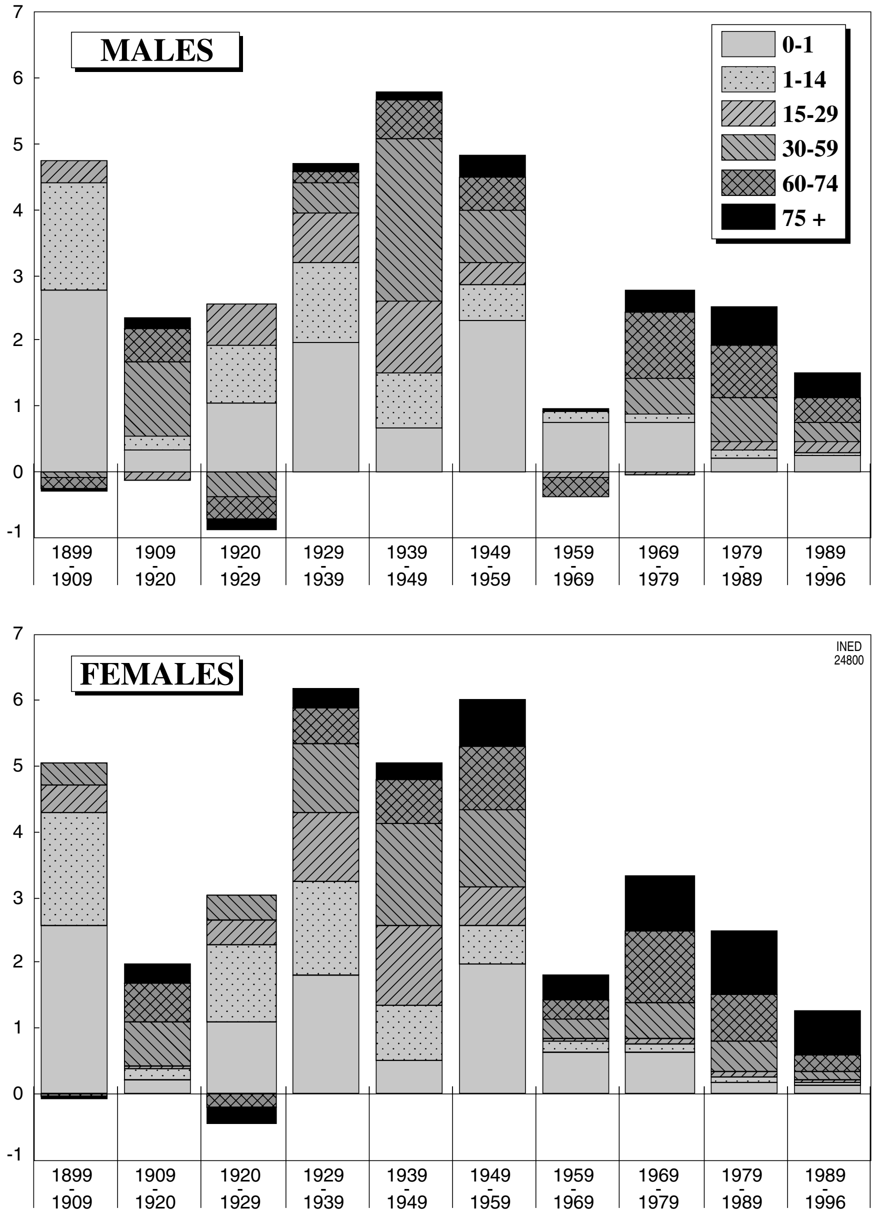


FIGURE 16–1 Contributions of changes in mortality for different age groups to life-expectancy gains registered in France for 10-year periods since 1899.

not taken over since the 1970s: the massive fall in mortality at older ages (after 60 and even after 75).

II. TEMPO EFFECT AND INTENSITY EFFECT ON THE DIFFERENCE BETWEEN TWO TOTAL FERTILITY RATES

Turning now to fertility, if we compare two total fertility rates (TFR^1 and TFR^2), the question generally raised is of an entirely different nature. As a rule, demographers want to know how much of the difference is due to a tempo change and how much is due to an actual change in the intensity of fertility in the cohorts involved.

Let I be the contribution of the change in intensity and C that of the tempo change. Ryder (1960, 1980) proposes the following formula for measuring them (Santini, 1992):

$$TFR_{t2} - TFR_{t1} = \left[\frac{\alpha^1 + \alpha^2}{2} \cdot (TFRC_{t2} - TFRC_{t1})\right] + \left[\frac{TFRC_{t1} + TFRC_{t2}}{2} \cdot (\alpha^2 - \alpha^1)\right] = I + C$$

where α_x is the fertility timing, equal to:

$$\alpha_x = \frac{f_{x,t}}{\sum_{x=15}^{50} f_{x,t-x}}$$

and

$$TFRC_t = \frac{TFR_t}{\sum_{x=15}^{50} \alpha_x},$$

where, generally:

$$\alpha = \sum_{x=15}^{50} \alpha_x \neq 1$$

In Italy, for example, the TFR fell from 2.337 in 1952 to 1.187 in 1994, a difference of –1.150 children. To decompose this difference into the intensity-change effect and the tempo-change effect, we must start by calculating the α values. For this, we divide each age-specific fertility rate f_x by the completed fertility of the cohort that has reached its mean age at childbearing in the year examined. For 1952, we divide the rates by 2.414, the average number of children per female of the 1923 birth cohort; for 1994, we divide by 1.593, the average number of children per female of the 1963 cohort (Table 16–3).

Applying the formula given above, the fall of 1.15 children per female in the Italian period fertility rate is decomposed as follows:

$$1.187 - 2.337 = \left[\frac{0.745 + 0.968}{2} \cdot \left(\frac{1.187}{0.745} - \frac{2.337}{0.968}\right)\right] + \left[\frac{\frac{2.337}{0.968} + \frac{1.187}{0.745}}{2} \cdot (0.745 - 0.968)\right] = (-0.703) + (-0.447)$$

Of the total reduction, 61% (–0.70 children) is therefore due to the fall in intensity and 39% (–0.45 children) is due to an intergenerational shift in birth timing.

If the necessary historic data are available, we can thus track the effects of intensity and tempo on the change in the total fertility rate. To continue with the example of Italian data for 1952 to 1994, we can, for example, distinguish two periods on either side of the year 1967, already used as an example earlier (Chapter 14): The first (1952–1967) corresponds to a rise in fertility due—as we have already suggested—to an earlier completion of fertility by the generations involved; by contrast, the second period (1967–1994) reflects a gradual postponement of the timing (Table 16–4).

TABLE 16–3 Selection of Data Needed to Distinguish Between the Effects of Intensity Change and Tempo Change on the Difference Between 1952 and 1994 Italian TFR

Age	Age-specific fertility rate f_x		Tempo indicator α_x	
	1952	1994	1952	1994
15	0.002	0.000	0.001	0.000
16	0.005	0.003	0.002	0.002
17	0.013	0.006	0.005	0.004
18	0.025	0.009	0.010	0.006
19	0.041	0.015	0.017	0.009
20	0.061	0.020	0.025	0.013
. . .	. . .	. . .	. . .	. . .
25	0.141	0.066	0.058	0.041
26	0.144	0.075	0.059	0.047
27	0.143	0.084	0.059	0.053
28	0.140	0.088	0.058	0.055
29	0.134	0.090	0.056	0.056
. . .	. . .	. . .	. . .	. . .
40	0.044	0.012	0.018	0.008
41	0.036	0.008	0.015	0.005
42	0.026	0.005	0.011	0.003
43	0.019	0.003	0.008	0.002
44	0.011	0.002	0.005	0.001
. . .	. . .	. . .	. . .	. . .
Total	2.337	1.187	0.968	0.745

TFR, total fertility rate.

TABLE 16–4 Effects of Changes in Intensity and Tempo of Cohort Fertility on Period Fertility in Italy, 1952–1967 and 1967–1994

	1952–1967	1967–1994	1952–1994
Initial TFR	2.34	2.53	2.34
Final TFR	2.53	1.19	1.19
Difference	0.20	–1.35	–1.15
Intensity effect	–0.20	–0.59	–0.70
Tempo effect	0.40	–0.75	–0.45

TFR, total fertility rate.

Figure 16–2 shows that the rise in period fertility between 1952 and 1967 is entirely due to the lead taken by successive cohorts in completing their fertility. This phenomenon even offsets the mild decline in intensity of cohort fertility. Whereas the timing lead should have single-handedly driven up the TFR by 0.4 children per between 1952 and 1967, the loss of intensity in the cohort fertility halved that gain to 0.2 children. By contrast, while the acceleration of this decline in cohort fertility made a substantial contribution of –0.6 children to the fall in total fertility between 1967 and 1994, this effect is, in itself, weaker than the effect of the

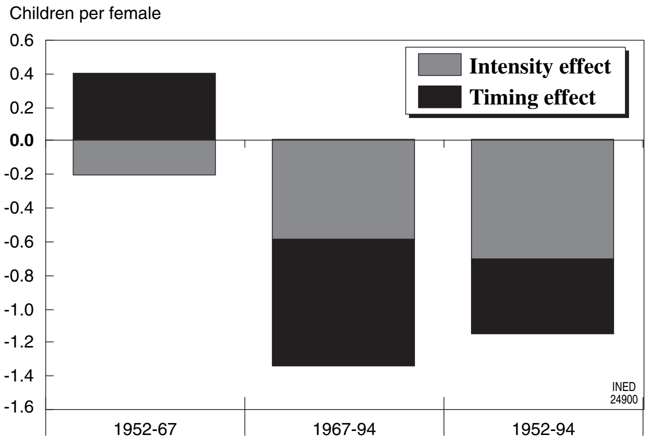


FIGURE 16–2 Effects of changes in intensity and tempo of cohort fertility on period fertility in Italy, 1952–1967 and 1967–1994.

delay by recent cohorts in completing their fertility (–0.8 children).

In fact, this attempt to interpret the difference observed between two total fertility rates is a first step toward the more general approach—known as the *translation* approach—aimed at interpreting period indicators in terms of cohort and vice versa, a topic discussed in the next chapter.

References

Andreev Evgueni, 1982. Method kompoenent v analize prichin smerti (components methods applied to the analysis of life expectancy), *Vestnik Statistiki*, no. 9, p. 42–47.

Arriaga Eduardo A., 1994. Measuring and explaining the change in life expectancies, *Demography*, February, no. 1, p. 83–96.

El Makrini Jamal, 1998. *Les méthodes de décomposition des tendances de la mortalité. Application à la Russie 1970–1994*. Louvain-la-Neuve, Institut de démographie, 76 p. (dissertation for Diplôme d'études avancées in demography).

Kortchack-Tchepurkovski Youri A., 1968. Vliyaniye smertnosti v raznih vozrastah na unelicheniye srednei prodoljitelnosti jizni (Influence of mortality at different ages on the increase in life expectancy), *in*: Iautcheniye vosproiaodstva naseleniya (Studies on population reproduction), p. 135–155. Moscow, Nauka.

Pollard John H., 1982. The expectation of life and its relationship to mortality, *Journal of the Institute of Actuaries*, vol. 109(442), Part 2, p. 225–240.

Pollard John H., 1988. Causes de décès et espérance de vie: quelques comparaisons internationales, *in*: Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches,* p. 291–313. Paris, INED, PUF (Travaux et documents, Cahier no. 119).

Pressat Roland, 1985. Contribution des écarts de mortalité par âge à la différence des vies moyennes, *Population*, vol. 40(4–5), p. 766–770.

Ryder Norman B., 1960. The structure and tempo of current fertility, *in*: NBER, *Demographic and economic change in developed countries*, p. 117–136. Princeton, Princeton University Press, 536 p.

Ryder Norman B., 1980. Components of temporal variations in American fertility, *in*: Robert W. Hiorns (ed.), *Demographic patterns in developed societies*, p. 15–54.—London, Taylor and Francis, 208 pp.

Santini Antonio, 1992. *Analisi demografica: fondamenti e metodi*. Scandicci (Florence), La Nuova Italia, 422 p.

Valkovics Emil, 1991. *An attempt of decomposition of the differences between the life expectancies at the age x on the basis of Belgian and Hungarian abridged life tables by causes of death*. Brussels, C.B.G.S. Werkdocument no. 71, 47 p.

CHAPTER

17

Demographic Translation

From Period to Cohort Perspective and Back

NICO KEILMAN

Department of Economics, University of Oslo, Oslo, Norway

I. THE NEED FOR BOTH COHORT AND PERIOD ANALYSIS

In his account of fertility levels and trends in England and Wales since the 1930s, Hobcraft (1996) noted that the mean age at birth had fallen rapidly after the Second World War. In 1972, women were 26.2 years old on average when they gave birth—3 years younger than in 1945 to 1946 (29.2 years). The decrease corresponds to more than 1 year per decade. This change in the timing of fertility had an upward effect on period total fertility rates (TFRs). Hobcraft argues that if successive cohorts of women have their children at progressively lower ages, such that the mean age falls every year by one tenth of a year, births that would have occurred during a period of 10 years without changes in the timing, are now squeezed into a period that is 1 year shorter. This pushes period fertility upward, and the period TFR will be inflated by 10%, compared to the TFR that would have occurred *without* changes in the timing.

Hobcraft's example demonstrates the point repeatedly stressed by Norman Ryder (1956, 1964, 1980): Even with constant completed cohort fertility, period fertility levels may vary. They are inflated in years in which childbearing is accelerated, and deflated when women postpone childbearing. Figure 14–11 and Figure 14–12 in Chapter 14 illustrate this point further: The increase in the period level of fertility in Italy in the 1960s was caused by a change in the tempo of childbearing among Italian women born in the 1930s. These women did *not* have more children than women born earlier, but their children were simply born at progressively younger ages. As a result, the period level of fertility was artificially high. This example shows that period fertility cannot be used as a reliable indicator for the level of cohort fertility: period fertility may be *distorted* in times of tempo changes in cohort fertility. At the same time, cohort fertility cannot be fully understood without studying periods.

The detailed interplay between period and cohort fertility, both its quantum (level) and tempo (timing) aspects, can be formalized mathematically. The resulting expressions constitute the core of what has become known as the *theory of demographic translation*, a term first used by Norman Ryder. The starting point is a table with age-specific birth rates for many calendar years. Since the period quantum (TFR) and the cohort quantum (completed cohort fertility [CFR]) are obtained on the basis of the same age-specific rates, but by summation in different directions (vertically for the TFR, diagonally for the CFR), there must be a relationship between the two quantum measures. In certain cases, when fertility changes show strong regularities (e.g., the TFR falls linearly, while the age pattern is constant), the resulting relationships are very simple mathematic expressions. The purpose of deriving such expressions is to gain insight into the degree of *translational distortion*, in other words to predict the quantum and tempo of cohort fertility, given trends in period fertility, and the other way round.

This chapter will give a brief overview of the theory of demographic translation. The theory can be applied to fertility and other demographic processes, such as first marriage and divorce. However, two conditions have to be fulfilled. First, the quantum and tempo indicators must develop sufficiently smoothly over a long period. Second, the sum of age-specific (or duration-specific) rates in either period or cohort dimension, or a transformation of this sum, must have a clear demographic interpretation. This must also be the case for the moments of the age schedule.

II. EARLY EXPRESSIONS BY RYDER FOR THE CASE OF AGE-SPECIFIC FERTILITY

Translation formulae can be used in two directions. From a theoretical point of view, one could be interested in expressions for the development of period quantum and tempo indicators, given certain trends in cohort indicators. Such expressions give insight in possible cohort mechanisms behind observed period developments, since period developments are considered a function of cohort developments. However, in practice, data for period developments are easier to obtain than those for cohort trends. Therefore, it is also useful to take the opposite point of view, and analyze cohort developments based on period developments.

1. From Cohort to Period

The starting point is a table with age-specific fertility rates for 1-year age groups and a number of calendar years. When we sum the rates over childbearing ages for calendar year t, the result is the TFR in that year, written as $TFR[t]$. It expresses the period quantum of fertility; see also Chapter 14. The cohort quantum is expressed by means of the CFR (Chapter 12), which is obtained by summing the rates diagonally. For example, for women born in year g, the $CFR[g]$ consists of the sum of rates for calendar year g at age 0, year $g+1$ at age 1, . . . year $g+48$ at age 48, and finally the rate in year $g+49$ at age 49. (Obviously, the rates below age 13 or 14 are zero.) Figure 14–11 in Chapter 14 plots the TFR for Italy for the years 1930 to 1994, and the CFR for the generations born in 1890 to 1962.

The TFR and the CFR express the *quantum* of fertility: To how many children does an average woman give birth in a period or a cohort perspective? The *tempo* aspect of fertility is expressed by the mean age at childbearing, which summarizes, in one number, the age at which women on average get their children. The period mean age for the year t is written here as $\bar{x}[t]$. Chapter 14 explains how it is computed based on a series of age-specific fertility rates for year t. The cohort mean age for a certain generation g, is written here as $\mu[g]$; it is computed analogously on the basis of a series of rates for generation g. Figure 14–12 in Chapter 14 plots the period and the cohort mean ages for Italian women, for the years 1953 to 1993 and the birth cohorts for 1900 1960. The mean age is one indicator that characterizes the age pattern of fertility. A second indicator is the variance, which tells us how wide the age pattern is for a certain birth cohort, or a given calendar year: A large variance implies that many women had their children at a much younger or much older age than the mean age.

Ryder derived mathematic expressions that predict the period quantum for a certain year as a function of cohort indicators for various cohorts: the CFR, the mean age, the variance, and other cohort indicators. Another expression gives the period mean age as a function of cohort indicators. The appendix to this chapter contains these expressions, together with their derivations. Under certain conditions, for instance, when the CFR is the same for all women born in successive generations, or when it varies linearly with time, the expressions become very simple. Here I shall study three such simplified cases.

First, I assume that both the cohort quantum and the age pattern of cohort fertility are constant, while only the mean age changes linearly. This would be the case when women born in successive generations on average would have the same number of children, whereas the curve of age-specific birth rates shifts progressively toward higher (or lower) ages. The second case is one with constant cohort tempo and a linear change in the cohort quantum. In this case, all rates grow or diminish from one cohort to the next with the same relative amount. The third case combines linear changes in both cohort quantum and mean age. Although these cases are clearly unrealistic, they provide nonetheless insight into the basic mechanisms of translation.

a. Constant Cohort Quantum and Linear Change in Cohort Mean Age

Constant cohort quantum implies that $CFR[g] = CFR$ for each cohort g. When the cohort mean age $\mu[g]$ follows a straight line, its increase or decrease is equal to $\Delta\mu$ years of age per birth year.[1] Finally I assume that other indicators of the fertility age schedule are constant. In that case we find that the period quantum TFR is constant, and that it equals (expression [A3] in appendix)

$$TFR = CFR \times (1 - \Delta\mu) \quad \text{(Eq. 1)}$$

When the mean age falls by one tenth of a year from one cohort to the next, the TFR is 10% higher than the CFR, and the translational distortion is 10%. In general, when childbearing is progressively concentrated in younger ages, the period TFR is inflated, other things being equal. However, the translational distortion in the mean age goes in the opposite direction (expression [A4] in appendix):

$$\bar{x}[t] = \frac{\mu[t]}{1-\Delta\mu} \quad \text{(Eq. 2)}$$

where $\mu[t]$ is the mean age for the cohort born in year t. When the cohort mean age falls by one tenth of a year from one birth cohort to the next, the period mean age is 10% *lower* than the cohort mean age, and *falls steeper*, namely by 0.1/0.9 = 0.11 years annually.

b. Constant Cohort Mean Age and Linear Change in Cohort Quantum

With constant cohort mean age ($\mu[g] = \mu$) and a linear change in the cohort quantum equal to ΔCFR per birth year, one obtains (see expression [A6] in appendix)

$$TFR[t] = CFR[t] - \mu \times \Delta CFR \quad \text{(Eq. 3)}$$

Thus, the TFR in year t equals the CFR for the cohort born in year t, minus the slope in the CFR times the mean age. In other words, the TFR changes linearly as well. With a mean age of 30 years and a CFR that falls by 0.05 children per woman each generation, the TFR is higher by 0.15 children, compared with the CFR. Because the CFR falls by ΔCFR = 0.05 child per generation, the decrease is $\mu \times \Delta$CFR = 0.15 children over μ = 30 cohorts. In other words, the TFR in year t equals the CFR for cohort t – 30. More generally,

$$TFR[t] = CFR[t - \mu] \quad \text{(Eq. 4)}$$

as expected. This simple relationship justifies common practice among demographers to shift the CFR curve between 25 and 30 years, when it is plotted in one graph together with the period TFR. See, for example, Figure 14–11 in Chapter 14, where the shift is 29 years.

[1] I use square brackets, i.e., [and], to denote the argument in a function. Parentheses, i.e., (and) indicate grouping of terms in an expression. For reasons of readability, I use the symbol Δ here to indicate annual change, both in terms of the mean age and the quantum. In the appendix, I have used a prime instead, indicating derivation with respect to time.

c. Linear Changes in Cohort Quantum and Cohort Mean Age

When both the CFR and the cohort mean age change linearly, one obtains (expression [A8]):

$$TFR[t] = CFR[t] \times (1 - \Delta\mu) - \mu[t] \times \Delta CFR \quad \text{(Eq. 5a)}$$

or the equivalent expression

$$TFR[t] = CFR[\gamma] - \Delta\mu \times CFR[t], \quad \text{with } \gamma = \gamma[t] = t - \mu[t] \quad \text{(Eq. 5b)}$$

Thus an increasing CFR and a fall in the mean age imply, according to expression (5a), an additional inflation of the TFR by an amount equal to $\Delta\mu \times CFR[t]$, compared with the situation in which only the CFR increases (expression [3]). This is close to the situation in England and Wales during the years of the baby boom: the CFR of cohorts born from 1905 to 1935 showed a near linear increase, from approximately 1.8 to 2.4 children per woman (Festy, 1979). As a result, the TFR from 1935 to 1965 rose also; see expression (4). However, the TFR was *extra* high (by an amount of $-\Delta\mu \times CFR[t] \approx (2/25) \times 2.1 = 0.17$) because the mean age for cohorts 1910 to 1935 fell by 2 years; see also the period indicators mentioned in the introduction. Denmark, Sweden, Finland, Norway, France, Netherlands, Switzerland, Canada, Australia, and the United States (white women) are other baby-boom countries in which an increase in CFR for women born in the first few decades of this century went together with a fall in the cohort mean age (see Festy, 1979).

2. From Period to Cohort

When cohort developments are considered a function of period developments, expressions similar to those given above can be derived. Assuming constant period quantum and a linear change in the period mean age gives the following result for the cohort quantum:

$$CFR[g] = CFR = TFR \times (1 + \Delta\bar{x}) \quad (1')$$

Thus a rise in the period mean age is reflected in a cohort quantum that is larger than the period quantum by a constant factor. Similarly, when the period mean age is constant, and the TFR varies linearly, one finds:

$$CFR[g] = TFR[g + \bar{x}] \quad (4')$$

The cohort quantum for women born in a certain year equals the period quantum $\bar{x}$ years later. Finally, when period quantum and period tempo are linear, the result is:

$$CFR[g] = TFR[g] \times (1 + \Delta\bar{x}) + \Delta TFR \times \bar{x}[g] \quad (5a')$$

or, equivalently

$$CFR[g] = TFR[\tau] + \Delta\bar{x} \times TFR[g] \quad (5b')$$

with

$$\tau = \tau[g] = g + \bar{x}[g].$$

It should be noted that assumptions of a linear trend in quantum or tempo indicators are primarily used for mathematic convenience because higher order derivatives in the relevant expressions vanish. When trends are nonlinear, polynomials of increasing order may be used instead. Although the mathematics still remain tractable, such polynomials cannot describe actual trends accurately over a long period, since they tend to result in unrealistically large positive or negative values in the long run. In practice, two solutions are followed. First, one may fit a low-degree polynomial over successive rather short intervals (Calot, 1992)—in other words, the polynomial holds only locally. A second solution is to use a bounded nonlinear function, for instance a logistic or a periodic curve (see De Beer, 1982; Foster, 1990).

III. EXPRESSIONS FOR NONREPEATABLE EVENTS

The expressions in the prior sections apply to age-specific fertility. They are based on the fact that when age-specific rates, either for a given calendar year or a given cohort, are summed over all ages, the sum reflects the quantum of the process. This is typically the case for repeatable events, such as childbearing regardless of parity. The rates are additive because a woman who gives birth to a child remains at the risk for a new birth (except for a short period immediately after delivery): The denominator of the rate is not affected by the event.

In contrast, we have *nonrepeatable* events such as childbearing broken down by birth order, or first marriage, or emigration. Events of this type are usually characterized by means of occurrence-exposure rates (o-e rates, *taux de première catégorie*).[2] Whelpton (1946) gives an early contribution for the case of age- and parity-specific fertility. An o-e rate expresses the risk of the event in question, for instance births of a certain order relative to all women of the corresponding parity, or first marriages relative to the number of never-married persons. When the o-e rates for such an event are age-specific, the sum of the rates over all ages does not reflect the quantum of the process in question.[3] For instance, whereas the quantum of first birth usually takes on values of between 80% and 95% (implying that 5% to 20% of the women remain childless), the rate sum is typically between 1.5 and 2.5.

Nevertheless, for some nonrepeatable events the translational distortion *can* be analyzed along the lines described, because the rate sum for these events, after an appropriate transformation, can be interpreted as the quantum of the process in question. As an example, take the case of births of order one. Consider the age-specific o-e rate for first births, and write the sum of these rates over all fertile ages for cohort g as $B[g]$. A traditional life table calculation shows that the cohort quantum of the process, expressed as the proportion of women in that cohort who at the end of the reproductive period ever had experienced a first birth, can be written as[4]

$$Q_c[g] = 1 - \exp(-B[g]) \quad \text{(Eq. 6)}$$

A similar expression holds for the period quantum of the process. Hence, the translation expressions given in the previous section can be applied to nonrepeatable events if one performs the exponential transformation of the predicted rate sum. For instance, in case the period rate sum of o-e rates for first births is constant, while the first moment follows a straight line, expression (1′) may be used to predict the cohort rate sum. Next, expression (6) predicts the cohort proportion of women who ever gave birth to a first child, see for instance Keilman (1994), Keilman and Van Imhoff (1995), and—for time-continuous expressions—Calot (1992). Empirical illustrations will be given in section IV. A number of points should be noted.

1. Because of the mathematics of the life table, the mean age of the process does not coincide with the first moment of the age schedule of o-e rates (Keilman, 1994). Translation formulas for the mean age or other tempo indicators are not known—only quantum expressions have been derived for nonrepeatable events. However, it will often be reasonable to assume that the slope of the mean age may be approximated by the slope of the first moment.

[2] Sometimes, incidence rates (epidemiology), also called frequencies (Coale), or *taux de deuxième catégorie* are used; see Chapter 8. Incidence rates are additive, but they exaggerate period distortions. See Chapter 14 and section IV this chapter.

[3] In certain cases the rate sum is a reasonable approximation for the quantum, see below.

[4] Expression (6) assumes piecewise constant intensities.

2. Although the procedure sketched in this section may be applied to such nonrepeatable events as birth of the first child, first marriage, emigration, and many others, some events cannot be analyzed this way. Examples are births of order two or higher by age of the mother and remarriage of divorced persons by age. The risk population of such processes may decrease and increase (for instance caused by births of the previous order or divorce). Such an increase is impossible in the case of first births or first marriage. As a consequence, the multistate life table that traces the fertility or nuptiality history of the real or synthetic cohort over its life course results in intractable matrix expressions for the quantum of these events. Another example is age-specific mortality. Because everyone dies, the quantum for mortality is 100%, and the interest is solely in the tempo aspects. As noted in the previous point, expressions for tempo indicators are unknown.
3. For low-intensity processes, the rate sum in expression (6) is small. In such cases, the rate sum itself approximates the quantum reasonably well, and all the expressions derived in the prior sections apply. Rate sums up to 0.2 are up to 10% higher than the corresponding quantum values. Examples of low-intensity processes are long distance migration, out-migration from large areas, and divorce in Mediterranean countries (broken down by marriage duration).

IV. THE BONGAARTS—FEENEY METHOD FOR TEMPO ADJUSTMENT OF PERIOD FERTILITY

Bongaarts and Feeney (1998) have proposed a method that corrects age- and parity-specific period fertility for distortions caused by tempo changes (see also Bongaarts, 1999). The purpose of the method is to obtain a tempo-free TFR (i.e., a TFR that would have been observed in year t if the age pattern of fertility had been the same as that in year $t - 1$). The starting point is the birth-order—specific TFR for year t defined as $TFR_p[t] = \Sigma f_p[t,x]$, where the fertility rate $f_p[t,x]$ expresses the number of births of order p by mothers aged x in year t, relative to the number of women aged x regardless of parity. The sum is taken over all fertile ages x. Assuming a constant shape of the age schedule of fertility (i.e., women of all ages defer or advance their births to the same extent), an adjusted TFR is computed as $TFR_p'[t] = TFR_p[t]/(1 - \Delta M_p[t])$, where $\Delta M_p[t]$ is the annual change in the mean age at childbearing for parity p. Summing the results over different birth orders gives the overall tempo-free total fertility $TFR'[t] = \Sigma TFR_p'[t]$.

Note the differences with Ryder's approach. First, Ryder did not include birth order. Second, he saw the tempo distortion in period fertility as caused by changes in cohort tempo. In contrast, Bongaarts and Feeney (B&F) (1998) assume that all changes are period driven. They do not attempt to predict cohort fertility because they assume that period-by-period changes are independent of age and cohort. In spite of these differences, for small changes in the mean age, the B&F approach (given birth order) gives the same result as Ryder's expression (1′),[5] although the interpretations of the two results are entirely different.

The B&F adjustment procedure is attractive because it is based on period data only. However, it has two major weaknesses (Van Imhoff, 2000; Van Imhoff and Keilman, 2000; Kohler and Philipov, 1999; Lesthaeghe and Willems, 1999). First, the method is based on fertility rates unsuitable for the purpose of tempo adjustment. The incidence rates used by B&F express the number of births of order p by mothers aged x in year t, relative to the number of women aged x *regardless of parity*.[6] The use of such rates in a period perspective introduces *extra* tempo distortions, compared to o-e rates. Whelpton noted this in 1946. When age-specific incidence rates for first births (for example) are summed for a given year, one erroneously assumes that the share of childless women at the end of one age interval is equal to that share at the start of the next interval. This is not necessarily the case because the age intervals refer to different cohorts. The stronger the tempo changes between cohorts, the more the shares for subsequent ages differ, and the stronger the effect of tempo distortions is exaggerated (Chapter 14, Figure 14–17). Quantum measures based on o-e rates of the type used in the prior section do not display this kind of bias. Therefore, Smallwood (1999; p. 40) concluded about the B&F adjustment that it ". . . provides a way of making a synthetic measure even more synthetic." The second problem is that the constant shape assumption underlying the B&F method is not supported by the data for a number of European countries: The age schedule is not only shifted toward higher or lower ages, but its shape changes as well. Hence period-by-period changes are not independent of age

[5] The formulas are different, because Bongaarts and Feeney assume a constant shape of the age-specific fertility schedule, whereas Ryder assumes constant period moments of orders two and higher. These assumptions are not equivalent (see the appendix, and, for a general treatment, Van Imhoff 2001, p. 62–64).

[6] The same disadvantage applies to the use of the method described by Pressat (1983), when applied, as Pressat proposes, to incidence rates.

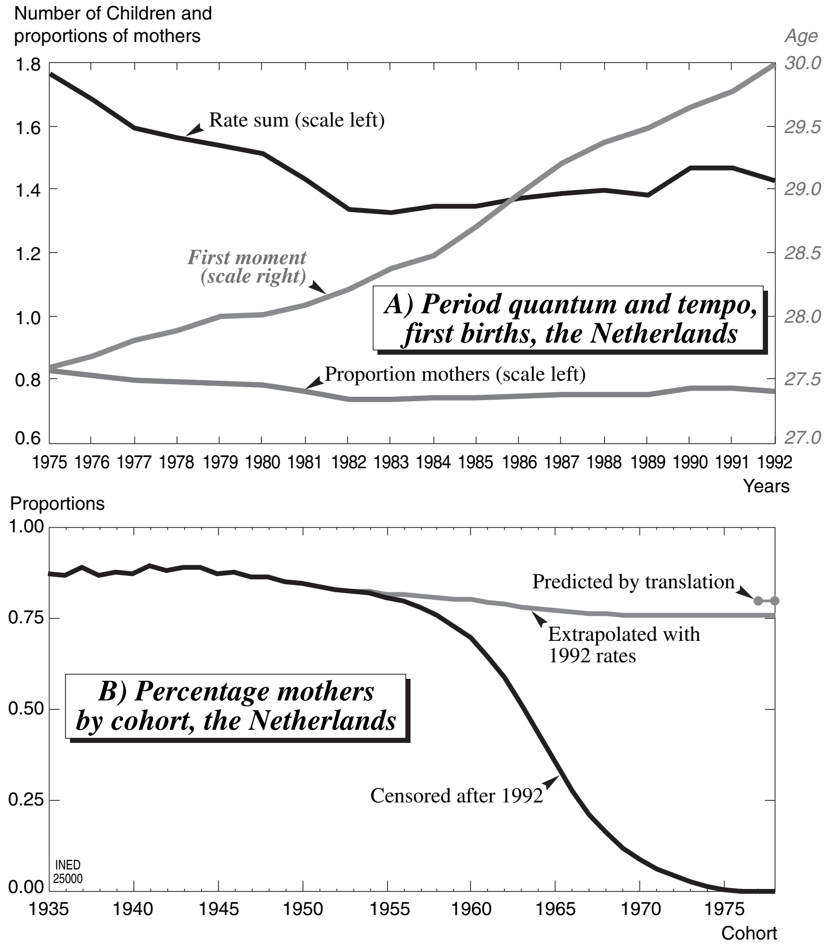


FIGURE 17–1 First illustration: first births in The Netherlands. (Source: BOSVELD WILLY, 1996. *The ageing of fertility in Europe: a comparative demographic-analytic study*. Amsterdam, Thesis Publishers, 285 p.)

and cohort, contrary to what B&F assume.[7] Thus period changes *are* dependent on cohort, and a pure *period quantum* is an untenable concept: cohort data are necessary to understand period effects fully (just as cohort effects cannot be fully understood without studying periods).

V. NUMERIC ILLUSTRATIONS

1. Childlessness in The Netherlands

Dutch population forecasts from the mid 1990s assumed a level of childlessness among women born in the 1970s and 1980s equal to 25% (De Beer, 1997). Indeed, the period level of first births suggested a proportion of women having at least one child equal to 75% (Figure 17–1A and 1B). For each year, this proportion was computed on the basis of the period sum of age-specific first birth rates of the occurrence-exposure type (Bosveld, 1996) and expression (6) was applied to that sum. The fall in the proportion of mothers between cohort 1945 and cohort 1955 in Figure 17–1B is real, but the further decline for later cohorts depends on the extrapolation method used. Compare the two lines *extrapolated with 1992 rates* and *censored after 1992*. Since 1981, the period sum of first birth rates (ages 15–39 years) in Figure 17–1A fluctuated around a level of 1.4 (with a slight tendency to increase), which led to an essentially constant proportion of mothers from that year onward. However, the period first moment (which reflects the mean age computed as $V_1[t]/V_0[t]$; see appendix) rose sharply during the years 1975 to 1992, with a slope equal to 0.14 years per year. Dutch women postponed the birth of their first child, and this depressed the period quantum.

[7] Kohler and Philipov (1999) extend the B&F approach to allow a changing variance of the fertility schedule. This reduces the empirical problem connected to the constant shape assumption, but not the methodologic one caused by the use of unsuitable fertility rates. Moreover, the adjusted TFR that they obtain is as unclear as that of Bongaarts and Feeney. In later work on fertility by age and parity, Kohler and Ortega (2001a, 2001b) introduce occurrence-exposure rates in an attempt to deal with the problem caused by the use of incidence rates, but the problem mentioned as point 2 in section III here is not solved.

Although the forecasters acknowledged this postponement effect, it was stronger than they believed in the mid 1990s. Indeed, the forecast published in 1999 assumes lower childlessness than previously: 20%, instead of 25% (De Beer, 1999). Translation theory could have predicted this in 1993 already. Assuming a constant period rate sum from 1981 onwards, and a linear first moment, expression (1′) predicts the cohort rate sum for the cohorts born in 1981 and later as $1.4 \times (1 + 0.14) = 1.60$, which implies a cohort quantum equal to 80% (expression [6]), and thus 20% childlessness.

2. Divorce in Norway

Figure 17–2A shows period proportions divorced in Norway. The underlying data are o-e rates for divorce, broken down by marriage cohort and marriage duration (Mamelund *et al.*, 1997). The year of divorce equals marriage year plus marriage duration. For each year, divorce rates were summed across marriage durations (up to 60 years in the original data); expression (6) resulted in period proportions divorced. The period proportion for year t is interpreted as the proportion of a fictitious marriage cohort that has experienced divorce by duration 60, given the divorce rates for the year t. Until the beginning of the 1970s, proportions divorced were very low in Norway, and the rate sum in Figure 17–2A is only slightly higher than the proportion, as expected.

The first moment fluctuated between 13 and 15 years since 1950, whereas the rate sum was more or less linear between 1970 and 1993. When we assume a constant first moment equal to 14 years, and a linear rate sum, expressions (4′) and (6) predict the proportion divorced for marriage cohort g as the period proportion in year $g + 14$. Figure 17–2B shows that the fit is remarkably good. Thus although no Norwegian

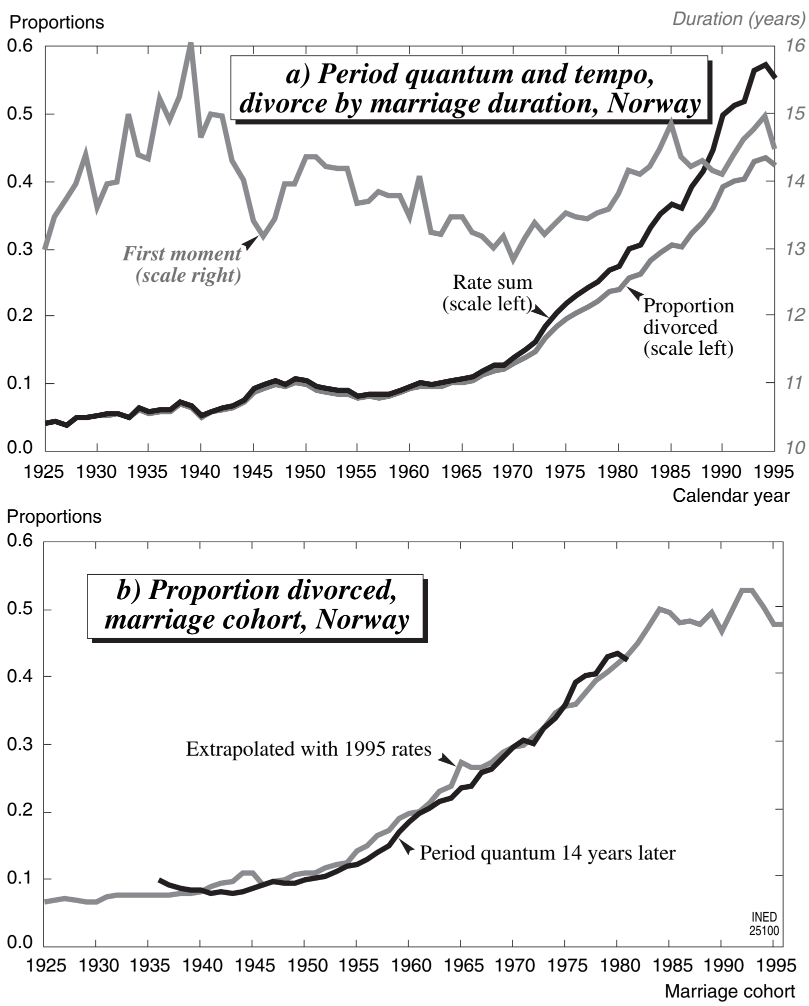


FIGURE 17–2 Second illustration: divorce in Norway. (Source: Svenn-Erik Mamelund, Helge Brunborg, and Turid Noack, 1997. *Skilsmisser i Norge 1886–1995 for kalenderàår og ekteskapskohorter*. Oslo, Statistics Norway, 115 p. [Rapporter 97/19].)

marriage cohort has ever had a proportion divorced that exceeds 25%, it is not unlikely that couples married in the last half of the 1960s will be the first ones to experience such a high share.

CONCLUSION

Translation theory provides expressions for the relationships between period and cohort quantum and tempo. The expressions can be applied to age-specific fertility, spanning several years and birth cohorts. With a slight modification, they can also be used to study other events, such as first marriage by age, divorce by marriage duration, emigration by age, or the birth of the first child by mother's age.

One of two perspectives can be adopted. First, the interest may be in period trends, and in tracing the effects of changing cohort behavior on those period trends. Applied to the case of age-specific fertility, the formulae show that when the CFR is constant, a fall in the cohort mean age at birth results in an inflated period TFR because women accelerate childbearing and births are *squeezed* into shorter periods. The TFR is pushed upward even more strongly when the CFR rises, in addition to the decrease in the mean age. In contrast, the TFR falls when women postpone childbearing (i.e., when the cohort mean age rises together with a constant or falling CFR). When the CFR and the cohort mean age move in the same direction, there are two opposite forces, and it is an empirical question whether the TFR is inflated or deflated. The second perspective is to take observed period trends as given, and to use translation theory to infer cohort developments. For example, the theory predicts, as expected, that the CFR is lower than the TFR by a 10% factor in case the period mean age falls by 1 year per decade and the TFR is constant.

From a mathematic point of view, the two perspectives are symmetric. However, there are strong empirical differences for age-specific fertility, and probably for other phenomena as well. The TFR shows much larger annual fluctuations than the CFR. Thus it is relatively easy to predict the CFR based on period quantum and tempo indicators: The model is more parsimonious than one that explains the TFR on the basis of cohort indicators. For instance, Calot (1992) analyzes age-specific fertility in France for the years 1900 to 1980 and cohorts in 1870 to 1950. He finds that period moments that develop linearly with time describe the CFR accurately. However, when he takes the opposite perspective and fits cohort moments to polynomials of increasing order when explaining the TFR, the accuracy he obtains is much less, even with fourth-degree polynomials (and the accuracy does not improve for higher order polynomials). Similarly, Foster (1990) concludes that period-based models provide a more parsimonious description of the observed patterns of age-specific fertility in eight countries in Europe and Northern America than cohort-based models do.

The primary use of translation theory is to improve the formal demographic analysis of historical developments.[8] Consider, for instance, the baby boom in Western countries (i.e., an increase in the TFR starting in the 1930s, and next a plateau in the 1950s and 1960s). Translation theory explains this trend by the fact that cohorts born in the first three or four decades of this century had increasingly larger numbers of children, and, at the same time, they accelerated childbearing. When quantum and tempo indicators develop according to a straight line over several decades, the resulting translation expressions are simple. This is seldom the case for period indicators. This makes translation theory less useful for predicting the behavior of cohorts, which the theory sees as a function of trends in period indicators.

References

Bongaarts John, 1999. The fertility impact of changes in the timing of childbearing in the developing world. *Population Studies*, vol. 53(3), p. 277–289.

Bongaarts John, and Feeney Griffith, 1998. On the quantum and tempo of fertility. *Population and Development Review*, vol. 24(2), p. 271–291.

Bosveld Willy, 1996. *The ageing of fertility in Europe: a comparative demographic-analytic study*. Amsterdam, Thesis Publishers, 285 p.

Calot Gérard, 1992. Relations entre indicateurs démographiques longitudinaux et transversaux. *Population*, vol. 47, p. 1189–1240.

De Beer Joop, 1982. Translation analysis as a device for extrapolating fertility rates. Voorburg, The Netherlands Central Bureau of Statistics (unpublished paper).

De Beer Joop, 1997. Bevolkingsprognose 1996. *Maandstatistiek van de Bevolking*, vol. 45(1), p. 6–12.

De Beer Joop, 1999. Bevolkingsprognose 1998–2050. *Maandstatistiek van de Bevolking*, vol. 47(1), p. 8–19.

Festy Patrick, 1979. *La fécondité des pays occidentaux de 1870 a 1970*. Paris, INED, Presses Universitaires de France, 399 p. (Travaux et Documents no. 85).

Foster Andrew, 1990. Cohort analysis and demographic translation. *Population Studies* vol. 44(2), p. 287–315.

Hobcraft John, 1996. Fertility in England and Wales: A fifty-year perspective. *Population Studies*, vol. 50(3), p. 485–524.

Keilman Nico, 1994. Translation formulae for non-repeatable events. *Population Studies*, vol. 48(2), p. 341–357.

Keilman Nico, and van Imhoff Evert, 1995. Cohort quantum as a function of time-dependent period quantum for non-repeatable events. *Population Studies*, vol. 49(2), p. 347–352.

[8] See also Chapter 16, where essentially the same ideas are used to analyze the difference between the TFR at two points in time. One part of the difference can be attributed to changes in cohort quantum, and the remaining part to tempo changes.

Kohler Hans-Peter and Philipov Dimiter, 1999. *Variance effects and non-linearities in the Bongaarts-Feeney formula*. Rostock, Max Planck Institute for Demographic Research. (Working Paper 1999–001).

Kohler Hans-Peter and Ortega José Antonio, 2001a. *Period parity progression measures with continued fertility postponement: a new look at the implications of delayed childbearing for cohort fertility*. Rostock, Max Planck Institute for Demographic Research. (Working Paper 2001–001). Available at http://www.demogr.mpg.de.

Kohler, Hans-Peter and Ortega José Antonio, 2001b. *Tempo-adjusted period parity progression ratios: assessing the implications of delayed childbearing for cohort fertility in Sweden, the Netherlands, and Spain*.—Rostock, Max Planck Institute for Demographic Research. (Working Paper 2001–035). Available at http://www.demogr.mpg.de.

Lesthaeghe Ron and Willems Paul, 1999. Is low fertility a temporary phenomenon in the European Union? *Population and Development Review*, vol. 25(2), p. 211–228.

Mamelund Svenn-Erik, Brunborg Helge, and Noack Turid, 1997. *Skilsmisser i Norge 1886–1995 for kalenderà6r og ekteskapskohorter*. Oslo, Statistics Norway, 115 p. (Rapporter 97/19).

Pressat Roland, 1969. *L'Analyse Démographique*. Paris, Presses Universitaires de France, 321 p. (2nd ed.).

Ryder Norman B., 1956. La mesure des variations de la fécondité au cours du temps. *Population*, vol. 11(1), p. 29–46.

Ryder Norman B., 1964. The process of demographic translation. *Demography*, vol. 1(1), p. 74–82.

Ryder Norman B., 1980. Components of temporal variations in American fertility, *in*: Robert W. Hiorns (ed.), *Demographic patterns in developed societies*. London: Taylor & Francis, p. 15–54.

Smallwood Steven, 1999. A comparison of four tempo adjusted fertility measures applied to England and Wales data, 1940–1997. London: London School of Economics. Unpublished MSc dissertation.

un, 2000. *Below replacement fertility. Population Bulletin of the United Nations*, Special Issue 40/41 1999. New York: United Nations.

Van Imhoff Evert, 2001. On the impossibility of inferring cohort fertility measures from period fertility measures. *Demographic Research*, vol. 1, article 2. Available at www.demographic-research.org.

Van Imhoff Evert and Keilman Nico, 2000. On the quantum and tempo of fertility [Comment]. *Population and Development Review*, vol. 26(3), p. 549–553.

Whelpton Pascal, 1946. Is family size increasing? *Journal of the American Statistical Association*, vol. 41(236), p. 501–516.

Wunsch Guillaume J. and Termote Marc G., 1978. *Introduction to Demographic Analysis*. New York, Plenum Press, 274 p.

Yntema Luitzen, 1977. *Inleiding tot de demometrie*. Deventer, Van Loghum Slaterus, 281 p.

APPENDIX 17–1 MATHEMATIC DERIVATIONS

Norman Ryder (1956, 1964) gave the first formal derivations. Alternative treatments of some special cases are contained, among others, in Pressat (1983) and Wunsch and Termote (1978). I will use the short and comprehensive general analysis of Yntema (1977).

Denote by $m[t,x]$ the fertility rate in year t at age x. Denote further the k-th *period* moment of these rates by $V_k[t] = \Sigma x^k \cdot m[t,x]$. Similarly, define the k-th *cohort* moment of those rates by $W_k[g] = \Sigma x^k \cdot m[g+x,x]$. In both cases, the sum is taken over all fertile ages. $V_0[t]$ is the TFR, or the period sum of rates in year t. Also, $W_0[g]$ is the Completed Cohort Fertility for cohort g.[9] Taylor series approximation of $m[t+x,x]$ about t gives

$$m[t+x,x] = \sum_i \frac{x^i}{i!} m^{(i)}[t,x].$$

Differentiation applies to time. Inserting this in the expression for W_k leads to

$$W_k[g] = \sum_i \frac{1}{i!} V_{k+i}^{(i)}[g].^{10} \qquad \text{(Eq. A1)}$$

Similarly, we find for period moments the following expression

$$V_k[t] = \sum_i \frac{(-1)^i}{i!} W_{k+i}^{i}[t]. \qquad \text{(Eq. A2)}$$

Expressions (A1) and (A2) are the general translation equations, which express cohort moments in terms of period moments, and vice versa. I will investigate a few special cases on the basis of the assumptions that quantum ($k = 0$) and mean age ($k = 1$) indicators are independent, or linearly dependent of time.

1. From Cohort to Period

a. Constant Cohort Quantum and Linear Cohort Mean Age

Assume that the cohort quantum W_0 is independent of time, and that the cohort mean age $\mu[g] = W_1[g]/W_0$ is a linear function of time. Assume further that all second and higher order moments $W_k[g]$ are independent of time. Thus, for $i > 1$, the terms in (A2) vanish and we find

$$V_k[t] = W_k[t] - W'_{k+1}$$

A prime indicates the first derivative with respect to time. This expression implies that the TFR ($k = 0$) is independent of time, and that it equals

$$TFR[t] = TFR = CFR - CFR \cdot \mu' = CFF \cdot (1 - \mu') \qquad \text{(Eq. A3)}$$

where μ' represents the slope in the mean age. The period mean age $\bar{x}[t]$ appears to be linear:

[9] Note that these moments are not normalized, in the sense that $V_0[t]$ and $W_0[g]$ do not equal one. Normalized moments are obtained as $V_k[t]/V_0[t]$ and $W_k[g]/W_0[g]$, k = 0, 1, 2, . . . For instance, the cohort mean age $\mu[g]$ equals $W_1[g]/W_0[g]$.

[10] Here and elsewhere, I shall write $V_k[g]$ to indicate the k-th moment for the calendar year $t = g$, and similarly for the derivatives of these moments. Likewise, $W_k[t]$ indicates the k-th moment for the cohort born in year t.

$$\bar{x}[t] = \frac{V_1[t]}{V_0[t]} = \frac{\mu[t]\cdot CFR}{CFR\cdot(1-\mu')} = \frac{\mu[t]}{1-\mu'} \quad \text{(Eq. A4)}$$

In some analyses (i.e., the Bongaarts–Feeney method) it is assumed that the shape of the age-specific fertility curve remains the same, but that this curve is shifted progressively toward higher (increasing mean age) or lower (falling mean age) ages. Clearly, this implies that the quantum is constant. But, as Evert van Imhoff pointed out to me, it does *not* imply time-independent moments of order two and higher, as assumed in this section. For instance, constant shape means that the variance is constant. Since the variance of any distribution equals the second (normalized) moment minus the square of the mean, a linear mean age must result in a quadratic second moment. Similar arguments hold for higher order moments. I will now treat the case of a linear mean age and a constant shape of the age-specific fertility curve.

Assuming linear mean age and constant shape, and writing the variance of the age schedule as σ^2, the second moment $W_2[g]$ equals $CFR\cdot\{\sigma^2 + (\mu[g])^2\}$, so that its second derivative is $2CFR\cdot(\mu')^2$. Similarly, the third moment $W_3[g]$ equals $CFR\cdot\{\gamma\sigma^3 + 3\sigma^2\mu[g] + (\mu[g])^3\}$. The symbol γ represents the coefficient of skewness, which is independent of time by assumption. The third derivative of $W_3[g]$ equals $6CFR\cdot(\mu')^3$. In general, the k-th derivative of the moment $W_k[g]$ equals $k!CFR\cdot(\mu')^k$ (see Van Imhoff, 2001, p. 62–64) for a general proof. This means that for $k = 0$, expression (A1) leads to

$$\begin{aligned} TFR[t] &= TFR \\ &= CFR\left(1-\mu'+(\mu')^2-(\mu')^3+\ldots\right) \\ &= \frac{CFR}{1+\mu'},\quad |\mu'|<1. \end{aligned} \quad \text{(Eq. A5)}$$

When comparing expressions (A3) and (A5), we see that when the mean age increases linearly, both a constant moments assumption and a constant shape assumption imply a *TFR* smaller than the *CFR*, but by different amounts. For weakly increasing mean ages, for instance by less than 0.1 year per calendar year, the difference is small.

b. Linear Cohort Quantum and Constant Cohort Age Pattern

A constant age pattern for cohort fertility, and a cohort quantum which depends linearly of time with slope W_0', implies that $V_0[t] = W_0[t] - W_1'[t] = W_0[t] - \mu W_0'$. In other words,

$$\begin{aligned} TFR[t] = CFR[t] - \mu\cdot CFR' = CFR[t-\mu],\\ \text{or } TFR[t+\mu] = CFR[t]. \end{aligned} \quad \text{(A6)}$$

Furthermore, $V_1[t] = W_1[t] - W_2'[t]$, so that we find for the period mean age $\bar{x}[t] = V_1[t]/V_0[t]$ that

$$\bar{x}[t+\mu] = \mu\cdot\frac{CFR[t-\sigma^2/\mu]}{CFR[t]}. \quad \text{(Eq. A7)}$$

The period mean age equals the cohort mean age times an adjustment factor (*distortion factor*), which depends on two quantities: the development of the cohort quantum and the relationship between the mean age and the variance of the cohort fertility age schedule. Assume that the variance is zero: All women get their children at the mean age. In this extreme case, the distortion factor equals one and the period mean age is constant, equal to the cohort mean age. In practice, the variance is larger than zero, and the distortion factor is larger/smaller than one, when cohort fertility falls/increases. Thus under the assumptions stated, the period mean age in year t is *higher* than the cohort mean age for women born in year t when the *CFR falls*, and vice versa. When women spread their births over a longer period in the life course, the variance of the fertility age schedule increases, and the distortion becomes stronger.

c. Linear Cohort Quantum and Linear Cohort Mean Age

Assume that both the cohort quantum and mean age are linear, whereas all the other moments are independent of time. This implies that

$$TFR[t] = CFR[t]\cdot(1-\mu') - CFR'\cdot\mu[t]. \quad \text{(A8)}$$

As expected, expressions (A3) and (A6) are special cases of (A8). The expression for the mean age is a complicated formula, which does not provide much insight. It is left out here.

2. From Period to Cohort

Expressions (A3)–(A8) for period quantum and mean age are based on expression (A2), assuming that cohort indicators are given. In the reverse situation, when the interest is in cohort quantum and mean age based on known period developments, expression (A1) is the starting point. The following results are easily verified.

3. Constant Period Quantum and Linear Period Mean Age

$$CFR[g] = CFR = TFR\cdot(1+\bar{x}') \quad \text{(Eq. A9)}$$

$$\mu[g] = \frac{\bar{x}[g]}{1+\bar{x}'}. \quad \text{(Eq. A10)}$$

$\bar{x}[g]$ is the period mean age in the year $t = g$.

a. Linear Period Quantum and Constant Period Age Pattern

$$CFR[g] = TFR[g + \bar{x}] \qquad \text{(Eq. A11)}$$

$$\mu[g - \bar{x}] = \bar{x} \cdot \frac{TFR[g + s^2/\bar{x}]}{TFR[g]} \qquad \text{(Eq. A12)}$$

where s^2 represents the variance of the period age pattern.

b. Linear Period Quantum and Linear Period Mean Age

It is assumed that second- and higher order period moments are independent of time.

$$CFR[g] = TFR[g]\cdot(1 + \bar{x}') + TFR'\cdot\bar{x}[g]. \qquad \text{(A13)}$$

CHAPTER

18

Age-Period-Cohort Models In Demography

JOHN R. WILMOTH

Department of Demography, University of California, Berkeley, California

Demographic events such as birth, death, marriage, and migration mark life's most important transitions and, not surprisingly, are influenced by a multitude of factors. These influences are not constant across time or space, and thus demographic rates display considerable variation. A crucial task of demographic analysis is to identify the factors that account for this variability.

One particular approach to this quite general problem is to study the variation in demographic rates along three critical dimensions: age of the event, year (or period) of its occurrence, and cohort of the individuals involved. Within this framework, *cohort* may refer to an individual's year of birth or to the time of some other important transition in the life course (e.g., a marriage cohort in a study of divorce). Likewise, *age* may mean a person's chronologic age at the time of the event or the time elapsed since another important transition (e.g., duration of employment in a study of retirement).

The underlying assumption of age-period-cohort (APC) analysis is that all factors influencing a demographic event or its rate of occurrence can be grouped meaningfully into these three categories. Of course, there may be interactions or overlap between the three sets of factors, but it is assumed, implicitly, that these influences are less important than the main effects associated with each dimension. The typical analytic strategy of APC studies, therefore, is first to identify patterns of change in demographic rates that are associated with these three dimensions or sets of causal factors. Once this elaborate description is available, the analyst attempts to identify the specific influences (biologic processes, historic trends, etc.) that are responsible for the observed patterns in terms of age, period, and cohort.

Based on this description, the logic of APC analysis seems simple and relatively straightforward. There is, however, a fundamental problem with this strategy related to the fact that, mathematically,

$$\text{cohort} + \text{age} = \text{period}.$$

This relationship is illustrated by means of a Lexis diagram in Figure 18–1. In the view of some researchers, the exact mathematic connection between age, period, and cohort renders all forms of APC analysis meaningless, since changes in a demographic process along one of the three dimensions cannot be distinguished statistically from changes along the two remaining dimensions (Glenn, 1976; Goldstein, 1979; Rodgers, 1982). Other researchers have proposed a variety of solutions for overcoming this problem and claim that valid and useful results can still be derived within an APC framework (Fienberg and Mason, 1979; Clogg, 1982; Caselli and Capocaccia, 1989; Wilmoth, 1990). All researchers would agree, however, that the identification problem that plagues APC analysis is a fundamentally difficult and perplexing problem.

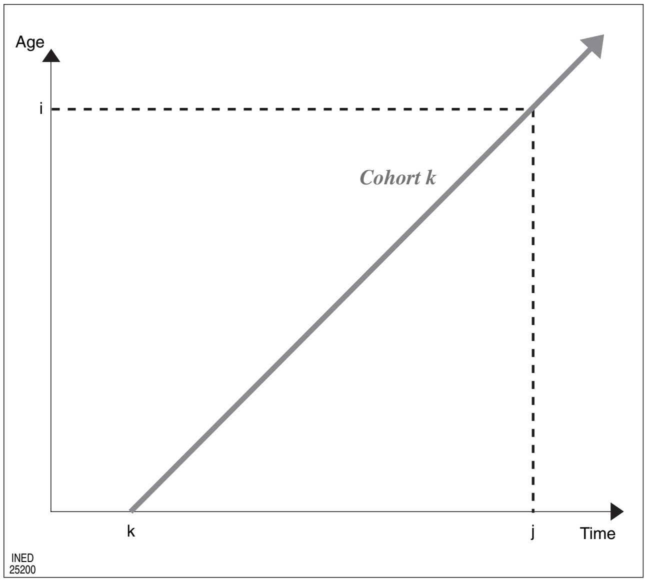


FIGURE 18–1 Lexis diagram showing the relationship between age, period, and cohort of a demographic event. The lifeline of individuals born at time *k* (cohort) is represented by the diagonal arrow. For such persons, a demographic event at age *i* occurs at time $j = k + i$ (period).

In this chapter, we first discuss the theoretical motivation of the age-period-cohort framework. We then provide a formal definition and discussion of the identification problem that affects APC modeling. Although we do not provide a complete literature review, we discuss strategies for estimating APC models that have been proposed in particular contexts. We also describe a small number of empirical applications of APC modeling that seem particularly informative. In conclusion, we offer some general remarks about the applicability of the APC framework.

I. THEORETICAL MOTIVATION

We begin our theoretical discussion by asking, "Why is the age-period-cohort framework useful?" The period of an event, for example, has no direct influence on its outcome but is merely a marker for other factors that may affect a demographic process. Periods may be associated with wars, epidemics, political and legal changes, famines, inventions, scientific discoveries, and changing fashions. It is these occurrences or situations, not the period itself, that account for the variation in demographic rates along the period dimension. Likewise, it is not cohort membership and chronologic age themselves that influence demographic rates, but rather the characteristics associated with those two dimensions.

It is worth considering, however, whether all three dimensions should be important in all situations. In the case of the classical demographic events (birth, death, marriage, and migration) it seems obvious that age, or duration, should be associated with sharp differences in outcomes. Due to biologic mechanisms, both the risk of death and the chance of giving birth vary enormously with age. Owing mostly to social and psychological mechanisms, the risk of divorce varies depending on the duration of marriage and the probability of migration changes over the life cycle. The age pattern of marriage is a complex function of both biologic maturation and social convention.

The immediacy of period-specific events makes their connection to demographic processes self-evident. The role of cohort-specific influences, on the other hand, is less obvious. In an influential article, Ryder (1965) laid out the theoretical justification for the use of cohorts in the study of social change. According to Ryder, fresh cohorts enable the process of social change, since young people are less constrained by history and more capable of adapting to or creating new modes of living. Furthermore, individuals experience certain critical events (e.g., infancy and childhood, education, military service, first employment) alongside peers from their particular cohort and the after-effects of these situations may mark them for life. The imprint of life-defining events may be biologic or social, but in either case we should expect cohorts to represent an important dimension of variation in subsequent demographic processes.

The direction and magnitude of the influence of cohort membership on demographic phenomena have been a matter of considerable speculation, although few firmly established generalizations have emerged. Perhaps the most well-known theory about the influence of cohort membership on demographic events is Easterlin's explanation of the American baby boom (Easterlin, 1961, 1978). According to this theory, smaller cohorts enjoy significant advantages as they enter the job market, which facilitates family formation and thus raises fertility levels. With regard to the baby boom, however, it is difficult to distinguish the effects of cohort size from the effects of an expanding postwar economy, both of which may have influenced life chances and family behavior around the 1950s.

In mortality studies, the discussion of the role of cohorts has focused on the after-effects of adverse events in early life on subsequent probabilities of death or survival (for a review, see Elo and Preston, 1992). Theoretically, it is possible that adverse events in early life could result in either higher or lower mortality in later life. On one hand, the survivors of a traumatic event or situation (e.g., famine, war, poverty) may be weakened by the experience and thus demonstrate unusually high levels of mortality later on. On the other hand, it is possible that the survivors in these situations may be an especially robust subset of the original cohort and thus could display unexpectedly low levels of mortality in subsequent years. These competing explanations are often referred to as processes of *debilitation* and *selection*, respectively. It is possible, of course, that both processes are operating at the same time. In these situations, it is their relative magnitudes that should determine whether an affected cohort experiences unusually high or low mortality in later years.

In summary, there is a sound theoretical basis for believing that age, period, and cohort may all be important dimensions of variation in demographic processes. It is this theoretical interest that has encouraged the development of statistical models incorporating all three components. We now turn our attention to such models.

II. AGE-PERIOD-COHORT MODELS

The standard age-period-cohort model has the following form:

$$f(r_{ijk}) = \mu + \alpha_i + \beta_j + \theta_k + \varepsilon_{ijk}$$

In this formulation, r_{ijk} is an observed demographic rate for some event that occurs at age i in year j for cohort k (thus, $k = j - i$); the function, $f(\cdot)$, is some transformation applied to the observed rates; the parameter, μ, establishes an overall level for $f(r_{ijk})$; the parameters, α_i, β_j, and θ_k, describe patterns of change in $f(r_{ijk})$ by age, period, and cohort, respectively; and the last term, ε_{ijk}, represents error (either in the specification of the model or in the original data) and random fluctuations.

Although the standard APC model is quite simple in form, its parameters are not easily estimated. The central difficulty is that there is no obvious means of identifying a unique set of parameter estimates that provide an optimal fit to the observed data. This identification problem affects not only the statistical estimation of the parameters but also their subsequent interpretation.

1. Identification Problem

The identification problem of the standard APC model is easily illustrated by considering the model in its estimated form:

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_k$$

Thus, the predicted value of the transformed rates, $\hat{y}_{ijk}$, equals the sum of the four estimated parameters, $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\theta}_k$.

The first task in identifying a unique solution for this model involves constraining the levels of the four parameters. A common approach is to require that

$$\sum_i \hat{\alpha}_i = \sum_j \hat{\beta}_j = \sum_k \hat{\theta}_k = 0$$

With this solution, $\hat{\mu}$ gives some average level of $f(r_{ijk})$, and the other three sets of parameters describe changes in the transformed rates with respect to the average. Alternatively, we could require that $\hat{\alpha}_i = \hat{\beta}_j = \hat{\theta}_k = 0$ for some specific choice of i, j, and k. In this case, $\hat{\mu}$ gives the predicted value of this particular $f(r_{ijk})$, and the other parameters describe changes with respect to this point of reference.[1]

Constraints of this sort are common in other statistical applications (analysis of variance, regression, etc.). Obviously, they involve an arbitrary choice about how to obtain a unique solution, but this choice affects only the level of the various parameters and thus does not fundamentally confuse their interpretation. APC models, however, present an additional, more perplexing identification problem involving not only the level of the various parameters but also their slope.

For example, suppose that $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\theta}_k$ were adjusted such that

$$\alpha_i^* = \hat{\alpha}_i + \lambda \cdot i$$
$$\beta_j^* = \hat{\beta}_j - \lambda \cdot j$$
$$\theta_k^* = \hat{\theta}_k + \lambda \cdot k,$$

where λ is some real number. Now, note that the predicted values in the model are the same for any choice of λ, since $k = j - i$:

$$\begin{aligned}\hat{y}_{ijk} &= \hat{\mu} + \alpha_i^* + \beta_j^* + \theta_k^* \\ &= \hat{\mu} + \hat{\alpha}_i + \lambda \cdot i + \hat{\beta}_j - \lambda \cdot j + \hat{\theta}_k + \lambda \cdot k \\ &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_k + \lambda \cdot (i - j + k) \\ &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_k.\end{aligned}$$

[1] Perhaps the best choice, for reasons of mathematic convenience, is to require that

$$\sum_i w_i \hat{\alpha}_i = \sum_j w_j \hat{\beta}_j = \sum_k w_k \hat{\theta}_k = 0,$$

where w_i, w_j, and w_k are weights corresponding to the number of observations for each particular age, period, and cohort. In a rectangular age-by-period array, for example, w_i and w_j would be constant (equaling the number of columns and rows, respectively) and thus would fall out of the equation. In this situation, only w_k would vary, equaling the number of elements in a diagonal of the matrix corresponding to an individual cohort.

Since different sets of age, period, and cohort parameters yield identical predicted values, there is no means of choosing between alternative solutions on the basis of goodness-of-fit. As with the earlier problem of uniquely identifying the level of these parameters, a unique solution demands an arbitrary choice. In this case, however, alternative solutions may have vastly different implications, since the speed and even direction of change in the age, period, and cohort components of the process are affected. In fact, by careful manipulation of λ, the linear trend in the three sets of parameters can be altered *ad libitum*.[2]

Typically, discussions of APC models have focused on the problem of uniquely identifying the linear trend in the age, period, and cohort parameters. Higher order components of these parameters (quadratic, cubic, etc.) and interactions between the three sets of factors have generally been ignored because they make identification of the model even more difficult (Fienberg and Mason, 1985; Clogg, 1982). However, second- and higher order components of the main effects are not uniquely identified if interaction terms are added to the standard APC model (Wilmoth, 1990).

2. Estimation Strategies

In this chapter, we do not attempt to provide a complete review of the literature on APC modeling. For that purpose, we refer the reader to excellent discussions by other authors (Hobcraft, Menken, and Preston, 1982; Mason and Fienberg, 1985). Instead of a review of the literature, we provide an overview of the main strategies that have been proposed for estimating and/or modifying the standard APC model in light of the identification problem discussed earlier. We divide these approaches into three groups: arbitrary assumptions, interaction terms and demographic translation, and direct measurement. See also the excellent discussion of these topics by Holford (1991).

[2] Note that the transformed parameters, α_i^*, β_j^*, and θ_k^*, as defined here, would not normally sum to zero. Alternatively, define them as follows:

$$\alpha_i^* = \hat{\alpha}_i + \lambda \cdot (i - \bar{i})$$
$$\beta_j^* = \hat{\beta}_j - \lambda \cdot (j - \bar{j})$$
$$\theta_k^* = \hat{\theta}_k + \lambda \cdot (k - \bar{k})$$

where $\bar{i}$, $\bar{j}$, and $\bar{k}$ are defined as the means of i, j, and k. It can be shown that the sums of these transformed parameters equal zero if the data matrix is rectangular. If the data array is not rectangular, a convenient solution is to use weights, as defined in Footnote 1, both for defining the parameter constraints and for computing the means of i, j, and k in the above formulas.

a. Arbitrary Assumptions

The first strategy for identifying an APC model is to make some arbitrary assumption about the linear trend in one of the three dimensions. Specific applications differ, however, depending on whether the assumption is viewed by the analyst(s) as an accurate reflection of an underlying reality or merely as a convenient strategy for estimating an arbitrary statistical model.

Unique estimates of the standard APC model (with three main effects, as described in prior sections) can be obtained by assuming that the parameters pertaining to some pair of cohorts (or ages or periods) are equal. For example, one could assume that $\theta_{k'} = \theta_{k'+1}$ for a specific cohort k'. This restriction permits estimation of the model, although at the expense of assuming that the cohort-specific effects for cohorts k' and $k' + 1$ are equal. Arguably, such an assumption is justified in specific cases based on prior knowledge.

In a widely cited paper, Fienberg and Mason (1979) illustrated the application of the standard APC model to the analysis of educational attainment in the United States. Here, the dependent variable consisted of proportions of individuals ages 20 and above who had, at some previous moment in their lives, advanced from one educational level to the next. These proportions were computed using decennial census data from 1940 to 1970. The authors noted correctly that educational attainment should under most circumstances be constant or slightly increasing over the adult life of a cohort (this expectation could be violated only by selective migration or differential mortality). In other words, age effects are of little importance in this model and can reasonably be set equal over some age range. Similarly, period effects can have little impact on the educational attainment of adults (most of whom have completed their schooling prior to the period of observation). Indeed, the authors justify the inclusion of period effects in the model based solely on concerns about changes in the recording of educational data in decennial censuses.

Due to the relative unimportance of these two sets of components (age and period) *a priori*, the choice by Fienberg and Mason to use educational attainment as an illustration of APC modeling is somewhat puzzling. Their resolution of the standard identification problem in this particular case was relatively straightforward and unproblematic, since the authors merely equated age effects for age groups 30 to 34 through 55 to 59 years.[3] This example illustrates the sensible use of

[3] Obviously, the model would have been fully identified if age effects for only two age groups had been set equal. Fienberg and Mason (1979: 53) argue, however, that there are "*a priori* grounds for interpreting age effects primarily at the tails of the age distribution" and thus prefer to equate the coefficients throughout the middle age range.

prior information to resolve the APC identification problem. It provides little guidance, however, about how to resolve the identification problem in situations where there is no strong basis for asserting the equality of at least one pair of effects *a priori*.

In another application of the APC model, Clogg (1982) examines rates of labor force participation in the United States. Here, there are strong reasons for expecting that the dependent variable is influenced by factors related to age, period, and cohort, but there is only a very weak basis for supposing, *a priori*, a particular relationship between some pair of effects. Nevertheless, the author constrains the last two cohort effects to differ by one or two percent (depending on the specific example) as a means of obtaining a unique set of parameter estimates. The substantive justification of this choice can fairly be termed *weak*, although arguably it has little impact on the conclusions of the study.

In general, even if prior knowledge supports the choice of a specific identifying restriction, it is important to recognize the sensitivity of the results to minor violations of the chosen assumption. If $\theta_{k'}$ and $\theta_{k'+1}$ are close but not equal, for example, the error that results from equating them will be slight for adjacent cohorts, but much larger for distant ones (Clogg, 1982). The importance of this imprecision will depend on the application, and the sensitivity of the results to any particular assumption can often be assessed quite easily. Nevertheless, the validity of results derived from APC models in which this last identifying constraint is, allegedly, a reflection of underlying reality remains quite dubious in most cases.

Alternatively, rather than asserting the reality of identifying constraints, certain parameter restrictions can be adopted as a matter of convenience merely to obtain a unique solution. For example, Pullum (1980) fit the standard APC model by equating two adjacent cohort parameters but avoided interpreting the resulting estimates literally. Pullum's main purpose in fitting such a model was merely to determine whether the fit (in a model of American fertility rates) is improved more by the addition of period or cohort parameters. He concluded that cohort parameters provide less explanatory power (per additional parameter) than period parameters. Since he was interested only in the overall fit of each model, not in the parameter estimates, the choice of identifying constraints was irrelevant.

Pullum also noted, however, that the second (and higher) differences of the parameter estimates in the three-factor APC model are invariant to the choice of an identifying assumption. In our notation, first differences depend on the choice of λ, although second differences do not. For example,

$$\begin{aligned}\Delta\alpha_i^* &= \alpha_{i+1}^* - \alpha_i^* \\ &= \hat{\alpha}_{i+1} + \lambda\cdot(i+1) - [\hat{\alpha}_i + \lambda\cdot i], \\ &= \hat{\alpha}_{i+1} - \hat{\alpha}_i + \lambda\end{aligned}$$

whereas,

$$\begin{aligned}\Delta^2\alpha_i^* &= \Delta\alpha_{i+1}^* - \Delta\alpha_i^* \\ &= \hat{\alpha}_{i+2} - \hat{\alpha}_{i+1} + \lambda - [\hat{\alpha}_{i+1} - \hat{\alpha}_i + \lambda]. \\ &= \hat{\alpha}_{i+2} - 2\hat{\alpha}_{i+1} + \hat{\alpha}_i\end{aligned}$$

First differences are related to the linear trend, and thus their dependence on λ is a reminder that the slope of the three sets of parameters cannot be uniquely identified in the standard APC model. Second differences are measures of deviation from the linear trend, and thus these quantities can be accurately estimated in the simple model with three main effects (but no interaction terms).

An analogous solution is to require that the slope of one of the three sets of parameters be constrained to equal zero. Following Wilmoth (1990), a simple three-factor model applied to a rectangular age-by-period array of demographic rates could be estimated by using the following constraints:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k w_k\theta_k = \sum_k k w_k\theta_k = 0$$

These constraints on θ_k are equivalent to requiring that both the *level* and the *slope* of a weighted least-squares regression line fitted to the cohort parameters should equal zero.[4]

Thus, one solution to the APC identification problem is to choose, arbitrarily, one dimension whose slope is constrained to equal zero (see also Caselli and Capocaccia, 1989). As a matter of convenience, the chosen dimension may correspond to the diagonals of a rectangular array of demographic rates. The resulting parameter estimates for this third factor have a simple interpretation: They represent deviations from the long-term pattern of change in that dimension. An analysis of such deviations is logically similar to an analysis of second derivatives, since both emphasize deviations from the long-term trend rather than absolute effects.

b. Interaction Terms and Demographic Translation

An obvious shortcoming to the standard APC model, aside from its problems of identifiability, is the absence of interaction terms. To address this inadequacy, Wilmoth (1990) proposed models with the following form:

[4] Here, as in Footnotes 1 and 2, the weights equal the length of the diagonals corresponding to individual cohorts.

$$f(r_{ijk}) = \mu + \alpha_i + \beta_j + \sum_{m=1}^{\rho} \phi_m \gamma_{im} \delta_{jm} + \theta_k + \varepsilon_{ijk}$$ [5]

This model differs from the standard APC model by the addition of ρ (usually one or two) multiplicative terms. The use of these interaction terms is motivated by the fact that the pace of change (over time) in demographic rates often differs by age. For example, mortality rates have typically fallen much more rapidly at younger than at older ages. This component of differential change by age cannot be expressed well in a model with no age-period (or age-cohort) interactions.

The above model is appropriate for an analysis of an age-period array of demographic rates, where the rates of change in r_{ijk} differ by age. The multiplicative terms capture these differential trends by age. The cohort parameters, θ_k, constrained to have zero level and slope, express deviations from the overall pattern of change by age and period. In short, the θ_k captures residual patterns that seem truly to lie along diagonals of the age-period array. Without the multiplicative term(s), however, estimates of θ_k may be heavily influenced by age-period interactions. The introduction of interaction terms by no means resolves the identification problem in the standard APC model. Their use is consistent, however, with the philosophy that the proper means of including all three sets of parameters in the model is to focus the description on two dimensions only and to treat the third dimension as a sort of residual.

The multiplicative interaction terms also help us to understand the notion of demographic translation (Ryder, 1964, 1980; see also Chapter 17). In the above model, the cohort term, θ_k, has been reduced to a residual component, and thus the linear or long-term cohort trend has been translated into the age and period dimensions. Even this residual component of the cohort trend, θ_k, can be expressed as age-period interactions, if a sufficient number of multiplicative terms are added to this model (Wilmoth, 1990). In this sense, the choice to assign any portion of the variability in demographic rates to one of the three dimensions is arbitrary, because it is always possible to re-express the pattern in terms of the other two dimensions by the addition of a sufficient number of interaction terms.

The only resolution of this logical dilemma is to acknowledge, as a matter of plausibility, that certain patterns must be due to changes in one specific dimension. For example, the residual cohort effects that emerge in a complete analysis of an age-by-period array (including at least some form of age-period interaction) cannot plausibly be attributed to age- and period-specific factors. These residual cohort effects tend to vary with high frequency and thus are not easily translated into the age and period dimensions. It is worth noting that Foster (1990) reached a similar conclusion (i.e., that only high-frequency cohort effects are plausibly identifiable) by an entirely different method of analysis.

c. Direct Measurement

Several analysts have noted that the only true resolution of the identification problem affecting APC analysis would be to measure directly the factors whose effects are summarized by the coefficients of the standard APC model. These factors would not typically be linearly related (although they might be strongly collinear), and thus a unique solution for the model could be found. An obvious difficulty with this strategy is the problem of defining and then measuring the proximate variables for which age, period, and cohort are, admittedly, only rough approximations.

In a study of mortality in Italy, for example, Caselli and Capocaccia (1989) use the probability of death in infancy or during the first 15 years of life as a measure of long-term cohort influences, whose magnitude may vary over the life course. In our notation, their model can be written

$$f(r_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_i \cdot Q_k + \varepsilon_{ijk}$$

where Q_k equals either q_0 or ${}_{15}q_0$. This model requires no special identifying restrictions. Because the estimated γ_i's are positive below age 45 and negative above this age, the authors conclude that "higher mortality early in life is associated with higher mortality up to age 45 and lower levels at later ages" (Caselli and Capocaccia, 1989, p. 152).

An alternative interpretation of these results, however, is that the age-cohort term in their model is, in reality, capturing the same sorts of patterns that were treated as age-period interactions by Wilmoth and colleagues (Wilmoth *et al*., 1989; Wilmoth, 1990). Indeed, the multiplicative age-cohort term in the model of Caselli and Capocaccia (1989) is remarkably similar, in both its form and its estimated values, to the multiplicative age-period term in the models by Wilmoth and collaborators. In other words, the age-cohort term in the above model may serve merely to document the relatively more rapid rate of change in mortality at younger than at older ages. Since infant-child mortality has declined almost monotonically for the cohorts in question, its function in the model may be simply to provide a marker for the second dimension of temporal mortality change.

[5] For a full discussion of the constraints required for fitting this model, see Wilmoth (1990).

The innovative applications and clear explanations of the age-period-cohort model by Caselli and Capocaccia (1989) are certainly valuable. Their attempt to measure age, period, or cohort effects directly, however, illustrates some of the statistical difficulties inherent in the task. As noted by Clogg (1982),

> The selection of the proper causal variables to be considered in a modeling procedure is a most difficult theoretical task, one that is at least as difficult as applying the age-period-cohort accounting framework (p. 460).

In other words, although it may seem in principle that direct measurement is an obvious solution for the APC identification problem, the practical application of this technique is perhaps more fraught with difficulties than the problem we are trying to overcome.

3. Empirical Examples

Two specific examples of the application of APC models to demographic data are reviewed here. These examples were chosen to illustrate two kinds of conclusions that can legitimately be drawn from APC models applied to time series of age-specific demographic rates. The first example, Pullum's (1980) study of American white fertility, illustrates the use of the standard APC model to compare the variability by cohort versus period. The second example, Wilmoth's (1990) analysis of French male mortality, employs a more complicated APC model with interaction terms to document the unusually high or low levels of mortality for certain cohorts.

a. American White Fertility, 1920 to 1970

Pullum (1980) analyzed an age-by-period array of age-specific fertility rates for American white women aged 15 to 44 during years 1920 to 1970. He fit all three possible two-factor models and the standard three-factor APC model. Table 18–1 shows χ^2 measures of fit for each model in four time periods, each spanning two decades (lower values indicate a better fit). The period *x* cohort model is clearly inferior to the other two-factor models, reflecting the primary importance of age as a dimension of variation in this case. It is not clear from the comparison in Table 18–1, however, whether age should be paired with period or cohort to obtain the best two-dimensional description of these data.

Table 18–2 reveals the greater importance of periods than cohorts as a dimension of variation in these data by comparing each two-factor model to the full APC model. The values in this table (which can be derived from those in Table 18–1) can be thought of as the improvement in fit when age, period, or cohorts is added as the third and final dimension of the standard APC model or as the loss of fit when one of these dimensions is removed from the full three-factor model. In each case, the improvement or loss of fit is measured relative to the change in degrees of freedom.

Pullum (1980) applied these same four models to an age-by-cohort array of age-specific fertility rates for American white women aged 15 to 44 years born during years 1905 to 1926 (thus including all cohorts whose reproductive lives fell entirely within the observation period) (Table 18–3). He notes that

> In contrast to the previous applications to rectangular [age-by-period] data, the inclusion of period effects appears *much* more useful than the inclusion of cohort effects (p. 236, emphasis in original).

For example, using measures like those presented here in Table 18–2, the loss of fit caused by removing all period parameters from the full APC model is 13.1 (increase in chi-square per degree of freedom) compared to only 0.4 if the cohort dimension is eliminated.

In conclusion, Pullum stated (p. 241)

TABLE 18–1 χ^2 Measure of Fit for Four Models Applied to Age-specific Fertility Rates, American White Women, Aged 15–44 Years, 21-year Time Intervals

Model	*df*	Interval of time			
		1920–1940	1930–1950	1940–1960	1950–1970
Age *x* period	580	148.1	330.2	411.3	93.3
Age *x* cohort	551	90.2	397.8	411.6	800.2
Period *x* cohort	560	13,306.8	14,952.4	20,603.5	21,130.1
Age *x* period *x* cohort	532	21.6	133.1	98.0	40.8

Note: Data for each time interval were in age-by-period format. χ^2 estimate is based on an artificial case base of 630,000 cases for each time interval.

Source: Pullum Thomas W., 1980. Separating age, period, and cohort effects in white U.S. fertility, 1920–1970. *Social Science Research*, vol. 9(3), p. 227–244. New York, Academic Press, Table 2.

TABLE 18–2 Reductions in χ^2 Per *df* When Age, Period, or Cohort is Added as Final Term of Standard APC Model of Age-specific Fertility Rates, American White Women, Aged 15–44 Years, 21-year Time Intervals

Effect	Interval of time			
	1920–1940	1930–1950	1940–1960	1950–1970
Age	474.5	529.3	732.3	753.2
Period	3.6	13.9	16.5	40.0
Cohort	2.6	4.1	6.5	1.1

Note: See note for Table 18–1.

Source: Pullum Thomas W., 1980. Separating age, period, and cohort effects in white U.S. fertility, 1920–1970. *Social Science Research*, vol. 9(3), p. 227–244. New York, Academic Press, Table 3.

TABLE 18–3 χ^2 Measures of Fit for Four Models Applied to Age-specific Fertility Rates, American White Women, Aged 15–44 Years, Cohorts Born 1905–1926

Model	*df*	χ^2
Age *x* period	580	96.1
Age *x* cohort	609	730.4
Period *x* cohort	588	15,846.8
Age *x* period *x* cohort	560	87.6

Note: Data were in age-by-cohort format. χ^2 estimate is based on an artificial case base of 660,000 woman-years.

Source: Pullum Thomas W., 1980. Separating age, period, and cohort effects in white U.S. fertility, 1920–1970. *Social Science Research*, vol. 9(3), p. 227–244. New York, Academic Press, Table 4.

> Despite the great theoretical appeal of the notion that continuities exist in the behavior of cohorts, we have found that the explanatory gain per cohort parameter is far less than the gain per period parameter. . . . The implications for our understanding of U.S. fertility are that, as a set, changes in those temporal variables which cut across cohorts, such as economic cycles, appear to be more important than changes in those variables which distinguish cohorts, such as shared socializing experiences.

Similar explorations of fertility patterns for other populations and time periods seem warranted. Pullum also suggested that these models could be usefully applied to fertility rates by parity.

b. French Male Mortality, 1946 to 1981

Wilmoth (1990) fit an APC model with two age-period interaction terms to age-specific mortality rates for French males during years 1946 to 1981. The full model, as stated earlier, was as follows:

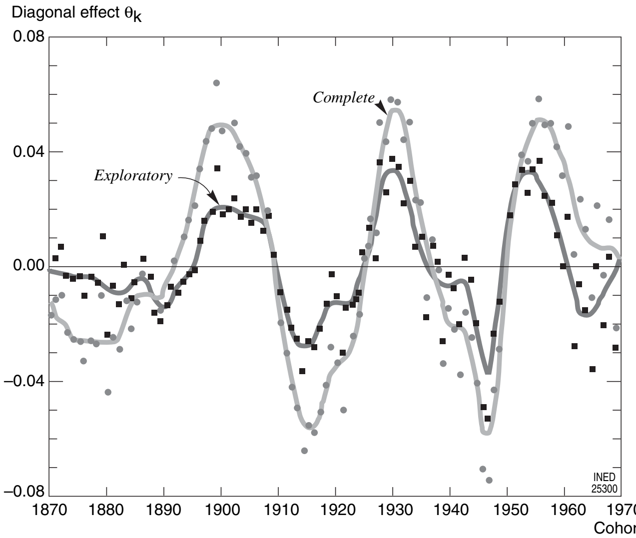


FIGURE 18–2 Residual cohort effects estimated by two methods in age-period-cohort model with interaction terms applied to age-specific mortality rates, French males, aged 0–89, years 1946–1981. Note: Data analyzed were in age-by-period format. The values depicted here are measures of the average level of *excess* or *deficit mortality* for each cohort (as a proportion of predicted levels) over the observation period. (Source: Wilmoth [1990], p. 307).

$$f(r_{ijk}) = \mu + \alpha_i + \beta_j + \sum_{m=1}^{\rho} \phi_m \gamma_{im} \delta_{jm} + \theta_k + \varepsilon_{ijk}.$$

The interaction terms in this model describe variations in the pattern of mortality decline by age: The first term captures the relatively faster decline at younger compared to older ages, while the second term reflects the pattern of increasing (or slowly decreasing) levels of mortality in late adolescence and early adulthood.

The cohort parameters in this model, θ_k, reflect residual levels of mortality relative to the thorough description by age and period. As shown here in Figure 18–2, Wilmoth's model can be used to document the persistence of relatively high or low levels of mortality for certain cohorts. Two sets of estimates for θ_k were derived using different methods for fitting the model, though the *complete* series is considered more reliable.

Wilmoth demonstrated the presence of substantial levels of *excess* or *deficit mortality* for various cohorts, successfully rejecting alternative explanations, for example, that the patterns were artifacts of bad data or of the model itself. Nevertheless, his explanation of their historical or biologic causes was less satisfactory. He acknowledged the inadequacy of existing explanations of these patterns when a comparison of results from different populations is made.

For example, it had been suggested that excess mortality for groups of male cohorts born around 1900 and 1930 may be related to early combat experiences near the end of the World Wars or to nutritional deprivation in adolescence resulting from the social and economic dislocations at these times (see Horiuchi, 1983; Wilmoth *et al*., 1989). However, these explanations fail to account for the presence of similar patterns of excess mortality among Japanese cohorts, both male and female, born around 1900, since Japan was only nominally involved in World War I (see Wilmoth, 1988, p. 107–113). Furthermore, the presence of similar patterns for females suggests that combat experience in the wars may not be the crucial explanatory variable. Finally, the apparent recurrence of a similar pattern for cohorts born after World War II suggests that the phenomenon may operate through mechanisms at least partially unrelated to the two wars.

Although the failure to explain these findings does not invalidate them, it does illustrate the difficulties of interpreting cohort-specific variations in demographic rates.

CONCLUSION

The studies by Pullum (1980) and Wilmoth (1990) illustrate that valid results *can* be derived within an APC framework if one accepts that there is no magic solution to the *identification problem*. Instead, it is necessary to seek findings that are invariant to the choice of identifying constraints (e.g., measures of fit, second differences) or that acknowledge the fundamental arbitrariness of these constraints (e.g., residual cohort effects with zero slope).

It is important to remember that the identification problems affecting APC analysis are not some statistical aberration. Rather, they reflect a fundamental lack of information in the data being analyzed. The first identification problem, involving the level of the various parameters, is not problematic in most situations, since our purpose in estimating these sorts of models is to analyze patterns of change over age or time, which are unaffected by an arbitrary choice of level. Furthermore, this kind of identification problem is familiar and affects other statistical methods as well.

The unusual aspect of APC analysis is that, even for the simplest model, identifying constraints are required in order to obtain unique estimates of the linear trend in the parameters. This situation arises directly from the relationship linking the three sets of causal factors:

$$\text{cohort} + \text{age} = \text{period}$$

In regression analyses, of course, it is standard practice to avoid using independent variables that are highly collinear (either individually, or in some combination). By their very nature, some variants of APC analysis ignore the sound logic of this practice and attempt, by some statistical trick, to perform the impossible, namely, to extract more information than what is contained in the data. Because of this fundamental lack of information, the results of an APC analysis reflect both the underlying patterns in the data and the assumptions adopted by the analyst. This reality does not invalidate the method in all situations, but it serves as an important reminder of the limitations that affect all models used with the purpose of separating age, period, and cohort effects.

References

Caselli Graziella, Capocaccia Riccardo, 1989. Age, period, cohort and early mortality: an analysis of adult mortality in Italy. *Population Studies* vol. 43(1), p. 133–153.

Clogg Clifford C., 1982. Cohort analysis of recent trends in labor force participation. *Demography* vol. 19(4), p. 459–480.

Easterlin Richard A., 1961. The American baby boom in historical perspective. *American Economic Review* vol. 51(5), p. 869–911.

Easterlin Richard A., 1978. What will 1984 be like? Socioeconomic implications of recent twists in age structure. *Demography*, vol. 15(4), p. 397–421.

Elo Irma T., Preston Samuel, 1992. Effects of early-life conditions on adult mortality: A review. *Population Index* vol. 58(2), p. 186–212.

Fienberg Stephen E., Mason William M., 1979. Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology*, vol. 10, San Francisco, Jossey-Bass, p. 1–65.

Fienberg Stephen E., Mason William M., 1985. Specification and implementation of age, period, and cohort models, *in*: William M. Mason and Stephen E. Fienberg (eds.), *Cohort analysis in social research*, p. 45–88. New York, Springer-Verlag, 408 p.

Foster Andrew, 1990. Cohort analysis and demographic translation: a comparative study of recent trends in age-specific fertility rates from Europe and North America. *Population Studies*, vol. 44(2), p. 287–315.

Glenn Norval D., 1976. Cohort analysts' futile quest: statistical attempts to separate age, period, and cohort effects. *American Sociological Review*, vol. 41, p. 900–904.

Goldstein Harvey, 1979. Age, period and cohort effects. A confounded confusion. *Bulletin in Applied Statistics*, vol 6(1), p. 19–24.

Hobcraft John, Menken Jane, Preston Samuel H., 1982. Age, period, and cohort effects in demography: a review. *Population Index*, vol. 48(1), p. 4–43.

Holford Theodore R., 1991. Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual Review of Public Health* vol. 12, p. 425–457.

Horiuchi Shiro, 1983. The long-term impact of war on mortality: old-age mortality of First World War survivors in the Federal Republic of Germany. *United Nations Population Bulletin* vol. 15, p. 80–92.

Mason William M., Fienberg Stephen E. (eds.), 1985. *Cohort analysis in social research: beyond the identification problem.* New York, Springer-Verlag, VIII + 400 p.

Pullum Thomas W., 1980. Separating age, period, and cohort effects in white U.S. fertility, 1920–1970. *Social Science Research*, vol. 9(3), p. 227–244. New York, Academic Press.

Ryder Norman B., 1964. The process of demographic translation, *Demography*, vol. 1(1), p. 74–82.

Ryder Norman B., 1965. The cohort as a concept in the study of social change. *American Sociological Review*, vol. 30(6), p. 843–861.

Ryder Norman B., 1980. Components of temporal variations in American fertility, *in*: Robert W. Hiorns (ed.), *Demographic patterns in developed societies*, p. 15–54. Londres, Taylor and Francis, 208 p.

Wilmoth John R., 1988. *On the statistical analysis of large arrays of demographic rates*. Princeton, Department of Statistics, Princeton University, 158 p. (Ph.D.Thesis).

Wilmoth John R., 1990. Variation in vital rates by age, period, and cohort, *Sociological methodology*, vol. 20, ed. by Clifford C. Clogg, p. 295–335. Oxford, Basil Blackwell.

Wilmoth John R., Vallin Jacques, Caselli Graziella, 1989. Quand certaines générations ont une mortalité différente de celle que l'on pourrait attendre. *Population* vol 44(2), p. 335–376. (Also in English: When does a cohort's mortality differ from what we might expect?, *Population: An English Selection*, vol. 2, 1990, p. 93–126.)

P A R T

IV

A TENTATIVE SYNTHESIS OF THE COMPONENTS OF DYNAMICS: FROM REPRODUCTION TO POPULATION MODELS

GRAZIELLA CASELLI AND JACQUES VALLIN

The components of population change, extensively examined in preceding chapters, must now be linked to population composition, and an attempt made to draw them together. The fact is that all the concepts, methods and measures so far defined form a coherent and relatively comprehensive corpus of analysis (notwithstanding some flaws and failings, as will be described) with which a *synthesis of dynamics and structure* can be made which will give insights into the underlying mechanisms of population development. Two useful approaches can be taken. The relatively simple one is to combine period or cohort fertility and mortality to measure the reproductive potential that a population would acquire from the consistent adoption of such demographic processes. This is what is called *population replacement* (Chapter 19). The other will enable us to go further: What would become of a population whose age-specific fertility, mortality, or migration rates became entrenched at a given period level? What in time would become of its total size, age-sex structure, crude birth, death, and migration rates, and, finally, its growth rate? The population models described in Chapter 20 were developed to address these issues. These will show that, in the final analysis, population structure is entirely determined by the history of its changes, because although structure has a strong duration-specific influence on change, the original structure carries within it the seeds of its own ultimate extinction.

CHAPTER

19

Population Replacement

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

At this point, our period measures have given us age-specific mortality and fertility functions that, in principle, express the incidence of two components of the natural increase in a *fictitious* (or *hypothetical or synthetic*) cohort subjected during its entire life to the conditions prevailing in the observation period. How do these functions translate into the overall population dynamics? Can we determine an *intrinsic force* that is always independent of the age-specific structure and strikes a balance between two given fertility and mortality functions? How do these two functions influence a population's size and structure? Demographers have sought to answer the first question by defining a reproduction rate (the subject of this chapter), then by constructing population models (discussed in Chapter 20).

Like life expectancy at birth or the average number of children per female, reproduction is a concept initially defined in cohort (longitudinal) terms and then transposed to a period (cross-sectional) analysis to summarize the actual situation at a particular time. In cohort analysis, demographers want to know how many children in the following generation will replace the men and women who procreated them. The indicator developed to answer this question for a cohort can be subsequently applied to period analysis for the purpose of assessing the potential dynamic inherent in the period fertility and mortality functions. In fact, as fertility rates are generally measured for females, reproduction will be initially defined for an imaginary population composed only of women giving birth to daughters, totally disregarding the male role in reproduction. We will then need to examine ways of addressing the issue for the male population and, most important, determine whether we can reconcile the female and male rates in a sexually reproducing population.

I. FEMALE REPRODUCTION

Since the aim here is to compare the number of daughters with the number of mothers, we must begin by eliminating males from the computation. This will lead us, first, to the concept of gross reproduction. Next, we will need to combine female fertility with female mortality to arrive at the concept of net reproduction, but we will see that the concept translates into different realities depending on whether we use it to describe reproduction of births, reproduction of women of childbearing age, or—more generally—reproduction of years lived.

1. Gross Reproduction

Robert R. Kuczynski introduced the concept of gross reproduction rate (R) in 1931. If we assume the absence of mortality, in a given birth cohort, the number of children to whom a woman gives birth is quite simply (as seen earlier) the sum of age-specific fertility rates, called completed fertility. As this aggregate comprises sons and daughters, we need only weight these offspring by the proportion of births that are female to define a *gross reproduction rate*. Let B be the births to this cohort of women, B^F the female births and $f_{(x,x+1)}$ the fertility rate at age x to $x + 1$. The gross reproduction rate R is written:

$$R = \frac{B^F}{B} \sum_{x=\alpha}^{\omega} f_{(x,x+1)}$$

It should be noted that, in a retrospective cohort observation, the completed fertility of women aged over 50 is given directly by the ratio of the number of actual births to the number of females surveyed; the gross reproduction rate equals the ratio of female births to the females surveyed.

The proportion of females (or males) at birth was one of the stablest demographic parameters until the advent of techniques for selecting fetuses based on their sex. Everywhere and at all times, the sex ratio at birth was about 105 males to 100 females, with very minor fluctuations around this average (between 103 and 107). This rule remains fairly widespread even today, except in certain populations with strong male preference and with access to early fetal sex diagnostics and selective abortion (Chapter 4): the gross reproduction rate can be obtained simply by multiplying completed fertility by 0.488, that is 100 divided by 205.

A gross reproduction rate equal to unity means that in the absence of mortality before age 50 (age taken as the end of the female childbearing period), the cohort of females studied has given birth to a number of daughters exactly equal to their own number. Consequently, to go back to the example of Jeanne Calment, the French women born in 1875 had a completed fertility of 2.6[1] and a gross reproduction rate of 1.28. In the absence of mortality, their fertility would thus have largely sufficed to ensure their replacement, since there would have been almost 30% more daughters than mothers.

In period analysis, we transpose the concept of gross reproduction to a fictitious cohort by weighting the total fertility rate (TFR) by the proportion of female births: this yields a measure of the theoretical reproduction capacity of a closed, mortality-free population that would indefinitely continue the same fertility pattern as the one reflected in a given year t in the period age-specific fertility rates:

$$R_t = \frac{B_t^F}{B_t} TFR_t = \frac{B_t^F}{B_t} \sum_{x=\alpha}^{\omega} f_{(x,x+1),t}$$

Of course, if we are working with age-group–specific fertility rates, we need to allow for age intervals a:

$$R_t = \frac{B_t^F}{B_t} \sum_{x=\alpha}^{\omega} a f_{(x,x+1),t}$$

[1] This is the result obtained for the aggregated 1871–1880 birth cohorts, which include that of Jeanne Calment (Festy, 1979), but the finding is not of great importance because the cohort synthetic indicators evolve slowly and without fluctuations.

This calculation method gives a TFR of 3.44 children per female: The renewal potential inherent in the age-specific fertility pattern of Jeanne Calment's birth year (1.68) was far greater than the actual one for her birth cohort. At that level of fertility, the population of daughters would have been over 70% larger than that of the mothers—in the absence of mortality—whereas we have just seen that the potential increase in the 1875 cohort was 28%.

2. Net Reproduction at Birth

Gross reproduction, however, gives only a very theoretical reproduction potential, as it does not make allowance for mortality. Actually, R is a more strictly female fertility measure than that customarily given by completed fertility or TFR. To measure net reproduction, we must take into account the fact that the mother cohort's initial population at birth is diminished by mortality even before reaching childbearing age. This proportionally reduces its capacity to effectively produce the female births that will replace it in the overall population.

a. Cohort Analysis

Paradoxically, the concept of net reproduction was introduced well before that of gross reproduction: in the late 19th century, by Richard Böckh (1886, 1890). It was then promoted in the United States by Robert R. Kuczynski (1931) and in France by Pierre Depoid (1941). When the gross rate is computed, the age-specific fertility rates are weighted by the probabilities of (female) survival at the same ages. This yields a *net reproduction rate at birth*, R_0:

$$R_0 = \frac{B^F}{B_t} \sum_{x=\alpha}^{\omega} f_{(x,x+1)} \cdot l_{x+0.5}$$

where $l_{x+0.5}$ is the probability of survival from birth to age $x + 0.5$.

In fact, if mortality is low and varies little at peak fertility ages, we need only multiply the gross reproduction rate by the probability of survival at the mean age at net childbearing[2]:

[2] This expression denotes the mean age computed for births actually occurring in the cohort, allowing for mortality. The mean age at net childbearing, $\bar{x}'$, is calculated, like net reproduction, from fertility rates weighted by survival rates:

$$\bar{x}' = \frac{\sum_{x=\alpha}^{\omega} (x+0.5) f_{(x,x+1)} \cdot l_{x+0.5}}{\sum_{x=\alpha}^{\omega} f_{(x,x+1)} \cdot l_{x+0.5}}$$

When mortality is low, there is actually little difference between the mean ages at gross and net childbearing.

$$R_0 \cong l_{\bar{x}'} \cdot R$$

A rate equal to unity means that, in a closed population, a cohort of mothers will have actually given birth to a number of daughters strictly equal to its own initial number.

For example, in Jeanne Calment's birth cohort (1875), we obtain a net reproduction rate at birth of 0.83. The lesson is that, despite its large replacement potential (1.28), the cohort's mortality between birth and reproductive ages made it fall far short of replacing itself by an equal number of daughters: The deficit is nearly 20%.

This below-unity net reproduction rate implies that if all cohorts behaved like that of Jeanne Calment, the population would be decreasing—but by how much and at what pace? The answer obviously depends on the mean interval between a cohort of mothers and that of its daughters, in other words, the mean age at childbearing. As explained in Chapter 3, the annual average growth rate r that causes the population size to increase from P_0 to P_t in a time t is defined by:

$$P_t = P_0 e^{rt}$$

in other words

$$r = \log \frac{P_t}{P_0} \cdot \frac{1}{t}$$

The annual average rate of increase ρ from a population of mothers to a population of daughters can therefore be estimated[3] by replacing (1) the P_t/P_0 ratio by the net reproduction rate, and (2) the duration t by the mean age at net fertility, $\bar{x}'$:

$$\rho \cong \frac{\log R_0}{\bar{x}'}$$

This rate resembles the intrinsic growth rate or Lotka rate, discussed in the next chapter.

The actual experience of Jeanne Calment's cohort thus translates into a Lotka rate of increase of −0.3% per year.

b. Period Analysis

As with gross reproduction, we can transpose the findings to a period analysis to define the net reproduction of a fictitious cohort in a given period:

[3] This formula is only approximate because the distribution of ages at childbearing has a nonnil variance. In this case, one can use the second-order approximation (Brouard, 1982):

$$\rho_{(2)} \cong \frac{\log R_0}{\bar{x}'} + \frac{\sigma^2 (\log R_0)^2}{2\bar{x}'^3}$$

$$R_{0,t} = \frac{B_t^F}{B_t} \sum_{x=\alpha}^{\omega} f_{(x,x+1),t} \cdot l_{x+0.5}$$

or, if we are working with age groups:

$$R_{0,t} = \frac{B_t^F}{B_t} \sum_{x=\alpha}^{\omega} a f_{(x,x+a),t} \cdot l_{x+0.5a}$$

As before, if mortality is low and varies little in the peak fertility years, we simply multiply the gross reproduction rate by the probability of survival at the mean age at childbearing:

$$R_{0,t} \cong l_{\bar{x},t} \cdot R_t$$

This rate measures the actual reproductive capacity of the population inherent in the age-specific fertility and mortality conditions of the period. In the conditions observed in 1875, this period net reproduction rate was 1.08 daughters per mother, a figure significantly above the level strictly needed for replacement and distinctly higher than the final rate for the women born that year. The reason is that between the birth of Jeanne Calment's cohort and the late 1920s, when it ceased to be fertile, the impact of the fall in fertility on reproduction came to outweigh the impact of the decline in mortality.[4] The period net reproduction rate fell from 1.08 in 1875 to 1.02 in 1889, when Calment's cohort entered its reproductive period (15 years of age). The rate dipped to 0.96 in 1904, when the cohort reached its mean age at childbearing (28.6 years) and finally to 0.93 in 1925, when the cohort reached the end of its reproductive period (50 years). In other words, the conditions prevailing at the cohort's birth seemed to offer it a fairly high net replacement potential, which would have allowed a mild increase (8%) in the number of daughters over the number of mothers. When the cohort reached its peak reproductive ages, the period conditions no longer ensured a one-for-one replacement of the population. However, the actual reproductive capacity of the 1875 cohort (0.83) eventually turned out to be far below each of these period measures taken in the course of its reproductive period. The reason is that, for this cohort, we also need to factor in the deficit of births due to World War I. Between 1916 and 1918, the period net reproduction rate was under 0.5.

This net reproduction measure can be expressed as a Lotka average annual rate of increase characterizing the period fertility and mortality rates:

$$\rho_t \cong \frac{\log R_{0,t}}{\bar{x}'_t}$$

[4] Including, at the end of the reproductive period, a measurable effect of the birth deficit of the war years.

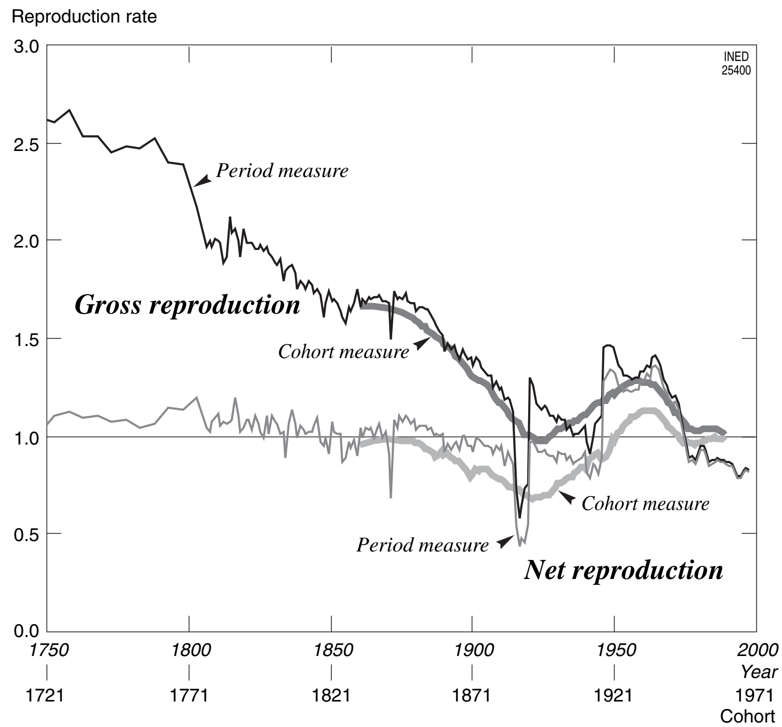


FIGURE 19–1 Long-term changes in gross and net reproduction rates, by birth cohort and according to period conditions. France, 1750–1997.

Between 1875 and 1925, the rate fell from +0.11% to –0.11%, with some sharply negative values during the war (as low as –1.22% in 1916).

c. Illustration

Figure 19–1 compares the long-term changes in the gross and net period and cohort reproduction rates in France. Period rates are shown from 1750 on.[5] Cohort rates are plotted beginning with the women born in 1830. For a meaningful comparison between the cohort-indicator changes and the period-indicator changes, we would need to position each cohort in the year in which it reached its mean age at childbearing. This has varied over time, declining from slightly over 30 years in the 18th century to 27 years in the mid 20th century, then rising to 29 years. For consistency, we have plotted each cohort in the year when it reached age 29.

Predictably, the gross reproduction rates exhibit the same contrast already emphasized in our analysis of TFRs (Chapter 14) between the very steady pattern of the cohort indicator and the highly fluctuating profile of the period indicator. The contrast is less pronounced for net reproduction, as it is subject to the fluctuations in cohort survival, which are rather sensitive to the swings in infant mortality. Again, however, the period indicator displays far more fluctuations.

The most interesting finding here is the spectacular convergence, in two centuries, between net reproduction and gross reproduction, in both the period indicators and cohort indicators. In earlier times, mortality between birth and reproductive ages drastically reduced the replacement potential inherent in the fertility rate. Today, that measure of mortality has collapsed to the point of creating only an insignificant difference between gross and net reproduction.

The figures also clearly show how consistently France's period net reproduction rate stayed close to unity throughout the period of study, with a single noteworthy exception: the baby boom of the 1950s and 1960s.

One final observation: For gross reproduction, the cohort and period indicators move reasonably in step

[5] Data for the 18th century are drawn from Louis Henry's historical reconstitution based on a representative sample of parish registers (Festy, 1979, for fertility; Blayo, 1975, for mortality). Our 19th century fertility data are from Depoid (1941), Festy (1979), and INSEE's *Annuaire Rétrospectif* (1966, 1990), while the mortality data are based on the reconstitution by France Meslé and Jacques Vallin (1989). For the 20th century, our fertility data are taken from INED (1975 and successive official reports annually published in *Population* by INED), INSEE yearbooks, and unpublished computations by Gérard Calot; the mortality data were computed by France Meslé and Jacques Vallin (Vallin, 1971; Vallin and Meslé, 2001).

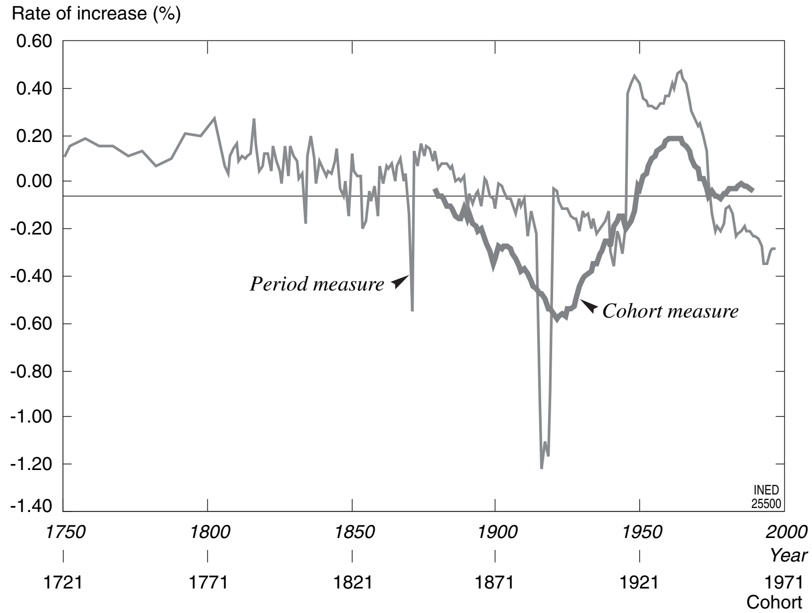


FIGURE 19–2 Long-term changes in average annual rates of increase in period and cohort net reproduction. France, 1750–1997.

(disregarding the period-indicator fluctuations); the net-reproduction indicators display a very different pattern. From the mid 19th to mid 20th centuries, the net reproduction of each cohort runs well below the net reproduction rate of the year in which the cohort reaches age 30. We find the same result, of course, when we compare the changes in the average annual rates of increase corresponding to these two net-reproduction indicators (Figure 19–2). This contradiction is due to the intrinsic difference between the period and cohort measures of net reproduction.

3. Reproduction at Age 15

The difference pointed out at the end of the previous paragraph can be explained as follows. In the cohort analysis, the expression of the net reproduction concept used until now, *net reproduction at birth* (hence the notation R_0), covers only one of several possible meanings. It tells us how many births of daughters will replace a cohort of births of mothers. As we have just seen, however, current mortality conditions evolve between a mother's birth and the time when she reaches reproductive age. Is it not more logical, therefore, to compare how many surviving daughters at age 15 would replace the cohort of mothers entering their reproductive period at the same age? Rather than measuring the replacement of births by births, we would thus measure the replacement of females of childbearing age by females of childbearing age. To obtain this *net reproduction rate at age 15*, R_{15}, for a generation, we should not weight fertility rates at each age by the mothers' probability of survival between birth and childbearing age. Instead, we must weight them by (1) the mothers' probability of survival between age 15 and childbearing age and (2) the daughters' probability of survival between birth and age 15:

$$R_{15} = \frac{B^F}{B} \sum_{x=\alpha}^{\omega} f_{(x,x+1)} \cdot \frac{l^m_{x+0.5}}{l^m_{15}} l^d_{15} = \frac{B^F}{B} \cdot \frac{l^d_{15}}{l^m_{15}} \sum_{x=\alpha}^{\omega} f_{(x,x+1)} \cdot l^m_{x+0.5}$$

where l^m_x and l^d_x are the probabilities of survival between birth and age x in the cohort of mothers and in the cohorts of daughters. As a rule, we lack the necessary data for a precise computation of the survival rate of daughters born to these mothers (as distinct from other daughters of the same cohorts born to mothers of different cohorts). We can, however, estimate the value from the survival rate in the cohort of daughters born in the year in which the mothers have reached their mean age at childbearing.

Moreover, as in the earlier exercise, we can generally simplify the calculation by using the mean age at childbearing to adjust gross reproduction for the probability of survival of mothers between age 15 and their mean age at childbearing and the probability of survival of daughters at age 15:

$$R_{15} \cong R \cdot \frac{l^m_{\bar{x}}}{l^m_{15}} l^d_{15}$$

For Calment's cohort, this net reproduction rate at age 15 was 0.91—significantly above the net reproduction rate at birth, but still well below unity. This sit-

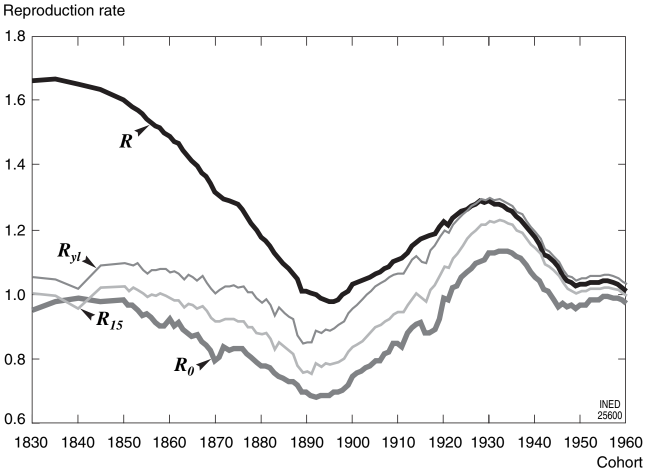


FIGURE 19–3 Net reproduction rate at birth and at age 15 and years-lived reproduction rate, in French cohorts born since the early 19th century.

uation is characteristic of a population whose mortality at early ages is very high but on a steep downtrend. The substitution of the daughters' infant and child mortality for that of mothers in the formula means a milder cut in the gross reproduction rate. Indeed, all French cohorts born since the mid 19th century display a net reproduction rate at age 15 that is substantially higher than their reproduction at birth (Figure 19–3).[6]

We can, of course, compute the net reproduction rate of a cohort at other ages, relative to their reproductive period, for example at their mean age at childbearing, provided there is no major difference in this age between mothers and daughters. In fact, as mortality is rather low at young adult ages, this additional refinement will not have a significant impact.

4. Reproduction of Years Lived

By contrast, to fully measure a population's capacity to ensure its own effective replacement, we should also take into account mortality beyond age 50. Suppose that mothers live 50 years and daughters 80 years. A net reproduction rate (at birth or at age 15) below unity would be enough to ensure the replacement of the total population. Historically, that is what occurred in France for many cohorts in which—thanks to the sharp rise in their life expectancy over age 50—not only the net reproduction rate but also the reproduction rate at age 15 were below unity even as the nation's population was increasing by more than net immigration would have allowed.

Consequently, to gain more insight into the concept of replacement, Louis Henry (1965) set out to calculate a *reproduction rate of years lived*, R_y, by weighting the net reproduction rate by the ratio of the daughters' life expectancy at birth to that of the mothers:

$$R_y = R_0 \cdot \frac{e_0^d}{e_0^m}$$

In Calment's cohort, the net years-lived reproduction rate was 1.02, more than 10% higher than the net rate at age 15, which was already 10% above the net rate at birth. On this measure, therefore, the cohort did succeed in replacing itself.

In France, the years-lived reproduction rate has consistently exceeded the rate at age 15 for all birth cohorts since the relevant data became available (1830). As a result, unlike net rates at birth or at age 15, the years-lived reproduction rate significantly exceeded unity for the cohorts born between 1830 and 1880. On the other hand, it fell below unity in the cohorts born between 1880 and 1900 owing to the enormous deficit of births caused by World War I (Figure 19–3).

Later, thanks to the *baby boom*, all three indicators topped unity, but the years-lived reproduction rate did so 20 cohorts before the net rate at birth. From that point on, in fact, the years-lived reproduction rate caught up swiftly with the gross reproduction rate. Remarkably, since the 1925 cohort, the years-lived indicator has overtaken the gross reproduction rate. The daughters' life-expectancy gains on their mothers have increasingly outweighed the fertility loss due to mortality before the reproductive period.

Obviously, if we calculate these replacement indicators not for a cohort but using period fertility and mortality functions applied to a fictitious birth cohort, all the variants of the net reproduction rate (R_0, R_{15}, R_y) coincide, since the mothers and daughters of the fictitious cohort are assumed to have the same mortality.

This observation suggests another one. For a cohort displaying the same fertility and mortality functions throughout its lifetime, the combination of the two functions into a reproduction rate gives us the exact measure of its presumed renewal capacity. In consequence, if we define the two functions as two constant laws applying to all cohorts, can we not *ipso facto* deduce from the functions the population's capacity to increase (or decrease) in the years ahead? In other words, to what extent does the Lotka rate of increase that we can deduce from the functions tell us the population's real rate of increase tomorrow? If, for

[6] By contrast, the cohorts born in 1840 to 1844 had a net reproduction at age 15 that was below their net reproduction at birth: This is due to the rise in infant and youth mortality in France between 1850 and 1880, plus an exceptional excess mortality in 1870 to 1871 caused by the Franco-Prussian war and the Paris Commune insurrection.

example, the fertility and mortality conditions of 1997 persisted indefinitely in Italy—with a net reproduction rate well below unity—the Italian population would surely decline. But there is little likelihood that the conditions of a given year will endure: 1997 conditions, for example, are subjected to the timing effects described earlier (see Chapter 14). Even if the conditions persisted, we would still have to factor in the role of the population's age structure. As we saw in Chapter 4, that structure would translate the fertility and mortality functions into births and deaths, which, in turn, would cause the structure to change. This issue will be explored in greater detail in the next chapter, devoted to population models.

Beforehand, however, we must go back to the initial simplification of the problem: what is the role of men in the process we designate here as reproduction?

II. MALE REPRODUCTION AND THE SEX-RECONCILIATION ISSUE

The concept of reproduction, which we have hitherto discussed in terms of childbearing, could be just as well examined from the standpoint of paternity, if we had male fertility statistics at our disposal. Unfortunately these are rare, for two reasons. The most important one is practical: paternity outside marriage is hard to observe, and even data on paternity in marriage are scarce. The second reason, no doubt, is a cultural one: most human societies have long regarded the reproduction of the species as the pre-eminent role of women. This is regrettable, as the analysis of male fertility—in those circumstances where it can be carried out—offers a different perspective on the reproduction issue. Incidentally, this raises a question that still awaits a satisfactory answer: how should demographers incorporate male-female cooperation into the measure of a population's fertility and reproduction?

1. Male Fertility and Male Reproduction: the Case of a Monogamous European Population

Male fertility is seldom measured. Some authors in the 1940s attempted to incorporate male fertility indices into the theoretical study of population (Karmel, 1947, 1948, 1949; Hajnal, 1950), but there are hardly any historical descriptions. Nicolas Brouard (1977) tried to reconstitute the evolution of male fertility and male reproduction in France over nearly a century.

Figure 19–4 shows the difference between male and female fertility in France in the late nineteenth century and the mid 1970s. The difference is significant. Men have a later fertility than women. However, in a population such as France's, where marriage is monogamous and the age difference between spouses fairly small, the similarities between the two fertility curves (same overall profile and closeness of the heights of the two modes) outweigh their differences (mostly due to the age difference between spouses). Moreover, these differences have narrowed sharply in a century owing to the sizable reduction in the age difference between spouses.

The fertility differences, however, are sufficient to induce significant differences in gross reproduction between sexes (Figure 19–5). For the late 19th century (1899), Nicolas Brouard finds a male TFR of 3.1 chil-

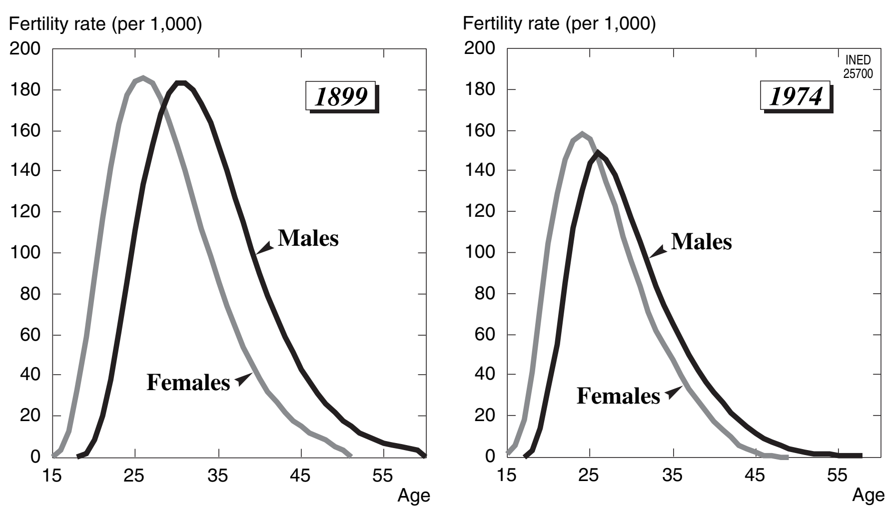


FIGURE 19–4 Age-specific male and female fertility rates in France, 1899 and 1974 (source: Brouard, 1977).

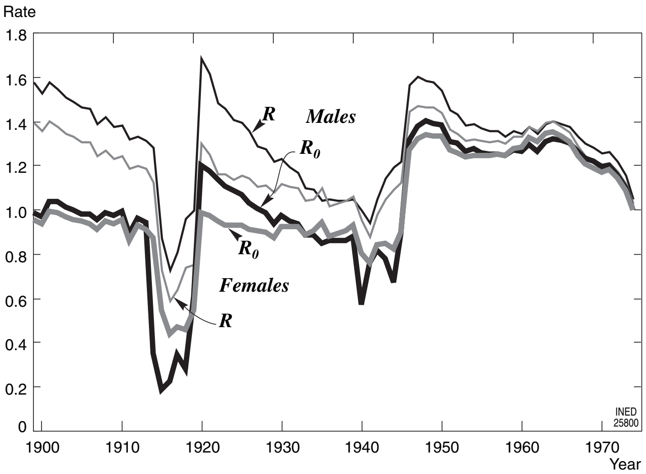


FIGURE 19–5 Comparative changes in female and male gross reproduction rates (R) and net reproduction rates (R_0). France 1899–1974 (source: computed from total fertility rates published by Nicolas Brouard, 1977).

dren versus a female TFR of 2.9 children. The relative gap between gross reproduction rates (1.6 for males and 1.4 for females) was even wider since there were more male than female births. However, as male mortality was greater and the mean age at paternity higher than the mean age at childbearing (33.8 versus 29 years), the sex differential was much smaller in terms of net reproduction: 0.99 for males and 0.95 for females (Table 19–1). With time, these differences have narrowed to near-insignificance in the 1970s.

TABLE 19–1 Indicators of Male and Female Reproduction in France, 1899 and 1974

Indicators	1899		1974	
	Male	Female	Male	Female
Total fertility rate (TFR)	3.08	2.86	2.05	2.11
Proportion of births of each sex	0.512	0.488	0.512	0.488
Gross reproduction rate (R)	1.58	1.40	1.05	1.03
Mean age at procreation ($\bar{x}$)	33.8	29.0	29.7	26.8
Net reproduction rate (R_0)	0.99	0.95	1.00	1.00
Lotka growth rate (first order)	–0.01	–0.07	0.00	0.00

Source: Computations from findings in Brouard 1977.

TABLE 19–2 Indicators of Male and Female Reproduction for Bandé Fulani

Indicators	Male	Female
Total fertility rate (TFR)	11.7	6.72
Proportion of births of each sex	0.512	0.488
Gross reproduction rate (R)	5.99	3.28
Mean age at gross procreation ($\bar{x}$)	44.3	27.6
Mean age at net procreation ($\bar{x}$)	40.1	27.0
Net reproduction rate (R_0)	1.79	1.49
Lotka growth rate (first order)	1.46	1.48
Standard deviation	11.2	8.3
Lotka growth rate (second order)	1.48	1.51

Source: Pison, Gilles, 1982. *Dynamique d'une population traditionnelle: les Peul Bandé (Sénégal oriental).* Paris, INED, PUF, 278 p. (Travaux et Documents series, Cahier no. 99).

2. Male Fertility and Reproduction: the Case of a Polygamous African Population

The pattern is very different in a population where polygamy is widespread and the age difference between spouses is large. An eloquent example is provided by Gilles Pison's study of the Bandé Fulani in eastern Senegal (1982). The age-specific male and female rates of this population are extremely divergent. Male fertility is later, more dispersed, and higher than female fertility (Figure 19–6). Although female fertility peaks at age 20 to 25 and is concentrated between the ages of 15 and 45, male fertility peaks at age 45 to 49 and is widely spread between the ages of 25 and 70. This does not prevent it from reaching 400 per thousand versus only 300 per thousand for female fertility. As a result, the TFR is much higher among men (nearly 12 children per male) than among women (6.7 children per female).

Such differences are made possible only by polygamy, which seems to allow men to have more children because they beget them later. Indeed, polygamy cannot become a significant phenomenon unless the age difference between spouses is wide.

At the same time, however, the massive differences observed in TFRs and—even more so—in gross reproduction rates (6 for males versus 3.3 for females) are substantially narrowed when we look at net reproduction: the figures are 1.8 for males versus 1.5 for females (Table 19–2). Even though male mortality is barely higher than female mortality in Africa, it exerts a greater downward effect on gross reproduction capacity for men, because it has to be factored in for a period of more than 40 years of mean age at net paternity, compared with only 27 years for female mean age at net childbearing. Last, when we compute the Lotka rate, the fact that reproduction requires 40 years of male life versus only 27 years of female life means that the average annual rates of increase implied by these

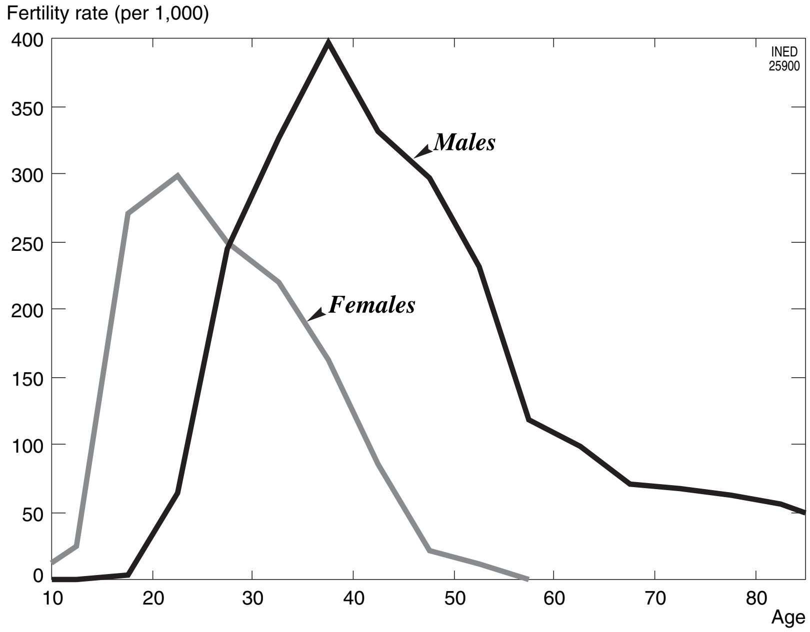


FIGURE 19–6 Age-group-specific male and female fertility rates among the Bandé Fulani of eastern Senegal (source: Pison, 1982).

reproduction figures are—in the end—virtually the same. In this manner, two theoretical gross reproduction capacities, one twice as large as the other, are reconciled in a single rate of actual procreation.

References

Blayo Yves, 1975. La mortalité en France de 1740 à 1829. *Population*, vol. 50, Démographie historique special issue, p. 123–142.

Brouard Nicolas, 1977. Évolution de la fécondité masculine depuis le début du siècle, *Population*, vol. 32(6), p. 1123–1158.

Brouard Nicolas, 1989. *Mouvements et modèles de population*. Yaoundé, IFORD, 184 p.

Böckh Richard, 1886. Statistik des Jahres 1884, *Statistisches Jahrbuch der Stadt Berlin*, vol. 12.

Böckh Richard, 1890. Die statistische Messung der ehelichen Fruchtbarkeit, *Bulletin de l'Institut International de Statistique*, vol. 5, section 1, p. 165–166.

Depoid Pierre, 1941. *Reproduction nette en Europe depuis l'origine des statistiques de l'état civil*. Paris, Statistique Générale de la France, 42 p. (études Démographiques series, no. 1).

Festy Patrick, 1979. *La fécondité des pays occidentaux de 1870 à 1970*. Paris, INED, PUF, 392 p. (Travaux et Documents series, Cahier no. 85).

Hajnal John, 1950. Births, marriages and reproductivity, England and Wales, 1938–1947, *in: Papers of the Royal Commission on Population, Section A. Marriages*, vol. II, p. 307–322. London, HMSO, 422 p.

Henry Louis, 1965. Réflexions sur le taux de reproduction, *Population*, vol. 20(1), p. 53–76.

INSEE, 1966. *Annuaire statistique de la France. Résumé rétrospectif*. Paris, INSEE, 566 + 134 p. (vol. 72, new series, no. 14).

INSEE, 1990. *Annuaire rétrospectif de la France, 1948–1988*. Paris, INSEE, 658 p.

Karmel Peter H., 1948. The relations between male and female reproduction rates, *Population Studies*, 1947–1948, vol. 1, p. 249–274.

Karmel Peter H., 1949. Le conflit entre les mesures masculine et féminine de la reproduction, *Population*, vol. 4(3), p. 471–494.

Kuczynski Robert R., 1931. *Fertility and reproduction: methods of measuring the balance of birth and death*. New York, Macmillan, 150 p. (reprinted 1982, Berlin, Akademie Verlag, 147 p.).

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIX^{e} siècle. *Population*, vol. 44(6), p. 1121–1158. (English version in *Population: An English Selection* no. 3, 1991, p. 33–62).

Pison Gilles, 1982. *Dynamique d'une population traditionnelle: les Peul Bandé (Sénégal oriental)*. Paris, INED, PUF, 278 p. (Travaux et Documents series, Cahier no. 99).

Pollard John, 1977. The two-sex problem, *in: IUSSP International Population Conference, Mexico, 1977*, vol. 2, p. 291–310. Liège, 528 p.

Vallin Jacques, 1971. *La mortalité par génération en France depuis 1899*. Paris, INED, PUF, 484 p. (Travaux et Documents series, Cahier no. 63).

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX^{e} et XX^{e} siècles et projections pour le XXI^{e} siècle*. Paris, INED, 102 p. + CDrom. (Données statistiques, no. 4, 2001).

C H A P T E R

20

Population Models

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

With the device of the hypothetical cohort, we have sought to estimate the fertility, mortality, and migrations in a cohort that would be subject throughout its life to the conditions prevailing in the period studied. Let us go one step further and examine the changes in a population whose age-specific fertility, mortality, and migration rates would be frozen at the levels observed at a given moment. In the long run, what would become of its size, its age and sex structure, its crude birth, death, and migration rates, and, ultimately, its growth rate?

In Chapter 4, we discussed the interaction between the population structure and the determinants of its changes. At this stage, however, we have not examined the instruments for measuring the phenomena involved, which will be discussed in the following chapters. For the moment, these instruments enable us to formalize the relationships, to understand how they work, and to attempt an answer to our opening question.

The first demographer to define the equation linking a population's age structure, fertility, and mortality was Alfred Lotka in 1939. In so doing, he demonstrated that, whatever its initial age structure, a closed population subjected indefinitely to constant age-specific fertility and mortality tends toward a constant age structure and a constant growth rate. Lotka's study gave rise to the concept of *stable population*.

To do this, however, Lotka doubly simplified the problem by restricting his model to the female population (the yardstick by which fertility is customarily measured) and specifying a closed population (with no migration). Other authors have since tried to extend the model. Their generalization effort has led them to analyze not only the changes in populations with constant growth rates, but also the changes in populations with variable growth rates.

In its generalized form, Lotka's equation may also provide an answer to a new question. Instead of exploring possible changes in the age structure of a population whose fertility, mortality, and migration rates were frozen at their present levels, we will ask what factors in past fertility, mortality, and migrations have led to the age structure observed at a given moment. For this purpose, we will use the cohort measures of fertility, mortality, and migrations.

I. LOTKA'S THEORY

Provided that mortality and fertility at each age consistently exhibit the same values whatever the initial age structure, birth rate, and death rate, the growth rate will ultimately approach the asymptotic value ρ—the intrinsic value—which expresses the population's fundamental capacity to multiply when freed of the disruptive influence of an arbitrary age structure.

This aspect of the intrinsic rate of natural increase ρ deserves our very close attention. The rate is of interest to us less as a forecasting instrument than because it expresses a current property of a population, its fundamental capacity to multiply. The crude rate of natural increase—i.e., the excess of the birth rate over the recorded death rate—does not give us the true measure of this capacity, as both the latter rates are influenced by the population's age structure, and this structure itself depends on what we can describe as fortuitous factors that have been at work in the past. There can be a sizable gap between the crude rate and the intrinsic rate.[1] (Lotka, 1939)

When we take a closed population in any given initial situation and arbitrarily make its mortality and fertility functions constant, we commit the population to a process of convergence toward a population whose growth rate and age structure will also be constant. The growth rate and age structure will be exclusively determined by the combination of fertility and mortality functions—in other words, they will be totally independent of the initial situation. The theoretical population reached at the end of this convergence (stabilization) process is called a *stable population*. In reality, according to the theory,

> This limit is never reached, as the population oscillates indefinitely around the stable state but the oscillations are increasingly weaker. If we set a limit to the amplitude of the oscillations, the limit is reached at the end of a period T that depends on the [mortality and fertility functions] in t_1 and the age structure of the population in t_1.[2] (Bourgeois-Pichat, 1994).

1. The Concept of Stable Population

The concept of stable population is not new. The Swiss mathematician Leonhard Euler defined its principle as far back as 1760 (Lombardo, 1976; Smith and Keyfitz, 1977).[3] In fact, it is Euler's studies that served as a basis, 50 years later, for the entry on *Population* by Thomas Malthus in the fourth edition of the *Encyclopædia Britannica* (Coale, 1979; Bourgeois-Pichat, 1994). From Euler, Malthus drew his conclusion that human populations expand geometrically whereas the production of food increases only arithmetically, which is why Lotka describes a population whose growth rate is constant as Malthusian.

But the international scientific community had forgotten the concept of stable population until Lotka's studies on the eve of World War II revived interest in it.[4] Lotka's contribution, in turn, took time to receive the full attention it deserved. Only Giorgio Mortara, working in Brazil in the 1940s and 1950s, realized its value for understanding population dynamics in the developing countries (Mortara, 1949). Not until the work of the Office of Population Research in Princeton[5] and the Population Division of the United Nations[6] in the early 1960s did stable-population theory pave the way for a proliferation of theoretical research and practical applications, which are now central to the analysis of population dynamics.

2. The Fundamental Equations

As Lotka reminds us (1939), in a closed population subject to a constant mortality regime defined by the probabilities $p(a)$[7] of survival between birth and age a, and displaying a constant age structure $c(a)$, the total size and the number of births increase (or decrease), according to Malthus's law, at a constant rate r, and the birth and death rates—known as intrinsic rates—are themselves constant.

The intrinsic birth rate b is therefore equal to

$$b = \frac{1}{\int_0^\omega e^{-ra} p(a)da}$$

and the proportion $c(a)$ of the population at a given age is equal to

$$c(a) = be^{-ra} p(a)$$

Naturally, if B denotes births, the population $N(a)$ aged a is at all times equal to

$$N(a) = Be^{-ra} p(a)$$

[1] Pourvu que la mortalité et la fécondité á chaque âge aient constamment les valeurs données, quelle que soit la répartition initiale par âge et les taux de natalité et de mortalité, en fin de compte le taux d'accroissement s'approche de plus en plus à la valeur asymptotique ρ, valeur intrinsèque, qui exprime la capacité fondamentale de multiplication de la population, libérée de l'influence perturbatrice d'une répartition par âge arbitraire.

C'est là un aspect du taux intrinsèque d'accroissement ρ qui mérite notre attention toute particulière. Ce taux nous intéresse moins comme instrument de prévision que par le fait qu'il exprime une propriété actuelle d'une population, sa capacité fondamentale de multiplication. Le taux brut d'accroissement—excédent du taux de natalité sur le taux de mortalité enregistré—ne nous donne pas la mesure véritable de cette capacité, car ces taux sont tous les deux influencés par la répartition par âge de la population, et cette répartition elle-même dépend de facteurs pour ainsi dire adventices qui ont agi dans le passé. Entre le taux brut et le taux intrinsèque il peut y avoir un écart importan.

[2] *Cette limite n'est jamais atteinte, la population oscille indéfiniment autour de l'état stable mais les oscillations sont de plus en plus faibles. Si l'on se fixe une limite pour la grandeur des oscillations, cette limite est atteinte au bout d'un temps T qui dépend des* [fonctions de mortalité et de fécondité] *au temps t_1 et de la composition par âge de la population à l'instant t_1.*

[3] An English translation of Euler's text by Nathan and Beatrice Keyfitz was published in *Theoretical Population Biology* (1970; p. 307–314). In the year following its original publication, Euler's study was discussed by Johann Peter Süssmilch (1761): His commentaries, as well, have been translated (Smith and Keyfitz, 1977).

[4] Lotka's 1939 publication, however, had been preceded by preliminary studies (Lotka 1907, 1922; Sharpe and Lotka, 1911) that had led him to this general statement.

[5] Particularly at the behest of Ansley Coale, Paul Demeny, Alvaro Lopez, and, later, Samuel Preston.

[6] Particularly at the behest of Bourgeois-Pichat.

[7] We adopt most of Lotka's original symbols, now standard in all the current literature on stable populations: We use a to denote age, rather than the x used previously.

which is the fundamental function of the stable population (or *Malthusian population*, as Lotka calls it).

However, as Lotka notes, the first two relationships are not independent and are thus inadequate for completely determining the quantities b and r. The first formula, which expresses a relationship between the birth rate and the growth rate, does not suffice to set their values once and for all. For this purpose, we need a second relationship that is independent of the first and incorporates fertility. This brings us to the difficult problem of finding a complete expression of fertility that allows for the sexuated mode of reproduction inherent in the human species. Like Lotka, we will have to restrict our analysis, at least initially, to a single sex—for convenience, the female sex.

Girls born in year t ($B(t)$) are born to mothers who, that year, have an age a falling within the reproductive period (say, 15–49 years). These mothers, in turn, were born in year $(t - a)$, when the number of births of girls was $B(t - a)$. As a result, in an infinitesimal time interval dt, there were $B(t - a)dt$ births of girls, of whom $B(t - a)p(a)dt$ survived in year t, at age $(a + dt)$, equal to $(a + da)$. Let $m(a)$ be the rate of girls per mother[8] aged a. By definition, this function $m(a)$ is, like the function $p(a)$, independent of time. We can thus define the annual number of births of girls in t as a function of the number of births of women in $(t - a)$:

$$B(t) = \int_0^{\omega} B(t-a)p(a)m(a)da$$

To find the growth rate of the stable population, Lotka solved this equation with an exponential series:

$$B(t) = \sum_{i=0}^{\omega} Q_i e^{r_i t}$$

He showed that each r_i satisfies the relationship:

$$1 = \int_0^{\omega} e^{-r_i a} p(a)m(a)da$$

and that the r_i coefficients are accordingly determined as the roots of the equation:

$$1 = \int_0^{\omega} e^{-ra} p(a)m(a)da$$

independent of the initial conditions, whereas the Q_i coefficients depend on those conditions.

[8] We use the expression rate of *girls per mother* to specify that we are talking not about a female fertility rate in the sense we used earlier (i.e., ratio of births of both sexes to females only) but about fertility rates obtained by calculating the ratio of female births only to females.

The roots of Lotka's fundamental equation are not all real, but obviously the function $p(a)m(a)$, product of a probability and a rate, can assume only real, positive values. In consequence, this equation can have only one real root, ρ, which is positive if the net reproduction rate R_0 is greater than 1:

$$R_0 = \int_0^{\omega} p(a)m(a)da > 1$$

negative for $R_0 < 1$, and null for $R_0 = 1$. In other words, its sign depends on whether the number of births increases, decreases, or stays constant from one generation to another.

Moreover, we can define the number of individuals $N(a)$ living in t as the sum of all those who, born in $(t - a)$, survive at age a in t. In t, the size of the population is therefore equal to

$$N(t) = \int_0^{\omega} B(t-a)p(a)da$$

In the case of the real root ρ, the relationship $B(t) = Q_\rho e^{\rho t}$ makes it possible to transform the previous relationship as follows:

$$N(t) = Q_\rho e^{\rho t} \int_0^{\omega} e^{-\rho a} p(a)da$$

in other words:

$$N(t) = Ke^{\rho t}$$

K is a constant. Thus ρ is, quite simply, the population's intrinsic rate of natural increase (or intrinsic growth rate).

As a result, this single real root ρ of Lotka's fundamental equation enables us to completely define the characteristics of the stable population (Lotka, 1939):

$$b_\rho = \frac{1}{\int_0^{\omega} e^{-\rho a} p(a)da}$$

$$c_\rho(a) = b_\rho e^{-\rho a} p(a)$$

$$1 = \int_0^{\omega} e^{-\rho a} p(a)m(a)da$$

3. From the Determination of *r* to the Calculation of the Stable Population

To obtain a concrete estimate of these characteristics from a given mortality function $p(a)$ and a given fertility function $m(a)$, we need to resolve the last of these three relationships, which will give us a measure of the intrinsic growth rate. Different methods have been proposed for this purpose.

Lotka (1939) suggested a three-level approximation. As a first approximation, he obtained

$$r \cong \frac{\log_e R_0}{\mu_1}$$

where R_0, the 0-order moment of the net maternity function $p(a)m(a)$, and μ_1, the order 1 semi-invariant, are in reality the net reproduction rate and the mean age at childbirth. His second approximation relies on the resolution of a quadratic equation, and his third on the analysis of the complex roots of the equation. John Pollard (1973) took Lotka's complex approach and expanded it on certain points.

Other authors have come up with more advanced solutions (see Coale, 1957; Keyfitz, 1968, 1972). However, matrix calculus offers an easier solution to the problem by placing the overall question of the shift to a stable state in another perspective: that of population projections. William Feller understood this advantage as far back as 1941.

II. MOVING TOWARD A STABLE STATE VIA THE PROJECTION MATRIX

It is fairly easy to deduce the properties of the stable population from a population's *projection matrix* (i.e., the matrix whose rows reproduce the overall female fertility rates[9] and whose subdiagonals reproduce the probabilities of survival in the course of each projection stage). This approach was introduced by Harro Bernardelli in 1941 and extensively described by Paul H. Leslie (1945). Given the constancy of the fertility and mortality rates of a stable population, the *projection matrix* remains identical over time. To make a demographic projection,[10] we simply multiply a vector comprising the number of females by age by the projection matrix for each stage of the perspective. As we are dealing with a stable population, the successive multiplication of the projection matrices is equivalent here to multiplying the matrix by itself and therefore to raising it to successive powers. Because of the strong convergence of stable populations, the projection should yield the characteristics of the stable population if we raise the projection matrix to a sufficiently high power.

[9] This is the net overall fertility rate taking into account survivals in the course of each projection stage (see Keyfitz, 1968).

[10] The issue of demographic projections will be discussed as such in several later chapters of this Treatise (see especially Chapter 72).

Let us take an example provided by Nathan Keyfitz many years ago (1964). It consists of a 3 × 3 projection matrix for U.S. females in the three 15-year age groups of 0–14, 15–29, and 30–44 years for the period 1940 to 1955. If we consider that U.S. fertility is complete at age 45, this submatrix is the only one that matters in terms of population dynamics, because women older than 45 years do not contribute to demographic reproduction. We therefore define the initial matrix

$$A = \begin{pmatrix} 0.32167 & 0.68154 & 0.12110 \\ 0.98610 & 0 & 0 \\ 0 & 0.97203 & 0 \end{pmatrix} = \begin{pmatrix} g(0) & g(1) & g(2) \\ p(0) & 0 & 0 \\ 0 & p(1) & 0 \end{pmatrix}$$

The first row gives the female fertility rates and the subdiagonal the probabilities of survival of the 0–14 age group to ages 15–29, and of the latter group to ages 30–44 years. Let us now raise matrix A to successive powers. By A^3 there are no null terms left. Let us now compare the m_{ij} elements of matrices A^{32} and A^{33}. The respective values of these matrices are

$$A^{32} = \begin{pmatrix} 3.53733 & 2.64654 & 0.440336 \\ 3.29245 & 2.46332 & 0.376344 \\ 3.02079 & 2.26008 & 0.345292 \end{pmatrix}$$

and

$$A^{33} = \begin{pmatrix} 3.74761 & 2.80386 & 0.428371 \\ 0.48877 & 2.60975 & 0.398716 \\ 3.20036 & 2.39442 & 0.365818 \end{pmatrix}$$

Let us start by taking the ratios of elements m_{ij}/m_{1j} of one of the two matrices, say A^{33} (we would obtain the same result with A^{32}). We obtain for $j = 1$, 2, and 3, respectively:

1.00000	1.00000	1.00000
0.93077	0.93077	0.93077
0.85397	0.85397	0.85397

We see, therefore, that the columns (as well as the rows, incidentally) of A^{32} and A^{33} are proportional and identical. We also have the size of the U.S. female population projected to periods 32 and 33, using the constant projection matrix A. The projection shows the following population sizes (to within a multiple of 10)**:**

Age group	Population size in period 32	Population size in period 33
0–14	96,131	101,845
15–29	89,476	94,795
30–44	82,093	86,973

Dividing these numbers by the size of first age groups, we obtain the relative structure of the population in the stable state.

Age group	Period 32	Period 33
0–14	1.00000	1.00000
15–29	0.93077	0.93078
30–44	0.85397	0.85397

The *column-vectors* of matrix A^n (large n) are therefore proportional to the relative structure in the stable state, considering that the stable state is nearly reached at the end of 33×15 projection years.

Let us now calculate the ratio of the elements of A^{33} to those of matrix A^{32}. We obtain the following results:

$$\begin{matrix} 1.05945 & 1.05944 & 1.05944 \\ 1.05945 & 1.05944 & 1.05945 \\ 1.05944 & 1.05944 & 1.05945 \end{matrix}$$

We obtain the same constant, 1.05944, nearly each time. It is the factor by which all the population numbers by age are multiplied to shift from one projection stage to the next in 15-year periods. The annual growth rate of the stable population is therefore equal to $1/15 \ln 1.05944$ since a constant growth rate leads to an exponential increase in the population, here 3.8% a year.

Let us now show that these results can be calculated from the eigenvectors of the projection matrix A associated with the main root (or eigenvalue) λ (see Keyfitz, 1968; Wattelar, 1971). Linear algebra shows us how to calculate the vertical (u) and horizontal (v) eigenvectors, respectively, so that:

$$\mathrm{A}u = \lambda u$$

$$v\mathrm{A} = \lambda v$$

associated with the main root λ.

We must begin by estimating the main root λ corresponding to the ratio of the sizes of the stable population at two different moments in time (Keyfitz, 1968). This is obviously equivalent to the ratio of the elements of matrix A^{n+1} to those of the matrix A^n, for a large n. Taking the previous $A(3 \times 3)$ matrix as an example, the ratio of the elements, as we have seen, yields the value $\lambda = 1.05944$ for $n = 32$. This value is unique, once stability is reached. It corresponds to the intrinsic growth rate of the stable population.

We can also calculate the same result from the characteristic equation of the projection matrix

$$\begin{vmatrix} g(0)-\lambda & g(1) & g(2) \\ p(0) & -\lambda & 0 \\ 0 & p(1) & -\lambda \end{vmatrix} = 0$$

which gives

$$\lambda^3 + \lambda^2\, g(0) + \lambda\, g(1)\, p(0) + g(2)\, p(0)\, p(1) = 0$$

This equation has only one real positive root that corresponds to the stable population's intrinsic growth rate. The uniqueness of the solution is easily demonstrated (Goodman, 1968). Let us write the characteristic equation in the general form:

$$\lambda^{k+1} - \sum_{y=0}^{k} h(y)\lambda^{k-y} = 0$$

where $h(y)$ represents the products $g(y)s(y)$, with $s(y) = \prod_{j=0}^{y} p(j)$, for example, in the characteristic equation:

$$h(2) = g(2)\, p(0)\, p(1)$$

We can also write:

$$\sum_{y=0}^{k} h(y)\lambda^{k-y} = \lambda^{k+1}$$

Dividing the two members by λ^{k+1}, we obtain:

$$\sum_{y=0}^{k} \frac{h(y)}{\lambda^{y+1}}\lambda = 1$$

As the left-hand term decreases from infinity to 0 when λ increases from 0 to infinity, the characteristic equation will admit only one real positive root λ.

Returning to the projection matrix:

$$A = \begin{pmatrix} g(0) & g(1) & g(2) \\ p(0) & 0 & 0 \\ 0 & p(0) & 0 \end{pmatrix}$$

where $g(k)$ are the fertility rates and $p(k)$ the probabilities of survival, and applying the theory of homogeneous linear equation systems with null determinant, we obtain the solution:

$$u(0) = 1$$

$$u(1) = p(0)/\lambda$$

$$u(2) = p(0)p(1)/\lambda^2$$

for the vertical eigenvector u. The horizontal eigenvector v will be equal to

$$v(0) = \lambda$$

$$v(1) = [\lambda - g(0)]/u(1)$$

$$v(2) = \frac{\lambda - [g(0)u(0) + g(1)u(1)]}{u(2)}$$

With the data of the U.S. projection matrix used earlier, we obtain the following for the vertical eigenvector:

$$u = \begin{pmatrix} 1 \\ 0.93077 \\ 0.85398 \end{pmatrix}$$

and for the horizontal eigenvector:

$$v = (1.05944 \ 0.79264 \ 0.12110).$$

We see that the vertical eigenvector u provides the relative stable structure, obtained earlier by raising A to a sufficiently high power. As for the horizontal eigenvector v, it is the discrete expression of the female *reproductive value*, a concept developed by Ronald Fisher (1958), which represents the present value of future births for females at each age i; the present value is obtained by applying the discount rate r, equal to the intrinsic growth rate of the stable population.

Note that Andrei Rogers (1966) generalized this approach to open populations. He divided the population into subsets subject to internal migrations, in addition to their specific fertility and mortality rates. The principle of strong convergence (or ergodicity), specific to the stable model, is also applicable here if the interregional projection matrix remains constant over time.

III. A SPECIAL CASE: THE STATIONARY POPULATION

There are as many stable populations as there are possible combinations between the different fertility and mortality functions imaginable. Depending on the resulting equilibrium between the force of mortality and the fertility rates, the intrinsic growth rate can be—as we have seen—positive, null, or negative.

1. The Stationary Population

When r is null, not only is the structure constant, but the population sizes, the number of births and the number of deaths are as well. The population is said to be *stationary*. It is entirely determined by its forces of mortality. This stationary population is, therefore, quite simply the L_x function of the life table (i.e., the *life-table population*. As we know (Chapter 11) that life expectancy is equal to

$$e_0 = \frac{\int_0^{\omega} l(a)da}{l_0}$$

and that the crude death rate is equal to

$$m = \frac{l_0}{\int_0^{\omega} l(a)da}$$

we see immediately that, in this particular case, life expectancy at birth is equal to the inverse of the intrinsic death rate and, since the growth rate is null, it too is equal to the inverse of the intrinsic birth rate

$$m = b = \frac{1}{e_0}$$

And, at any age x, life expectancy at this age is equal to the inverse of the crude death rate above age x:

$$e_x = \frac{1}{{}_{(\omega-x)}m_x}$$

Thomas Greville (1943; see also Chapter 14) uses this property to close a life table.

2. Growth Potential of the Population

One of the possible applications of the stationary model is to measure what Paul Vincent (1945) has called the *potential for increase* and various English-language authors have used the expression *population momentum*, with a fuller, more general mathematical elaboration (Keyfitz, 1971, 1985; Preston and Guillot, 1997). The idea is simple: By how much will the population vary between the moment when its fertility and mortality functions freeze at the levels strictly needed for cohort replacement (R_0 = 1) and the moment when it reaches the stationary state? The question was first raised in connection with the decline in fertility in the developing countries. Would a very swift decline indicate the end of the population explosion often denounced as an obstacle to development? No, because the populations of those countries had accumulated a high growth potential in their very young age structures. We can also ask the same question today in the developed countries, where the fall in fertility and greater longevity have produced increasingly older age structures: by how much would the population decline if, by miracle, fertility were to recover and settle at the 2.08 children per female needed—in present mortality conditions—to attain the stationary state?

Samuel Preston and Michel Guillot (1997), using Vincent's concept and the work of Keyfitz, have generalized the mathematical definition of the potential and arrived at the following formula:

TABLE 20–1 Growth Potential Observed in 1997 in Main Regions of World and Selected Countries

Region or country	Growth potential
Africa	1.56
West Asia	1.56
Southeast Asia	1.48
Latin America	1.48
South Central Asia	1.47
East Asia	1.22
North America	1.10
Europe	0.98
Austria	0.96
Russia	0.94
Italy	0.91
Germany	0.88
World	1.35

Source: Preston and Guillot, 1997. Population dynamics in an age of declining fertility synthesis, *Genus*, vol. 53(3–4), p. 15–31.

$$M = \int_0^{\omega} \frac{c(a)}{c_s(a)} \cdot w(a)da$$

where $c(a)$ is the age structure of the initial population, $c_s(a)$ that of the population in the stationary state, and $w(a)$ the ratio of net reproduction above age a to the mean age at childbirth in a stationary regime.[11]

It is thus clear that a population's potential for increase is a function of the age-structure disparities between the initial situation and the stationary state. Table 20–1 shows the potentials calculated by Preston and Guillot (1997) for the main regions of the world and selected countries of interest. If the African population had reached stationarity conditions by 1997, it would nevertheless have continued to grow by 56% before becoming stationary. Conversely, if Germany's fertility had returned by the same date to the level needed for strict generation replacement, its population would nevertheless have had to fall 12% before stabilizing.

IV. SOME APPLICATIONS OF THE STABLE POPULATION MODEL

The stable population model has been highly successful and has been applied in many ways in a wide range of fields, from correcting imperfect data.[12] (Coale and Demeny, 1967) to modeling the acquired immunodeficiency syndrome (AIDS) epidemic (Brouard, 1987, 1994). We will limit ourselves to two very different examples here.

1. The Coale–Demeny Models

The most popular tool based on the theory of stable populations is unquestionably the network of Coale–Demeny models (1966, 1983). Coale and Demeny began with four families of model life tables[13] built from real data,[14] and a set of age-specific fertility functions. They derived a vast set of stable populations defined by different combinations of mortality and fertility levels and structures. This set comprises four subsets designated as *West model, East model, North model*, and *South model*, according to the origin of the real data used, although this does not impart any geographic significance to the models. By way of illustration, Table 20–2 lists some population-dynamics indicators (crude birth and death rates, growth rate, and mean age of the population) obtained for selected combinations of mortality and fertility described by the West model.

As expected, the higher the gross reproduction rate, the higher the birth rate (Table 20–2 and Figure 20–1). However, for an identical gross reproduction rate, as the table shows, the birth rate varies according to life expectancy at birth. This variation, in turn, depends on the fertility level: It is near zero when the reproduction rate is below 2, and gradually widens as fertility rises. But the variation is not truly significant except at implausible fertility levels, when the gross reproduction rate exceeds 4 or 5.

For an identical life expectancy, the crude death rate depends, in turn, on fertility. Whatever the life expectancy at the lowest fertility levels, the death rate decreases when fertility increases (Figure 20–2). But when the gross reproduction rate exceeds 2, the death rate varies according to life expectancy. If the latter is below 50 years, the crude death rate increases with fertility, whereas if life expectancy exceeds 50 years, the rate stagnates or actually keeps declining. The mean age of a stable population depends very little on life expectancy, even if some differences appear when the gross reproduction rate exceeds 2 (Figure 20–3). For all gross reproduction rates above 1, the mean age rises with life expectancy. For a gross reproduction rate of 2.5, the mean age falls from 28.5 years with a 20-year life expectancy to 24 years for an 80-year life expectancy.

[11] Or $w(a) = \dfrac{\int_a^{\omega} p(x)m(x)dx}{\bar{x}}$

[12] On this point see also Chapter 127.

[13] The concept will be discussed in greater detail in Chapter 128.

[14] See Chapter 128.

TABLE 20–2 Selected Indicators Characterizing the Stable Populations of the Coale–Demeny West Model: Females

Parameters[a]	Gross reproduction rate								
	0.8	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**4.0**	**5.0**	**6.0**
					$e_0 = 20$ y				
b	9.6	13.3	22.6	31.5	39.9	47.7	61.4	73.1	83.3
m	56.9	53.1	48.7	47.7	48.2	49.5	52.9	56.5	60.0
r	–0.5	–0.4	–0.3	–0.2	–0.1	0.0	0.1	0.2	0.2
$\bar{x}$	44.5	41.5	35.7	31.6	28.5	26.2	22.7	20.3	18.5
					$e_0 = 30$ y				
b	9.7	13.4	22.5	31.0	38.8	45.9	58.1	68.5	77.4
m	43.7	39.8	35.0	33.5	33.3	33.9	35.7	37.8	39.9
r	–0.3	–0.3	–0.1	0.0	0.1	0.1	0.2	0.3	0.3
$\bar{x}$	43.8	40.5	34.2	29.9	26.7	24.4	21.0	18.7	17.0
					$e_0 = 40$ y				
b	9.7	13.4	22.3	30.5	37.9	44.5	55.9	65.4	73.5
m	34.9	30.9	25.8	23.9	23.4	23.4	24.3	25.5	26.7
r	–0.3	–0.2	0.0	0.1	0.1	0.2	0.3	0.4	0.5
$\bar{x}$	43.5	39.9	33.2	28.8	25.6	23.3	20.0	17.8	16.2
					$e_0 = 50$ y				
b	9.7	13.3	22.1	30.0	37.1	43.4	54.2	63.1	70.7
m	28.4	24.4	19.1	16.9	16.0	15.7	15.9	16.5	17.2
r	–0.2	–0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.5
$\bar{x}$	43.3	39.5	32.6	28.1	24.9	22.5	19.3	17.2	15.6
					$e_0 = 60$ y				
b	9.7	13.3	21.9	29.7	36.5	42.6	52.9	61.4	68.5
m	23.4	19.3	13.9	11.4	10.3	9.7	9.4	9.5	9.7
r	–0.1	–0.1	0.1	0.2	0.3	0.3	0.4	0.5	0.6
$\bar{x}$	43.1	39.1	32.0	27.5	24.3	21.9	18.7	16.7	15.2
					$e_0 = 70$ y				
b	9.6	13.1	21.7	29.3	36.0	41.9	51.8	60.0	66.9
m	19.4	15.3	9.7	7.1	5.8	5.1	4.4	4.2	4.1
r	–0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.6
$\bar{x}$	43.3	39.2	31.9	27.2	24.0	21.6	18.4	16.4	14.9
					$e_0 = 80$ y				
b	8.9	12.4	20.9	28.5	35.2	41.1	51.0	59.1	65.9
m	16.8	12.6	7.0	4.4	3.0	2.3	1.5	1.2	1.1
r	–0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.6
$\bar{x}$	45.6	41.1	33.0	27.9	24.4	21.9	18.6	16.4	14.9

[a] *b*, birth rate; *m*, death rate; *r*, natural growth rate; $\bar{x}$, mean age of population.
Note: This table merely reproduces some of the combinations of mortality and fertility supplied by the Coale–Demeny models.
Source: Coale, Demeny, and Vaughan, 1983. *Regional model life tables and stable populations*, 2nd ed. New York, Academic Press, 496 p.

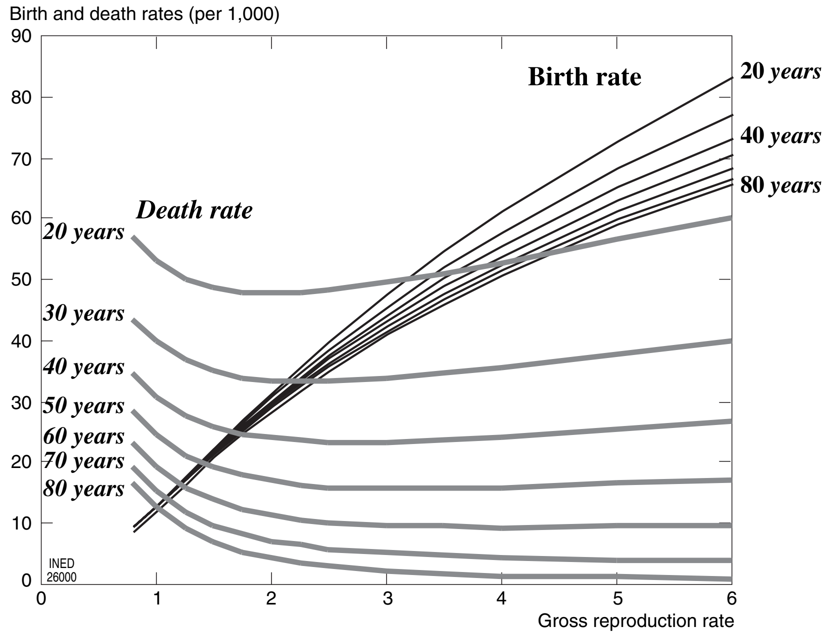


FIGURE 20–1 Crude birth and death rates as a function of gross reproduction rate for selected life expectancies at birth, based on stable populations of Coale–Demeny West model (females).

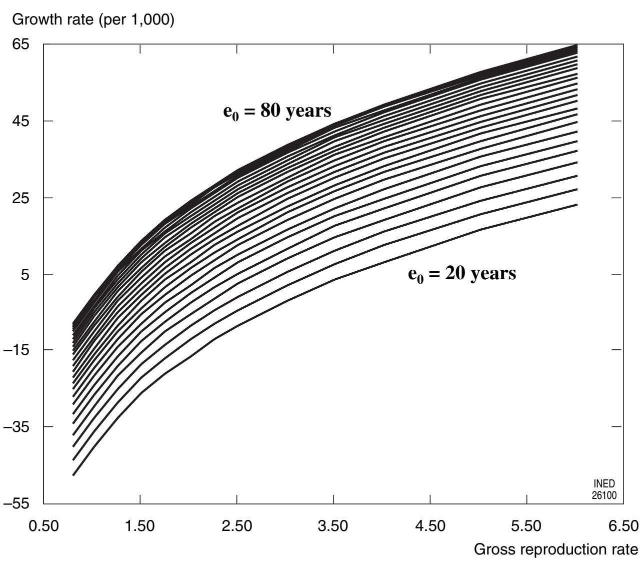


FIGURE 20–2 Growth rate as a function of gross reproduction rate for selected life expectancies at birth, based on stable populations of Coale–Demeny West model (females).

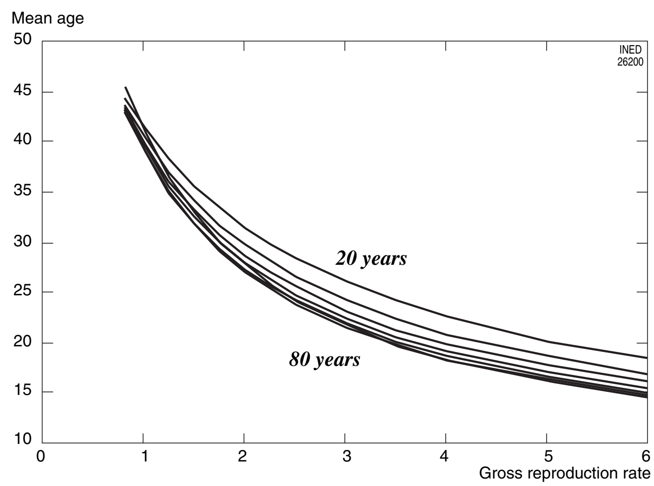


FIGURE 20–3 Mean age of population as a function of gross reproduction rate for selected life expectancies at birth, based on stable populations of Coale–Demeny West model (females).

The mean age is much more sensitive to fertility. Whatever the life expectancy, the mean age literally collapses when the gross reproduction rate rises from 0.8 to 2.5 or 3 (Figure 20–3).

Figure 20–4 gives another perspective on the same relationships by plotting birth and death rates as a function of life expectancy at birth, for selected values of the gross reproduction rate. As long as the gross reproduction rate is below 2, the birth rate remains virtually identical regardless of life expectancy. Even at higher fertility levels, the variation is modest. By contrast, the death rate varies very sharply, as one might expect, with life expectancy. However, this variation depends to a significant degree on fertility: the decline in the crude death rate with the rise in life expectancy is all the steeper as fertility is higher.

Stable-population models are a powerful tool for analyzing real data, in many ways. First, they allow a

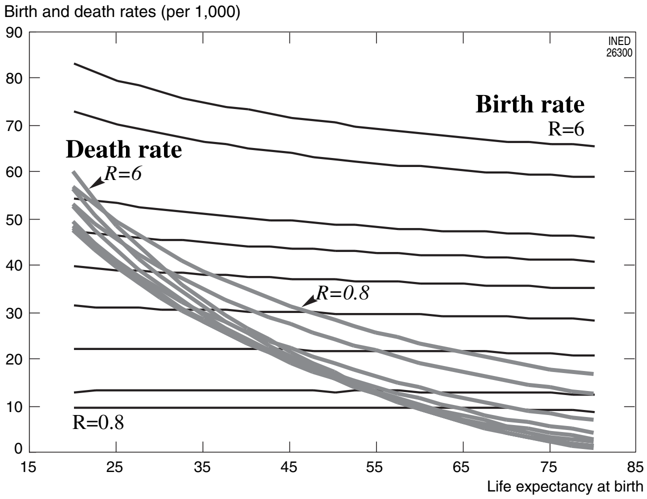


FIGURE 20–4 Crude birth and death rates as a function of life expectancy at birth for selected gross reproduction rates, based on stable populations of Coale–Demeny West model (females).

test of the internal consistency of the data gathered. Some combinations of results are perfectly impossible, and the comparison with models is often the best way to determine this. Better, the models and the underlying theory serve as the basis for building special tools to adjust imperfect data or fill gaps.[15]

Stable-population models are also a starting point for examining the sequence of events that constitute the real history of populations. We will give only one example here, supplied by Gustavo De Santis (1997), who compares (1) the nomograms of the population growth rate for different fertility–mortality combinations proposed by Ansley Coale and Susan Watkins (1986) working from Coale–Demeny models and (2) the values of that rate at the same levels of fertility in Italy since 1861 and in the world since 1950 (Figure 20–5).

The chart shows that with a nearly constant fertility rate of about five children per female from 1861 to 1911, and despite a major increase in life expectancy, the Italian population experienced an initial destabilization due to the progressive shift in the isogrowth line to higher growth zones. Since 1911, the Italian population has backtracked: In the early 1970s, it even seemed to be settling around zero growth, despite continuous gains in life expectancy; it then changed course abruptly in the mid 1980s, sinking into negative-growth zones. Meanwhile, the pattern in the world population, actually observed from 1950 to 1995 and projected through 2050, effectively illustrates two phenomena: first, the growth-acceleration phase combining high fertility and increased life expectancy (1950–1970); second, the rapid deceleration that seems to justify the United Nations projection based on the hypothesis of a shift to zero growth toward the middle of this century.

[15] See especially Chapter 127 on indirect estimation methods.

2. Example of Transition to Stationary State

One of the strengths of the stable-population theory is that it provides a way to rapidly illustrate the longer term evolution of a population whose current fertility and mortality conditions are held constant indefinitely. Its structure and growth rate in the stable state are known in advance, but we still need to determine how long it will take to reach that state, and with what fluctuations. This involves experimentation on concrete cases—a very standard exercise in population projection. A later chapter (Chapter 78) discusses the very–long-term projections of the world population, where different types of transition to the stable state will be examined. For the moment, we shall simply borrow an older example from Jean Bourgeois-Pichat and Si-Ahmed Taleb (1970), but reversing their approach. In other words, we explore what fertility and mortality conditions need to be specified for a population to take it to a given stable state.

In the days when the third world population boom was at its peak, the focus on the goal of stabilizing population size—in the everyday sense of the term—drew special attention to the specific case of stationary populations. What was needed to bring a fast-growing population to zero growth as soon as possible? The stable-population models instantly provide the fertility target needed for a given mortality level. However, the shift from the initial state to the stationary state requires a certain period of time, whose length depends on the difference between initial fertility and the fertility leading to the stationary state. Can this interval be shortened? Certainly not. But there is nothing, at least in theory, to stop us from anticipating the final result by immediately freezing the growth rate at zero. We will not reach the stable state earlier. On the contrary, the growth-rate freeze will trigger massive swings in all the other parameters, and these fluctuations will take time to absorb.

In 1970—following a call by General Draper, U.S. Representative to the United Nations (UN) Population Commission, for a shift to zero growth by the end of the 20th century—Bourgeois-Pichat and Si-Ahmed Taleb took the example of Mexico to illustrate the wide gap between the fulfillment of General Draper's wish and a less unrealistic process that would lead Mexican women to adopt, by the early 21st century, the fertility rate needed to reach stationarity in the long run. For simplicity's sake, the authors confined their discus-

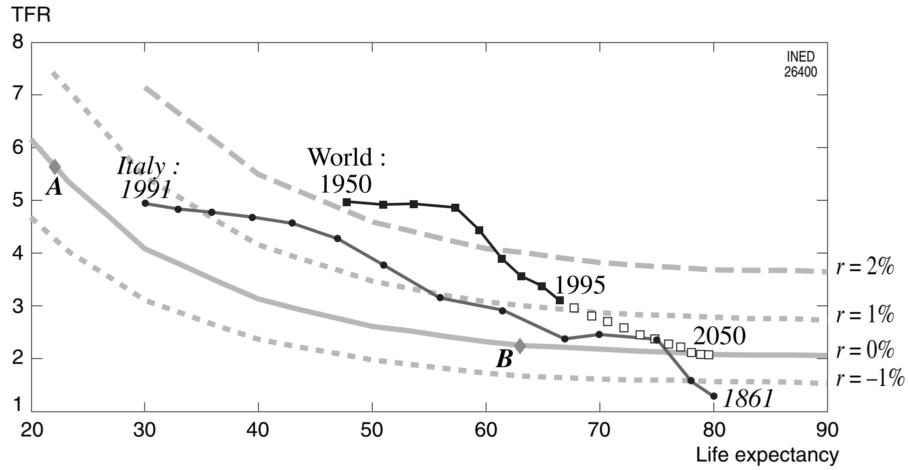


FIGURE 20–5 Long-term changes in growth rate of Italian female population since 1861 and of world female population since 1950 as a function of life expectancy and fertility, compared with Coale–Watkins growth-rate nomogram. (Source: De Santis, 1997. *Demografia ed economia*. Bologne, Il Mulino, 354 p. [Coll. Stuci e Ricerche].)

sion—as in stable-population theory—to the female population.

Figure 20–6 shows, as one might have expected, that the first solution would stabilize the Mexican population much sooner and at a much lower level than the second. In the first scenario, the population would level off, by definition, by 1995 to 1999, at twice its 1960 level; in the second scenario, the population would keep rising until ca. 2045 to 2049, reaching four times its initial size. For neo-Malthusians in a hurry, the result is tempting—but at what price?

Figure 20–7 shows the swings to which the crude birth and death rates would be subjected in an abrupt transition to zero growth (Chart a). In fact, the chart shows only the first of these swings—followed by a long series until the stationary state is reached. If instead, we assume that the fertility conditions obtaining in that stationary state will be realized in the near future, the birth rate and death rate will—on the contrary—gradually converge (Chart b). The equalization of the rates requires more time but quickly produces near-stability.

Figure 20–8 illustrates the violent swings in the size of a special age group—the school-age population—triggered by the realization of the first hypothesis. One can imagine the headaches for planners! The second hypothesis admittedly leaves room for an initial phase of rapid expansion of the school-aged population, but the population reaches a ceiling fairly quickly, after which the swings are very mild.

This example has taken us some distance from the pure theory of stable populations, but gives us a clear picture of the gulf between the theoretical conditions of stability and its practical fulfillment.

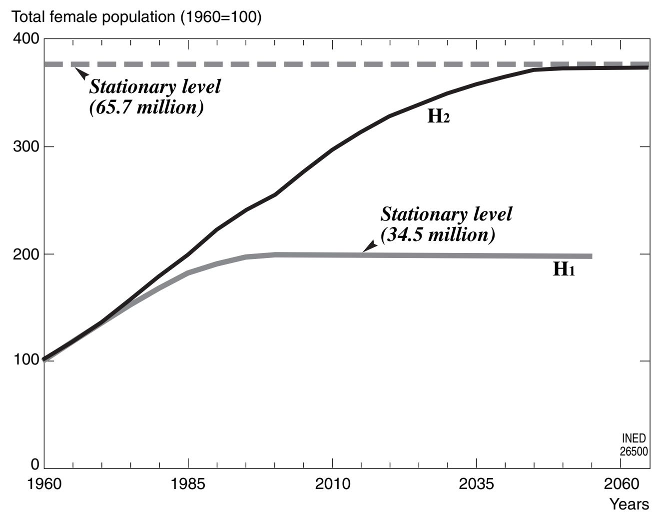


FIGURE 20–6 Century projection of total female population of Mexico in scenarios of Bourgeois-Pichat and Taleb (1960 = 100 = 17.5 million). (Source: Bourgeois-Pichat and Taleb; 1970. Un taux d'accroissement nul pour les pays en voie de développement en l'an 2000: rêve ou réalité? *Population*, vol. 25(5), p. 957–974.)

V. THE PROBLEM OF THE TWO SEXES

While highly useful for many applications, the theory of stable populations unfortunately suffers from a serious flaw: it is based on the simplifying assumption that the population consists of a single sex that reproduces itself, as it were, by parthenogenesis. For the sake of convenience, of course, demographers choose the female sex, since the general practice is to compute age-specific fertility rates for females. This is because the necessary information is easier to compile:

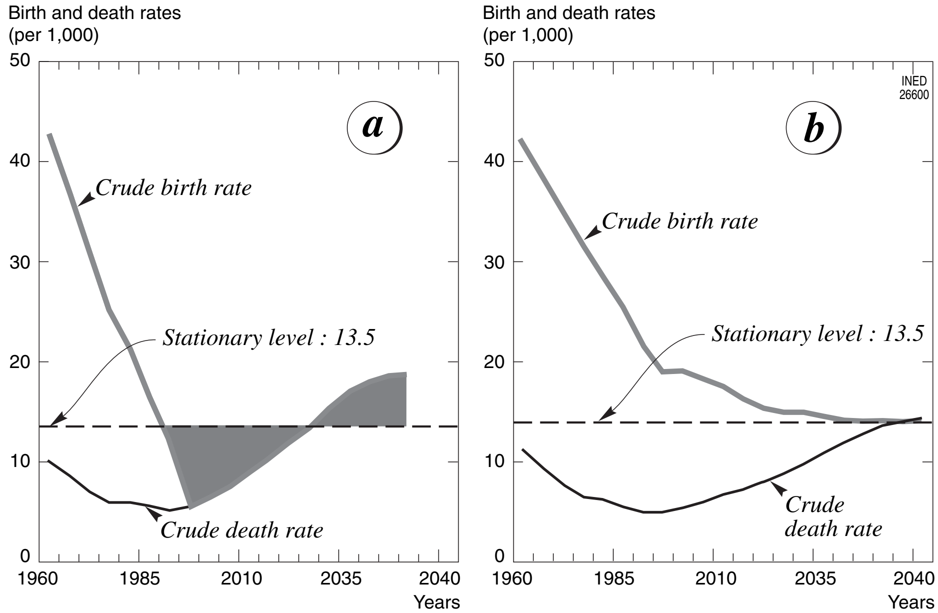


FIGURE 20–7 Changes in crude birth rate **(A)** and crude death rate **(B)** of Mexican female population, on basis of Bourgeois-Pichat and Taleb's two projections (1960 = 100 = 17.5 million). (Source: Bourgeois-Pichat and Taleb, 1970. Un taux d'accroissement nul pour les pays en voie de développement en l'an 2000: rêve ou réalité? *Population*, vol. 25(5), p. 957–974.)

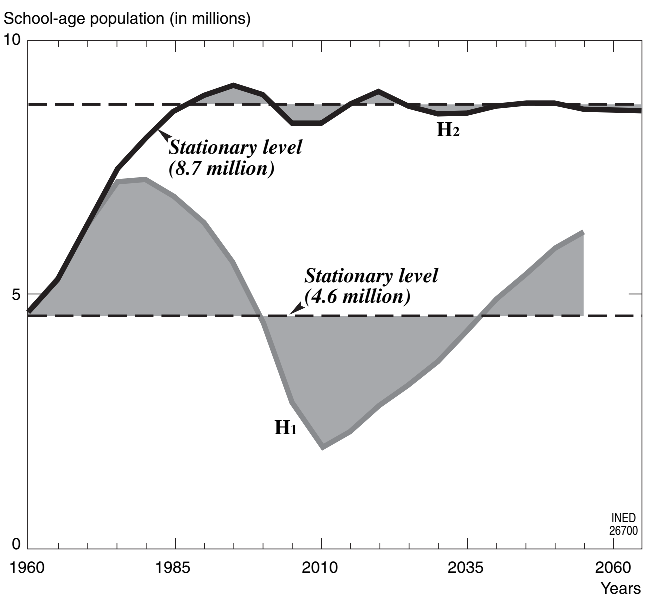


FIGURE 20–8 Change in school-aged population (5–14 years) in Mexican female population, on basis of Bourgeois-Pichat and Taleb's two projections (1960 = 100 = 17.5 million). (Source: Bourgeois-Pichat and Taleb, 1970. Un taux d'accroissement nul pour les pays en voie de développement en l'an 2000: rêve ou réalité? *Population*, vol. 25(5), p. 957–974.)

the reproductive period is shorter, and filial relation is easier to identify for births out of wedlock.

If male fertility rates are available—and they are seldom calculated even though it is often possible to do so (Brouard, 1977)—one could obviously repeat the process with an exclusively male population. On this point, it is interesting to note that, in 1907, when F. R. Sharpe and Alfred Lotka tried for the first time to model the growth of a human population, they used a male population and assumed that the growth of the female population simply had to be compatible with the hypothesis of constant male fertility and mortality.

In a very summary manner, demographers typically skirt the issue by linking a male population to the female population analyzed, on the assumption of a constant sex ratio at birth h (Bourgeois-Pichat 1994). This gives $B_f(t) = b_f N_f(0)e^{rt}$ for female births and $B_h(t) = hb_f N_f(0)e^{rt}$ for male births. The associated male population is therefore a Malthusian population with the same rates as the female population.

Clearly this procedure does nothing to solve the basic problem, which is that a population's fertility cannot be reduced to that of males or females only, but depends on the joint action of both sexes in couples: if we separate the sexes in the same initial population, we will end up with two stable models characterized by two different intrinsic growth rates. For example, Keyfitz (1968), reporting results for several countries, shows that, with the U.S. fertility and mortality of

1964, the intrinsic rate would have been 15.7/1000 using the female population and 17.5/1000 using the male population.

The problem is to determine whether we can refine the model to obtain the true intrinsic rate of the total population taking into account the behavior of each sex.

David Kendall (1949) was probably the first to raise the issue and attempt to solve it. He began by demonstrating the absurdity of a population projection based on an application to each sex of its own intrinsic rate. This would necessarily cause a divergence between the two fractions of the population, with a continuous shift in the sex ratio. Starting, for example, with the U.S. population of 1964, after 100 years, the number of males would be 19% greater than that of females, and after 200 years it would be 43% greater. Such a projection would ultimately eliminate one of the two sexes. Admitting different growth rates for the two sexes not only divorces us from the stable-population model from the outset, but also ultimately yields absurd results.

Leo Goodman (1953, 1967) has examined two approximate solutions, one with female dominance, the other with male dominance, depending on whether one applies female or male fertility. In either case, each sex is assigned its own mortality but, in the first case, male and female births depend on the proportion of females in the population; in the second, they depend on the proportion of males. In either case, the growth rates of the male and female populations converge and the sex ratio stabilizes, but at significantly different levels. The author then proposes a compromise solution, where he adjusts each of the two previous approaches by a factor that depends on the square root of the products of the proportions of males and females. In this case the difference between the two approaches is eliminated but the intrinsic sex ratio does not necessarily lie somewhere in between the two preceding ones; indeed, it can even diverge from these values sharply.

Another solution was proposed by Alfred Pollard (1948), in which the sons depend on the fathers and girls on the mothers. The intrinsic growth rate is given by:

$$r = \frac{T^*\rho^* + T\rho}{T^* + T}$$

where $T\times$ and T are the male and female intervals between two successive generations and $\rho\times$ and ρ the intrinsic male and female rates. This model, which is very ingenious as it eliminates all internal conflict, was expanded by Luitzen Yntema (1952).

These different hypotheses have also been discussed in the context of stochastic models (Goodman, 1967; Keyfitz, 1977). However, a full description of the human reproduction process would have to include the formation of couples and the measurement of their fertility. In societies where marriage is the sole locus of expression of fertility, one can imagine a model based on nuptiality[16] and marital fertility.[17] Indeed, some authors have tried to incorporate these two components into stable-population models (Pollard, 1948; Karmel, 1949; Keyfitz, 1977, 1985). But nuptiality poses two problems: the first, which is analogous to the one discussed earlier in this chapter, is that nuptiality is measured separately for the two sexes; the second is the fundamental difficulty of achieving consistency between the measurement of male and female behavior (Henry, 1968).

VI. FROM SEMISTABLE AND QUASI-STABLE POPULATIONS TO GENERALIZED STABLE EQUATIONS

Lotka's theory shows that if we apply constant fertility and mortality to any initial population, the population converges toward the stable state and a constant age structure independent of the initial age structure.

In the late 1960s, Bourgeois-Pichat showed that the stable-population equations could also apply, in some cases, to variable-growth populations. In particular, he distinguished between what he called semistable populations and quasi-stable populations (Bourgeois-Pichat 1966, 1971; United Nations 1968, 1972).

Subsequently, Neil Bennett and Shiro Horiuchi (1981), followed by Preston and Coale (1982), have shown that Lotka's model could be generalized to cases in which an initial population is assigned time-varying mortality and fertility functions: here as well, the age structure—which varies over time in this case—will nevertheless eventually be redefined independently of the initial population's age structure. This generalized model, often called the *r*-variable model, can also be defined with or without migrations.

1. Semistable and Quasi-stable Populations

In a posthumous volume, INED published the manuscript chapters that Bourgeois-Pichat had drafted to summarize his work on stable populations in preparation for a book that he unfortunately did not live

[16] See the Chapters 25, 26 and 27.
[17] See Chapter 29.

to complete. The gathering notably includes several chapters on semistable and quasi-stable populations (Bourgeois-Pichat, 1994).

a. Semistable Populations

The semistable population is a mathematical concept. It is a population whose age structure stays constant despite the variation in its growth parameters. When its fertility and mortality freeze, it therefore reaches the stable state immediately, whereas an ordinary population will only reach it—under the identical hypothesis—after some time. Bourgeois-Pichat (1994, p. 169) defines the semistable population's fundamental property as follows:

> the age-specific survival function *p(a,t)*, the age-specific fertility *f(a,t)* [*m(a,t)* in Lotka's notation], and the rate of change *r(t)* depend on time *t*, but are linked in a given instant t by relationships that are those of the stable population in *t*.

After demonstrating the concept's mathematical validity, Bourgeois-Pichat identified its usefulness as follows:

> [i]n reality, we observe many populations that, over time, preserve age structures subject to little if any variation. We may therefore qualify these as semistable populations. In consequence, we can apply to these real populations the formulas defined for the stable populations[18] (Bourgeois-Pichat, 1994, p. 116).

b. Quasi-stable Populations

The quasi-stable population is, on the contrary, an experimental concept that rests on the following observational fact: If we project a population with a constant age-specific fertility function but a mortality whose fluctuations are bounded by a network of model life tables, its age structure will vary little. In the same way as we can speak of a stationary population linked to a life table, we can speak of a semistable population linked to a network of model life tables. In fact, experience shows that a network of stable populations calculated on a universe of model tables can itself be virtually equated with the different stages of the quasi-stable population linked to this network (Bourgeois-Pichat, 1994). This is, therefore, another concrete field of application for stable-population equations.

[18] La fonction de survie selon l'âge a, *p(a,t)*, la fécondité par âge *f(a,t)* [*m(a,t)* dans la notation de Lotka] et le taux de variation *r(t)* dépendent du temps *t*, mais sont liés à un instant donné *t* des relations qui sont celles de la population stable à l'instant *t*.

dans la réalité, on observe beaucoup de populations qui conservent, au cours du temps, des compositions par âge, sinon invariables, du moins peu variables. On peut donc les assimiler à *des populations semi-stables. On peut alors appliquer* à *ces populations réelles les formules établies pour les populations stables.*

2. Generalized Equation in a Closed Population

In a period approach (Preston and Coale, 1982; Wunsch, 1989), with $D(a,t)$ and $N(a,t)$ as the deaths and the population aged a in t, we can write that $D(a,t) = -dN(a,t)$. According to the theory of partial derivatives, the total differential $dN(a,t)$ is equal to:

$$dN(a,t) = \frac{\partial N(a,t)}{\partial a} da + \frac{\partial N(a,t)}{\partial t} dt$$

which corresponds to the increases in age and time da and dt. The growth rate $r(a,t)$ of the population $N(a,t)$ as a function of time is therefore:

$$r(a,t) = \frac{1}{N(a,t)} \cdot \frac{\partial N(a,t)}{\partial t} dt$$

The instantaneous death rate in t, $D(a,t)/N(a,t)$, is equal to (Bennett and Horiuchi, 1981):

$$\begin{aligned}\mu(a,t) &= \frac{D(a,t)}{N(a,t)} \\ &= \left[\frac{1}{N(a,t)} \cdot \frac{\partial N(a,t)}{\partial t} da + \frac{1}{N(a,t)} \cdot \frac{\partial N(a,t)}{\partial t} dt\right]\end{aligned}$$

hence

$$\mu(a,t) = -\left[\frac{1}{N(a,t)} \cdot \frac{\partial N(a,t)}{\partial t} da + r(a,t)\right]$$

and last,

$$\frac{1}{N(a,t)} \cdot \frac{\partial N(a,t)}{\partial a} da = -[\mu(a,t) + r(a,t)]$$

Turning to the inverse functions and integrating from a to x, we obtain:

$$N(x,t) = N(a,t)e^{-\int_a^x \mu(x,t)dx - \int_a^x r(x,t)dx}$$

It will be recalled (Chapter 11) that $\exp\left[-\int_a^x \mu(x,t)dx\right]$ is the probability $_{(x-a)}p_a(t)$ of surviving from a to x measured on a period basis for year t. Hence:

$$N(x,t) = N(a,t)_{x-a}p_a(t)e^{-\int_o^a r(x,t)dx}$$

But, for a = 0, $N(a,t) = N(0,t)$, in other words, births $B(t)$ in year t. As a result:

$$N(a,t) = B(t)_a p_0(t)e^{-\int_o^a r(x,t)dx}$$

Within a birth cohort, we therefore have $N(a,t) = N(0,t-a)_a p_0(a)$, where $_a p_0(a)$ is the probability of survival from birth to age a in the cohort aged a in t and therefore born in $t - a$. In this case we must have the equality:

$$N(a,t) = B(t)_a p_0(t) e^{-\int_o^a r(x,t)dx} = N(0,t-a)_a p_0$$

and as

$$N(0,t-a) = N(0,t) e^{\int_{t-a}^{t} r(0,z)dz}$$

allowing for the variation in births between the instants t and $t - a$ (Arthur and Vaupel 1984), we obtain the generalized stable equation in the absence of migration:

$$N(a,t) = N(0,t)_a p_0(a) e^{-\int_{t-a}^{t} r(0,z)dz}$$

The demographic characteristics of the populations subject to this equation are

Crude birth rate:

$$b(t) = \frac{B(t)}{\int_0^\omega N(a,t)da} = \frac{1}{\int_0^\omega p(a,t) e^{-\int_0^a r(x,t)dx}}$$

Age structure:

$$c(a,t) = \frac{N(a,t)}{\int_0^\omega N(a,t)da} = b(t) e^{-\int_0^a r(x,t)dx} \cdot p(a,t)$$

and, introducing the age-specific fertility rates $m(a,t)$ in t, we obtain the *characteristic equation* that enables us to determine the current crude rate of increase:

$$1 = \int_0^\omega e^{-\int_0^a r(x,t)dx} p(a,t) m(a,t) da$$

These three relationships are the generalization of the properties prevailing in the stable state. It is easy to find these properties if we remember that, in the stable state, the growth rate $r(x,t)$ is constant over time. The characteristics of the stable population are thus merely a special case of these general relationships that apply to any closed population.

3. Extension to Migrations and Generalization to all Attrition or Accession Processes

We can take generalization one step further by incorporating the effects of migration flows.

a. General Formula

For this purpose, we must introduce a new function, the force of net emigration (Preston and Coale, 1982). Let $e(x,t)$ be this force at age x and in t. We have

$$N(a,t) = N(0,t)_a p_0(t) e^{-\int_a^x r(x,t)dx} e^{-\int_a^x e(x,t)dx}$$

In fact, as Preston and Coale (1982; p. 221, 227) write

> [N]othing limits us to recognizing only one form of *migration* or even one form of mortality. Any form of attrition or accession can be introduced into [the last equation] simply by recognizing that it must act analogously to migration or mortality from all combined causes. [The equation] is the basis of a surprisingly general set of relations. In particular, one can see that the age composition of any population at any moment (assuming only that age composition and its change through time are continuous) is completely determined by the rate of increase in the number at each age at the given moment, together with the rate of attrition (including negative attrition) at each age from each of a number of independently operating factors. To be more specific, if the rate of increase *r(x)* is known for each age x from 0 to the highest age attained, and if the values of i different attrition factors *i(x)* are also known, the age composition is completely determined and can readily be calculated; conversely, if the age distribution and all but one of the attrition factors are known, the rate of attrition for the omitted factor can readily be calculated.
>
> This set of relations is known in demography for particular instances, and the basic equation in differential form is familiar in mathematical biology and actuarial work, but the full (though simple) generalization seems to have escaped attention.

The general fundamental equation is:

$$N(a,t) = N(0,t) e^{-\int_a^x r(x,t)dx} e^{-\sum_i \int_a^x \mu_i(x,t)dx}$$

Applying this formula to the Swedish population of 1976, Preston and Coale (1982) reconstructed the population by year of age for 1976 from the year's births, the growth rate of the population at each age in 1976, the net age-specific emigration in 1976, and the period life table. The comparison with the real population shows that the error committed never exceeds 1 per thousand except at the oldest ages (over 95).

b. A Cohort Application to the Study of Relationships Between Movement and Structure

Horiuchi (1991) proposed a cohort application of the generalized equation to an open population to separate the effect on the change in the number of persons at each age observed at two given dates from the cohort-specific change in the number of births and in mortality and migration functions. The growth rate $r(a,t)$ of the population aged a in t may be expressed as a function of the rate of increase in births from one cohort to another, $r_B(t-a)$ and the effect of combined changes in the forces of mortality and net emigration, as follows:

$$r(a,t) = r_B(t-a) - \int_0^a \frac{\partial \mu(x,u)}{\partial u}dx - \int_0^a \frac{\partial e(x,u)}{\partial u}dx$$

where $\mu(a,t)$ and $e(a,t)$ are the forces of mortality and net emigration at age a in year t and where the derivatives must be established for $u = t - a + x$.

Using this formula, we can estimate the share of each of the three components (births, force of mortality, and force of net emigration) in the growth rate of the population at a given age, and we can therefore measure the impact of their evolution on the observed changes in population size.[19] Since the model rests on the age-specific forces of mortality and migration, it also enables us—in theory—to break down the total effects of the mortality and migration changes into the age-specific or age-group-specific effects of the changes in the two phenomena. As a rule, however, we do not know the age-specific and cohort-specific force of net emigration. But, if the age-specific population growth is known, we need only calculate two elements (here, the effect of births and the effect of mortality) in order to obtain the third as the difference between them. Here, of course, it is the third element that combines the specific effect concerned (the effect of migrations) and the interactions.

If we multiply each term of the equation by the population $N(a,t)$ aged a in t, we can distribute the absolute difference in population (between the two cohorts at age a) across the three factors. Consolidating the results obtained from age a to age b, we obtain the distribution for the absolute difference observed for the age group (a–b).

To assess the weight of each of the three factors on the relative change in the population's age structure $c(a,t)$, we must solve the following equation, given the formula for the age-specific composition in the r-variable model (Caselli and Vallin, 1990):

$$\frac{\partial c(a,t)}{\partial t} = \left[\frac{\partial N(a,t)}{\partial t}\cdot\frac{1}{N(t)} - \frac{N(a,t)}{N(t)}\cdot\frac{dN(t)}{dt}\frac{1}{N(t)}\right]$$

where $N(t)$ is the total population, or, more simply:

$$\frac{\partial c(a,t)}{\partial t} = \frac{1}{N(t)}\left[\frac{\partial N(a,t)}{\partial t} - c(a,t)\cdot\frac{dN(t)}{dt}\right]$$

[19] Except that the order in which the factors work is not taken into account (net emigration, for example, could have produced births and deaths). Only the direct effect is examined, not the interactions.

Since we have already used the equation yielding $r(a,t)$ to distribute the differences in population size at each age between the three effects (births, mortality, migrations), the last equation enables us to measure the role of each of the three factors in the age-structure variation.

We applied this model to French and Italian data to determine the influence of the natural-increase and migration components on the change in the proportion of elderly persons between 1952 and 1986. In Italy, for a total increase of 8 percentage points in the proportion of women aged 60 and older, we estimate that only 2 points were due to the steep fall in the number of births, more than 5 points were due to the decline in mortality, and a mere 0.7 to migration. In France, the contrast is even sharper, since the change in births has sharply slowed aging: this has almost entirely offset the aging effect of the decline in mortality (Table 20–3).

CONCLUSION

The principles of stable-population theory and their generalization have paved the way for a vast range of applications in two areas: the reconstruction of incomplete or defective data and the analysis of past and future population change. We will return to both issues throughout this treatise.

The reconstruction of incomplete data may involve compensating for the lack of detailed information on populations of the past or even of certain countries or regions today. Alternatively, we may want to assess the

TABLE 20–3 Respective Contributions of the Cohort-specific Changes in Births, Deaths, and Migrations to the Increase in the Proportion of Persons Aged 60 and Over, France and Italy, 1952–1986

Factors	France		Italy	
	Male, %	Female, %	Male, %	Female, %
Proportions of persons aged 60 and over				
1952	13.43	18.77	11.15	13.12
1986	15.34	21.14	16.15	21.14
1986–1952 difference	1.91	2.37	5.00	8.02
Contribution of change				
Births	–4.13	–5.04	1.48	1.99
Deaths	5.26	6.32	2.34	5.30
Migrations	0.78	1.09	1.17	0.73

quality or make up for the deficiencies of data collected in insufficiently reliable conditions. In all these applications, population models have yielded many procedures, often highly efficient. For demographers the world over, the stable populations defined by Coale and Demeny from their families of model life tables now play a rather similar role to what the log table used to represent for mathematicians. Many demographers have helped to develop methods based on population models for solving specific problems (Coale and Trussell, 1996). The UN (1967, 1983), realizing the value of these methods for improving our knowledge of the international demographic situation, has developed an array of instruments for use by countries with incomplete or deficient statistics.

Stable-population models have also powerfully contributed to the analysis of the demographic past and future. The theory of demographic transition—which will be abundantly referred to in the subsequent chapters of this treatise—rests on the following postulate: All populations undergo a crucial, historical phase of modernization that takes them from a quasi-stationary *old regime* characterized by high mortality and fertility to a modern regime, also quasi-stationary, in which these two natural-increase factors have fallen to distinctly lower levels. There is obviously a special link between the discussions on this theory and the use of population models. And both these discussions and these models underlie most of the population-projection studies that seek to forecast and explore the scenarios for future population change.

References

Arthur W. Brian and Vaupel James, 1984. Some general relationships in population dynamics. *Population Index*, vol. 50, p. 214–226.

Bennett Neil G. and Horiuchi Shiro, 1981. Estimating the completeness of death registration in a close population. *Population Index*, vol. 47, p. 207–221.

Bennett Neil G. and Horiuchi Shiro, 1984. Mortality estimation from registered death in less developed countries. *Demography*, vol. 21, p. 217–234.

Bernardelli Harro, 1941. Population waves. *Journal of the Burma Research Society*, vol. 31(1), p. 1–18.

Bourgeois-Pichat Jean and Taleb Si-Ahmed, 1970. Un taux d'accroissement nul pour les pays en voie de développement en l'an 2000: rêve ou réalité? *Population*, vol. 25(5), p. 957–974.

Bourgeois-Pichat Jean, 1958. Utilisation de la notion de population stable pour mesurer la mortalité et la fécondité des populations des pays sous développés. *Bulletin de l'Institut International de Statistique*, no. 36, p. 94–121.

Bourgeois-Pichat Jean, 1966. *The concept of stable population. Application to the study of populations with incomplete demographic statistics.* New York, United Nations, 237 p. (Doc. ST/SOA/Série A139)

Bourgeois-Pichat, 1971. Stable, semi-stable populations and growth potential, *Population Studies*, vol. 25(2), p. 235–254.

Bourgeois-Pichat, 1994. *La dynamique des populations. Populations stables, semi-stables, quasi-stables.* Paris, INED, 296 p. (Travaux et Documents, Cahier no. 133)

Brouard Nicolas, 1977. Évolution de la fécondité masculine depuis le début du siècle, *Population*, vol. 32(6), p. 1123–1158.

Brouard Nicolas, 1987. Sida: durée d'incubation, taux de croissance, taux de reproduction nette, *Population*, vol. 42(6), p. 797–818.

Brouard Nicolas, 1994. Aspects démographiques et conséquences de l'épidémie de sida en Afrique, *in*: Jacques Vallin (ed.), *Populations africaines et sida*, p. 119–178. Paris, La Découverte/CEPED, 224 p. (Recherche).

Caselli Graziella and Vallin Jacques, 1990. Mortality and population aging. *Revue européenne de démographie/European Journal of Population*, vol. 6, p. 1–25.

Coale Ansley J. and Demeny Paul, 1966. *Regional model life tables and stable populations*. Princeton (New Jersey), Princeton University Press, 871 p.

Coale Ansley J. and Demeny Paul, 1966. *Regional model life tables and stable populations*. Princeton (New Jersey), Princeton University Press, 871 p.

Coale Ansley J. and Demeny Paul, 1967. *Methods of estimating basic demographic measures from incomplete data*. New York, United Nations, 126 p. (Coll. Manuals on methods of estimating population, Manual IV, doc. ST/SOA/Series A/42).

Coale Ansley J. and Guo G., 1989. Revised regional model life tables at very low levels of mortality. *Population Index*, vol. 55(3), p. 367–609.

Coale Ansley J. and Trussell James, 1966. The development and use of demographic models. *Population Studies*, vol. 50(3), p. 469–484.

Coale Ansley J. and Watkins Susan Cotts, 1986. The decline of fertility in Europe. Princeton (New Jersey), Princeton University Press, 484 p.

Coale Ansley J., 1956. The effects of changes in mortality and fertility on age composition. *Milbank Memorial Fund Quarterly*, no. 34, p. 79–114.

Coale Ansley J., 1957. A new method for calculating Lotka's r the intrinsic rate of growth in a stable population. *Population Studies*, vol. 11(11), p. 92–94.

Coale Ansley J., 1968. Convergence of a human population to a stable form. *Journal of the American Statistical Association*, no. 63, p. 395–435.

Coale Ansley J., 1972. *The growth and structure of human populations: a mathematical investigation*. Princeton (New Jersey), Princeton University Press, 227p.

Coale Ansley J., 1979. The use of modern analytical demography by T. R. Malthus. *Population Studies*, vol. 33(2), p. 329–332.

Coale Ansley J., Demeny Paul and Vaughan Barbara, 1983. *Regional model life tables and stable populations*, 2nd ed. New York, Academic Press, 496 p.

De Santis Gustavo, 1997. *Demografia ed economia*. Bologne, Il Mulino, 354 p. (Stuci e Ricerche)

Euler Leonhard, 1760. Recherches générales sur la mortalité et la multiplication du genre humain. *Mémoires de l'Académie royale des sciences et des belles lettres*, no. 16, p. 144–165.

Feller William, 1941. On the integral equation of renewal theory. *Annals of Mathematical Statistics*, no. 12, p. 243–267.

Fisher Ronald A., 1958. The fundamental theorem of natural selection, *in: The genetical theory of natural selection*, 2nd ed., p. 25–30. New York, Dover, 291 p.

Goodman Leo A., 1953. Population growth of sexes. *Biometrics*, vol. IX, p. 212–225.

Goodman Leo A., 1968. An elementary approach to the population projection-matrix. *Demography*, vol. 5(1), p. 382–409.

Greville Thomas, 1943. Some methods of constructing abridged life tables, *Records of the American Institute of Actuaries*, vol. 32, première partie, p. 29–43.
Henry Louis, 1968. Problèmes de la nuptialité: Considérations de méthode. *Population*, vol. 23(5), p. 835–844.
Henry Louis, 1972. *Démographie: analyse et modèles.* Paris, Larousse, 341 p.
Horiuchi Shiro, 1991. Assessing the effects of mortality reduction on population ageing. *Population Bulletin of the United Nations*, no. 31–32, p. 38–51.
Karmel Peter H., 1949. Le conflit entre les mesures masculine et féminine de la reproduction, *Population*, vol. 4(3), p. 471–494.
Karmel Peter H., 1948. The relations between males and females in stable populations. *Population Studies*, vol. 1, p. 353–387.
Keyfitz Nathan, 1964. The population projection as a matrix operator, *Demography*, vol. 1(1), p. 56–61.
Keyfitz Nathan, 1968. *Introduction to the mathematics of population*. Reading (Massachusetts), Addison-Wesley, 450 p.
Keyfitz Nathan, 1971. On the momentum of population growth, *Demography*, vol. 8(1), p. 71–80.
Keyfitz Nathan, 1985. *Applied mathematical demography*, 2nd edition. New York and Berlin, Springer-Verlag, 442 p.
Keyfitz Nathan, 1977. *Applied mathematical demography*. New York, John Wiley & Son, 388 p. (Reedited in 1985 by Springer-Verlag, New York, 442 p).
Leslie Paul H., 1945. On the use of matrices in certain population mathematics. *Biometrika*, vol. 33(Part 3), 183–212.
Lombardo Enzo, 1996. Euler: un précurseur des modèles mathématiques en démographie. *Genus*, vol. 32(1–2), p. 129–139.
Lopez Alvaro, 1961. *Some problems in stable population theory*. Princeton, Office of Population Research, 107 p.
Lotka Alfred J., 1907. Studies on the mode of growth of material aggregates. *American Journal of Science*, vol. 24, p. 199–216.
Lotka Alfred J., 1922. The stability of the normal age distribution. *Proceedings of the National Academy of Sciences*, no. 8, p. 339–345.
Lotka Alfred J., 1934. *Théorie analytique des associations biologiques. Première partie: principes*. Paris, Herman & Cie, 150 p. (Coll. Actualités Scientifiques et Industrielles, no. 187)
Lotka Alfred J., 1939. *Théorie analytique des associations biologiques. Deuxième partie: analyse démographique avec application particulière à l'espèce humaine*. Paris, Herman & Cie, 139 p. (Coll. Actualités Scientifiques et Industrielles, no. 780)
Mortara Giorgio, 1949. *Utilisation des statistiques du recensement pour le calcul des tables de survie et autres données démographiques. Application á la population du Brésil*. New York, United Nations, 61 p. (Population Studies, no. 7).
Pollard Alfred, 1948. The measurement of reproductivity, *Journal of the Institute of Actuaries*, vol. LXXIV, p. 288–305.
Pollard John, 1973. *Mathematical models for the growth of human populations*. Cambridge, Cambridge University Press, 186 p.
Preston Samuel H. and Coale Ansley J., 1982. Age structure, growth, attrition, and accession: a new synthesis, *Population Index*, vol. 48(2), p. 217–259.
Preston Samuel H. and Guillot Michel, 1997. Population dynamics in an age of declining fertility synthesis, *Genus*, vol. 53(3–4), p. 15–31.
Preston Samuel H., 1983. An integrated system for demographic estimation from two age distribution, *Demography*, vol. 20(2), p. 213–226.
Preston Samuel H., 1986. Relation between actual and intrinsic growth rates, *Population Studies*, vol. 40(3), p. 343–352.
Preston Samuel H., 1987. Estimation of certain measures in family demography based on generalized stable population relations *in*: John Bongaarts, Thomas K. Burch and Kenneth Wachter (eds.) *Family demography: methods and their application*, p. 40–62. Oxford, Clarendon Press, 365p.
Rogers Andrei, 1966. The multiregional matrix growth operator and the stable interregional age structure, *Demography*, vol. 3(2), p. 537–544.
Sharpe F. R. and Lotka Alfred J., 1911. A problem in age distribution, *Philosophical Magazine*, vol. 6(21), p. 435–438.
Smith David and Keyfitz Nathan, 1977. *Mathematical Demography. Selected papers*. Berlin, Heidelberg and New York, Springer-Verlag, 514p.
Süssmilch Johann Peter, 1761. *Die göttliche Ordnung*, vol. 1, p. 291–299. Berlin, s.n., 722 p.
United Nations, 1967. *Manual IV. Methods of estimating basic demographic measures from incomplete data*. New York, United Nations, 126 p. (Population Studies, no. 42).
United Nations, 1973. *World population prospects as assessed in 1968*. New York, United Nations, 168 p. (Population Studies, no. 53).
United Nations, 1974. *Les perspectives d'avenir de la population mondiale évaluées en 1968*. New York, United Nations, 172 p. (études démographiques, no. 53).
United Nations, 1983. *Manual X. Indirect techniques for demographic estimation*. New York, United Nations, 304 p.
Vincent Paul, 1945. Potentiel d'accroissement d'une population stable. *Journal de la Société de Statistique de Paris*, vol. 86, p. 16–29.
Vincent Paul, 1945. *Potentiel d'accroissement d'une population*. Paris, Berger-Levrault, 26 p.
Wattelar Christine, 1971. Représentations matricielles du mouvement naturel et du mouvement migratoire d'une population, *Recherches Economiques de Louvai* no. 4, p. 379–432.
Wunsch Guillaume, 1989. Relation générale entre mouvement et structure démographique. Synthèse des méthodes et des applications récentes, *Revue Européenne de Démographie / European Journal of Population*, vol. 5(1), p. 71–89.
Yntema Luitzen, 1952. *Mathematical models of demographic analysis*. Leiden, J. J. Groen & Zoon, 78p.

APPENDIX 20–1 Key Functions in Stationary, Stable, and Ordinary Populations

Function	Notation	Formula: Stationary population	Formula: Stable population	Formula: Ordinary population
Age structure	$c(a)$	${}_np_a$	$be^{-ra}p(a)$	$b(t)e^{-\int_0^a r(x,t)dx}\,p(a,t)$
Ratio of populations at two ages	$\frac{c(a+n)}{c(a)}$	${}_np_a$	$e^{-rn}\,{}_np_a$	$e^{-\int_a^{a+n} r(x,t)\,dx}\,{}_np_a$
Birth rate	b	$b=\frac{1}{\int_0^\omega p(a)da}$	$b=\frac{1}{\int_0^\omega e^{-ra}p(a)da}$	$b(t)=\frac{1}{\int_0^\omega p(a,t)e^{-\int_0^a r(x,t)dx}}$
Distribution of births by mother's age	$v(a)$	$p(a)m(a)$	$p(a)m(a)e^{-ra}$	$p(a,t)m(a,t)e^{-\int_0^a r(x,t)dx}$
Life expectancy at birth	e_0	$\frac{1}{b}$	$\frac{\int_0^\omega c(a)e^{ra}da}{b}$	$\frac{\int_0^\omega c(a,t)e^{\int_0^a r(x,t)dx}}{b}$
Net reproduction rate	R_0	1	$\int_0^\omega p(a)m(a)e^{-ra}da$	$\int_0^\omega p(a,t)m(a,t)e^{-\int_0^a r(x,t)dx}da$
Life expectancy in a state G[1]	e_0^G	$\frac{\int_0^\omega g(a)c(c)da}{b}$	$\frac{\int_0^\omega g(a)c(a)e^{ra}da}{b}$	$\frac{\int_0^\omega g(a,t)c(a,t)e^{\int_0^a r(x,t)dx}}{b}$
No. persons aged a^* in terms of deaths after age a^*	$N(a^*)$	$\int_{a^*}^\omega D(a)da$	$\int_{a^*}^\omega D(a)e^{r(a-a^*)}da$	$\int_{a^*}^\omega D(a,t)e^{\int_{a^*}^a r(x,t)dx}da$
No. persons aged a^* in terms of deaths before age a^*	$N(a^*)$	$N(0)-\int_0^{a^*}D(a)da$	$e^{ra^*}\left[N(0)-\int_0^{a^*}D(a)e^{ra}da\right]$	$e^{\int_0^{a^*} r(x,t)dx}\left[N(0,t)-\int_0^{a^*}D(a,t)e^{\int_0^a r(x,t)\,dx}da\right]$
Probability of survival from a^* to a^*+1, in terms of deaths	${}_np_a$	$\frac{\int_{a^*+n}^\omega D(a)da}{\int_{a^*}^\omega D(a)da}$	$\frac{\int_{a^*+n}^\omega D(a)e^{r(a-a^*)}da}{\int_{a^*}^\omega D(a)e^{r(a-a^*)}da}$	$\frac{\int_{a^*+n}^\omega D(a,t)e^{\int_{a^*}^a r(x,t)dx}da}{\int_{a^*}^\omega D(a,t)e^{\int_{a^*}^a r(x,t)dx}da}$

a. *g(a)* is the incidence at age *a* of state *G*.
b. Notation *a* gives exact age; notation *a** gives age in complete years.
Adapted from table in Preston and Coale, 1982. Age structure, growth, attrition, and accession: a new synthesis, *Population Index*, vol. 48(2), p. 217–259.

PART

V

FROM HOMOGENEITY TO HETEROGENEITY

New Ways Forward for Demographic Analysis

GRAZIELLA CASELLI AND JACQUES VALLIN

To illuminate the fundamental mechanisms of population dynamics, we were obliged in the foregoing chapters, to oversimplify the realities by making two *fundamental assumptions*: (1) all the individuals in a population are, if not identical, at least exposed to completely equal mortality, fertility, and migration risks (the continuity or homogeneity hypothesis); (2) the probabilities of dying, procreating, and migrating are completely independent (independent events hypothesis). This is obviously not the case in real life. So the analysis needs to be fine-tuned to clarify how the *heterogeneity of a population* can distort the measurement of changes in it, and how to go beyond the independent events hypothesis.

One presumptive advantage of using the birth cohort as the unit of analysis, as we have done, is specifically that it enables us to work on more homogeneous sets of individuals than those of a hypothetical cohort constructed from data from as many different cohorts, with their different life experiences, as of ages observed. There is not even complete homogeneity within birth cohorts, in fact. As well as being genetically and biologically diverse, the constituent individuals have a different cultural baggage—either transmitted from their parents or self-acquired during life—belong to different social classes, live in salubrious or unsanitary conditions, and so forth—that tend to refute the argument that all are exposed to exactly the same risks of death, procreation or migration. It is clear that internal heterogeneity (within the same birth cohort or within a same-age population) can be at least partly offset by distinguishing subpopulations by attributes or characteristics relevant to the purpose of analysis, like socio-professional categories for mortality, educational levels for fertility, and so forth. However, the data necessary to correlate all the attributes or characteristics capable of influencing the phenomena studied are rarely available. Taking this to its logical extreme would run the considerable risk of having to take each individual in isolation from all others, which would completely negate any statistical approach. Identifying homogeneous groups of individuals at a given time does not mean that these groups will remain so throughout the period during which their demographic behaviors are to be measured. The degree of heterogeneity may increase or decrease as the individuals live their lives, constantly changing states for a particular attribute or characteristic. Chapter 21 first takes a general look at the *concealed population heterogeneity* as illustrated by the analysis of mortality. Chapter 22 goes beyond the limits of traditional migration analysis to redefine the concepts of mobility associated with different levels of spatial organization and addresses the linkages between internal mobility and the spatial distribution of a population. *Life history analysis* methods (Chapter 23) will take the process a step further by simultaneously countering the homogeneity and independent event hypotheses. Finally, it will be seen that the life history analysis used to account for interaction of phenomena leads to multiple possible levels of aggregation. As a result, the final chapter of this volume (Chapter 24) will consider *multilevel analysis*. Part V, therefore, marshals new approaches intended to overcome the problems of traditional analysis in preparation for moving on from a descriptive analysis of demographic phenomena to explaining the determinants of change, which will be looked at in Section II of Volume I and in Volume II of this treatise.

C H A P T E R

21

Unobserved Population Heterogeneity

JAMES W. VAUPEL AND ANATOLI I. YASHIN
Max Planck Institute for Demographic Research, Rostock, Germany

I. POPULATION HETEROGENEITY

All populations are heterogeneous. In demographic analyses, two dimensions of individual differences (age and sex) generally are observed. Many other characteristics may be observed, including date and place of birth, urban versus rural residence, marital status, nationality, religious affiliation, number of children, number of siblings, age of mother and father at an individual's birth, household structure, socioeconomic status, educational achievement, occupation, spouse's occupation, smoking behavior, diet, height, environmental quality at current residence, health status, cognitive and physical functioning, genotype, and so forth. In even the most thorough study, however, most attributes of individuals are not measured. Indeed, most studies focus on only a handful or two of the multitudinous dimensions of differences that distinguish one individual from another.

Observed heterogeneity creates various analytic opportunities for demographers. Multiple regression analysis, logit and probit analysis, survival analysis, and other statistical methods have been developed to estimate the impact of observed covariates. These methods are treated in other chapters and in standard textbooks, and will not be reviewed here.

Unobserved heterogeneity creates analytic problems rather than analytic opportunities: unobserved heterogeneity is a nuisance, a headache, a bête noire. Unobserved heterogeneity creates difficulties for demographers because demographers study how population characteristics change over age and time and place (and unobserved heterogeneity distorts observed patterns of change).

II. COMPOSITIONAL CHANGE

The root of the problem is that the members of population cohorts gradually die off or drop out. Animals and plants die, machines break down, bachelors marry, the married divorce, the childless give birth, those with one child have a second, children leave parental homes, students complete their education, the unemployed find jobs, the well get sick, and the ill recover. Much of demographic analysis focuses on the transition rates associated with such changes. In many instances, demographers are interested in how transition rates vary with age: They study, for instance, age-specific death rates and marriage rates. In other cases, duration matters, as in studies of recovery rates from an illness or divorce rates as a function of the duration of a marriage. In analyses of first, second, and subsequent births, birth rates by parity and time since last birth are of interest.

Hence, much of demographic analysis concerns the estimation and comparison of drop-out rates in cohorts that are changing because their members are dropping out. The problem is that those who drop out probably have a greater tendency to drop out than those who do not. People who die at some age tend to be frailer or more susceptible or at higher risk than those who survive to an older age. Couples who conceive after a month or two of trying may be more fecund than those who first conceive after many months. Marriages that quickly end in divorce may have been shaky marriages from the start. Thus, the composition or structure of a heterogeneous cohort changes as the cohort dies off. The frail or susceptible tend to die first, leaving a more robust surviving cohort.

III. THREE LEVELS OF EXPLANATION

Age- or duration-specific changes in birth, marriage, death, and other transition rates can be interpreted in three alternative ways that might be called level 0, level 1, and level 2 explanations. A level 0 explanation is that the data are erroneous. A level 1 explanation is that the observed change is produced by a corresponding change at the individual level. A level 2 explanation is that the observed change is an artifact of a change in the structure of the population (i.e., a change in the composition of a heterogeneous cohort).

Consider the report that the increase in mortality with age slows at the oldest ages (Vaupel *et al*. 1998; Thatcher, Kannisto, and Vaupel 1998). A direct, level 1 explanation would be that for individuals at advanced ages the probability of death increases relatively gradually with age. A level 0 explanation (bad data) would be that death rates at advanced ages are distorted by age-misreporting problems and that the apparent deceleration of mortality is a consequence of age exaggeration. Finally, a level 2 explanation would be that the leveling off of death rates after age 100 might be "caused by decreases in the average frailty of a population cohort at later ages as frailer members are removed by mortality" (Vaupel *et al*., 1979).

Observed patterns of mortality deceleration in different populations are almost certainly due to a mix of these three levels of explanation, with the importance of the different explanations differing from population to population. In almost all populations there are problems with age-misreporting at advanced ages and in many populations such misreporting is very severe (Jeune and Vaupel, 1999). All populations are heterogeneous, so level 2 explanations must have some validity, although it is currently unclear how much of the deceleration can be explained by compositional change. The level 1 explanation that individuals age more slowly at advanced ages may be partially right—or completely wrong. There is some suggestive evidence that for individuals the chance of death actually rises faster than exponentially at advanced ages, even though population death rates are increasing slower than exponentially (Yashin and Iachine, 1997).

At least since Edmund Halley (1693), demographers have recognized the importance of level 0 and level 2 explanations as alternatives to direct level 1 explanations. All careful demographers are aware of the prevalence of bad data and all well-trained demographers know that demographic rates can differ because of differences in population composition. Nonetheless, level 1 explanations—that what is observed on the population level also holds on the individual level—seem so natural that even careful demographers often find themselves naively and uncritically slipping into direct interpretations of population changes and differences (Vaupel and Carey, 1993).

IV. FRAILTY MODELS

Demographers try to distinguish between type 1 and type 2 explanations using frailty models (Vaupel *et al*., 1979) and the statistical methods of survival analysis (Cox and Oakes, 1984). In this approach, the trajectory of a cohort's rate of death or exit is usually captured by either the survival function $s(x)$ or the hazard function $\mu(x)$. Demographers call this hazard function the *force of mortality* when they are studying death rates and in some contexts the term *intensity* is used instead of hazard. The survival function and the hazard function are interrelated by the following two formulas:

$$\mu(x) = -\frac{ds(x)/dx}{s(x)} \qquad \text{(Eq. 1a)}$$

and

$$s(x) = e^{-\int_0^x \mu(t)dt} \qquad \text{(Eq. 1b)}$$

In the simplest case there is no information about the characteristics of the individuals in the cohort except age (and whatever characteristics describe the cohort as a whole, such as "males born in France in 1948").

Because all populations are heterogeneous, it makes sense to model the population as a mix of homogeneous subpopulations (which might each consist of a single individual). Let $s(x,z)$ be the survival function for the subpopulation with *frailty z*, where frailty in this context simply refers to the susceptibility or liability of the subpopulation to the hazard. In general, frailty models are designed such that the greater an individual's frailty, the greater the individual's susceptibility or liability to the hazard of interest.

Let $\bar{s}(x)$ be the survival function for the population as a whole, such that

$$\bar{s}(x) = \int_0^\omega s(x,z)g(z)dz \qquad \text{(Eq. 2a)}$$

in the continuous case, where $g(z)$ is the probability distribution of z at age zero and

$$\bar{s}(x) = \sum_z \pi(z)s(x,z) \qquad \text{(Eq. 2b)}$$

in the discrete case, where $\pi(z)$ is the proportion of the cohort in subpopulation z at age zero. This general frailty model can be more specifically formulated in several ways.

1. Relative-Risk Models

One specification is the *proportional-hazards* or *relative-risk* model

$$\mu(x,z) = z\mu(x) \qquad \text{(Eq. 3a)}$$

or, equivalently,

$$s(x,z) = s(x)^z \qquad \text{(Eq. 3b)}$$

which was suggested by Vaupel *et al*. (1979). In this model, $\mu(x)$ is the baseline, standard, underlying hazard for individuals of frailty one and $s(x)$ is the corresponding survival function. Vaupel *et al*. (1979) show that

$$\bar{\mu}(x) = \bar{z}(x)\mu(x) \qquad \text{(Eq. 4)}$$

where $\bar{z}(x)$ is the average frailty of those alive at age x. Because z is fixed and does not vary with age, $\bar{z}(x)$ declines with age as the frail drop out of the cohort. Hence, $\bar{\mu}(x)$ increases more slowly than $\mu(x)$ does. Indeed, $\bar{\mu}(x)$ can be declining even though $\mu(x)$ is rising.

For this model z is often taken to be gamma distributed with mean 1 and variance σ^2, because this gamma distribution leads to convenient mathematic relationships. In particular, for gamma-distributed frailty

$$\bar{z}(x) = \left(1 + \sigma^2 \int_0^x \mu(t)dt\right)^{-1} = \bar{s}(x)^{\sigma^2} \qquad \text{(Eq. 5)}$$

where $\bar{s}(x)$ is the survival function for the population as a whole. It follows from Equation 5 that

$$\bar{s}(x) = (1 - \sigma^2 \ln s(x))^{-1/\sigma^2} \qquad \text{(Eq. 6)}$$

As a specific example of this kind of gamma-frailty relative-risk model, suppose that mortality on the individual level follows a Gompertz trajectory:

$$\mu(x) = ae^{bx} \qquad \text{(Eq. 7)}$$

It follows from Equations 4 and 5 that the population trajectory of mortality will follow the logistic pattern

$$\bar{\mu}(x) = \frac{ae^{bx}}{1 + \frac{a\sigma^2}{b}(e^{bx} - 1)} \qquad \text{(Eq. 8)}$$

leveling off at a value of b/σ^2.

2. Accelerated-aging Models

Another specification is the accelerated-aging model

$$\mu(x,z) = \mu(xz) \qquad \text{(Eq. 9)}$$

which is analogous to the accelerated-failure model used in reliability engineering. In the special case where $\mu(x)$ follows the Weibull trajectory

$$\mu(x) = ax^b \qquad \text{(Eq. 10)}$$

where a and b are parameters, this model is equivalent to the relative-risk model, because

$$a(zx)^b = z^b ax^b = z'\mu(x) \qquad \text{(Eq. 11)}$$

In the special case where $\mu(x)$ follows the Gompertz trajectory, given in Equation 7, the accelerated-aging model is of the form ae^{zbx} whereas the relative-risk model is of the form zae^{bx}. Small changes in the slope parameter b can have larger effects on mortality at older ages than big changes in the level parameter a. Hence, much less heterogeneity is needed in an accelerated-aging Gompertz model than in a relative-risk Gompertz model to produce substantial differences between $\mu(x)$ aand $\bar{\mu}(x)$ at older ages. This is illustrated in Figure 21–1.

3. Discrete Frailty Models

The discrete frailty model is also a useful specification of the general frailty approach, as discussed by Vaupel and Yashin (1985). In this case,

$$\mu(x,z) = \mu_z(x) \qquad \text{(Eq. 12)}$$

Thus, z is now an index for the different subpopulations, each of which has a hazard function. Let π_z be the proportion of the population in subpopulation z at age zero. Then

$$\bar{s}(x) = \sum_z \pi_z s_z(x) \qquad \text{(Eq. 13)}$$

and

$$\bar{\mu}(x) = \sum_z \pi_z s_z(x)\mu_z(x) \Big/ \sum_z \pi_z s_z(x) \qquad \text{(Eq. 14)}$$

If it is assumed that z is a relative-risk factor, then $\mu_z(x) = z\mu(x)$ and $s_z(x) = s(x)^z$. Heckman and Singer (1984) suggested that this specification be used to

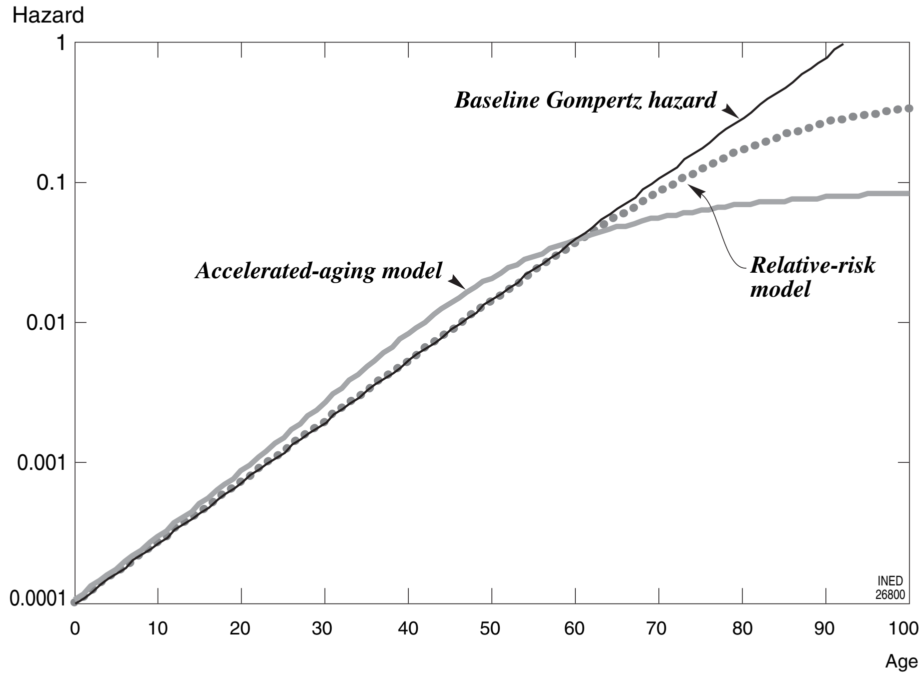


FIGURE 21–1 The accelerated-aging model can produce greater mortality deceleration with less heterogeneity than the relative-risk model baseline Gompertz hazard with $a = 0.0001$ and $b = 0.1$ compared with population hazard in relative-risk model with $\sigma^2 = 0.25$ and in accelerated-aging model with $\sigma^2 = 0.05$. Note that hazards are shown on a log scale.

control for the effects of hidden heterogeneity when fitting models to data. More generally, however, $\mu_z(x)$ can take on a different functional form for each value of z.

A simple example of discrete frailty models is the mover–stayer model (Blumen *et al.*, 1955) in which one group in the population is susceptible to emigration, marriage, divorce, or some disease and the other group is immune. Let π be the proportion of the population that is susceptible. Thus it follows from Equation 14 that

$$\bar{\mu}(x) = \pi s(x)\mu(x)/(\pi s(x) + 1 - \pi) \qquad \text{(Eq. 15)}$$

Even if $\mu(x)$ steadily increases, $\bar{\mu}(x)$ will eventually decline as $s(x)$ approaches zero. Figure 21–2 shows an illustrative example.

Divorce rates in some countries and periods follow the kind of rising–falling pattern shown in Figure 21–2. Does this imply that marriages are shakiest after a few years of marriage? Not necessarily, as Figure 21–2 illustrates. The same general effect could be produced if the second group were not immune but simply at low risk. Indeed the rising–falling pattern could be produced if the hazard steadily increases for the high-risk group but steadily decreases for the low-risk group. For one group marriages strengthen with duration, whereas

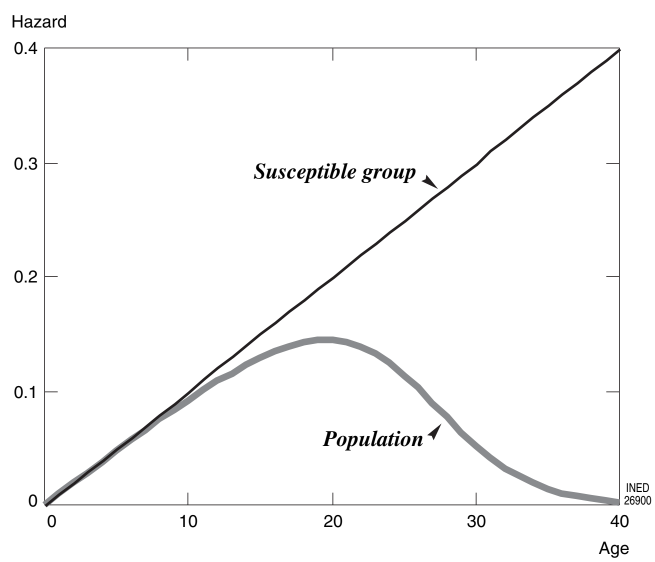


FIGURE 21–2 The population hazard may increase and then decline if the hazard rate for one group is increasing and the other group is immune. The hazard for the susceptible group is $\mu(x) = 0.01x$. It was assumed that 95% of the population is susceptible.

for the other, marriages weaken (despite the appearance of the curve for the entire cohort, there is no divorce hump).

4. Changing Frailty Models

As Box (1979) asserted, all models are wrong, but some models are useful. It is often useful to define an individual's frailty as fixed, at least after some age, and to classify individuals into groups depending on their frailty at that age. Alternatively, it may sometimes be useful to develop models in which an individual's frailty can change with time or age.

In one simple model of this kind, all individuals start out with frailty one. They suffer a hazard of death of $\mu_1(x)$ at age x. They also are subject to the hazard $\lambda(x)$ that their frailty will change from one to two, in which case their hazard of death changes to $\mu_2(x)$. The second state might be associated with some morbid event, such as having a heart attack or losing the ability to walk. Alternatively, the hazard of *death* could be the hazard of divorce and the event could be having a baby. Let $s_1(x)$ denote the proportion of the cohort that is alive with frailty one at age x and let $s_2(x)$ similarly denote the proportion of the cohort that is alive with frailty two at age x. In the simplest case when the three hazards functions are constant, it is not difficult to show that

$$s_1(x) = e^{-(\mu_1+\lambda)x} \quad \text{(Eq. 16a)}$$

and

$$s_1(x) = \frac{\lambda}{\mu_1+\lambda-\mu_2}\left(e^{-\mu_2 x} - e^{-(\mu_1+\lambda)x}\right) \quad \text{(Eq. 16b)}$$

The population hazard is given by

$$\bar{\mu}(x) = \frac{\lambda\mu^2 e^{(\mu_1+\lambda-\mu^2)x} + (\mu_1-\mu_2)(\mu_1+\lambda)}{\lambda e^{(\mu_1+\lambda-\mu_2)x} + (\mu_1-\mu_2)} \quad \text{(Eq. 17)}$$

At age zero, $\bar{\mu}(0) = \mu_1$ and as x approaches infinity, $\bar{\mu}(x) = \mu_2$ if $\mu_2 \leq \mu_1 + \lambda$ and $\bar{\mu}(x) = \mu_1 + \lambda$ otherwise.

Hervé Le Bras (1976) and Leonid Gavrilov and Nathalia Gavrilova (1991) proposed generalizations of this model. Instead of two states of frailty, suppose that frailty z can equal any nonnegative integer. Initially everyone has frailty zero. People with frailty z face a hazard of death of $\mu_o + z\mu$ as well as a hazard of $\lambda_o + z\lambda$ that their frailty will change to $z + 1$. Although μ_o, μ, λ_o and λ are constants and do not vary with age or time, the population hazard $\bar{\mu}(x)$ follows a logistic trajectory. Anatoli Yashin, James Vaupel, and Ivan Iachine (1994) show that this trajectory is identical to the trajectory obtained if frailty is fixed and gamma distributed and the baseline hazard is of the Gompertz form $\mu(x) = ae^{bs}$ with $\mu(x, z) = z\mu(x) + c$, where c is the constant Makeham term. Without ancillary information it is impossible to tell whether frailty is fixed or frailty is changing.

Instead of only taking on discrete values, frailty can be modeled to vary continuously. Vaupel, Yashin, and Kenneth Manton (1988), for instance, develop a changing-frailty model based on a stochastic differential equation. They apply the model to clarify the interaction of debilitation, recuperation, selection, and aging. The model yields various insights about lingering mortality consequences of disasters such as wars, famines, and epidemics that may weaken the survivors. A key result is that debilitation and selection are interdependent: debilitation that increases population heterogeneity will result in subsequent mortality selection; selection, by altering the distribution of frailty, will influence the impact of debilitating events. The basic equation of the model is

$$\mu(x,z) = \mu_o(x) + z(x)\mu_*(x) \quad \text{(Eq. 18)}$$

where $\mu_o(x)$ is the baseline hazard, $\mu_*(x)$ determines the additional hazard, and $z(x)$ is the frailty of the individual at age x as given by

$$z(x) = Y^2(x) \quad \text{(Eq. 19)}$$

where $Y(0)$ is normally distributed and

$$dY(x) = [a_o(x) + (a_1(x) - a_2(x))Y(x)]dx + b(x)dW(x) \quad \text{(Eq. 20)}$$

where W is a Wiener process with $W(0) = 0$. The functions a_o and a_1 represent the effects of debilitation whereas a_2 represents homeostatic healing and recuperation; the function b determines the importance of the Wiener process term.

5. Correlated-Frailty Models

Because of shared genes and a shared childhood environment, two twins may have similar frailties. More generally, relatives or people who live in the same environment may have similar frailties. As discussed by Vaupel (1990 and 1991), shared-frailty models can be used to analyze such situations, but a more appropriate and powerful approach involves the correlated-frailty models developed by Yashin and colleagues and explained in Yashin, Vaupel, and Iachine (1995) and Yashin and Iachine (1997).

A simple variant of this kind of model involves pairs of twins, with one twin having fixed frailty z_1 and the other twin having fixed frailty z_2 and with the hazard of mortality given by $\mu(x, z_i) = z_i\mu_o(x)$, $i = 1,2$. The correlation between the two frailties is modeled as follows. Let

$$z_1 = y_o + y_1 \quad \text{(Eq. 21a)}$$

and let

$$z_2 = y_o - y_2 \quad \text{(Eq. 21b)}$$

where the y_i, $i = 0,1,2$, are three independent random variables that are gamma distributed with the same scale parameter. The gamma distributions of y_1 and y_2 have the same shape parameter, but this parameter may differ for y_o. The frailties z_1 and z_2 are constrained to have a mean of one and they have the same variance σ^2. The values of σ^2 and ρ, the correlation coefficient between the two frailties, are simple functions of the scale and shape parameters.

As shown by Yashin, Vaupel, and Iachine (1995), the bivariate survival function for the population of twins is given by $\bar{s}$

$$\bar{s}(x_1,x_2) = \bar{s}(x_1)^{1-\rho}\,\bar{s}(x_2)^{1-\rho}(\,\bar{s}(x_1)^{-\sigma^2} + \bar{s}(x_2)^{-\sigma^2} - 1)^{-\rho/\sigma^2} \quad \text{(Eq. 22)}$$

Because the survival of adult twins is very similar to the survival of adult singletons, in studies of adult mortality the function $\bar{s}(x)$ can be taken from demographic life tables for the general population. In this case the bivariate survival function depends only on σ^2 and ρ; no assumptions have to be made about the shape of a baseline hazard function. Using Equation 22 values of σ^2 and ρ can be estimated using the kind of maximum-likelihood estimation described below. Then the baseline survival function can be calculated by rearranging Equation 6 as follows:

$$s(x) = \exp((1 - \bar{s}(x)^{-\sigma^2})/\sigma^2) \quad \text{(Eq. 23)}$$

and $\mu(x)$ can be calculated from Equation 1. Applying this model to survival data on Danish twins born between 1870 and 1900, Yashin and Iachine (1997) found that the baseline hazard of mortality increases faster than exponentially after age 30 even though the population hazard of mortality decelerates at advanced ages. That is, it is possible that the observed leveling off of mortality may be entirely accounted for by a level 2 explanation (compositional change due to mortality selection) and the actual trajectory of mortality for individuals may rise more rapidly than a Gompertz curve.

V. EMPIRICAL DATA

The survival or duration data used in fitting frailty models is often of the following form. There are n individuals in some cohort, with observed ages at death X_i, $i = 1, \ldots, n$. More generally, X_i stands for age at some event, such as marriage, or some duration, such as time from marriage to divorce. For simplicity, we will refer to X_i as age at death.

Age at death may not be known for all individuals: it may only be known that the individual survived at least until some age. These are called *right-censored observations*. They can arise if some individuals never "die" (e.g., some women never give birth, some people never marry, some married people never divorce). They can also arise if some individuals are still alive at the end of a study or if some individuals drop out of the study and are lost to follow-up.

It may also be the case that it is only known that individual i died between age x_1 and age x_2. And it may be the case that an individual is not followed from age zero but from some age x_o, so that the individual is only at risk of dying after age x_o. These are called *left-censored* or *left-truncated observations*.

Various covariates may be observed: We will let w_{ij} denote the value of the j-th covariate for individual i.

VI. METHODS OF PARAMETER ESTIMATION

Various methods may be used to fit frailty models to empirical survival or duration data. In an important article on deceleration in the age pattern of mortality at older ages, Shiro Horiuchi and John Wilmoth (1998) estimate the parameters of a Gompertz–Makeham model with period effects and unobserved frailty by a weighted-least-squares procedure. They carefully document the method they use, so their article is of pedagogic value as well as being of substantive interest.

It is more common, however, to use maximum-likelihood methods in analyses of survival or duration data in general and in fitting frailty models in particular. We recommend a textbook such as that by David Cox and David Oakes (1984) for discussion of this approach. Here we adumbrate a few points of particular relevance to the estimation of parameters of frailty models.

The likelihood of an observation X_i can be thought of as the probability of observing this value given a particular model with specific parameter values. More generally, the likelihood can be proportional to the probability instead of being equal to the probability, because any parameter values that maximize the probability of the data will also maximize any quantity that is proportional to the probability. Let $s_i(x)$ be the probability of surviving from age 0 to age x, for some individual with a vector of covariates w_i and with some unobserved frailty z. Then, if age at death X_i is observed and if the individual is followed from age zero, the likelihood of the observed age at death is $\mu_i(X_i)s_i(X_i)$. If it is only known that the individual survived at least to age X_i, the likelihood of this observation is $s_i(X_i)$. If it is known that the individual died between ages X_{1i} and X_{2i}, then the likelihood of this datum is $s(X_{1i}) - s(X_{2i})$. If the individual is first observed at age X_{1i} and then dies at age X_{2i}, then the likelihood is $\mu(X_{2i})s(X_{2i})/s(X_{1i})$.

It is customary in survival analysis to make calculations in terms of the logarithm of the likelihood, the log-likelihood, because the likelihood of a data set is often extremely small. Let $L(X_i)$ denote the log-likelihood of the observation. The log-likelihood of the entire data set is given by the sum of all the $L(X_i)$'s. The maximum-likelihood estimate of the parameter values in a model is the estimate that maximizes the likelihood or, equivalently, the log-likelihood of the data.

The effect of observed covariates on survival can be modeled in many ways. Because our focus here is on hidden heterogeneity and not on general methods of survival analysis, we restrict our attention to the simple case in which the covariates are fixed (rather than changing over time). Furthermore, we will assume that the impact of the covariates on an individual's hazard is given by $W_i\mu_i(x, z)$, where W_i is the net relative-risk imposed by the vector of observed covariates. Often in survival analyses, W_i is modeled by $W_i = e^{\sum_j b_j w_j}$, where the b_j's are coefficients that are estimated.

For ease and conciseness of exposition we will consider only the relative-risk frailty model with gamma-distributed frailty. For the relative-risk model (Equation 3b),

$$s_i(x) = s(x)^{W_i z} \qquad \text{(Eq. 24)}$$

where $s(x)$ is the baseline or standard survival function for individuals of frailty 1 and with an estimated relative-risk W_i of 1. In the special case of gamma-distributed frailty, it follows from formula (6) that

$$\bar{s}_i(x) = (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma_2} \qquad \text{(Eq. 25)}$$

where $\bar{s}_i(x)$ is the probability that an individual with estimated relative-risk W_i will survive to age x and where σ^2 is the variance of frailty. The bar over the s indicates that $\bar{s}_i$ is an average: unobserved frailty z is removed from the formula by taking the expected value of s with respect to z. The corresponding value of $\bar{\mu}_i(x)$ can be calculated by Equation 1 and the log-likelihood $L(x)$ can then be calculated as indicated above.

A remaining issue is how to estimate the baseline survival function $s(x)$. There are two approaches to this. First, a parametric form can be assumed. For instance, it might be assumed that $s(x)$ [and $\mu(x)$] are of the Gompertz or Weibull form. Kenneth Manton, Eric Stallard, and James Vaupel (1986) provide an example of this kind of analysis.

Alternatively, $s(x)$ can be estimated nonparametrically. That is, values of $s(x)$ can be estimated for a sequence of ages over some age range without imposing any assumptions on the shape of the trajectory of $s(x)$. Several different methods of nonparametric estimation have recently been proposed and research in this area is rapidly developing. Here we sketch one method, to illustrate the general idea of nonparametric estimation.

Suppose that the survival data that are available for analysis are based on a large random survey of some population. Further suppose that survival in the population is known, perhaps from vital-statistics data. Let $\bar{\bar{s}}(x)$ be the survival curve for the population. As above, let $\bar{s}_i(x)$ be the survival function for the individual i in the survey. For a large random survey, the following equation might approximately hold:

$$\bar{\bar{s}}(x) = \sum_{i=1}^{n} s_i(x)/n \qquad \text{(Eq. 26)}$$

If so, the following method could be used. For the relative-risk gamma-frailty specification, Equation 25 can be substituted in Equation 26, yielding

$$\bar{\bar{s}}(x) = \sum_i (1 - \sigma^2 W_i \ln s(x))^{-1/\sigma^2} \Big/ n \qquad \text{(Eq. 27)}$$

For any specific set of estimated values for W_i and σ^2, one and only one value of $s(x)$ will satisfy this equation. Hence, σ^2 and the coefficients that determine W_i can be determined by maximum-likelihood estimation under the constraint that Equation 27 holds.

The theoretical and practical properties of this algorithm still need to be investigated. Many other estimation procedures that do not require parametric estimation of $s(x)$ are being developed and various imputation methods, EM algorithms, and other concepts might be used (see, e.g., Andersen, Borgan, Gill, and Keiding, 1993). The statistical estimation of frailty models is currently a hot topic of statistical research and the coming decade is likely to produce major advances in the development of powerful, practicable procedures.

CONCLUSION

The frailty models and parameter-estimation methods sketched above are not yet available as part of easy-to-use computer software packages. Several software packages, however, include convenient options for fitting other kinds of models to survival and duration data. In particular, David Cox (partial likelihood) regression can be readily applied to empirical observations. Why bother with frailty models when Cox regression can be used to estimate the coefficients of covariates in hazards models? There are three main reasons.

First, Cox regression yields coefficient estimates that tend to be biased toward zero. As epidemiologists have often observed, most risk factors that raise the chance of death appear to become less important with age or duration. The reason generally is that high-risk individuals who survive often have unobserved strengths or advantages, whereas many of the apparently low-risk individuals who survive may be relatively weak or unhealthy along unobserved dimensions. Consequently, at older ages or longer durations, the high-risk group differs in composition from the low-risk group: the high-risk group has lower unobserved frailty. If unobserved frailty is not included in the model, this effect will result in a convergence with age of the hazard functions for the two groups, as discussed by Vaupel *et al*. (1979) and Vaupel and Yashin (1985). The proportional-hazards assumption used in Cox regression does not allow for such convergence: The estimated relative risk is a measure of the average relative risk over the entire age range. The implication of this is that Cox regression tends to result in underestimates of risk factors: The estimates are biased toward zero. More generally, any method that ignores hidden heterogeneity will tend to underestimate risk factors at older ages or longer durations.

Second, frailty models permit estimation of underlying (or baseline) hazards (i.e., the hazards that govern the trajectory of risks at the individual level). It may be of interest, for instance, to know whether the underlying hazard is monotonically increasing even though the observed population hazard first rises and then declines. More generally, demographers are concerned about whether observed trajectories of demographic rates over age or duration can be explained by level 1 accounts or level 2 accounts. Does the trajectory observed for a population also hold for the individuals who comprise the population—or is the trajectory attributable to compositional change? Frailty models are designed to address this question.

Third, frailty models permit the use of ancillary vital statistics data in the analysis. For example, as briefly discussed above, it is possible to analyze detailed data on some subset of a population (e.g., twins or those who participate in a survey) together with the vital statistics data on the survival of the general population. The combination of detailed data on a subpopulation with survival data on the entire population leads to more accurate statistical estimates. This combination seems natural and highly appropriate for demographers.

References

Andersen Per Kragh, Borgan Ørnulf Gill Richard, Keiding Niels 1993. *Statistical Models Based on Counting Processes*. New York, Springer-Verlag, 768 p.

Blumen Isadore, Kogan Marvin, McCarthy Philip J., 1955. *The industrial mobility of labor as a probabilité process*. Ithaca (New York), Cornell University Press, 163 p. (*Cornell Studies of Industrial Labor Relations*, vol. 6).

Box George E. P., 1976. Science and statistics, *Journal of the American Statistical Association* vol. 71(356), p. 791–802.

Cox David R., Oakes D., 1984. *Analysis of survival data*. London, Chapman & Hall, VIII + 201 p.

Gavrilov Leonid A., Gavrilova Nathalia S., 1991. *The biology of life span*. New York: Harwood, 392 p.

Halley Edmond, 1693 [republished *in:* Smith David, Keyfitz Nathan (eds.)], 1977. *Mathematical demography*. Berlin, Heidelberg et New York, Springer, 514 p.

Heckman James J., Singer Burton, 1984. Econometric duration analysis, *Journal of Econometrics*, vol. 24, p. 63–132.

Horiuchi Shiro, Wilmoth John R., 1998. Deceleration in the age pattern of mortality at older ages, *Demography*, vol. 35, p. 391–412.

Jeune Bernard, Vaupel James W. (eds.), 1999. *Validation of exceptional longevity*. Odense (Danemark), Odense University Press, 249 p.

Le Bras Hervé. Lois de mortalite et age limite, *Population*, vol. 31, p. 655–662.

Manton Kenneth G., Stallard Eric, Vaupel James W., 1986. Alternative models for the heterogeneity of mortality risks among the aged, *Journal of the American Statistical Association*, vol. 81, p. 635–644.

Thatcher Roger, Kannisto Väinö, Vaupel James, 1998. *The force of mortality at ages 80 to 120*. Odense, Odense University Press, 104 p. + fig. and tabl. (Coll. Odense Monographs on Population Aging, vol. 5).

Vaupel James W. *et al*., 1998. Biodemographic trajectories of longevity, *Science*, no. 280, p. 855–860.

Vaupel James W., Carey James R., 1993. Compositional interpretations of medfly mortality, *Science*, no. 260, p. 1666–1667.

Vaupel James W., Yashin Anatoli I., 1985. Heterogeneity's ruses: some surprising effects of selection on population dynamics, *American Statistician*, vol. 39, p. 176–185.

Vaupel James W., 1990. Kindred Lifetimes: Frailty models in population genetics *in:* Julian Adams *et al*., (eds.), *Convergent questions in genetics and demography*, p. 155–170. London, Oxford University Press, XII + 361 p.

Vaupel James W., 1991. Relatives' risks, *Theoretical Population Biology*, vol. 37, p. 220–234.

Vaupel James W., Manton Kenneth G., Stallard Eric, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, vol. 16, p. 439–454.

Vaupel James W., Yashin Anatoli I., Manton Kenneth G., 1988. Debilitation's aftermath, *Mathematical Population Studies*, no. 1, p. 21–48.

Yashin Anatoli I., Iachine Ivan A., 1997. How frailty models can be used for evaluating longevity limits, *Demography*, vol. 34, p. 31–48.

Yashin Anatoli I., Vaupel James W., Iachine Ivan A., 1994. A duality in aging: the equivalence of mortality models based on radically different concepts, *Mechanisms of Ageing and Development*, vol. 74, p. 1–14.

Yashin Anatoli I., Vaupel James W., Iachine Ivan A., 1995. Correlated individual frailty, *Mathematical Population Studies*, no. 5, p. 1–15.

CHAPTER

22

Mobility and Spatial Heterogeneity

DANIEL COURGEAU
Institut national d'études démographiques (INED), Paris, France

Up to Chapter 20 of this treatise, discussion has been concerned with the population of a country or region whose individual members possess the same patterns of behavior, with the same levels of mortality and fertility and with equal propensities to exit or enter the population. This assumption of population homogeneity, taken with the condition of independence between demographic phenomena, forms the basic premise of the traditional analytical approach to demographic processes (Henry, 1959).

As was seen in Chapter 21, the assumption of individuals' social and spatial homogeneity needs to be abandoned so that the heterogeneity of the populations under study can be included in the analysis. Later it will be seen that the assumption of independence between phenomena must also be abandoned (Chapter 23). We begin, however, by introducing individuals whose demographic behavior is dependent on their region of residence, and we observe the changes in their spatial distribution over time, produced by their mobility. Analysis needs to consider behavior patterns with respect to fertility, nuptiality and mortality for the territory under observation and the migration flows between its component regions.

I. CONCEPTS OF MOBILITY

Demographers have traditionally treated the events that make up their field of study as being located primarily in time while completely ignoring their spatial aspect.[1] This oversight has long been a handicap to the analysis of mobility, whose very definition is impossible without reference to space. Also, construe the space in question not as an immutable entity that is external to the population but as continuously produced and transformed through individual and collective activity, a contextual element being fundamental to human groups and social institutions. Linking these different social spaces is mobility—by the migrants exchanged between them but also by all manner of spatial flows that vary continuously over time (tourism, cross-boundary commuting, professional traveling, and so forth).

We can see from this why a definition of mobility cannot be based on the multiplicity of flows that by their nature are transient and ever changing. A clear definition can, however, be based on the social space in which this mobility occurs. In this case it can be said that the study of spatial mobility involves identifying the changes occurring over time in the relations between an individual or social group, and space (Courgeau, 1988). This space-focused definition expresses the basic principle behind the approach—individuals are constantly mobile, so the aim is to describe and analyze this mobility and its changes over time. This definition also avoids the search for a definitive and universally valid typology of migrations, which is nonexistent. Such a classification would

[1] The revival of interest in this aspect is recent, thanks in particular to developments in multilevel analysis (Courgeau and Baccaïni, 1997, 1998).

in effect fix the phenomenon in a rigid framework that would be unable to accommodate either the appearance of new forms of mobility or the disappearance of existing forms. The advantage of the approach as proposed lies precisely in permitting analysis of these new forms of mobility, which can take completely unexpected forms.[2]

However, the temporal dimension of mobility is also an important element and must be incorporated in analysis. It is particularly important for relating this mobility to other demographic or economic phenomena and for identifying the very close interdependencies that exist between the events in the individual's life course (Courgeau, 1985). Event history analysis (Chapter 23) is an effective tool for illustrating these interactions.

A definition along the lines indicated also allows integration of the various concepts of migration or spatial mobility into a systematic analytic framework. We will show how these concepts can be defined in such a way that they combine to form a consistent and nested conceptual structure.

1. Change of Residence

The notion of residence was introduced to provide a relatively straightforward way of determining the position of each individual in a particular country. As used in demography, the term individual residence means the place where the person lives, in the sense of *domicile*, denoting the place where a person is legally deemed to reside. The latter criterion must be sufficiently unambiguous for us to be able to attribute a single place of residence to each individual at a specified time. We will see later that definitional changes can hamper comparability over time, while definitional differences between countries are an obstacle to comparability in space.

Assuming these problems to have been resolved, at least for the members of a particular subpopulation, we can say that migration

> . . . involves a change in usual place of residence. The place left by the migrant is the place of origin or place of departure; the place to which the migrant goes is the place of destination or place of arrival.

If places of origin and places of destination are equated with actual places of residence, all migratory moves will be included: this concept enables international comparisons to be made, on condition that place of residence is defined in the same way in all countries.

[2] See in particular the new forms of mobility identified by Hervé Domenach and Michel Picouet (1987) and Ronald Skeldon (1990), in the developing countries, and more generally by Wilbur Zelinsky (1971).

Failing this, migration that occurs inside the regions that define the places will not be enumerated as such, thereby making comparison impossible, due to the differences in geographical divisions between different countries.

Also to be considered are the individuals and subpopulations for whom the concept of residence is either meaningless or problematic. A more detailed examination is required for these particular cases. Groups and individuals such as nomads and seminomads, sailors, and tramps, cannot be attributed a fixed place of residence, and so for them the concept of change of residence is irrelevant. These subpopulations must be clearly identified from the outset and alternative concepts employed.

In other cases it is defining a single place of residence that is problematic. For example, should soldiers in garrisons, patients in hospitals, boarders in schools, and so on, be treated as residing at these locations or at their personal place of residence? Such cases call for the adoption of classificatory criteria, which, although usually arbitrary, should if possible be kept constant over time so that comparisons can be made. This last condition is not always respected, creating discontinuities in statistical series. Thus in France until 1954 the usual place of residence of servicemen, school boarders, and so forth was where they lived on a more or less temporary basis. Since the 1962 census, for those who also possess a personal residence, it is this that is counted. By contrast, for prison inmates and individuals in communal accommodation (i.e., members of religious communities), the place of communal residence is used, whether or not they also have a personal residence.

Thus we can see that although the concept of residence is useful in census-type operations to determine the population of a particular place, it is less accurate when applied to the study of spatial heterogeneity, where movements of a more temporary nature are also important for an understanding of societal change. In these conditions new concepts have to be used.

2. Changes of Dwelling Unit

Given the impossibility of distinguishing clearly between temporary mobility and migration, one solution is to consider both forms of mobility simultaneously. The concept of dwelling or dwelling unit, defined as denoting housing accommodation appropriate for occupation by one household, appears preferable to that of residence, since even when parts of the population live in nonconventional housing—in tents or on boats, for example—the term *dwelling unit* is still comparable in meaning. This concept is much more flexible than that of residence and means that at any given time an individual can be associated with a specific dwelling. Nomadic or seminomadic populations can thus be tracked over time, because they are associated with a dwelling unit.

With the details of this concept made clear, it is straightforward to define migration as a change of dwelling. Such a definition both extends and encompasses the concept of change of residence. Criteria can be selected, which vary between countries and over time, with which to decide whether a change of dwelling is in fact also a change of residence. The population registers for Hungary, for example, record both migratory and temporary moves and show the latter to be twice as numerous as the former.

The dwelling unit concept is becoming essential for a correct analysis of spatial mobility. The developing countries, in particular, are witnessing a rapid growth of temporary migratory movements with a preeminent role in economic and other areas of national life. But in the industrialized countries too, new spatial structures are now emerging in which temporary mobility has an increasingly important role.

A further advantage of this concept is its formal equivalence with the concept of migration. In both cases an individual is tracked over his or her lifetime and at each point in time is associated with a single location, which may be a place of residence or a dwelling unit. Both types of mobility can thus be approached using similar methods of analysis.

3. Change of Life Space

If the identification of an individual with a specific place at a specific time is to be satisfactory it must give an effective summary of the spatial area within which he or she interacts. Although this was certainly the case in the past, notably for much of the rural population, the space in which individuals nowadays lead their lives is expanding greatly. Substituting the place of residence or dwelling unit for this space is no longer satisfactory and the definition must be enlarged to include a new dimension.

This dimension is provided by the life space concept, which encompasses all the places with which individuals interact at any given time. Included in this space, as well as the places they actually visit or stay in, are those to which they are relationally linked. To facilitate quantitative analysis, however, this space has to be limited to a smaller number of places that can be precisely defined.

In an initial simplified form, life space is the individual's spatial work-residence pattern. The commuting journeys thus defined can be studied in relation to the migratory moves made by individuals. It is interesting to note that although at certain times in the life course this may be two distinct places, as when people are economically active and do not work at home, it is the same place when people work at home or are retired. This new concept thus corresponds to a space whose dimensions can vary over time.

Other forms of life space could include holiday homes or the places of residence of an individual's family and friends. On condition that these different places and the individuals associated with them are specified sufficiently carefully, the life space can be made more complex yet remain quantifiable.

From a formal perspective, the life space concept differs significantly from the two previous concepts. It involves monitoring along spatial and temporal dimensions not one but a varying number of locations. A spread or extension may occur through the addition of new points of spatial attachment. An example of this is when individuals start employment yet retain all their existing points of attachment. Conversely, elimination of some existing points causes the life space to contract or shrink. This is frequently observed when a person retires. When some existing points are lost but new ones acquired, we could speak of a shift in the life space, such as accompanies a change of workplace, for example. Finally, this shift could become transplantation, in the case that contact is broken with all former points. This results in a complete change in the person's spatial implantation and the occupation of a new territory.

Collection and analysis of information on these life space changes requires innovative survey designs and new methods of measurement and investigation. Though still in its early stages this work is rich in potential for future research.

II. MEASUREMENT ISSUES

The concepts of mobility are thus much more numerous and complex than those associated with other demographic phenomena, and they also imply use of sharply contrasted measurement procedures. Each of these measurement procedures is in addition associated with a specific type of data source that must be reviewed at the same time.

III. MIGRATORY MOVES, REGISTERS, AND SURVEYS

The first approach considers all the migratory moves undertaken by an individual in the population being studied. Individuals are counted as many times as they move. Working from this standpoint, therefore, what is being counted is a number of events, whether these are changes of residence or of dwelling. This is the measure used in population registers or in prospective and retrospective surveys, already mentioned in Chapters 13 and 14.

The most satisfactory method for completeness of coverage is certainly that of population registers, when they are well maintained. Migratory moves are then recorded as and when they take place. Unfortunately, only a few countries, mainly in northern and Eastern Europe, operate and maintain correctly such a system of continuous registration. Furthermore, they are not especially reliable for measuring international migration, compared with migratory movements within national boundaries. They record only a limited amount of personal information, such as civil status, nationality, and occupation, which is unsuitable as a basis for very detailed analysis, compared with information collected by surveys. As operated in some countries, however, the registers record changes both of residence and of dwelling (Hungary and Poland, for example) and in some cases even changes in the life space, as measured by residence, dwelling and work place (Denmark), thus allowing a more detailed measurement of population mobility.

Prospective surveys record moves as they occur and yield high quality information when carried out in good conditions. Unfortunately, almost insurmountable problems arise with data collection when following up such a sample in the national territory and, to an even greater extent, abroad, because the sample is likely to become biased as a result of selective loss of individuals. People who have been through a difficult experience or who have made a migratory move without leaving an address at their place of origin form special subsamples of the population; they will either refuse to be reinterviewed for the survey or will be impossible to trace for the follow-up. Also, the long interval between initiating a survey and analyzing the results leads many investigators to prefer retrospective surveys. It is true that a good-quality analysis requires an observation period of at least 10 years and often longer.

Thus, investigators usually opt for a retrospective survey in which a sample of individuals are questioned on their past mobility. By selecting cohorts of a fairly advanced age at the time of the survey, information is obtained that relates to much of their lifetime. With this procedure the problem of sample attrition does not arise since all individuals are questioned on a single occasion. In addition, the investigator conducting the survey is able to analyze the results immediately, since the totality of retrospective information is collected. However, there can be serious memory recall problems given that individuals are required to remember events that may have occurred long ago and to date them as accurately as possible.

These recall problems were tested empirically with a retrospective survey conducted in Belgium, a country that also possesses a population register (Poulain *et al.*, 1991, 1992). Although the retrospectively collected information is of significantly poorer quality for migration compared with other demographic phenomena (marriage or birth of children), we were able to verify that the material errors in the retrospective collection did not generate large errors in the longitudinal analysis (Chapter 13) or event history analysis (Chapter 23) of these moves (Courgeau, 1991a, 1992). It seems that the errors introduce a random noise in the exact dating of events but do not create bias in the analysis results. Memory would thus appear to be reliable on the points that are crucial for the analysis.

Another source of bias in these surveys is the fact that interviewing is restricted to individuals who have survived up to the survey and who have not emigrated abroad definitively. The assumption has to be made that, for as long as they are in the territory being studied, their migration behavior is the same as those present at the time of the survey. The hypothesis is made that the sample is noninformative (Hoem, 1985) with respect to the behavior we wish to study. The validity of this hypothesis can be tested for countries that possess population registers. This verification has been undertaken for the study of other demographic phenomena (Lyberg, 1983), and the hypothesis does seem to be supported, but we are not aware of any studies that have attempted it for migratory movements.

Surveys, finally, provide information suitable for the study of life spaces, ranging from the simplest to the most complex. They allow us to show the changes that occur over the lifetimes of individuals and to relate these to their personal characteristics and to the characteristics of the sociospatial environment in which their lives are led.

1. Migrants, Latest Migration and Censuses

Exhaustive measurement of migration activity in the census, even if limited to the intercensal period, is impractical for reasons of cost. Consequently we have to make do with less detailed information, derived from just one or two questions, to provide a general view of population mobility.

If we compare the places of residence of individuals at the beginning and end of the period, we obtain the number of migrants when these two places are different. The focus in this approach is the individual, ignoring any intervening moves that may have occurred in the period and comparing only the origin and destination points. The period considered is usually of 1 or 5 years, producing annual or 5-year rates, respectively, for use in analysis. In some countries, such as France, the migration question in the census concerns the intercensal period. The reason for this is that knowing the populations of a region at two successive censuses, its natural increase over the period, number of internal and international immigrants, and number of internal emigrants, allows the level of international emigration can be deduced. Unfortunately, this estimation by comparison of two successive censuses usually generates a very large error term that makes the method unreliable (Baccaïni, 1999).

The *place of birth* question asked in most censuses can be used to measure what are known as lifetime migrants (i.e., whose place of birth was in a different administrative unit to their current residence). Their number is a useful general indicator of migration activity in countries with very low levels of mobility but is of limited interest for most modern countries where individuals undertake a great many migratory moves over their lifetime.

Another approach is to ask only about individuals' last residential migration and previous place of residence. With this we measure the number of latest migrations or more exactly the number of migrants by latest migration. In contrast to the single question on migrants, measurement of latest migrations precludes two analytical possibilities. First, whereas the number of migrants can be measured on any of a country's territorial subdivisions, by asking for the previous address in as much detail as possible the latest migrations in effect impose a particular migration-defining division. Reconstituting the numbers of migrants that would be observed using different subdivisions is thus impossible. Let us assume, for example, the latest migrations are measured between departments. If we wish to estimate the most recent migrations between communes, those whose most recent move was inside a department and who had previously made an interregional migration, will not be captured. Conversely, if we wish to estimate the most recent interregional migrants, those whose previous move was between departments in the same region will not be identified either. An additional drawback is that the most recent migration, the date of which will vary between individuals, cannot be used to make population projections with migration, as this requires the numbers to be measured over fixed 1- or 5-year periods.

For these reasons the questions on the latest migration, although often still included in censuses, should not be used. Measurement of migrants avoids these disadvantages and is strongly recommended.

IV. MOBILITY OVER TIME

In traditional longitudinal analysis of age-specific all-orders migration, the rates are calculated in exactly the same way as the age-specific emigration rates discussed in Chapters 13 and 14. These rates are seldom used, because they are not amenable to a simple modelling of migration flows. However, if mobility is analyzed by order of move and duration of residence, we obtain rates and probabilities that are much less volatile and easier to apply in modelling attempts. Modelling with a small number of parameters is in fact essential in the case of migration activity, given the complexity of the phenomenon and the large number of flows to be introduced, especially when working with a division into regions.

Let us see first of all how these rates and probabilities can be estimated, before applying them in the modelling process and comparing the numbers of migrants and migratory moves obtained with the various measurement methods.

1. Mobility by Order of Move and Duration of Residence

We proceed for first migrations in the same way as for first births. Data from retrospective surveys are unlike data drawn from prospective surveys and population registers because they do not include deaths. If we denote $M_x^{i,1}$ the number of first migratory moves observed between ages x and $(x+1)$ in the birth cohort i, and $P_x^{i,0}$ the nonmover population of age x, the migration probability mg of rank order 1 is written, when working with retrospective data:

$$mg_x^{i,1} = \frac{M_x^{i,1}}{P_x^{i,0}}$$

When working with prospective data the mortality of nonmovers between ages x and $(x + 1)$, $D_x^{i,0}$, has to be introduced, which gives the following migration probability:

$$mg_x^{i,1} = \frac{M_x^{i,1}}{P_x^{i,0} - 0.5D_x^{i,0}}$$

For higher than first-order migrations, numerous studies have established that it is much better to work on the duration of residence between successive moves rather than on the age of the individual. This is because with the latter the populations at risk obtained for the young ages are very small, resulting in rates or probabilities with substantial variance that are liable to lead to completely wrong estimations if used for distributions or survival functions. Such problems are avoided if we work on duration of residence, and the much more uniform curves that result can be used for a simple modelling of successive migrations.

In the case of migration of order n, we denote $M_t^{i,n}$ the number of migrations of order n occurring in an interval between t and $(t + 1)$ of the previous migration in cohort i, $P_t^{i,n-1}$ the population of individuals who have made $(n - 1)$ migrations and not yet made the next order migration, and $O_t^{i,n-1}$ the losses from observation affecting this population. The latter have to be included because we are working on the duration of residence since the last migration. The migration probability for rank order n is then written, if we have retrospective data:

$$mg_t^{i,n} = \frac{M_t^{i,n}}{P_t^{i,n-1} - 0.5O_t^{i,n-1}}$$

If we have prospective data, we need to introduce the deaths in the population of individuals having made a migratory move of rank order $(n - 1)$ at time t, $D_t^{i,n-1}$:

$$mg_t^{i,n} = \frac{M_t^{i,n}}{P_t^{i,n-1} - 0.5O_t^{i,n-1} - 0.5D_t^{i,n-1}}$$

The analysis can be further developed by calculating the probabilities by duration of residence and age at previous migration. To do this we obviously need to be able to work on large numbers or combine ages to form age groups.

For illustrative purposes let us examine migrations by the Norwegian 1948 birth cohort, followed from age 15.[3]

Figure 22–1*a* shows that first-order migration is clearly differentiated from higher order migrations.

The migration probabilities for rank order 1 rise to a maximum at age 23 whereupon they fall rapidly. In contrast, the migration probabilities for higher order moves are all very similar and follow the same pattern of a continuous sharp fall without reaching a maximum. It can also be noted that, at least for durations up to 6 years, the probabilities are higher the lower the migration rank order, although the differences are never great. If the age group in which the previous migration was made is also introduced, we obtain for rank orders of move 2, 3, and 4 the results in Figures 22–1*b*, *c*, and *d*. These show clearly that these probabilities depend on the age at the previous migration—the younger the age at this migration the higher the probabilities—but they are now almost independent of the order of move being considered. These results show that migrations of orders higher than one can be modelled using a small number of parameters.

2. Time-dependent Models

These sets of migration probabilities by order of move and duration of residence can be conveniently summarized by means of simple models. The various migration models encountered in the literature cannot be discussed in detail here and we will simply consider the most commonly used.

The oldest and most widely used formal model is certainly the so-called *mover–stayer* model (Blumen *et al.*, 1955). Its basic hypothesis is that only a fraction of a population that has made a previous migration is at risk of making a subsequent move. The first step in the modelling procedure is to disaggregate the population that has made a migration of order $n - 1$ into a subpopulation that will make no further move and, on the other hand, a subpopulation that will move again. For this second subpopulation the hypothesis is then made that the probability of making a new move is independent of the duration between the move of order n and move of order $n - 1$. If we assume that the proportion of potential migrants is K, the proportion of stayers at time t may be written as

$$S(t) = (1 - K) + K\exp(-\mu t)$$

where μ is the instantaneous migration hazard in the population of potential migrants. We verify that when $t \to \infty$ the proportion of stayers tends to $(1 - K)$, as predicted by the model. The curve representing the hazard rates of order higher than 1 does indeed have the same form as the curves in Figure 22–1.

The relatively constant scale of these parameters can be confirmed by working on the sample of women in the INED *Triple Biography* (3B) survey from which respondents under age 50 have been excluded. Table 22–1 gives the parameter estimate values of K for migrations of successive orders and involving changes of dwelling, commune, or department. The proportions of potential migrants for any given type of migration are seen to be virtually independent of migration rank order. However, this constant declines slightly as

[3] We would like to thank the Norwegian statistical services for allowing us to use the data files produced from their population registers and censuses.

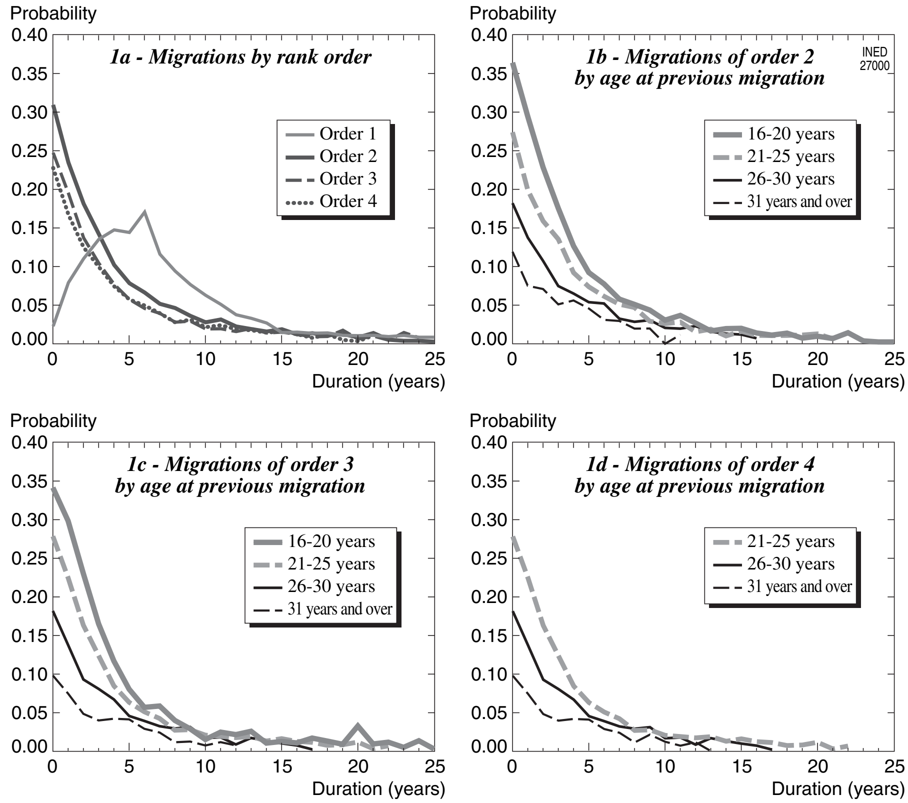


FIGURE 22–1 Migrations in the Norwegian 1948 birth cohort. (Source: Population Register of Norway).

TABLE 22–1 Probability of a Subsequent Residential Migration in France, for Successive Migrations and Different Types of Move

Change	Individuals aged 50 or over, %					No. observed
	*K*2	*K*3	*K*4	*K*5	*K*6	
Dwelling	0.89	0.86	0.87	0.91	0.84	1954
Commune	0.77	0.75	0.76	0.76	0.74	1610
Department	0.72	0.64	0.70	0.71	0.67	1028

Source: INED, "3B" survey.

we proceed from changes of dwelling to changes of commune and to changes of department.

Let us now consider changes of dwelling and see how parameters K and μ vary as a function of age at the previous migration and of the order of move. The results appear in Table 22–2.

This shows a slight decline in parameters K and μ as mean age at the previous migration rises. On the other hand, for a given age at the previous migration, these two parameters are only slightly dependent on order of move. This result is very similar to what we determined by empirical observation for the Norwegian migrations.

Modelling of successive migrations is also possible using the *Gompertz* model. This model postulates that the variation in the instantaneous hazard rate is proportional to the population at risk at a particular point in time. Under these conditions the proportion of stayers is written:

$$S(t) = (1 - K)^{1-\exp(-\mu t)}$$

where K again represents the proportion of individuals who will migrate, and μ is a parameter expressing this relationship between hazard rate change and population at risk (for more details see Courgeau and Lelièvre, 1989, 1992, 2001). When $t \to \infty$, we verify that the proportion of stayers tends to $1 - K$ as predicted by the model.

Ginsberg (1979) conducted a comparative assessment of these two main models with several others using data from the Population Register for Norway. From this it emerged that the fit obtained with these two types of model was on the whole equally good for all the flows under consideration (by age, marital status, income, region of origin, etc., representing 308 different flows) and considerably better than that obtained with the other distributional models (exponential, Weibull, log-normal, etc.). However, the Gompertz model is more flexible, because it allows both

TABLE 22–2 Parameters K and μ for Changes of Dwelling of Successive Orders, Pre-1931 Birth Cohort, France

Mean age at move of order $(n-1)$, x	K_2	μ_2	No. observed	K_3	μ_3	No. observed	K_4	μ_4	No. observed x
17.5	0.91	0.19	820	0.92	0.20	309	0.91	0.18	113
22.5	0.78	0.16	563	0.84	0.18	420	0.87	0.17	242
27.5	0.67	0.15	114	0.70	0.16	222	0.80	0.16	192

Source: INED, "3B" survey.

increasing and decreasing hazards, and can thus accommodate cases in which the flows follow distributions that are nonmonotonic decreasing.

3. Migrants and Migrations

We are now in a position to model the differences between measurements of migrations derived from population registers or surveys and census-based measurements of migratory flows. Such modelling is necessary because census data are often only available over periods of varying length, while for making comparisons across time and across countries it is preferable to work with results relating to flows and to annual migration rates.

The model we present here is in a simplified form, and while it can be made more realistic by adopting more complex hypotheses, this does not fundamentally modify the theoretical argument or the results.

The first assumption made by the model is that the instantaneous migration probability, p in the total population is time invariant. This hypothesis can be considered to be valid for a short period of time though it may well not hold over longer periods. Thus for the United States the probability remained at around 0.19 from the end of the World War II to the start of the 1970s, when it fell slightly to 0.17 in subsequent years. This shows that the hypothesis is satisfied for this country, and the same is true for many other countries (Long and Boertlein, 1976). The second assumption is that a mover–stayer model adequately describes the migrations observed in the country. As noted, this hypothesis is often verified correctly. In addition, parameters K and μ are assumed not to change as a function of order of move or of age at previous migration. This hypothesis is not fully verified for migration in France and Norway, but it can serve as a satisfactory first approximation of behavior. For ease of calculation the assumption is also made that the population does not change during the interval under consideration: It is equal to P.

Over a very short interval $(\theta, \theta + d\theta)$ this population will make $Ppd\theta$ migrations. Under the mover–stayer model, only a proportion of these movers will go on to make a new migration: $PpKd\theta$. Let us consider the distribution over time of these additional moves. During the time interval $(t, t + dt)$ these new migrations will satisfy the following formula:

$$\frac{d[M_n(t)]}{PpKd\theta - M_n(t)} = \mu dt$$

where $M_n(t)$ represents these new migrations occurring between θ and t. Integrating between θ and t, we obtain this number:

$$M_n(t) = PpKd\theta[1 - \exp(-\mu\{t-\theta\})]$$

By varying θ, between an initial time ($\theta = 0$) and a final time ($\theta = t$) we will have counted all the new migrations that occur in this period:

$$\int_{\theta=0}^{\theta=t} PpK[1-\exp(-\mu\{t-\theta\})]d\theta$$
$$= PpK\left[t - \frac{1}{\mu}(1-\exp\{-\mu t\})\right]$$

Lastly, if there are no return migrations, by calculating the difference between the total migrations that will be recorded during the period ($M(t) = Ppt$) and these multiple migrations, we obtain the number of migrants that would be given by a census at time t in the place of residence at the initial time:

$$m(t) = Pp\left[(1-K)t + \frac{K}{\mu}(1-\exp\{-\mu t\})\right]$$

Introduction of a geographical division into units such as communes and departments, for example, means that return migrations also have to be included. If we assume that these return migrations $r(t)$ form a constant proportion of the migrations of order higher than 1 that occur in a year, this hypothesis is written:

$$r(t) = l\sum_{i=2}^{\infty} m_i(t)$$

where $m_i(t)$ represents the migrations of order i that occur during the year t. This hypothesis is verified for France (Courgeau, 1973a, 1979) and allows us to write the number of migrants as:

$$m(t) = Pp\left[(1 - K(1+l))t + \frac{K(1+l)}{\mu}(1 - \exp\{-\mu t\})\right]$$

Applying this model to the data for the United States and France produces the parameter estimates reported in Table 22–3.

With these parameter estimates we can compare the main characteristics of migrations in the two countries. We note first that the instantaneous migration hazard p in the United States is almost twice what it is in France (0.183 compared with 0.104). This greater mobility is not associated with a smaller proportion of nonmovers, since in both the United States and France approximately 80% of individuals who have made a residential migration subsequently make another one. The essential difference, however, is the time that elapses before these new migrations occur. In the United States, nearly half the population at risk makes this new migration in the year following the previous migration, compared with less than one fifth in France.

A comparison can also be made between changes of dwelling and changes of county in the United States. These are characterized primarily by a much lower rate of intercounty mobility, whereas this time the proportions of nonmovers and the interval between moves are very similar in each case. It must be noted, however, that the data do not enable us to distinguish authentic nonmovers ($1 - K$) from return migrants l, since we can only estimate a coefficient that measures nonmovers corrected for returns, $K(1 + l)$. For changes of dwelling, the estimated parameter obviously relates only to nonmovers, because moves back to a previous dwelling are negligible. Figure 22–2 displays the curves observed for migrants, compared with the curves estimated by the model, for changes of dwelling and of county in the United States.

TABLE 22–3 Parameter Estimates of the Migrants–Migrations Model, United States and France

Coefficients	Change of dwelling		Change of county United States (1960–1970)
	United States (1960–1970)	France (1968–1970)	
μ	0.490	0.180	0.520
p	0.183	0.104	0.075
$K(1 + l)$	0.810	0.780	0.770

Source: Courgeau, Daniel, 1982. Comparaison des migrations internes en France et aux États-Unis, *Population*, vol. 37(6), p. 1184–1188 and Long, John F. and Boertlein, Celia G. 1981. *Using migration measures having different intervals*. Washington DC, US Bureau of the Census, 39 p. (Population Division working paper).

We note first the near-perfect match between the curves showing observed and estimated data. The corresponding migrations show the large size of the differences between migrants and migrations, of approximately 1:2, after 5 years.

This model can of course be made more realistic by introducing a variation over time in the probability of moving and dependencies between parameters K or μ and age at the previous migration. This involves no fundamental change to the formulas used.

V. MOBILITY, SPACE, AND SPATIAL MODELS

The next step in the analysis is to incorporate the spatial context in which these migrations occur. We define a territorial division into r regions between which flows of migrations or of migrants are observed. Let us begin by looking at the different rates and indices that can be defined to describe these flows.

1. Interregional Mobility Rates

Let us consider the totality of inflows and outflows affecting a particular region. The most favorable situation is when the analyst has data from population register or survey sources. For region i, we have already defined the annual age-specific emigration rates, x, e_x^i, and the rates for age groups of n years ${}_ne_x^i$ (see Chapter 13). We also explained that age-specific immigration rates are harder to define, since the population at risk is no longer

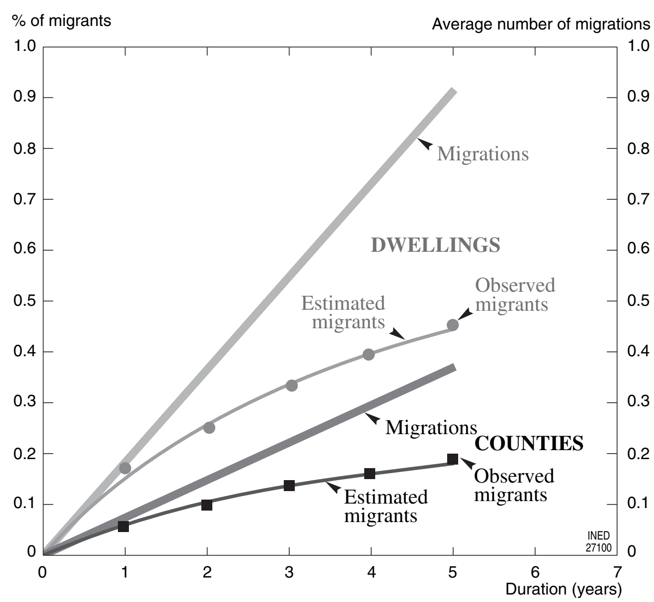


FIGURE 22–2 Proportions (%) of estimated and observed migrants and average number of migrations, United States).

that of region i but of the rest of the world. In this case the preferred solution is to use the same denominator as for the emigration rate, though by so doing we no longer have a conventional demographic rate.

The situation is more complex when census data are available for migrant numbers over a duration of n years. Calculation of a corresponding annual rate is now made impossible by the multiple moves and return migrations that occur during this period (Courgeau, 1973a, 1979). Thus all we can calculate is a *probability of emigrating* over n years, or *proportion of migrants*, by dividing this age-specific number by the population of individuals surviving at the census and who resided there at the start of period P_x^i:

$$_n\varepsilon_x^i = \frac{_nE_x^i}{P_x^i}$$

In this case, the proportion of immigrants will also use the same denominator, even although this term no longer represents the population at risk.

If we consider now the flows between two regions, two types of rates can be calculated. The first relates the number of migrations or migrants only to the average population of the area of origin, considered to be at risk. Thus the rate of emigration from i to j at age x can be written

$$e_x^{ij} = \frac{E_x^{ij}}{0.5(P_x^i + P_{x+1}^i)}$$

The second type of rate includes in addition the population of the area of destination. Numerous empirical studies have established that the flow depends as much on the population of the destination area as on that of the origin area (Courgeau, 1970; Poulain, 1981). We are then able to define an *index of migration intensity* between two regions, i and j, in the period (t, $t + n$), for migrants of age x at the start of the period, over n years:

$$mg_x^{ij} = \frac{E_x^{ij}}{P_x^i P_{x+n}^j}$$

This index can be interpreted as the probability that two individuals alive at the end of the period and selected at random, one from the population living in i at the start of the period, the other from the population living in j at the end of the period, will be identical (Courgeau, 1980). The effect on the migrants of the origin and destination populations is thus eliminated. It can be demonstrated that when two or more regions are combined, as origins or destinations, the intensity index is a mean, weighted by the products of the populations, of the intensity indices for each pairing. It follows that when the indices between each pairing are equal, the index for their combination will always take this shared value.

2. Spatial Models

For a country divided into r regions and a population containing y age groups, $yr(r-1)$ indices must be calculated in order to characterize all the flows affecting the individuals. It is not hard to see the advantage of condensing this information into a smaller number of summary indices with which to characterize mobility in the country in question. This can be done by means of explanatory models in which interregional distance—in its spatial and more sociological conceptualization—is introduced as a variable. This distance is known to be a deterrent to migration, a phenomenon related to the sharp reduction in relational networks that occurs as the distance between individuals increases. In the discussion that follows, we present only the most commonly used model (for a more comprehensive review of the other types of model in use see Courgeau, 1980).

Let us suppose that the distance between two regions i and j is d^{ij} and that the index of migration intensity between these two regions is mg_x^{ij}. A Pareto-type distribution can then be fitted to the set of flows observed:

$$mg_x^{ij} = \frac{k}{(d^{ij})^n}$$

where parameters k and n are estimated from the data. These models give a generally good account of the migration flows and yield coefficients of determination of between 0.8 and 0.9 (Poulain, 1981).

This kind of model can be used to describe the distribution of migrants across various divisions of the territory. Using different spatial units and running the model with different exponents, we have given a theoretical demonstration (Courgeau, 1973b) that there is a simple relation expressing the migration rate observed as a function of the number of units in the division used, r, which is equal to the ratio between the total population of the country and that of an average unit:

$$mg_r = K \log r = K(\log P - \log P_r)$$

Empirical verification of this law is straightforward for countries for which migration rates by various territorial divisions are available. Table 22–4 presents the results observed in Holland and France.

The annual migration rates in France were estimated using the migrants—migrations model presented above and are therefore fully comparable with the rates observed using the Dutch Population Register. The model is found to give a good fit for both

TABLE 22–4 Relation Between Migration Rate and Average Population of a Spatial Unit, Holland and France

Territorial division	Annual migration rate, %	No. spatial units	Parameter K
	Holland (1982)		
Dwelling	10.59	5,310,945	0.0159
Municipality	3.90	774	0.0156
Economic region	3.30	129	0.0165
Corop region	2.23	40	0.0151
Province	1.54	12	0.0160
	France (1968–1975)		
Dwelling	10.37	17,744,985	0.0143
Commune	6.44	36,394	0.0142
Department	3.09	95	0.0156
Region	1.90	22	0.0142

Source: Courgeau, Daniel, van der Erf, Rob F., and Poulain, Michel, 1989. Mobilité et découpage géographique: comparaison entre Belgique et Pays-Bas, *L'Espace Géographique*, no. 1, p. 39–53 and Courgeau, Daniel, 1980. *Analyse quantitative des migrations humaines*. Paris, Masson, 225 p. (Coll. Anthropologie physique, no. XII).

Dutch and French data. The curves showing the percentages of migrants as a function of the logarithm of the average population of a spatial unit form straight lines almost parallel to each other (constant parameter K). Consequently, if a measure of residential migration was not available, good estimates could be made based on the other types of migrations.

This means that in the countries for which we possess no general measure of residential migration but only measures of mobility between various types of territorial units, a residential migration rate can be estimated that is suitable for comparison with other countries. The divisions for which data are available can also be used to check the validity of the model to ensure comparability. These models have been used to compare population mobility in the European countries (Rees and Kupiszewski, 1999).

VI. MULTISTATE MODELS

We again consider a country divided into r regions. The objective is to represent the probabilities of the various demographic events in matrix form, which we can use to project from the age-specific regional population at time t to that at time $(t + 1)$. This representation is much more complex than that considered in Chapter 20, which allowed only for mortality and fertility in a single population. Moreover, the matrices depend heavily on the hypotheses made about the migrations. In what follows we consider only the two main hypotheses.

1. Priority to Emigration Rates

For ease of exposition we will limit ourselves to a female population. Let us suppose that for each region $i(i = 1, \dots r)$ we know the survival probability between the age groups x and $x + 1$, p_x^i, and the number of girls born between t and $t + 1$ to a woman of age x at t, f_x^i. For the migration, we determine the probability of migrating from region i to region j, ε_x^{ij}. This probability will thus be calculated in relation to the population of the area of origin, like the proportions of emigrants presented earlier. To simplify the calculations we consider the female population, by year of age up to the limit age l and we assume that only one event can occur in the course of a single year.[4] Under these conditions we can project the population of age x of region i at t, $P_x^i(t)$, to its population of age $(x + 1)$ at $(t + 1)$:

$$P_{x+1}^i(t+1) = P_x^i(t)\left[p_x^i - \sum_{j\neq i}\varepsilon_x^{ij}\right] + \sum_{j\neq i}\varepsilon_x^{ji}P_x^j(t)$$

For the births in the course of the year we can write:

$$P_0^i(t+1) = \sum_{x=15}^{49} f_x^i P_x^i(t)$$

The set of probabilities can be represented in matrix form, M, while the age-specific populations of the regions are represented by a vector of $r(l + 1)$ elements, $P(t)$ and $P(t + 1)$, between which the following relation is verified

$$P(t+1) = MP(t)$$

This relation can be used to project the initial population, by making the assumption that matrix M remains constant over time, for example, or changes according to certain rules. We can show that when matrix M does not change, the population will tend towards a stable population structure. The first real element in this matrix tells us if the overall population will increase or decrease. The stable population structure will be reached in two stages: at the end of the first the populations of each region become stable; at the end of the second stability is reached in the regional age structure (Rogers, 1995).

It remains to estimate the probabilities, based on the annual emigration and mortality rates. At age x, for example, we will write the following matrix of the annual rates of emigration, e_x^{ij} and mortality (q_x^i)

[4] More realistic assumptions can easily be made, although presentation of the results then becomes more complex (Rogers, 1995).

$$M'(x) = \begin{bmatrix} \left(q_x^1 + \sum_{j\neq 1} e_x^{1j}\right) & -e_x^{21} & \cdots & -e_x^{n1} \\ -e_x^{12} & \left(q_x^2 + \sum_{j\neq 2} e_x^{2j}\right) & \cdots & -e_x^{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -e_x^{1n} & -e_x^{2n} & \cdots & \left(q_x^n + \sum_{j\neq n} e_x^{nj}\right) \end{bmatrix}$$

It can be shown that the matrix of survival and emigration probabilities can be written:

$$M(x) = \left[I + \frac{1}{2}M'(x)\right]^{-1}\left[I - \frac{1}{2}M'(x)\right]$$

where I is the unity matrix, whose diagonal terms are equal to 1, and all the others zero (Willekens and Rogers, 1978; Rees, 1986). It is of course possible to work on 5-year periods and age groups.

2. Priority to Migration Intensities

The previous solution is not very satisfactory because it is based on emigration rates, whereas as was seen earlier, many studies have confirmed that relations between two regions are better represented by indices of intensity. Thus, if we now let mg_x^{ij} be the migration intensity between i and j, the previous relation between populations at t and $t + 1$, can be rewritten in the following form:

$$P_{x+1}^i(t+1) = P_x^i(t)\left[p_x^i - \sum_{j\neq i} mg_x^{ij}\ P_{x+1}^j(t+1)\right] + P_{x+1}^i(t+1)\sum_{j\neq i} mg_x^{ji}\ P_x^j(t)$$

To identify the populations for estimation more easily, this system of equations can be written in the form:

$$P_{x+1}^i(t+1)\left[1 - \sum_{j\neq i} mg_x^{ji}\ P_x^j(t)\right] + P_x^i(t)\sum_{j\neq i} mg_x^{ij}\ P_{x+1}^j(t+1) = p_x^i\ P_x^i(t)$$

We see in this case therefore that these equations can still be represented in matrix form, but that this matrix will depend on the populations at time t:

$$M[P(t)]\,P(t+1) = P(t)$$

If matrix $M[P(t)]$ possesses an inverse, we can write:

$$P(t+1) = (M[P(t)])^{-1}\ P(t)$$

We still have a relation with which to project from $P(t)$ to $P(t + 1)$, but it is no longer a linear relation. If for example this nonlinear relation is used to project the initial populations under the assumption that the migration intensity values remain constant over time, we no longer obtain a stable population (Courgeau, 1991b; Keilman, 1993). Instead, some populations may disappear, while others may follow a sustained cyclical or even *chaotic* course, albeit one that is fully specified because the model contains no random variable.

CONCLUSION

We have seen that measurement and analysis of population mobility is extremely complex. The phenomenon must be considered in its temporal as well as spatial dimensions if all its aspects are to be grasped. The demographic analysis of migrations has developed new methods to improve our understanding of these flows, while making use of techniques originating in economics and human geography (Courgeau, 1976).

As part of this, linear regression methods can be used to explain the aggregated flows with respect to the characteristics of the origin and destination areas. This involves a generalization of spatial models, in particular that of Pareto, by introducing physical and social distance but also variables such as the effect of unemployment, the percentage of old people, and more generally the economic, social and political attributes of the regions of origin and destination (Puig, 1981).

An alternative approach to migration involves working at the individual level and explaining migration behavior by means either of logistic models, which do not introduce time, or event history models, which do (Chapter 23). Differences in migration behavior are explained here with reference to characteristics related to the occupation, family, and so forth of the individuals to identify the personal motivations in these migrations and the effect of time on the process (Courgeau, 1985).

Last, there is an approach that attempts a synthesis of the two previous strategies by means of multilevel models (Chapter 24). These seek to explain individual characteristics by introducing personal but also aggregated characteristics, and differentiate the behaviors with regard to the regions under examination. These models are a rich analytical tool for the study of migrations, situating the individuals under consideration in a multi-dimensional context that is both geographical and social.

References

Baccaïni Brigitte, 1999. Analyse des migrations internes et estimation du solde migratoire externe au niveau local à l'aide des données censitaires, *Population*, vol. 54(4–5), p. 801–816.

Blumen Isadore, Kogan Marvin and McCarthy Philip J., 1955. *The industrial mobility of labor as a probabilistic process*. Ithaca (New York), Cornell University Press, 163 p. (*Cornell Studies of Industrial Labor Relations*, vol. 6).

Courgeau Daniel and Baccaïni Brigitte, 1997. Analyse multiniveaux en sciences sociales, *Population*, 52, 4, p. 831–864, (also in English: Multilevel analysis in the social sciences, *Population. An English Selection*, 1998, vol. 10(1), p. 39–72).

Courgeau Daniel and Lelièvre Eva, 1989. *Analyse démographique des biographies*, éditions de l'INED, Paris, 270 p. (Also in English: *Event history analysis in demography*, Oxford, Clarendon Press, 1992, 226 p., and in Spanish: *Análisis demográfico de las biografías*, México, El Colegio deMéxico, 2001, 306 p.).

Courgeau Daniel, 1970. *Les champs migratoires en France*. Paris, INED, Presses Universitaires de France, 158 p. (Coll. Travaux et documents, Cahier no. 58).

Courgeau Daniel, 1973a. Migrants et migrations, *Population*, vol. 28(1), p. 95–129, (Also in English: Migrants and migrations, *Population, Selected papers*, no. 3, October 1979, 36 p.).

Courgeau Daniel, 1973b. Migrations et découpage du territoire, *Population*, vol. 28(3), p. 511–537.

Courgeau Daniel, 1976. Quantitative, demographic and geographic approaches to internal migration, *Environment and Planning A*, vol. 8, p. 261–269.

Courgeau Daniel, 1980. *Analyse quantitative des migrations humaines*. Paris, Masson, 225 p. (Coll. Anthropologie physique, no. XII).

Courgeau Daniel, 1982. Comparaison des migrations internes en France et aux États-Unis, *Population*, vol. 37(6), p. 1184–1188.

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life-cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel, 1988. *Méthodes de mesure de la mobilité spatiale: migrations internes mobilité temporaire, navettes,* Paris, INED, 301 p.

Courgeau Daniel, 1991a Analyse de données biographiques erronées, *Population*, vol. 46(1), p. 89–104 (Also in English: Impact of response errors on event history analysis, *Population. An English Selection*, 1992, vol. 4, p. 97–110).

Courgeau Daniel, 1991b. Perspectives avec migrations, *Population*, vol. 46(6), p. 1513–1530.

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, no. 1, p. 7–26. (Also in English: From the group to the individual: what can be learned from migratory behaviour, *Population. An English Selection*, 1995, vol. 7(1), p. 145–162).

Courgeau Daniel, van der Erf Rob F., and Poulain Michel, 1989. Mobilité et découpage géographique: comparaison entre Belgique et Pays-Bas, *L'Espace Géographique*, no. 1, p. 39–53.

Domenach Hervé and and Picouët Michel, 1987. Le caractère de réversibilité dans l'étude de la migration, *Population*, vol. 42(3), p. 469–484.

Ginsberg Ralph B., 1979. Timing and duration effects in residence histories and other longitudinal data, II: studies of duration effects in Norway, 1965–71, *Regional Science and Urban Economics,* vol. 9(4), p. 369–392.

Henry Louis, 1959. D'un problème fondamental de l'analyse démographique, *Population*, vol. 14(1), p. 9–32.

Henry Louis, 1981. Dictionnaire démographique multilingue, Liège, Ordina Editions, 179 p.

Hoem Jan, 1985. Weighting, misclassification and other issues in the analysis of survey samples of life histories, *in:* James Heckman and Burton Singer (eds.), *Longitudinal analysis of labour market data*, p. 249–293. Cambridge, Cambridge University Press, 425 p.

Keilman Nico, 1993. Emerging issues in demographic methodology, Alain Blum and Jean-Louis Rallu (eds.), *European population, II: demographic dynamics*, p. 483–508. Paris, John Libbey, INED, 526 p. (Coll. Congrès et colloques, no. 9).

Long John F. and Boertlein Celia G. 1981. *Using migration measures having different intervals*. Washington DC, US Bureau of the Census, 39 p. (Population Division working paper).

Long Larry, 1988. *Migration and residential mobility in the United States*. New York, Russel Sage Foundation, 398 p.

Long Larry, Boertlein Celia, 1976. *The geographical mobility of Americans: an international comparison*. Washington, Government Printing Office, 46 p.

Lyberg Ingrid, 1983. The effect of sampling and nonresponse in the estimates of transition intensities: some empirical results from the 1981 Swedish Fertility Survey, *Stockholm Research Report in Demography*, no. 14, 15 p.

Nam Charles, Serow William and Sly David, (eds.), 1990. *International handbook on internal migration*. New York, Greenwood Press, 438 p.

Poulain Michel, 1981. *Contribution à l'analyse spatiale d'une matrice de migration interne*, Louvain la Neuve, Cabay, 226 p.

Poulain Michel, Riandey Benoît and Firdion Jean-Marie, 1991. Enquête biographique et registre belge de population: une confrontation des données, *Population*, vol 46(1), p. 65–88 (also in English: Data from a life history survey and the Belgian population register: a comparison, *Population. An English Selection*, 1992, vol. 4, p. 77–96).

Puig Jean-Pierre, 1981. La migration régionale de la population active, *Annales de l'INSEE*, no. 44, p. 42–79.

Rees Philip, 1986. Developments in the modelling of spatial populations, *in:* Robert Wood and Philip Rees (eds.), *Population structures and models*, p. 97–125. London. Allen & Unwin, 439 p.

Rees Philip, Kupiszievski Marek, 1999. *Migrations internes et dynamique démographique régionale en Europe: une synthèse*, Strasbourg, éditions du Conseil de l'Europe, 126 p. (Also in English: *Internal migration and regional population dynamism in Europe: a synthesis*, Strasbourg, Council of Europe, 1999, 113 p).

Rogers Andrei, 1995. *Multiregional demography: principles, methods and extensions*. New York, Chichester, 236 p.

Skeldon Ronald, 1990. *Population mobility in developing countries: a reinterpretation*. London, Belhaven Press, 274 p.

Willekens Frans, Rogers Andrei, 1978. *Spatial population analysis: methods and computer programs*. Laxenburg, IIASA, 302 p.

Zelinsky Wilbur, 1971. The hypothesis of the mobility transition, *Geographical Review*, no. 61, p. 219–249.

C H A P T E R

23

Demographic Event History Analysis

DANIEL COURGEAU AND EVA LELIÈVRE

Institut national d'études démographiques (INED), Paris, France

As was seen in Chapter 10, traditional longitudinal analysis examines the occurrence of a single demographic event, and that event alone, by eliminating the effect of all other phenomena that are treated as interfering events. This form of analysis was developed before large amounts of individual-level data became available through specific surveys and before major advances in probability theory and statistical techniques. Longitudinal analysis studies the phenomena of interest isolated in their *pure state* and uses primarily civil registration or population register data to calculate series of probabilities or rates by age or by durations between two successive transitions. Obviously, this approach is purely descriptive.

The growth of survey-taking in the 1980s (Groupe R. A. B., 1999) and developments in probability and statistical techniques (Cox, 1972; Kalbfleish and Prentice, 1980; Courgeau and Lelièvre, 1989, 1992, 2001; Andersen *et al*., 1993; Lelièvre and Bringé, 1998) equipped demographers to switch their attention from aggregate-level data (the numbers of individuals experiencing an event in a particular year) to individual-level data (an individual with certain characteristics experiences various events at different times and in different contexts over the length of his or her life course). The result is an explanatory approach that attempts to disentangle the mass of factors to identify the interdependencies between phenomena and assess the influence on behavior over time of individual characteristics.

Studying these complex individual histories or trajectories requires the elaboration of an analytical and methodologic framework that generalizes the ideas outlined in the previous paragraphs.

I. FROM LONGITUDINAL ANALYSIS TO EVENT HISTORY ANALYSIS

Traditional longitudinal analysis is based on a double postulate: The demographer studies the occurrence of a single event in a population that remains *homogeneous* for as long as that phenomenon is present and from which the effect of the interfering phenomena, considered to be *independent* of the phenomenon being studied is eliminated (Henry, 1959).

Satisfaction of these conditions is the prerequisite for calculating the conventional demographic rates and probabilities, which are taken to express the risk of experiencing the event in question assuming no losses from observation or interfering events. Chapter 9 showed that the results thus obtained need to be treated with a degree of skepticism. We will now go further and show that these postulates have to be changed if the behaviors are to be analyzed in their true complexity.

Let us begin by considering the condition of independence. Empirical observation and simple reasoning tell us that there is not usually independence between processes (Henry, 1959). It is, for example, reasonable to expect migration to a metropolitan area to alter the fertility behavior of a female migrant from a rural region.

Accordingly, we need to discard this hypothesis and introduce the effects of the other phenomena into the analysis, not by excluding the individuals who experience them as in the traditional analysis, but by assessing their influence on the processes being studied. Also note that when analysis is limited to a single event, the restrictive assumption of independence makes it impossible to study losses from observation due to competing events, as with cause-specific mortality, for example. Nor can we work on a subgroup of the population that can be entered by several different events (as in the case of direct entry to a population of manual workers at the end of schooling or indirect entry after a series of previous experiences in the workplace). It would be unfortunate to forgo these analyses simply because they imply a heterogeneity of the cohort under investigation.

Consider now the condition of homogeneity. It is obvious from the sheer diversity of individuals that no human group will be homogeneous (Henry, 1959). One way of dealing with this heterogeneity is to break down the population being studied into subgroups, so that the population in each can be considered as homogeneous. The resulting multistate table will have an ever-larger number of cells, each containing fewer individuals. Large random errors result, as was seen in Chapter 8, and we can never be certain that all the factors of heterogeneity have been taken into account. Decomposition into subgroups is therefore an unsatisfactory solution and another way must be found to incorporate heterogeneity.

A further restriction is that for a population to remain homogeneous *over time*, the assumption must be made that new members immediately acquire the same characteristics and adopt the same behavior as existing members. Concomitantly, the same applies to those who move out of this subpopulation and join another, whose behavior they are assumed to adopt, while losing all recollection of their past conditions. This amounts to making a Markovian assumption, whereby experience of the present state is totally independent of the individuals' history. Also, the individuals remaining in the subpopulation are assumed to have a behavior that evolves in the same way over time. This assumption again bears little relation to what is actually observed, since real-life individuals obviously do remember their past and this does influence their future behavior.

1. Changing Paradigm

These problems highlight the need for a new paradigm that will enable us to analyze interdependent processes affecting heterogeneous populations. This requires using a different unit of analysis—instead of the single isolated event that was considered previously, analysis now focuses on the entire individual life course, which is treated as a complex stochastic process. The following postulate can then be stated: The trajectory an individual follows over his or her lifetime moves through a succession of different states, and the position occupied at any given moment depends on the individual's life course to date, all that he or she has experienced or acquired, contextual constraints, and the working of personal choice (Courgeau and Lelièvre, 1996). With this change in perspective, the bases of event history analysis can be formulated in terms of individual processes.

The first point to note is that losses from observation (also known as *censored information*) are no longer problematic because the date of the survey or—if population registers are being used—of the study, is not linked in any way to the individual's life. The condition of independence between loss from observation and the events being studied is thus satisfied. The sampling is said to be noninformative, and these censored observations can be taken into account when estimating the hazard rates (Courgeau and Lelièvre, 1989, 1992, 2001). Selection bias can arise because only the individuals who survive and who are present when a retrospective survey is conducted can be questioned. Such bias can be measured using data from prospective surveys or population registers and varies depending on the population being studied. It should be adjusted for, whenever possible (Hoem, 1985). However, it can be assumed to be of limited importance if the event being studied does not occur in an elderly population or one subject to high levels of emigration.

This approach makes it possible to study the dependence between the processes under study and the other phenomena, which are now considered not as interfering events but as interactive events. In the next chapter we will see how this can be achieved by developing multistate event history models. When working on competing events, we can use cumulative transition intensities (sum over time of the hazard rates of experiencing the events being studied), which, contrary to survival probabilities, are suitable for valid comparison. They can be analyzed as approximately independent processes when the actual risks are dependent (Aalen, 1976).

Last, the heterogeneity of individuals must be introduced to improve our understanding of their behavior. Here we assume that behavior patterns are not innate but change over the life course in response to all that the person experiences and acquires with time. It is clear that traditional longitudinal analysis, whereby the initial population is broken down into more homogeneous subpopulations, offers no solution to this problem because it fails to allow for continuous change in this accumulated knowledge and experience. By contrast, the introduction of methods that generalize regression techniques offers a more satisfactory way of including these characteristics, which can then be considered in their time-dependent dimension.

Describing the full complexity of these behavior patterns calls for detailed observation of the different phenomena at work over the life of the individual. Traditional observation using data from civil registration, census, and survey sources is not adequate for this task; more sophisticated data-gathering tools are required to capture the multiple and interlinked chronologies of the different processes. Of these instruments, we prefer the life history survey, which assembles data on the events or *milestones* in the family, occupational, mobility histories that make up a person's life. These critical transitions enable us to identify the dynamic of the interacting processes through comparison of the various histories. Thus, a simple birth history does not constitute a life history account, because it does not include the other events that are related to it. In the same manner, a record of only the *high points* in an individual trajectory, even when it extends over several life domains, cannot form a basis for an event history analysis because it fails to observe the temporal continuity of each process.

Although realistic in scope, a data-collection operation that enables the subject to situate his or her history in the broader temporal context enhances reporting quality as regards consistency between the different calendars, by favoring a better general recall of events in different domains relative to each other (see Groupe R. A. B., 1999).

The event history techniques for handling the data raise highly complex statistical problems that have been overcome largely through advances in the theory of martingales, stochastic integration and counting processes (Anderson *et al.*, 1993). A detailed exposition of these methods is not possible here, but after outlining the basic principles and main results, we will show examples illustrating their application in different areas of demographic research.

II. STATISTICAL FORMALIZATION

An initial formalization, referred to as *nonparametric*, enables us to illustrate the interactions between phenomena. Imagine a population comprising n groups between which transitions occur over time. Let $T_{i,j}$ be the positive random variables corresponding to the times of these transitions. If we have $P_i(t)$ individuals in state i just before state t and if $M_{i,j}(t)$ transitions are observed between states i and j at time t, the instantaneous transition rate between these states can be defined in the following form (Johansen, 1983):

$$h_{i,j}(t) = \lim_{dt \to 0} \frac{P(T_{i,j} < t + dt \mid T_{i,j} \geq t)}{dt} = \frac{M_{i,j}(t)}{P_i(t)}$$

The corresponding cumulative transition intensity is then written:

$$H_{i,j}(t) = \int_0^t h_{i,j}(u)du = \sum_{\theta} \frac{M_{i,j}(\theta)}{P_i(\theta)}$$

These transition intensities (or hazards) can be estimated from observations and compared with each other.

Consider the relationship between family formation and urbanization (Courgeau 1987, 1989). The fertility of French women born between 1911 and 1935 is compared according to their residential migrations between metropolitan areas (Paris, Lyon, and Marseille) and less urbanized nonmetropolitan areas. After second births, migration to the metropolitan areas is accompanied by a sharp reduction in fertility—approximately one third—whereas out-migration from the metropolitan areas increases it to the same extent, multiplying the cumulative intensity by 1.4 or even 2. However, migration to metropolitan areas is by women whose fertility before the move was already as low as that prevailing in the urban areas, resulting in a *selection* process. Conversely, migration to the less urbanized regions attracts women whose fertility before moving was the same as that of other women in the metropolitan areas who did not move. In this instance the reduction of the migrants' fertility leads to a process of behavioral *adaptation*.

An additional finding is that the birth of successive children influences women's residential migration, which combined with the previous results produces a *reciprocal dependence*. The propensity to move to a metropolitan area is reduced after each successive birth, whereas the same births have the opposite effect on the migration risks of those leaving metropolitan areas.

Complex dependencies can be identified using this analysis. For example, one process may be observed to influence another whereas there is no effect operating in the other direction. In this case we speak of *unilateral dependence* or *local dependence*. For instance, if we observe the interactions between marriage and departure from farming of men originally employed in this sector, leaving farming is seen to produce a sharp rise in their chances of getting married, whereas marrying while still in farming does not affect their chances of changing employment. We might also observe a *total independence* between these events, in the case that neither has any influence on the other. However, this case is seldom encountered in practice, thus illustrating the need for an event history analysis.

Analysis of this type can be applied to more complex situations, involving trivariate and multivariate relationships, for example. Thus we could study the interaction between fertility and separation in different types of union (cohabitation or marriage) or the interaction between two events with multiple occurrences (fertility and residential migrations).

To develop the analysis further and examine the effect of multiple characteristics on more than one process, we have to adopt more restrictive assumptions about the behavior being studied, by use of parametric or semiparametric proportional hazard models, for example.

Consider an individual k who makes a transition from state i to state j at time t. The instantaneous hazard is expressed in the following general form, which is a generalization of the Cox model (1972):

$$h_{i,j}(t|Z_k(t)) = h^0_{i,j}(t)\exp(\beta_{i,j}Z_k(t))$$

where the first term is the instantaneous transition rate from state i to state j, at time t, under the standard conditions, $Z_k(t) = 0$. The second term of the equation incorporates the individual's characteristics, which may be time dependent, in the form of a column vector $Z_k(t)$ and a vector of $\beta_{i,j}$ parameters. When the individual characteristics are binary, this exponential expresses their proportional effect on the baseline hazard.

When $h^0_{i,j}(t)$ is considered as a function whose form is fully specified in advance, the model is described as *parametric*. Many possible functions can be used for this purpose (Pareto, Gompertz, Weibull, log-normal, etc.) depending on the phenomenon being investigated. Their parameters and the values of $\beta_{i,j}$ can be estimated by the maximum likelihood method. However, this formulation implies the selection of a function, which is not always fully adapted to the phenomenon being studied.

For this reason it is preferable to leave this baseline hazard in a non parametric form. The result is a *semi-parametric* model (Cox, 1972), whose parameters $\beta_{i,j}$ can be estimated using partial likelihood methods, which can then be used to estimate the baseline hazard rates $h^0_{i,j}(t)$.

For example, looking again at third births and the influence on them of residential migration to or from metropolitan areas, the effect of many different characteristics can be demonstrated. Thus the fact of having a farming father influences this birth: before migration the effect is to delay third births while after migration the effect disappears. Other variables are not affected by this migration. For example, the more siblings a women has, the more likely she is to have a third child herself, irrespective of whether she lives in a metropolitan area or not, or has moved or not. When all these effects have been incorporated, moving to metropolitan areas remains associated with a reduced chance of having a third child, whereas moving out of metropolitan areas continues to be associated with a raised fertility, thus confirming the persistence of what was observed when the characteristics were not introduced.

Because, as was seen in Chapter 21, we can never be sure that all the variables with an effect on the phenomena of interest have been introduced, it is important to know how this unobserved heterogeneity affects the estimated risks.

If this unobserved heterogeneity is independent of the characteristics introduced, Bretagnole and Huber-Carol (1988) have established that when a proportional hazards model is used the result is to underestimate the absolute value of the effect of these characteristics, whereas in a linear regression it has no effect on these parameters. This means that an effect that appeared significant will remain so when the initially nonobserved characteristics are introduced. Conversely, an effect which initially appeared not significant, may become significant when the previously unobserved characteristics are introduced.

When the explanatory factors, observed or not, are correlated with each other, the choice of characteristics to introduce into the model is of crucial importance. This can be illustrated by an example. When studying duration of residence at a particular dwelling, age is usually taken to be a characteristic strongly correlated with the risk of moving. But it is reasonable to think that this characteristic is itself linked to other characteristics with a more fundamental role in the migration process, such as the stage of the individual's family life at the start of residence, the tenure status of the present dwelling, and so on. Many such characteristics can be introduced, after age, using data from the *Triple Biography* survey conducted by INED (Courgeau, 1985). Not only does this greatly improve the performance of the model for explaining these durations of residence, but more important, the age effect decreases sharply and, for some of the cohorts observed, even disappears completely. Age thus appears merely a proxy variable, attributable to the fact that it is strongly correlated with the onset of critical transitions in the life course.

The proportional hazards model is merely one of several alternative models that can be used. Others include the linear and logistic models, models introducing time-varying estimated $\beta_{i,j}$ parameters and so forth (Gill, 1992). Some of these models present the advantage of dispensing with some of the earlier restrictions relating to unobserved heterogeneity.

In the case of a proportional hazards model it is important to verify with a nonparametric model whether the effect of the various risks considered separately does in fact have a multiplicative effect on the hazard rates.

III. APPLICATIONS OF THESE METHODS IN VARIOUS RESEARCH FIELDS

These techniques have been applied in many research fields, ranging from econometrics to biostatistics and epidemiology. In this section, we give an overview of their demographic applications in several demographic areas. The studies are categorized by their primary field of application, although it should be emphasized that their methodology addresses the interaction between several fields. Attention will be drawn to the originality of the results obtained with these methods.

1. Mortality

Mortality analysis was an important field of application in the early development of these methods, hence the name *survival analysis*, by which they are sometimes known. Since the original article by Cox (1972), the spread of these techniques in medical research has been reviewed (see, for example, Andersen, 1991). But their use outside of clinical trials and medical epidemiology has not developed to the same extent, particularly in demography. One reason for this concerns the type of data needed for their implementation. But we begin by looking at two examples of studies based on survey data.

Use of proportional hazards models to analyze the effects of childbearing on the occurrence of Hodgkin's disease (Kravdal and Hansen, 1996) is a perfect illustration of the most important use of this method for prognostic purposes. An earlier article by these authors, using a more conventional methodology, concluded that childbearing conferred a protective effect against Hodgkin disease. However, that protective effect is called into question in this article (Kravdal and Hansen, 1996) by the more relevant results obtained using survival models.

Guan Guo and Lawrence Grummer-Strawn (1993) analyze twin survival using data on multiple births from 26 Demographic and Health Surveys (DHS). This identifies the factors of vulnerability for twins and indicates a general increase in the effects of the variables measured for twins compared with those for other children. The model used is derived from the bivariate model and takes into account the dependence between the survival chances of the two categories of children. This is a major advantage of these models.

In contrast to traditional demographic mortality analysis, the two previous studies do not use civil registration data. That source supplies too few individual characteristics for it to be suitable for in-depth event history analysis. A linkage is sometimes possible with census data, which provide more detailed information. Thus several authors (Sahli *et al.*, 1995; Wunsch *et al.*, 1996) have studied mortality in relation to characteristics from three successive censuses linked with Norway's register of deaths. These analyses take the form of a series of logistic regressions on time-ordered sequences, without introducing duration. This procedure, while entirely legitimate and effective, differs from the event history family of models by not including duration, which is a fundamental dimension of these models. However, it will be seen shortly that duration can be introduced for the analysis of interval-censored life history information of this kind.

2. Fertility

Fertility analysis has produced numerous surveys suitable for event history analysis. In a review article, Adrian Raftery *et al.* (1996) give a critical appraisal of the methods used to analyze the World Fertility Survey (WFS) data. With the exception of descriptive analyses, using conventional tabular presentations of fertility rates and linear regression analyses in which the explanatory variable is the number of live births, the authors indicate that although these data give detailed, reproductive histories combined with calendars of contraceptive use they have seldom been subject to event history analysis. After outlining the practical difficulties involved in conducting such an analysis and the innovative nature of the indicators obtained, they present an analytical strategy for event history modelling of WFS data, with a view to increase the use of this strategy. Using the number of live births as a regression variable encounters a certain number of problems, which can only be resolved within the much more appropriate framework of event history analysis. The first problem concerns incomplete information due to censoring at the time of the survey, something that cannot be allowed for in a conventional regression analysis. The second problem relates to the impossibility of distinguishing cohort and period effects. Finally, if the influence of explanatory variables varies with parity, these effects are *diluted* in a conventional regression because individual variables cannot be modified according to exposure duration (i.e., cannot be time-dependent). The authors then apply the proposed procedure to WFS survey data for Iran, achieving results that could not be obtained with the other procedures.

In an article on extramarital fertility, Eva Lelièvre (1993) reviews the methodologic problems encountered in such a study and how they can be overcome by means of event history analysis. Particular attention is given to the question of the validity of analyses where the results are conditioned by knowledge of the future (i.e., by the outcome of the process being studied). In the case of extramarital fertility, the processing of the period of cohabitation is determinant. If, for the evaluation of fertility differentials, unions are distinguished by whether they lead to a marriage or not, the rates are in effect conditioned relative to the future. These difficulties are avoided with an event history analysis that allows for past events and introduces union transformations at the time they occur.

3. Union Formation and Dissolution

Many analysts have come to favor event history analysis for the study of divorce, because the introduction of time-dependent variables is necessary to grasp the nature of the mechanisms at work. This also makes it easier to distinguish between period, age, and cohort effects.

Michael Bracher *et al*. (1993) have carried out a systematic examination of hypotheses advanced to explain the risk of marriage dissolution in Australia, using a Cox model. In constructing their model, they consider successively the introduction of historic time (characterized by period and cohort aspects) and biographic time (age at marriage, age and duration of marital union), the relevant characteristics. It is clear that only event history methods (and the data for running them) are suitable for conducting such an analysis, which is applied to empirical data.

Working on the plentiful and very precise data from the Swedish Population Register, Hoem (1997) has called into question the effects of educational attainment on divorce and has shown (as is only possible with event history models) that these effects are time-variant. These findings run counter to the generally accepted hypothesis (in the English-speaking and Scandinavian countries) of a decrease in the risk of divorce as educational attainment increases. This hypothesis was already at odds with the opposite or nonexistent effect of educational attainment observed in other countries, and thus appears very questionable.

In his study on the place of children in the history of couples, Laurent Toulemon (1994), responding to an earlier critical comment by Adrian Raftery *et al.* (1996), proposes simple indicators, drawn from event history analysis, to present the variations in the risk of marriage breakdown by number of children and age of last child. This study concludes that increased risks of marri-age breakdown are partly independent of fertility behavior.

4. Internal Migration

The French *Echantillon Démographique Permanent* (Demographic Longitudinal Survey), which links civil registration and census data since 1968, makes possible a new approach to interval-censored event histories. In this data file, the residence of individuals is recorded only at the time of censuses and family events. The dates observed are therefore those before and after the migration event to be studied.

Two simplifying assumptions have to be made to use these interval-censored data. The first is that only one of the migrations being studied can occur between two observation times. The second is that the observation scheme is non informative, i.e. that observation of individual trajectories occurs independently of their progression, at random points in time. The first of these conditions determines the choice of observation, using both census and civil registration data, to obtain the optimum estimation. For the second condition, it can be seen that while it is fully satisfied by census observation, civil registration observation can be a source of bias, since there is no reason to expect migrations to be independent of family events.

The test conducted with data from the Triple Biography survey, which were artificially censored for this purpose (Courgeau and Najim, 1995, 1996), established that bias was not large when working on changes of residence. The estimates are similar with census or civil registration data, or with both together, although when civil registration data alone are used significant overestimation of the hazards occurs related to the interactions between migration and fertility. On the other hand, when working on changes of region, serious errors arise due to return migration. A semiparametric analysis of residential migration after marriage has been conducted using the EDP data (Courgeau *et al*., 1998).

5. Occupational Histories

In the field of the study of individuals' occupational histories, Daniel Courgeau (1993) has shown how event history analysis can answer a perennial research question. The problem is that of how do we *allow for the interactions between occupational and migratory events, by introducing a continuous time during which these events occur?* Whereas traditional methods cannot deal with this problem, the general framework of event history analysis is effective for identifying the nature of the interactions in question and the heterogeneity of the population under consideration.

This general framework is also effective for handling far more complex problems, such as the interaction between the professional and migration behavior of the members of a couple (Courgeau and Meron, 1995). These interactions can be modelled, taking into account the effect of the couples' characteristics, which can be time-dependent, to obtain information about the close links between its residential mobility, occupation, and family life. In this way we are able to relate the three main role domains in the life of a couple.

This brief review shows the scope of application of event history analysis and indicated its value as an explanatory tool in demography.

IV. FUTURE DEVELOPMENTS

The event history approach is now widely accepted in the social sciences, including in demography (Keilman, 1993). We must nonetheless resist the temptation to explain a behavior by reference only to individual-level characteristics. The danger then is of committing the atomistic fallacy, because the analysis ignores the broader context in which human behavior occurs. Individuals actually operate in multiple domains (family and occupational, for example) and are linked to a geographic context (resident of a neighborhood, originating from a region, etc.), all of which influence their behavior.

On the other hand, when working at a given level of aggregation, only the conventional demographic rates can be related to the aggregate-level characteristics of the regions being considered, such as the regions' emigration rates to their percentages of farmers, labor force participation rates, and so forth. The danger here is of committing what is known as the *ecological fallacy*, which occurs when inferences about individual-level behavior are drawn from aggregate-level results.

Thus, we must find a way of integrating these different levels of aggregation in the same event history analysis. An initial approach is to examine the effect of both individual and aggregate characteristics on individual behavior, with the possibility of introducing a specific effect for each region at any given level of aggregation. This leads to contextual and multilevel event history models (see Chapter 24).

A second approach involves the analysis of more complex groups, when it is no longer possible to consider the group as forming a single entity but as composed of interacting individuals. This means that instead of studying the interactions between the processes that affect an individual, we study the interactions between individual processes affecting members of a group (Lelièvre *et al*., 1997, 1998). This shift from individuals to their entourage (or *contact circle*), whether for the purpose of data gathering or analysis, implies rethinking the units used for longitudinal observation.

Whereas an individual can be tracked over his or her entire lifetime, the practical impossibility of defining a household in longitudinal terms (Keilman and Keyfitz, 1988) has led to development of the concept of entourage or *contact circle* (Lelièvre and Bonvalet, 1994). It is still the individual who is tracked, but linked with all the members of the various households to which he or she has ever belonged, plus other close but not necessarily co-resident family members.

One approach involves an event history analysis that treats such groups as a composite individual who can be described by both collective and individual characteristics. But the events studied will be mainly collective in nature (change of membership, place of residence, etc.), whereas individual characteristics and the interactions between members are considered only implicitly. In these conditions, the techniques of event history analysis of the group considered as an individual can indeed be applied, on condition the definition of the group and the events it experiences are carefully specified.

Modelling on these lines is not perfect, however, because it overlooks the fact that the process being studied is multivariate, with each of the group trajectories resulting from the interacting trajectories of its members. Since the life histories of the individual members of the group will no longer be independent, a specific modelling strategy has to be developed. This problem has been explored from a statistical angle and these models are currently being adapted for application to demographic processes (Lelièvre *et al*., 1997, 1998).

From the standpoint of data collection strategies, all the events of union and reproductive histories are already included in an individual life history survey, and if the varying position of the individual within his or her contact circle is defined using these dimensions, no additional information is required. If data collection also includes the residential histories of children/parents (direct descent) when the respondent does not live with them, information on the role of kinship in individual decisions becomes available. By working on factual data, such as the location or occupation of a certain number of close friends and relatives, the existence of an influence can be established when there is residential or social proximity. Last, the structure of the individual's network, that is, the nature of his or her contact circle will be known for the entire life course of the individual. A survey along those lines, the *Biographies et entourage* (Event Histories and Contact Circle) survey, was conducted by the Institut national d'études démographiques at the end of 2000. Analysis of the results from this survey will represent a major advance for event history analysis.

References

Aalen Odd O., 1976. Nonparametric inference in connection with multiple decrement models, *Scandinavian Journal of Statistics*, no. 3, p. 15–27.

Andersen Per Kragh, Borgan Ørnulf, Gill Richard, Keiding, 1993. *Statistical Models Based on Counting Processes*. New York, Springer Verlag, 768 p.

Bretagnole Jean and Huber-Carol Catherine, 1988. Effects of omitting covariates in Cox's model for survival data, *Scandinavian Journal of Statistics*, n° 15, p. 125–138.

Courgeau Daniel and Lelievre Éva, 1986.—Nuptialité et Agriculture, *Population*, vol. 41, n° 2, p. 303–326.

Courgeau Daniel and Lelievre Éva, 1989.—*Analyse démographique des biographies*, Paris, coll. de l'INED, PUF, 268 p. (Also in English: *Event history analysis in demography*, Oxford, Clarendon Press, 1992, 226 p., and in Spanish: *Análisis demográfico de las biografías*, México, El Colegio deMéxico, 2001, 306 p).

Courgeau Daniel and Lelievre Éva, 1996. Changement de paradigme en démographie, *Population*, n° 3, p. 645–654. (Also in English 1997. Changing paradigm in demography, *Population: English Selection*, vol 9, p. 1–10).

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life-cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel, 1987. Constitution de la famille et urbanisation, *Population*, vol. 42(1), p. 57–82. (Also in English 1989. Family formation and urbanization, *Population: English Selection*, vol. 1, p. 123–146).

Cox David R., 1972. Regression models and life tables (with discussion), *Journal of Royal Statistical Society*, vol. B34, p. 187–220.

Gill Richard, 1992. Multistate life-tables and regression models, *Mathematical Population Studies*, no. 3, p. 259–276.

Groupe de réflexion sur l'approche biographique, 1999. *Biographies d'enquêtes. Bilan de 14 collectes biographiques*, Paris, INED/PUF, 340 p. (Coll. Méthodes et savoir, no. 3.)

Henry Louis, 1959. D'un problème fondamental de l'analyse démographique. *Population*, vol. 14(1), p. 9–32.

Hoem Jan, 1985. Weighting, misclassification and other issues in the analysis of survey samples of life histories, *in:* James Heckman and Burton Singer (eds.), *Longitudinal analysis of labour market data*, p. 249–293. Cambridge, Cambridge University Press, 425 p.

Kalbfleisch John D., Prentice Ross L., 1980. *The statistical analysis of failure time data*, New York, John Wiley & Sons, 322 p.

Keilman Nico and Keyfitz Nathan, 1988. Reccurent issues in dynamic household modelling, *in*: Nico Keilman, Anton Kuijsten and Ad Vossen (eds.), *Modelling household formation and dissolution*, p. 254–285. Oxford, Clarendon Press, 298 p.

Lelièvre Èva, Bonvalet Catherine and Bry Xavier, 1997. Analyse biographique des groupes, les avancées d'une recherche en cours, *in:* Daniel Courgeau (ed.), Nouvelles approches méthodologiques en démographie, *Population*, vol. 52(4), p. 803–830. (Also in English, 1998: Event history analysis of groups: the findings of an on-going research project, *Population: English Selection*, vol 10(1), p. 11–37).

Lelièvre Èva, Bonvalet Catherine, 1994. Construire les principes d'une analyse biographique du groupe familial, in *Ménages, familles, parentèles et solidarités dans les populations méditerranéennes*, Actes du séminaire international de l'AIDELF (1996), p. 103–116.

Lelièvre Èva, Bringe Arnaud, 1999. *Manuel pratique pour l'analyse statistique des biographies/Practical guide to event history analysis using SAS, TDA, STATA*, Paris, INED/PUF, 378 p.

Mortality

Andersen Per Kragh, 1991. Survival analysis 1982–1991: the second decade of the proportional hazards regression model, *Statistics in Medicine*, vol. 10, p. 1931–1941.

Guo Guan and Grummer-Strawn Lawrence, 1993. Child mortality among twins in less developed countries, *Population Studies*, vol. 47, p. 495–510.

Kravdal Øystein and Hansen Svein, 1996. The importance of childbearing for Hodgkin's disease: new evidence from incidence and mortality models, *International Journal of Epidemiology*, vol. 25, p. 737–743.

Lopez Alan, Caselli Graziella and Valkonen Tapani (eds.), 1995. *Adult mortality in developed countries*. Oxford, Clarendon Press, 361 p.

Salhi Mohamed, Caselli Graziella, Duchêne Josianne, Egidi Viviana, Santini Antonio, Thiltgès Évelyne and Wunsch Guillaume, 1995. Assessing mortality differentials using life history: a method and applications, *in:* Alan Lopez, Graziella Caselli and Tapani Valkonen (eds.), *Adult mortality in developed countries: from description to explanation, p. 57–79*. Oxford, Clarendon Press, 361 p. (Coll. International Studies in Demography).

Wunsch Guillaume, Duchêne Josianne, Thiltges Évelyne and Salhi Mohamed, 1996. Socio-economic differences in mortality. A life course approach, *European Journal of Population*, vol. 12(2), p. 167–185.

Union Formation and Dissolution

Bracher Michael, Santow Gigi, Morgan Philip S. and Trussell James, 1993. Marriage dissolution in Australia: Models and Explanations, *Population Studies*, vol. 47(3), p. 403–425.

Hoem Jan, 1997. Educational gradients in divorce risks in Sweden in recent decades, *Population Studies*, vol. 51, p. 19–27.

Leridon Henri, 1990. Cohabitation, marriage, separation: an analysis of life histories in French cohorts from 1968 to 1985, *Population Studies*, vol. 44, p. 127–144.

Toulemon Laurent, 1994.—La place des enfants dans l'histoire des couples, *Population*, vol. 6, p. 1321–1346.

Fertility

Lelièvre Eva, 1993. Extra-marital births occurring in cohabiting unions, *in:* Maire Nì Bhrolcháin (ed.), *New perspectives on fertility in Britain*, p. 109–120. Londres, OPCS, 169 p. (Coll. Studies on Medical and Population subjects no. 55).

Kravdal Ø and Hansen S. (1993). Hodgkin's disease: The protective effect of childbearing. *International Journal of Cancer*, 55: 56–61.

Raftery Adrian, Lewis Steven, Aghaganian Akbar and Kahn Michael, 1996. Event history modeling of World Fertility Survey Data, *Math. Population Studies*, 2, p. 129–154.

Internal Migration

Courgeau Daniel, Lelievre Eva, Wolber Odile, 1998.—Reconstruire des trajectoires de mobilité résidentielle. Éléments d'une analyse biographique de l'EDP, *Économie et Statistique*, no. 316–317, p. 163–176.

Courgeau Daniel, Najim Jamal, 1995. Analyse des biographies fragmentaires, vol. 50(1), p. 149–168. (Also in English 1996. Interval-censored event history analysis, *Population: An English Selection*, vol. 8, p. 191–208).

Work Careers

Courgeau Daniel and Meron Monique, 1995. Mobilité résidentielle, activité et vie familiale des couples, *Economie et Statistique*, no. 290, p. 17–31.

Courgeau Daniel, 1993. Nouvelle approche statistique des liens entre mobilité du travail et mobilité géographique, *Revue Economique*, vol. 44(4), p. 791–807.

C H A P T E R

24

Individuals and Context in the Multilevel Approach to Behavioral Analysis

DANIEL COURGEAU

Institut national d'études démographiques (INED), Paris, France

Analysis of multistate data aggregated by socio-occupational group, geographic area, administrative unit, and so on, is based on the assumption that the behavior of individuals is homogeneous in each of these subpopulations. If these conditions hold, it is sufficient simply to use the conventional demographic measures and rates and identify the relations that exist between them. For example, a study might relate the emigration rates for the various regions of a country to the unemployment rates, percentages of farmers, and so forth observed in each region. But we must be careful not to infer that these relationships between aggregate-level values are also valid for the corresponding individual-level characteristics. That would be making the error of reasoning known as the *ecological fallacy*—a positive association between a region's emigration rate and the proportion of farmers in its population in fact tells us nothing about the migration probability of farmers, for whom the correlation may be negative. Aggregate-level models can also be constructed in which fertility, mortality and interregional migrations are introduced simultaneously (see Chapter 20 on *r*-level population models), under the additional assumption of independence between the phenomena being studied.

The hypothesis of homogeneity of behavior within each group becomes unnecessary if we work on individual-level data, because an individual's behavior can then be related to his or her personal characteristics (see the previous chapter on event history analysis, and Chapter 131 on causal models in the final volume of this treatise). In this case, the fact that an individual has moved out of a particular region will be related to factors such as being unemployed, being a farmer, and so on. However, the danger here is committing the *atomistic fallacy*, because the analysis ignores the context in which human behavior occurs. In the real world, this context influences individual behavior and it is a mistake to consider individuals in isolation from the constraints imposed by the groups and settings in which they belong.

What is needed, therefore, is an analysis that works at different levels of aggregation simultaneously but whose objective remains that of explaining individual behavior. The pitfall of the ecologic fallacy is thus avoided, because the aggregated characteristic is used to measure a construction that is different from its equivalent at the individual level. It is introduced not as a proxy variable but as a characteristic of the subpopulation that influences the behavior of an individual member of that group. At the same time, the atomistic fallacy ceases to be a problem once the analysis accounts properly for the context in which the individual lives.

I. INDIVIDUAL AND AGGREGATE MEASURES: CONTEXTUAL ANALYSIS

The behavior to be analyzed here is still individual but the explanatory variables can be both individual and aggregate. Let us begin by looking at the different types of aggregation that can be employed.

1. Types of Measures

For any aggregation level, we can simply add together the individual characteristics and estimate the percentages and averages. Examples are the percentage of farmers, the average starting salary in an occupation, and so on. More complex analytical procedures can also be applied. Thus, as well as average income, the product of this and the binary variable for whether the individual is a farmer could be introduced, which, as will be seen shortly, gives a better characterization of the interaction between individual- and aggregate-level characteristics, or the correlation between this income and household size.

Other characteristics for a given level of aggregation are more general by nature and do not apply at the individual level. Examples are the number of hospital beds in a particular region, or the number of classes in a school. But although these do not correspond to any individual characteristic, they can nonetheless be aggregated at other levels. Thus the number of hospital beds in a region is the sum of the number of beds in each department of the region.

Another class of characteristic is defined for a single level of aggregation and cannot be aggregated at higher levels. The political orientation of a commune, as defined by the party affiliation of its mayor, for example, cannot be aggregated with those of neighboring communes, which may cover a broad spectrum. This characteristic therefore cannot exist at the individual level, because the individual may have voted for another candidate, nor does it exist at the level of the department, because the different communes of the department are not usually all of the same political orientation. Nonetheless, it may influence the behavior of the individuals who live in the commune.

2. A Regression Model

Having specified the characteristics to be evaluated, we will now consider the analytic models that can be used to measure their impact. We begin with an extension of conventional estimation procedures, such as multiple linear regression or logistic regression. This extension is usually termed *contextual analysis*.

This model explains individual behavior by reference to both individual and aggregate characteristics. Let us first of all see how it is formulated in the case of a regression that introduces variables considered at different levels. The example used here is taken from the 1992 Demographic and Health Survey (DHS) of Morocco (Enquête Nationale sur la Population et la Santé, or ENPS-II), for the analysis of fertility determinants in a rural environment (Schoumaker and Tabutin, 1999). Individual fertility is measured by the DRAT (duration ratio), obtained by dividing the number of children each married woman has had by the theoretical number of children she would have had (adjusted for her age and length of union) in a regime of natural fertility. These data are organized in the 72 sampling clusters, which are treated as separate groups.

We write as $y_{i,j}$ the DRAT at the time of the survey of a woman i living in group j. This fertility will first be related to the fact of her husband being a farmer or not ($x_{i,j}$):

$$y_{i,j} = a_0 + a_1 x_{i,j} + e_{i,j} \qquad \text{(Eq. 1)}$$

where $e_{i,j}$ is an individual random term.

The estimated parameter values are reported in the first column of Table 24–1, with their standard error in brackets. According to this conventional regression model, women married to a farmer have a higher fertility than the others. But these parameters may differ depending on the groups and we can estimate them separately for each group j, with the following regression models:

$$y_{i,j} = a_{0j} + a_{1j} x_{i,j} + e_{i,j} \qquad \text{(Eq. 2)}$$

where $e_{i,j}$ is a group random term.

Figure 24–1A displays the results of this estimation. The great diversity in the parameters is clearly apparent, but no conclusion is possible, because most of the parameters a_{1j} are not significantly different from zero. The numbers observed in each group are too small on which to base a conclusion: The differences between the slopes may be the result of purely random error. The solution is to work on the entire population and introduce new characteristics: the percentage of women married to a farmer in cluster $j(x_{.,j})$ and its interaction with the fact of the husband being a farmer, thereby introducing an interaction between an individual and aggregate characteristic. The new model is written

$$y_{i,j} = a_0 + a_1 x_{i,j} + a_2 x_{.,j} + a_3 x_{i,j} x_{.,j} + \varepsilon_{i,j} \qquad \text{(Eq. 3)}$$

where $\varepsilon_{i,j}$ is an error term representing the effects of all the implicit or unobserved variables. This model can of course readily be extended to include additional explanatory variables that are considered to act simultaneously on the female fertility index. Parameter estimation is by the least squares method, with the usual assumptions made for the residuals (i.e., a normal distribution of expectation zero and a constant variance σ^2 for all values of the explanatory variables). Let us

TABLE 24–1 Comparison of results from contextual and multilevel models (rural Morocco, 1992)

Characteristics	Contextual models Estimation (standard error)		Multilevel models Estimation (standard error)	
Fixed				
Constant	0.786 (0.009)	0.749 (0.016)	0.800 (0.014)	0.751 (0.027)
Husband farmer	0.060 (0.013)	0.047 (0.033)	0.040 (0.014)	0.067 (0.036)
% husbands farmer		0.112 (0.040)		0.122 (0.059)
Interaction husband farmer * % husbands farmer		–0.028 (0.061)		–0.072 (0.067)
Random at group level				
σ^2_{u0} (Constant)			0.0084 (0.0024)	0.0076 (0.0024)
σ_{u01} (Covariance)			–0.0017 (0.0020)	–0.0014 (0.0019)
σ^2_{u1} (Husband farmer)			0.0003 (0.0024)	0.0001 (0.0023)
Random at individual level				
σ^2_{e0}	0.1090 (0.0031)	0.1085 (0.0031)	0.1022 (0.0030)	0.1022 (0.0030)
–2 (log-likelihood)	1,530.20	1,519.04	1,455.49	1,451.23

Sources: ENPS-II survey, Morocco, 1992; Schoumaker and Tabutin, 1999.

examine the interpretation of the parameter estimates in more detail.

Calculating the mean of all the observations of group j, gives the following relation, which is valid for each group considered separately:

$$y_{.,j} = a_0 + (a_1 + a_2)x_{.,j} + a_3x^2_{.,j} + \varepsilon_j \qquad \text{(Eq. 4)}$$

where ε_j is the mean error term for each group. We then see that the estimations of the mean values are not distributed randomly but lie along a parabola or a line, if a_3 is zero, when they are plotted as a function of $x_{.,j}$. Again for group j, we can now observe the position of the individuals in this group relative to the mean values that have just been calculated. By subtracting relations (3) and (4), we obtain the following expression:

$$y_{i,j} - y_{.,j} = (a_1 + a_3x_{.,j})(x_{i,j} - x_{.,j}) + \varepsilon_{i,j} - \varepsilon_j \qquad \text{(Eq. 5)}$$

which shows that in each group the estimations lie broadly on lines of slope $a_1 + a_3x_{.,j}$. We can easily verify that these lines pass through a fixed point if a_3 is not zero or that they are parallel to each other in the opposite case.

For the case examined here, the various parameter estimations are given in the second column of Table 24–1. We see first of all that the lines corresponding to each group are parallel to each other because parameter a_3 is not significantly different from zero: There is no interaction effect between individual fertility and that of the group. On the other hand, fertility is no longer identical regardless of group but is strongly differentiated: As the percentage of farmers among men in the group rises, so fertility increases. However, when this increase is allowed for, the relationship between fertility and the fact of the husband being a farmer is reduced by a quarter (coefficient of 0.047 instead of 0.060). Figure 24–1B displays these lines, which can be compared with the regressions calculated on each group separately (Figure 24–1A). From this we can see that while contextual analysis does take into account the different levels at which the groups are situated, it is less effective at explaining the variation in the slopes corresponding to the various groups.

As it stands, Equation 3 introduces only a single error term at the individual level, which means that the results for individuals in a group are analyzed as if they are independent. In addition, the effects of the aggregate characteristics have a predetermined form—quadratic or linear for the mean, and convergent or parallel lines for the group effect. To dispense with these restrictions, the next step is thus to introduce the random variables specific to each level of aggregation.

II. INTRODUCING GROUP EFFECTS: MULTILEVEL MODELS

It is reasonable to think that the result for an individual in a group may depend on the results obtained by the other individuals in the same group. Because they overlook this intragroup dependence, contextual models produce biased variance estimates for the contextual and individual effects, resulting in confidence intervals that are too narrow. This problem of intragroup dependence can be handled by introducing random effects into the previous contextual models. Let us examine this for the regression discussed in Equation 2.

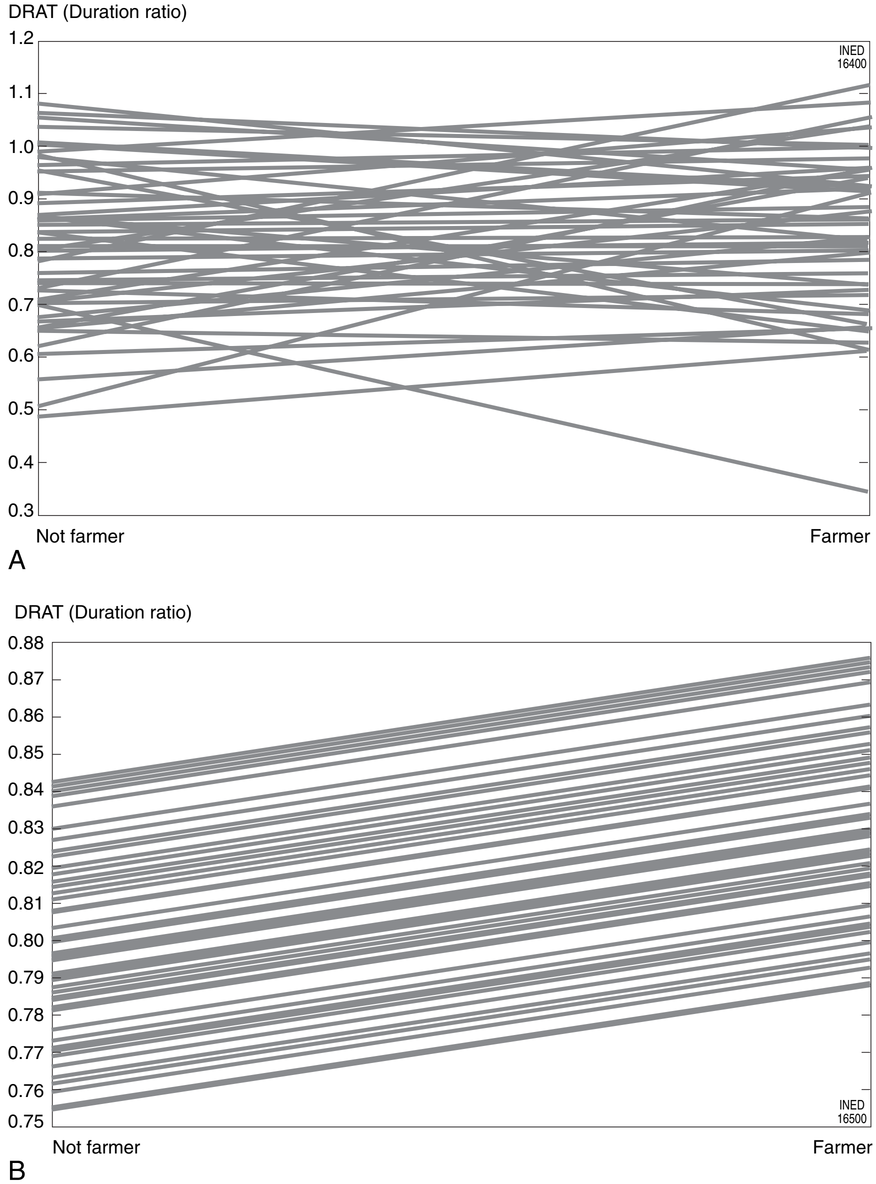


FIGURE 24–1 Fertility estimations by whether the husband is a farmer. **A:** Separate regression on each context. **B:** Contextual regression.

1. Reworking the Previous Regression Model

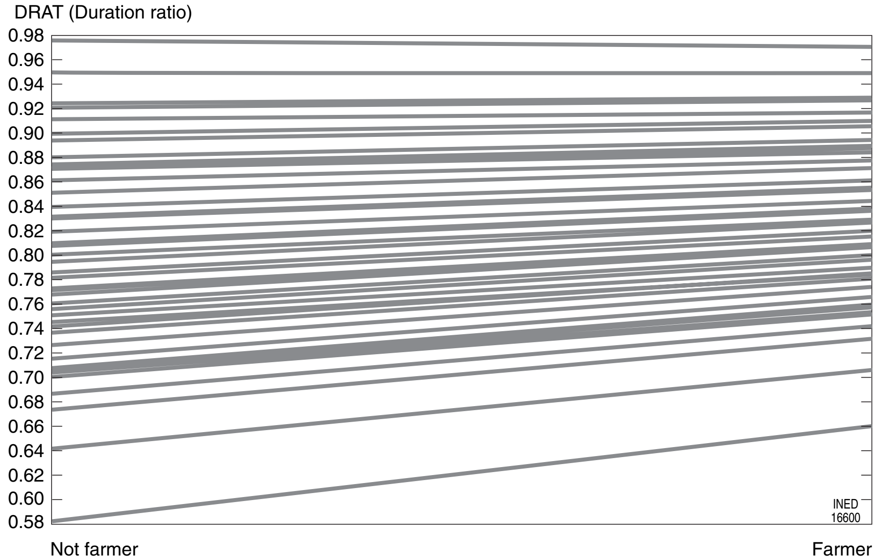
Because of the small number of individuals observed in each group, the regression parameters estimated earlier will be treated as a random sample from a larger set of parameters. This results in the parameters of Equation 2, a_{0j} and a_{1j}, becoming random variables at the level of the groups, for which we want to estimate the mean a_0 and a_1 and the variation around this mean. We can then write:

$$a_{0j} = a_0 + u_{0j},\ a_{1j} = a_1 + u_{1j} \qquad \text{(Eq. 6)}$$

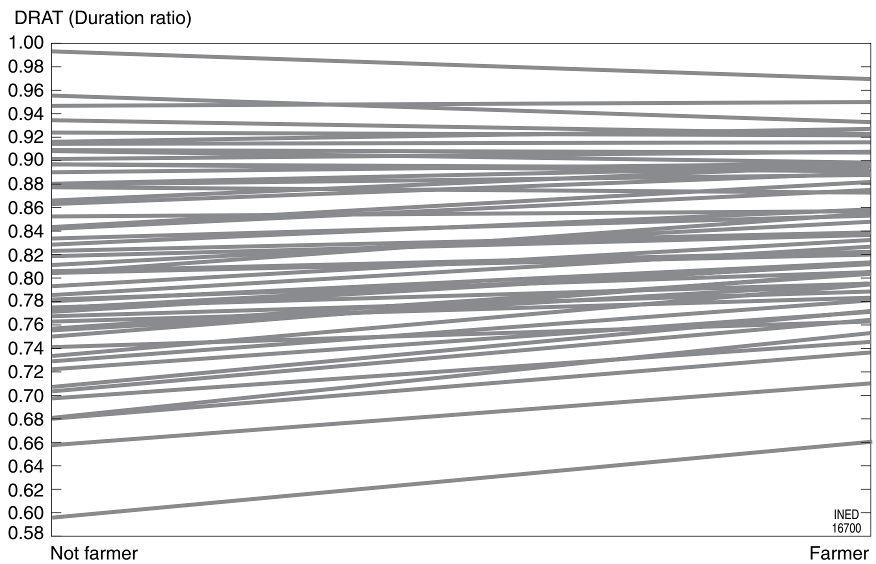
where u_{0j}, u_{1j} and $e_{i,j}$ are random variables of expectation zero and with the following variances and covariances:

$$\begin{gathered}\mathrm{var}(u_{0j}) = \sigma^2_{u0},\ \mathrm{var}(u_{1j}) = \sigma^2_{u1},\\ \mathrm{cov}(u_{0j},u_{1j}) = \sigma_{u01}\end{gathered} \qquad \text{(Eq. 7)}$$

these three terms corresponding to the level of the groups, and $e_{i,j}$ (Equation 1) to the individual level. This model is fully *multilevel*. The parameters of such a model can be estimated using generalizations of the least squares method (Goldstein, 1995, Courgeau, 2004).



C



D

FIGURE 1 (*Continued*) **C:** Multilevel model. **D:** Multilevel model with all variables. Source: Schoumaker Bruno and Tabutin Dominique, 1999. Analyse multi-niveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, *in:* UEPA-NSU (ed.), *La population africaine au 21e siècle*, vol. 1, p. 299–332. Dakar, UEPA, 630 p. (Proceedings of the Third African Population Conference, Durban).

Applying this model to the previous data gives the results set out in the third column of Table 24–1. The multilevel model, which introduces only the constant and the fact of the husband being a farmer, gives an effect for the latter characteristic very similar to that obtained with the contextual model that introduces all the other characteristics (0.040 compared with 0.047). This is because the introduction of the constant term in the random variables at the level of the groups confirms the effect of these groups on fertility, without the additional terms having to be introduced. Figure 24–1C shows the fertility predicted by this model, depending on whether or not the husband is a farmer. Although the results for the random variable corresponding to having a farmer husband are not significant, the strong negative correlation, close to –1, between the constant and the fact of having a farmer husband, is now reflected in the *fanning-in* configuration of the lines corresponding to each group—The higher the fertility of the nonfarmers, the greater the reduction in the slope of these lines. This result is not obtained with the contextual model, due to its overrestrictive specification—it is closer this time to the regression lines of Figure 24–1A. Also observed, as expected, is an increase in the standard errors of the fixed parameters, compared with the estimations from Equation 1.

The last column of Table 24–1 introduces the percentage of men who are farmers and the product from multiplying this with the fact that the individual is himself a farmer, again in a multilevel model. The variances of all the random terms are seen to decrease. This confirms that these aggregate characteristics account for part of the random terms at the group level as well as reducing the effect of the farmer husband. However, a significant effect is still associated with the random term corresponding to the constant. Figure 24–1D displays the results predicted by this model, which are very similar to what was obtained in Figure 1A, although some significant parameters from the multilevel model used are now also given.

We have discussed these results at length to demonstrate the utility of multilevel regression. This analysis must be continued with other characteristics, such as educational attainment, standard of living, and so forth of the wife or of her household to see whether this random effect persists and to show the effect of these variables on the fertility level being studied.

2. Analysis for Binary or Polytomous Data

Now let us look at how to construct, for example, a *multilevel logistic model*. For this purpose we use a practical example, drawn this time from the Norwegian population register.[1] We examine interregional migrations over a 2-year period (1980–1981) by women born in 1958 and resident in Norway in 1991.

Let us assume that we are working with the probability that a characteristic to be estimated, $y_{i,j}$, in this case, the fact of having migrated, is 1. Individual i is present in region j before migrating. We want to examine the relationship between this probability and an explanatory variable, $x_{i,j}$, which is assumed here to be binary. This probability, conditioned by the fact of the individual having the characteristic $x_{i,j}$, is written as follows:

$$P(y_{i,j} = 1|x_{i,j}) = p_{i,j} = [1 + \exp(-[a_0 + u_{0j} + (a_1 + u_{1j})x_{i,j}])]^{-1} \qquad \text{(Eq. 8)}$$

It follows that the answers $y_{i,j}$, are distributed according to a binomial distribution:

$$y_{i,j} \sim B(p_{i,j},1) \qquad \text{(Eq. 9)}$$

[1] We would like to thank the Norwegian statistical services for allowing us to use the data files produced from the Norwegian population registers and censuses.

In this case we have the following conditional variance:

$$\mathrm{var}(y_{i,j}|p_{i,j}) = p_{i,j}(1 - p_{i,j}) \qquad \text{(Eq. 10)}$$

The model then becomes a nonlinear model:

$$y_{i,j} = p_{i,j} + e_{i,j}z_{i,j}, \quad \text{where} \quad z_{i,j} = \sqrt{p_{i,j}(1-p_{i,j})}, \quad \text{and} \quad \sigma_e^2 = 1 \qquad \text{(Eq. 11)}$$

The individual level variance in this case is equal to unity, and we will work principally on the regional level variances and covariances.

In Table 24–2 we have estimated the parameters of the contextual logit and multilevel models, for studying the migration probabilities of the 19 Norwegian regions according to whether the women have at least one child before this migration.

The first contextual model yields a lower migration probability for women with at least one child (first column), close to that obtained for the fixed part of the multilevel model (column 3), although with a smaller variance for the first model. The random part of the multilevel model shows a variation by region and by whether or not the woman has at least one child. We show the joint influence of the fixed and random parameters at the regional level by calculating the logit functions for the probabilities of emigrating from region j, according to whether the woman is childless or not. The logit function of women without children, Π_{0j}, is simply the sum $a_0 + u_{0j}$; its between-region variance is equal to σ^2_{u0}. The logit function of women with children, Π_{1j}, is given by the sum $a_0 + a_1 + u_{0j} + u_{1j}$; so its between-region variance is equal to $\sigma^2_{u0} + 2\sigma_{u01} + \sigma^2_{u1}$.

The data in Table 24–2 can be used to calculate the between-region variance of women with at least one child (0.205), which is four times higher than that of women with no children (0.051). Figure 24–2 displays the values of Π_{0j} and Π_{1j}, corresponding to women with and without children, linked by a line for each region. These lines are characterized by a fan-shaped form, which is explained by the differences between the variances and by the positive correlation between the random variables, corresponding to women with and without children, close to unity (0.93).

When the aggregate characteristics are introduced, these are highly significant in the contextual model (column 2). Figure 24–3 presents the logits of the probability of migrating according to the regional percentage of women with children. This shows that the migration probabilities depend both on whether women have children and on the proportion of women with children living in the different regions. The higher this proportion, the more these probabilities are reduced by the fact of having children. A region's high

TABLE 24–2 Migrations in Norway, whether or not women have children

Characteristics	Contextual models Estimation (standard error)		Multilevel models Estimation (standard error)	
Fixed				
Constant	–1.464 (0.018)	–0.962 (0.062)	–1.503 (0.056)	–1.235 (0.027)
With children	–0.973 (0.048)	–0.125 (0.185)	–0.992 (0.080)	–0.222 (0.349)
Proportion with children		–0.020 (0.002)		–0.009 (0.009)
With children *proportion with children		–0.029 (0.002)		–0.027 (0.013)
Random at regional level				
σ^2_{u0} (constant)			0.051 (0.029)	0.049 (0.022)
σ_{u01} (covariance)			0.042 (0.027)	0.032 (0.015)
σ^2_{u1} (with children)			0.070 (0.068)	0.038 (0.029)
–2 (log-likelihood)	22,934	22,808	19,237	19,237

Source: Norwegian population register, Central Bureau of Statistics, Oslo.

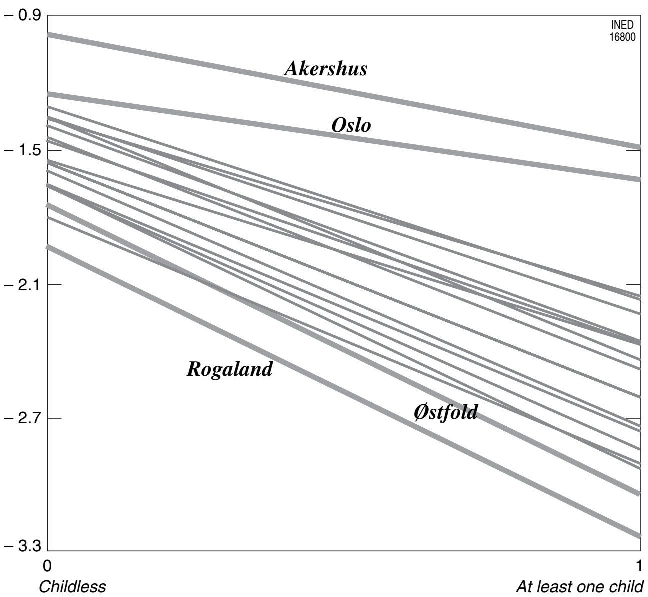


FIGURE 24–2 Logit model on Norwegian migrations: logit of the probability of migrating by number of children.

aggregate fertility is associated with a greater capacity to retain women with children.

These conclusions are completed by the multilevel model. Only the interaction effect between the individual and aggregate characteristic is significant on the fixed parameters, indicating as before that the higher the proportion of women with children in a region, the less likely these women are to migrate. At the same time, a reduction in variance is observed at the regional level for women with children (0.151 compared with 0.205), which confirms the effect of the aggregate variable. However, the regional effect for the constant is still significant and at the same level (0.05): this was not properly accounted for in the contextual model. Figure 24–3, which also displays the logits of

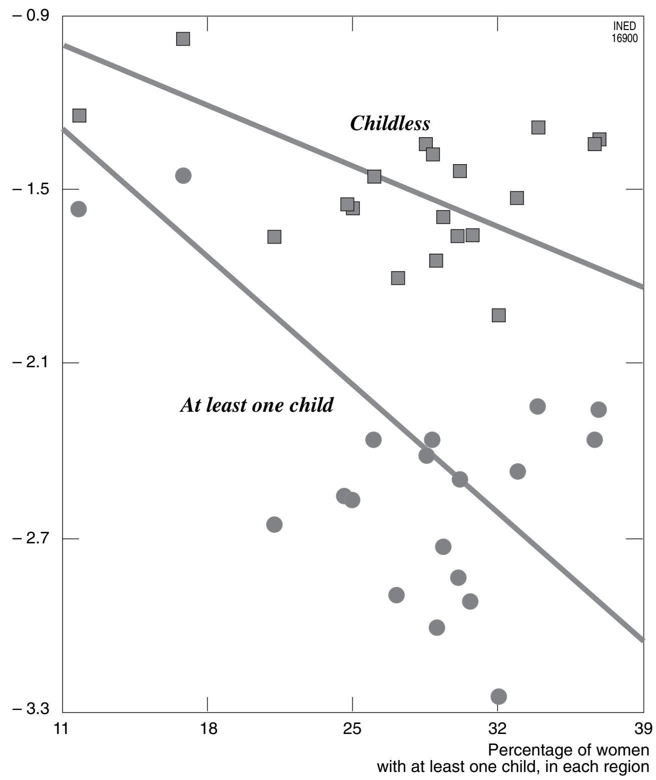


FIGURE 24–3 Logit model on Norwegian migrations: logit of the probability of migrating, relative to the percentage of women with at least one child.

the probability of migrating for each region as a function of the percentage of women with children, reveals a more complex situation than that obtained with the contextual logit regression: the points corresponding to the different regions are highly dispersed in relation to the regression lines plotted on the same diagram.

Various characteristics (marital status, labor market status, educational attainment, etc.) can of course be introduced, separately or simultaneously, to examine their impact on the probability of interregional migration. Their effects may differ from those we have identified by introducing the presence of children. Thus for men the fact of working in farming greatly reduces the probability of emigrating (Courgeau and Baccaïni, 1997). Conversely, it has no effect on the random variables at the regional level: In this case the lines, corresponding to those shown on Figure 24–2, will be parallel to each other. Introducing the percentage of farmers present in each region now produces two parallel lines that rise with this percentage: farmers have a much lower probability of migrating than other categories, but the higher the percentage of farmers in a particular region, the higher the probability of migrating for all categories. This result clearly illustrates the danger of inferring results for individuals from results obtained at a more aggregated level. Having a large number of farmers in a region results in a higher probability of migrating for all population categories, doubtless due to the increased scarcity of nonagricultural employment in such regions. However, this does not mean that farmers have a higher probability of emigrating than the other categories. Indeed, the opposite is observed.

Finally, such analyses can be extended to competing risk models in which individuals experience several types of events, such as death from a specific cause, emigration to different regions, and so forth. These models are a relatively straightforward generalization of the binomial model presented above and we will not discuss them in detail here.

3. Multilevel Event History Analysis

We now come to the most comprehensive analysis of behavior, which combines an event history approach (see Chapter 23 on *event history analysis*) and a multilevel analysis. Because individuals are followed throughout their occupancy of a given state, some of their own characteristics may change at certain points in time (they marry, change occupation, etc.) and the characteristics of the regions in which they live will be subject to continuous change over time (increase in the percentage of married people, changes in the percentages of occupational groups, etc.).

Conducting such an analysis calls for finely detailed survey data, of the kind obtained from retrospective designs, that record all the events in the life course of individuals. Population register data, even when linked to census data as in the Norwegian source used in the prior section, can be used to monitor only a small number of events. It is inadequate to sustain a full multilevel event history analysis. Also, for event history surveys to be usable they must be on large numbers. The data must be capable of revealing the differences between a large number of groups, geographic regions, and so forth.

To illustrate an analysis on these lines we will use data from the Youth and Careers survey conducted by INSEE on a sample of nearly 20,000 individuals. The event we examine is leaving home for the first time by young girls (9043). France has been divided into 19 regions.

The exposition uses a semiparametric model, also known as the Cox model. The objective of this model is to estimate an instantaneous hazard rate (in this case for leaving the parental home), at $t_{i,j}$, as a function of the column vector of the different characteristics $Z_{i,j}$ and the regions j in which the individual i lives. This hazard rate, $h(t_{i,j};Z_{i,j})$, can be written:

$$h(t_{i,j};Z_{i,j}) = h_0(t_{i,j})\exp(\beta_j Z_{i,j}) \qquad \text{(Eq. 12)}$$

In this case, some of the parameters for estimation of the row vector, β_j, will have a random part.

Let us suppose that the times at which a young woman leaves the parental home, or is lost to observation without yet having left, are time-ordered and that at each point in time we can determine the population exposed to the risk. At time $t_{i,j} = l$, the probability that a young woman i will leave her parents, conditioned by the population exposed to the risk and by a departure from the parental home occurring at this date, is equal to:

$$\frac{h_0(l)\exp(\beta_j Z_{i,j})}{\sum_{k\in R_i} h_0(l)\exp(\beta_j Z_{k,j})} \qquad (13)$$

where R_i is the complete population of individuals exposed to the risk at $t_{i,j} - 0$ in any group.

Cancelling out $h_0(l)$, which is in both numerator and denominator, we obtain a partial likelihood that no longer depends on this baseline hazard by multiplying these conditional probabilities at all the dates. What we have written here is a *multilevel event history model*.

So far we have assumed that a single event is observed at each point in time. In practice data tend to be more grouped and we observe $n_{i,j}$ departures from the parental home at date l. The parameters β_j can still be estimated by the maximum likelihood method. The values of $h_0(l)$ are then estimated using these values of $\hat{\beta}_j$. Estimation of the latter function can be simplified by fitting a polynomial distribution (Goldstein, 1995), thus giving in effect a parametric model. A third-degree function is used here.

The explanatory variables can be defined at any level of aggregation and can be considered time dependent. Those introduced here are the individual's birth cohort, whose effect is represented by a second-degree function; the number of brothers and sisters, up to a maximum of four; and the fact of the young woman being in employment before leaving the parental home, which is a time-dependent variable.

Table 24–3 presents the estimated parameter values depending on whether a simple or multilevel event history model is used. The fixed parameters are very similar in both cases. They indicate a slight fall in age at leaving the parental home, up to the cohorts born at the end of the 1950s, followed by a sharp increase, in parabolic form, of this age for subsequent cohorts. The more brothers and sisters there are, the earlier leaving the parental home occurs. Last, for the woman in employment, the probability of leaving the parental home is multiplied by 1.9 (exp [0.64]). The multilevel model shows that depending on the individual's region of residence there is substantial variation in the probabilities. Figure 24–4 shows how this multiplicative effect varies from region to region. The highly urbanized regions are those where leaving the parental home occurs latest (regions around Bordeaux, Marseille, Paris), whereas the opposite is observed in the rural regions (such as Brittany and Normandy). The cohort effect at the level of the random variables is not significant: The curves on Figure 24–4 are nearly all parallel to each other, even though they were estimated with this random effect.

The example given above is of course merely the initial stage of such an analysis. Many other characteristics will have an influence on the behavior of young people, including family and parental background, the economic, political and social conditions at the time of taking the decision, and a range of individual characteristics.

TABLE 24–3 Leaving parental home, France, by various characteristics

Characteristics	Simple event history model Estimation (standard error)	Multilevel event history model Estimation (standard error)
Fixed		
Cohort*	–0.109 (0.018)	–0.121 (0.022)
Cohort^2	–0.222 (0.028)	–0.234 (0.017)
Number of brothers and sisters	0.039 (0.008)	0.037 (0.007)
Working prior to departure	0.641 (0.026)	0.635 (0.039)
Random		
σ^2_{u0} (Constant)		0.012 (0.005)
σ_{u01} (Covariance)		0.006 (0.004)
σ^2_{u1} (Cohort)		0.005 (0.004)
–2 (log-likelihood)	42,797	22,684

*Cohort centred on 1964 and divided by 10.
Source: Youth and Careers survey, INSEE.

III. GENERALIZATION OF THE ANALYSIS

In the present exposition, we have deliberately restricted attention to relatively simple models, with the aim of illustrating the contribution that contextual and multilevel approaches can make to an exclusively individual approach. We will now indicate the extensions that can be given to these models.

So far we have considered hierarchical aggregation at a maximum of two levels: the individual and rural clusters; the individual and geographic regions. If we retain this *hierarchical classification* we can begin by considering a greater number of levels of aggregation. Thus the individual is a member of a household, which is resident in a neighborhood, which in turn is located in a town, and so forth. However, any other structure of classification can also be used. For example, pupils are grouped by classes, the classes are in schools, the schools may be public or private, and so forth. The

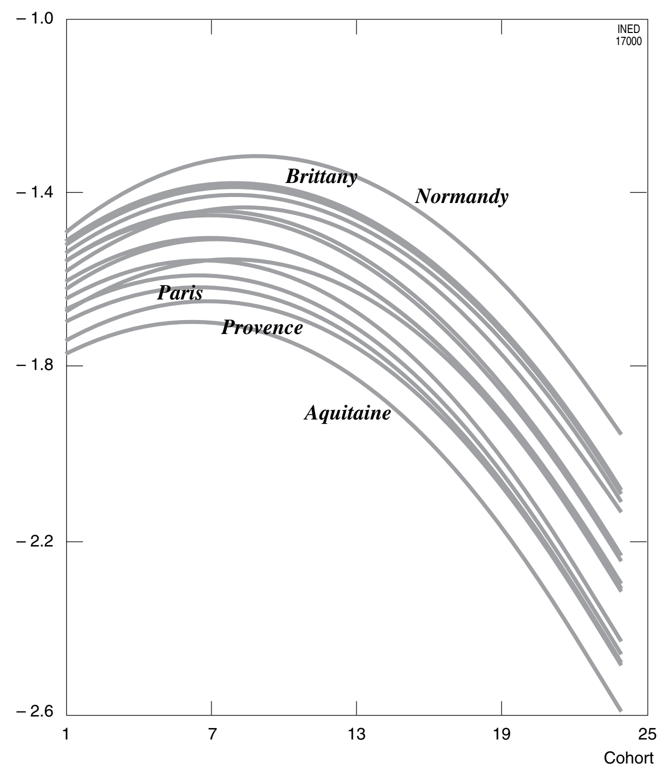


FIGURE 24–4 Logarithm of the probability of migrating by Region.

explanatory characteristics, individual or aggregate, will operate at these different levels, with the random variables specific to each. In each case, careful thought is required about how to interpret the action of a characteristic that operates at different levels of aggregation.

More complicated structures of classification can be realized, such as individuals resident in towns that are grouped by ascending size order but also by their role as centres of government, manufacturing, tourism, etc. The result here is a *cross-classification* in which towns are classified both by their size and by their function. The models presented earlier can be generalized to classifications like these (Goldstein, 1995; Courgeau, 2003).

It is of course possible to have data structures that are in part hierarchical and in part cross classified. For example, individuals can be classified by type of residential area (town centre, suburb, etc.) and type of workplace (industrial, commercial, etc.), which are themselves units in a hierarchical classification of departments and regions.

We can also try to dispense with the hypothesis that underlies the previous analyses, namely, that the groups have no structure. This hypothesis is untenable when working on small groups such as the family and the household. In this case we need to introduce a social structure for these groups, by distinguishing the behavior of parents from that of children and of other people living in the group, for example (Lelièvre *et al.*, 1997).

Looking still further ahead, there is also a need to advance beyond the individual approach used here, which explains behavior by characteristics measured at different levels of aggregation. The analysis must be extended to examine the elaboration and modes of operation of the various levels. How are these levels organized, and how is their action modified? For example, actions by isolated individuals in a given community may produce an awareness of problems and action to resolve them by those with influence at the community-wide level (new laws, acceptance of new forms of behavior, and so on). This could create new problems at both the societal and individual levels, for example.

In summary, multilevel analysis offers a solution to many problems encountered when demographic analysis is conducted at either the individual or a more aggregated level. It combines the different levels of aggregation in an entity that is more informative than when each level is considered separately. As such it enables us to assess the conjoint influences of individual and aggregated characteristics on demographic behavior. Further advances require its generalization to the study of the operation of the different levels of aggregation, thereby offering new insights into the nature of change in human societies in their full complexity.

References

Bressoux Pascal, Coustière Paul and Leroy-Audouin Christine, 1997. Les modèles multiniveau dans l'analyse écologique: le cas de la recherche en éducation, *Revue Française de Sociologie*, vol. 38(1), p. 67–96.

Courgeau Daniel and Baccaïni Brigitte, 1997. Analyse multiniveaux en sciences sociales, *Population*, vol. 52(4), p. 831–864. (Also in English: Multilevel analysis in social sciences, *Population: An English Selection*, 1998, vol. 10(1), p. 39–71).

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, vol. 1, p. 7–26. (Also in English: From the group to the individual: what can be learned from migratory behaviour, *Population: An English Selection*, 1995, vol. 7(1), p. 145–162).

Courgeau Daniel, 1999. De l'intérêt des analyses multi-niveaux pour l'explication en démographie, *in:* Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie*, p. 93–116. Louvain-la-Neuve, Academia-Bruylant/L'Harmattan, 670 p. (Chaire Quetelet, 1997).

Courgeau Daniel, (ed.), 2003. *Methodology and epistemology of multilevel analysis*, Methodos Series, no. 2. Dordrecht/Boston/London, Kluwer Academic Publishers, 236 p.

Courgeau Daniel, 2004. *Du groupe à l'individu: synthèse multiniveau*, Paris, coll. de l'INED, PUF, 242 p.

Diez-Roux Ana, 1998. Bringing context back into epidemiology: variables and fallacies in multilevel analysis, *American Journal of Public Health*, vol. 88(2), p. 216–222.

Entwistle Barbara and Mason William M., 1985. Multilevel effects of socio-economic development and family planning programs on children ever born, *American Journal of Sociology*, no. 91, p. 616–649.

Goldstein Harvey, 1995. *Multilevel Statistical models.* London, Edward Arnold, 178 p.

Lelièvre Éva, Bonvalet Catherine and Bry Xavier, 1997. Analyse biographique des groupes: les avancées d'une recherche en cours, *in:* Daniel Courgeau (ed.), Nouvelles approches méthodologiques en démographie, *Population*, vol. 52(4), p. 803–830 (Also in English: Event history analysis of groups. The findings of a on-going research project, *Population: an English Selection*, 1998, vol. 10(1), p. 11–37).

Loriaux Michel, 1989. L'analyse contextuelle: renouveau théorique ou impasse méthodologique, *in:* Josiane Duchêne, Guillaume Wunsch and Éric Vilquin (eds.), *L'explication en sciences sociales: la recherche des causes en démographie*, p. 333–368. Louvain-la-Neuve, Éditions Ciaco, 476 p.

Schoumaker Bruno and Tabutin Dominique, 1999. Analyse multiniveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, *in:* UEPA-NSU (ed.), *La population africaine au 21^{e} siècle*, vol. 1, p. 299–332. Dakar, UEPA, 630 p. (Proceedings of the Third African Population Conference, Durban).

Von Korff Michael, Koepsell Thomas, Curry Susan and Diehr Paula, 1992. Multilevel analysis in epidemiologic research on health behaviors and outcomes, *American Journal of Epidemiology*, vol. 135, p. 1077–1082.

SECTION

II

DETERMINANTS OF FERTILITY

Introduction

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

After population dynamics, considered in the first section of this treatise, the next three sections address a composite topic that we have split among two volumes purely to avoid ending up with a tome too voluminous and difficult to handle. Logically, the determinants of the three components of population dynamics—fertility, mortality, and migration—could have been considered together in a single section. This is because these three factors of population change and their trends depend on processes which may be very different but have the same rationale: Based on biologic preconditions specific to humankind, the propensities to procreate, migrate, and die are conditioned by an extremely complex set of variables, proximate or not, ecologic, geographic, economic, sociologic, political, cultural, philosophical, metaphysical, and so forth, which are time and space dependent. Through them, all populations at every moment of their history develop individual or collective faculties and behaviors that determine their dynamics. So, where Section I analyzed the internal mechanisms of these dynamics; measured fertility, mortality, and migration; and discussed the linkages between these three factors and the age–sex structure of population before taking them all into account in models that help us gain insight of the demographic impacts of a given process, the following three sections will attempt to show how the elements of these dynamics are shaped and change.

Because the task had to be divided between three sections, each section will be given over to one component of population change: fertility, mortality, and migration. This section considers fertility determinants.

Although recent developments in bioscience mean that human reproduction by cloning no longer lies entirely in the realms of science fiction, humans essentially remain sexual beings whose reproduction is possible only through the combination of male and female gametes. And what is a biologic necessity is also broadly true in sociologic terms: In most cases, gamete combination is preceded by the formation of a union, notwithstanding the rapid changes in the typology, and even the very nature, of unions in the past few decades.

One of the primary factors in human fertility therefore is union formation. In almost all societies, couples' union formation, lifestyle, and relations have been governed by social, ethical and religious rules that are often so strict that marriage was in many cases for long a precondition of fertility, nonmarital fertility being completely forbidden. That being so, the study of marriage should precede that of fertility, the latter being mainly composed of marital fertility in many societies. It is, admittedly, a foregone conclusion that rules are never applied to the letter, and nonmarital fertility may exist alongside marital fertility. But that is by definition illegitimate, since legitimate fertility can only take place within marriage. Demography, born in a highly puritanical Europe, for long had only these opprobrium-filled expressions to distinguish not only marital from nonmarital fertility, but also the issue of these two forms of fertility: legitimate births and illegitimate births. No longer so. Not only have demographers had to account for population dynamics where the distinction is more complex, less rigid, and less discriminatory, as in some African societies for example, but, in Europe itself, the picture has changed radically with the spread of nonmarital cohabitation and new patterns of union. The fact still remains that, whatever the form, the union of two individuals of opposite sexes is still in most cases a precondition of fertility. On a purely technical level, admittedly, nonsexual reproduction is now possible for a woman. But the overwhelming majority of such cases are couples seeking to overcome infecundity. A study of union formation, and an analysis of *nuptiality*, is therefore the necessary first step toward explaining fertility factors. That will be the focus of part one of this section.

But union formation is not the only precondition of fertility. The union so formed must be fecund and fertile. It must be biologically fecund (i.e., physiologically capable of conception). This means not only must neither the woman nor the man be infecund but also that the union they form must itself be fecund. Then, too, the ability to conceive must be transformable into a capability to give birth. That means the mother being able to carry her pregnancy to term and produce a live birth. Here again, biology plays an overriding role through the process of fetal development exposed to the risk of all causes of intrauterine mortality. Finally, it entails the woman—the union—being fertile (i.e., able to effectively manifest its fertility). Here, of course, far more than in other respects, individual and social behaviors also come into play. Sexual behaviors, contraception, and induced abortion are all proximate determinants of fertility. These interconnected proximate determinants of fecundity and fertility are examined in Part II.

However, before these proximate factors, sometimes called *intervening variables*, come into play there is a set of fundamental or nonproximate fertility determinants that stem from the economic, social, cultural, or ideologic context in which individuals and couples live. The complexity of these relations has led to the development of a constant flow of new explanatory models and theories of fertility trends. These will be considered in the third and final part of this section.

PART

I

A PRECONDITION TO FERTILITY

Union Formation

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Any self-respecting textbook on demographic analysis will deal with nuptiality on equal terms with fertility, mortality or migration. Rightly so, since marriage, like conception, childbirth, migration, or death is a major life event whose impact in a population merits the same demographic analysis as any other vital event. Rightly also on the grounds that, educationally speaking, nuptiality probably lends itself best to the most comprehensive explanation of the principles of analysis.

And yet, marriage was deliberately not considered in Section I of this treatise. It is true, as stated in the introduction to Section II that union formation is in most cases a precondition to fertility. However, it is only of several. Marriage *per se* has no direct influence on population size and development: Many marriages are celebrated in a given year, no new members will be added to the population if the newly formed unions are infecund or permanently infertile. The only difference is the exposure to the risk of conception and childbirth, assuming that out-of-wedlock fertility is lower than in-wedlock fertility. Consequently, marriage or, more generally, union formation had no place within the analysis of population dynamics as such, but should more appropriately be considered in connection with the factors of that element of the dynamic that it helps to determine: fertility.

The traditional approach focuses on the fertility differentiation that union formation creates between the individuals (especially women, since fertility is most often measured in relation to women). The first thing therefore is to measure nuptiality itself, which entails a breakdown between married and nonmarried couples, before measuring the fertility of married and nonmarried couples. This will give a breakdown of fertility into marital and nonmarital fertility, usable to gauge the influence of nuptiality patterns on fertility trends.

But no union is never-ending: Inherent within union formation is union dissolution. This means considering widowhood and, where relevant, divorce in the same way as nuptiality, which clearly complicates the model, especially as a dissolved union may be followed by a second, third, and other unions. Almost the first step, therefore, is to look at nuptiality by order of marriage. However, even that is not enough, for not only are there other forms of marriage beside monogamous marriage—not least polygamy, still very common in some societies—but the types of unions themselves, monogamous or otherwise, are increasingly varied, both in legal and social terms, and in terms of the effective commitment to a permanent union.

For that reason, no less than three chapters are given over to the measurement of nuptiality. Chapter 25, authored by Patrick Festy, gives an overview of the traditional analytical approach to union formation and dissolution, informed essentially by data from European countries. In Chapter 26, Philippe Antoine then explains how the biographic approach can inform that analysis, at the risk of somewhat begging the question of factors. Finally, in Chapter 27 the same author reviews the many forms that unions can take, referring to a case study of Africa.

Union formation and dissolution may be fertility factors, but also have their own determinants as the biographic analysis will have shown. Thérèse Locoh takes a more detailed look at these in Chapter 28.

Finally, in Chapter 29, Silvana Salvini and Antonio Santini look at measurements of marital and nonmarital fertility.

CHAPTER

25

Analysis of Couple Formation and Dissolution

PATRICK FESTY

Institut national d'études démographiques (INED), Paris, France

Strictly speaking, marriage is not a demographic event, for it does not contribute—as do births and deaths—to population replacement. Indeed, in several ways, it stands apart from the other topics of demographic study.

Marriage is an essential step in family formation and in childbearing. In technical terms, it is the first of the intermediate fertility variables (i.e., an essential factor in the exposure to the risk of birth). In other words, the likelihood of a couple having a child in the year ahead is usually very different depending on whether the partners are married or not. To treat married persons as a distinct category thus involves defining a group whose fertility behavior is more consistent than that of the total population.

However, married status is only an approximation of the intermediate variable that the demographer would like to define. It is a social status in which people are more likely to want children, or to agree to having children, than in the unmarried status. In Western societies, there has long been a desire for fertilizing intercourse to take place in marriage; fertility out of wedlock was stigmatized. But the frontier between the desirable and the nondesirable can shift. Today, in France, the married and the unmarried are not separated as sharply as before given society's ready acceptance that unmarried persons in stable partnerships should have children. The frontier now lies between (stable) couples and noncouples; the defining factor that marks the entry into family life—because it signifies an abrupt change in the risk of birth—is the formation of the couple, whether legalized by marriage or not.

The same pattern has long been observed in many non-Western societies as well. For example, in Latin America or in the Caribbean, marriage is only a formal and optional stage in a partner relationship, without significant impact on fertility. It has never been a major object of study for demographers. In such societies, as in Western countries today, the difficulty is to determine a marriage-equivalent event marking the entry into family life, whose frequency and other characteristics can be measured.

The privileged attention given to marriage is partly due to the existence of an easily available source that enables it to be identified and enumerated: the civil registration records. As this source is also the one that enables demographers to compile birth and death statistics, the study of marriage is typically linked to that of fertility and mortality. The source is, however, much older than civil registration and demographic studies, since the Catholic Church has long required baptisms, marriages, and burials to be recorded in parish registers. Better yet, marriage played a central role in the recording process, because the purpose was to make sure that the marriage was possible: The spouses had to be free—through widowhood—of any earlier marital ties, and had to prove they were not too closely related to marry.

As a result, France has a source dating back to the 17th century that captures the basic characteristics of marriages. By contrast, it is often hard to construct a life table spanning all ages of life for so remote a period. In the 19th century, France added a question on marital status to the census: The question—almost as old as modern censuses themselves—provides information that supplements the classic statistics and is based on principles discussed later.

However, marriage is not only a major element for understanding fertility that is well covered in current statistics. It is also a convenient educational tool for setting out the essential principles of classic analysis of demographic phenomena and addressing most of the difficulties encountered in the study of the main phenomena.

Marriage, like childbirth, is a repeatable event—or, as demographers say, a *renewable event*. However, unlike births, the repetition cannot occur unless another event has taken place: divorce or widowhood. Nuptiality is defined by all or part of the marriage-termination–remarriage sequence, which, as it unfolds, may become first marriage–divorce (or widowhood)–second marriage, and so forth.

Moreover, the stable union that is the relevant object of study in certain societies can take the legal form of marriage or be merely consensual. This option, offered to the partners when they begin to cohabit, is formally comparable to competing risks, which exposes an individual to death from cancer, cardiovascular disease, and so on.

Last, more than any other analysis, the study of marriage provides a tangible illustration of a fundamental difficulty in demographic measurement: How does one evaluate the sensitivity of the number of marriages to the number of males *and* females old enough to form partnerships? The same difficulty occurs in fertility studies and in all instances where the encounter of two persons is a prerequisite to the occurrence of an event.

I. PRINCIPLES OF COHORT ANALYSIS

As we have seen, nuptiality encompasses various events that can form a sequence. However, these successive events are governed by three common principles of analysis:

1. *The breakdown into successive events makes each of them unique* or, more precisely, nonrenewable in the life of a group: there is only one first marriage; a person can lose his or her partner only once, and so forth. To assess the importance of a phenomenon associated with these events, we must begin by measuring its frequency: What proportion of persons marries, at least a first time; what proportion of couples divorce; how many divorced women remarry; and so on.

   For each stage (first marriage, divorce, remarriage of divorcees, etc.) there is a frequency, and for each frequency there is a group in which we can measure it. Everyone can marry (a first time) but only married persons can divorce, and only divorced or widowed persons can remarry. The frequency of marriage will therefore be measured in the total population, that of divorce or widowhood in the group of married persons, and that of remarriage in the group of divorced or widowed persons.
2. *In each of these groups, a period of time must elapse* before we can determine the number of persons who have experienced a given event. The starting point of the period is identical to the date of formation of the group studied: the frequency of divorce will be measured 25 or 30 years after the marriage date, that of remarriage some 15 years after the divorce or widowhood, and so on. First marriage has no specific origin-event other than birth and no other specific duration than age.

   To characterize a nuptiality-related event—such as first marriage, divorce, or remarriage of divorcees—we determine not only its frequency but also its distribution in time. This distribution is often summed up by a baseline value such as mean duration. In the sequence of married-life events, each specific duration originates in the immediately preceding event. We therefore measure the mean duration of marriage at the time of divorce, the mean time since divorce at the remarriage of the divorcees, and so on. As the duration specific to first-marriage analysis is identical with age, we likewise compute a mean age at first marriage.

   Louis Henry (1966) summarizes these first two principles of analysis as follows: When event B cannot occur unless A has taken place, we define a cohort of persons having experienced event A in a given year and we measure, within this cohort, the proportion of persons who have subsequently experienced event B and the distribution of durations elapsed between A and B. The proportion is the quantum of phenomenon B, while the distribution is its tempo.
3. *Because time elapses between A and B, persons in group A may have left the group at the time when they would have liked to experience B*. In a birth cohort, persons who would have married at 25 are prevented from doing so by premature death or departure (emigration). The number of marriages taking place is inversely related to the number of

> deaths and migrations occurring in the cohort studied. It is not, therefore, a very satisfactory basis for measuring nuptiality *per se*. To arrive at that figure, we must determine how many marriages would have occurred at each age if there had been no mortality or migration until that age, that is in a *pure state* by analogy with the principles of chemical analysis that allow simple and pure elements to be isolated from mixtures.

The estimation of the number of marriages at a given age x absent earlier mortality or migration rests on the following hypothesis: Persons deceased or emigrated before age x would have had the same risk of marrying at age x as survivors present at that age. This hypothesis—known as the *independence hypothesis*—underlies the determination of quantum and tempo of demographic phenomena. In the area of nuptiality, it is applied to the measurement of first-marriage frequency as well as to the calculation of the mean duration of marriage at the time of divorce, and so on. We will examine its relevance and validity.

II. THE NUPTIALITY TABLE

Using the model of the probability of dying by sex and age, we can calculate a first-marriage probability by sex and age x, which measures the risk of contracting a first marriage in the interval x, $x + \Delta x$ for persons who have not yet been married at age x. If marriage is understood in the narrow sense of a legal marriage, these persons still have never-married status (if we are dealing with a first union, in a broader sense, the persons thus defined have always been free of any conjugal tie; we shall return to this issue). Let:

$$\Delta_x^{n_x} = M_c(x,x + \Delta x)/C_x.$$

At older ages, the number of single or never-marrieds C_x (C representing celibacy) is reduced not only by the number of first marriages M_c, but by the deaths and migrations of never-marrieds. If we compare the number of never-marrieds at age 50, C_{50}, with the number of never-marrieds at age 15, an age before which marriage is very rare in present western societies, the difference $C_{15} - C_{50}$ or the ratio C_{50}/C_{15} depends not only on nuptiality but also on the mortality and migration of never-marrieds.

To express the sole effect of nuptiality,[1] we compute a series c_x by combining, through multiplication, the probabilities of not marrying $(1 - \Delta_x^{n_x})$ at successive ages from C_{15} onward. For example, if the first-marriage probabilities have been calculated on annual intervals,[2] we have

$$c_{50} = C_{15}(1 - {}_1n_{15})(1 - {}_1n_{16}) \ldots (1 - {}_1n_{49}).$$

More generally, we can write:

$$c_x = C_{15}(1 - {}_1n_{15}) \ldots (1 - {}_1n_{x-1}).$$

$$c_{x+1} = C_{15}(1 - {}_1n_{15}) \ldots (1 - {}_1n_{x-1})\,(1 - {}_1n_x).$$

and therefore:

$$c_{x+1}/c_x = 1 - {}_1n_x.$$

$${}_1n_x = 1 - c_{x+1}/c_x = (c_x - c_{x+1})/c_x.$$

As the series c_x is reduced, from one age to the next, by the sole effect of nuptiality, the difference between c_x and c_{x+1} is due only to marriages of never-marrieds $m_c\,(x,x + 1)$. Hence:

$${}_1n_x = m_c(x,x + 1)/c_x = M_c(x,x + 1)/C_x.$$

The probability calculated on observation data (in a reality where the number of never-marrieds is determined by marriages, deaths, and migrations) is then incorporated into a nuptiality table, where only marriages reduce the number of never-marrieds. If, of 10,000 never-marrieds at age 15, there are 4000 remaining at age 30 who are still never-married, survivors, and present, there will be, perhaps, about 5000 persons of the same age in the nuptiality table—absent mortality and migration. The transition from age 30 to age 31, measured in reality on the 4000 survivors present, will then be applied to the 5000 never-marrieds in the table, under the implicit hypothesis that mortality and migration have not altered the representativeness of the 4000. This is known as the hypothesis of inde-

[1] Block capitals indicate real data, whereas lower-case letters represent data in the table.

[2] Marriages are not the only events occurring in the interval $\Delta\psi$: There are also deaths and migrations of never-marrieds. These last two events can prevent never-marrieds from marrying in the same interval. The number of never-marrieds exposed to the risk of marrying will be modified as a result. For example, if a never-married person dies during the interval, he or she is excluded from the risk for the remainder. The effect is negligible if the interval Δx is very short, and increases with the duration of the interval. If it becomes significant, we factor it in by making two simplifying hypotheses (see Chapter 8): (1) the deceased are excluded from the risk of marriage for the remaining half of the interval; (2) their risk of marriage is the same as that of the other never-marrieds.

The number of marriages is consequently reduced by $D_c(x,x + \Delta x) \times \frac{1}{2} \times {}_{\Delta x}n_x$.

Hence:

$$M_c(x,x + \Delta x) = C_x\,{}_{\Delta x}n_x - (D_c(x + \Delta x)_{\Delta x}n_x)/2$$

and

$${}_{\Delta x}n_x = M_c(x,x + \Delta x)/(C_x - D_c(x + \Delta x)/2)$$

Migrations of never-marrieds are similarly accounted for. The effects of emigration are comparable to those of mortality.

pendence between (1) the mortality and migration of never-marrieds and (2) the nuptiality of never-marrieds.[3]

We can thus construct a nuptiality table of never-marrieds, whose elements are comparable to those of a life table (i.e., at successive ages), three series, respectively, describing first-marriage probabilities, never-marrieds (taking $c_{15} = C_{15}$), and first marriages. With 1-year intervals between successive ages, we obtain:

Age	Probabilities	Never-marrieds	Marriages
15 y	${}_1n_{15}$	c_{15}	$m_c(15,16)$
16 y	${}_1n_{16}$	c_{16}	$m_c(16,17)$
17 y	${}_1n_{17}$	c_{17}	$m_c(17,18)$
—	—	—	—
x	${}_1n_x$	c_x	$m_c(x,x+1)$
$x+1$	${}_1n_{x+1}$	c_{x+1}	$m_c(x+1,x+2)$
—	—	—	—
49 y	${}_1n_{49}$	c_{49}	$m_c(49,50)$
50 y	${}_1n_{50}$	c_{50}	$m_c(50,51)$

$$c_{x+1} = c_x(1 - {}_1n_x).$$

$$c_x - c_{x+1} = m_c(x,x+1).$$

$${}_1n_x = m_c(x,x+1)/c_x.$$

As the number of marriages m_c generally becomes negligible after age 50, we admit that the ratio c_{50}/c_{15} measures the total proportion never-married. The mean age at first marriage is written:

$$\bar{x} = (15{,}5\ m_c(15{,}16) + \ldots + (x + 0{,}5)m_c(x,x+1) + \ldots + 49{,}\ 5m_c(49,\ 50))/(c_{15} - c_{50}).$$

III. PROXY MEASURES OF THE NUPTIALITY TABLE

Let us take, for simplicity's sake, a closed population (i.e., without migration). Its nuptiality table sets up a virtual universe that also lacks mortality. The number of never-marrieds accordingly moves from c_{15} to c_{50} as age rises, and the ratio c_{50}/c_{15} measures the resistance to nuptiality; it is equal to $1 - N_{50}$ with $1 - N_{50} = \prod_{15}^{49}(1 - {}_1n_x)$. In these same conditions of no mortality, the population, regardless of marital status, remains fixed: $p_{15} = p_{50}$. As the entire population is still never-married at age 15, c_{50}/c_{15} is equal to c_{50}/p_{50}.

In reality, never-marrieds aged 50 have resisted both nuptiality, and never-married mortality; and the population, regardless of marital status, has resisted mortality. Hence:

$$C_{50} = C_{15}\ (1 - N_{50})\ l_{c50}.$$

$$P_{50} = P_{15}l_{50}.$$

where l_{c50} denotes the probability of survival of never-marrieds between ages 15 and 50, and l_{50} the probability of survival of the total population between the same ages. This gives:

$$C_{50}/P_{50} = (C_{15}/P_{15})(1 - N_{50})\ (l_{c50}/l_{50}).$$

As the entire population is still never-married at age 15, $C_{15} = P_{15}$ and we can write:

$$C_{50}/P_{50} = (c_{50}/p_{50}) \times (l_{c50}/l_{50}).$$

The ratio C_{50}/P_{50}, easily calculable in a population census, is equal to the proportion never married in the nuptiality table—to within one factor that measures the difference in survival rates between never-marrieds and the total population. This factor is generally below unity, because never-marrieds have a lower survival rate than the total population.

In an open population, differential migration has a comparable effect. In some cases, its influence can greatly exceed that of the comparative survival of the two groups. Let us bear in mind, however, that the calculation of the proportion of never-marrieds in the total population of the same age in the censuses gives a proxy measure of the proportion never married. The proxy value is smaller than the actual value because of the excess mortality of never-marrieds. The absence of mortality, which characterizes nuptiality tables, also enables us to write:

$$m_c(x,x+1)/p_x = c_{15}(1 - N_x)_1n_x/p_{15} = (1 - N_x)_1n_x.$$

In reality, the equivalent computation on observed data gives:

$$M_c(x,x+1)/P_x = C_{15}(1 - N_x)\ l_{cx1}n_x/P_{15}l_x = (1 - N_x)_1n_x(l_{cx}/l_x).$$

Thus, the ratio of observed marriages of never-marrieds to the total population of the same age (i.e., *first-marriage frequency*; see Chapter 8) is equal to the marriages in the nuptiality table $[(1 - N_x)_1n_x]$, to within one factor. The latter expresses, as before, the difference in the survival rates of never-marrieds and of the

[3] In most cases, the independence hypothesis has little effect on the validity of nuptiality tables, given the modest numerical impact of mortality and migration among the never-married. In studying events other than the marriages of never-marrieds, we shall encounter situations where the outcome is different, and we shall examine the possible ways of dispensing with this simplistic but convenient hypothesis.

total population. The cumulated first-marriage frequencies, which are easy to compute from current civil-registration statistics and from estimates of the population by sex and age, therefore give a proxy measure of marriages in the nuptiality table. The measure is proxied by default in the frequent case where never-marrieds have a lower survival rate than the total population.

Some components of the nuptiality table—in particular the factors directly involved in the determination of the total proportion remaining single and of the mean age at first marriage—can be calculated on widely disseminated statistics. This gives considerable relevance to the construction of all or part of nuptiality tables.

IV. PERIOD MEASURES

The cohort measurements we have just made capture the age-by-age transition of a male or female cohort from never-married to married status. In Chapter 14, we saw that this type of process could also be described in a hypothetical (or synthetic) cohort on the basis of statistical information pertaining to a given year rather than to a birth cohort. The aim here is to construct a *period nuptiality table*. We begin by computing the first-marriage probabilities of never-marrieds ${}_1n_x$. Next, we combine them at successive ages to produce a series of never-marrieds c_x and a series of marriages $m_c(x,x+1)$. In this combination, we act *as if* the populations at successive ages were linked as closely as in the cohort analysis, whereas in fact they necessarily belong to distinct cohorts in a given year rather than to a single cohort.

The summary indices such as the proportion never married c_{50} or the mean age at first marriage, calculated from the marriages in table m_c, are generally interpreted as the quantum or the tempo of the nuptiality that would be recorded in a cohort if the period conditions persisted long into the future. The implicit hypothesis that makes the computation meaningful is not very different from the independence hypothesis on which cohort construction is based. For example, the probability at age 30 in 1995 is a fragment of the history of the generation born in 1965; its inclusion in the pseudo-history of the hypothetical cohort associated with the year 1995 supposes that the operation has a meaning, despite the fact that this pseudo-history before age 30 differs from the cohort's true history. We thus postulate that the fraction of never-marrieds who marry at a given age is independent of the number of never-marrieds who married before that age: each probability is cut off from its past.

The separation between the events or the risks of a year and the past of each cohort is so sharp that it is almost impossible to formalize the links between the period tables and the cohort tables or between the summary period indices and cohort indices derived from them (see Chapter 17). These links are easier to identify using other summary indices such as cumulated first-marriage frequencies.

We have seen that the first-marriage frequency is very nearly identical to the marriages in the nuptiality table of a cohort of never-marrieds. Their longitudinal combination enables us to construct a table. But we can also combine them at different ages in the same year, in which case we can compute total period indices: sum or mean age.

The ratio of the number of first marriages to the total population of the same sex and the same age—which defines first-marriage frequency—is of the same nature as the computation of the total fertility rate. The period sum of these values has the same relationship to the cohort sums as the sums of total period and cohort fertility rates studied by Nico Keilman (see Chapter 17). In particular, a fall in the age at first marriage in successive cohorts drives up the combined first-marriage frequency, whereas an increase in that age drags the frequency down. When the age at first marriage declines sharply, the upward pressure may actually drive the total period index over the 100% ceiling on cohort indices. The implausibility of such values is a reminder that period measures are only fictions, when it comes to expressing the frequency of behaviors. The lesson also deserves to be heeded in cases where the period index is low: Weak values do not necessarily signal a high never-married rate in cohorts measured in that period, but may be due to ever-later marriages in successive cohorts.[4]

1. Selected Illustrations

In Sweden, the number of marriages started declining before the late 1960s. The annual nuptiality index of never-married women, which was close to 100%, fell by one third in 4 years and has remained stable for more than 20 years at around 50% (Figure 25–1). What was once perceived as a reaction to the economic crisis

[4] Henri Leridon and Laurent Toulemon (1997) show, using the French example, that there are substantial discrepancies between series of total period indices based, respectively, on period nuptiality tables and cumulated first-marriage frequencies. They also measure the gaps between period indices and cohort indices, and show that they are narrower when one uses period tables than when one uses cumulated first-marriage frequencies. They conclude that period tables are a "better" approximation of cohort measures (but is that a sound validation criterion?).

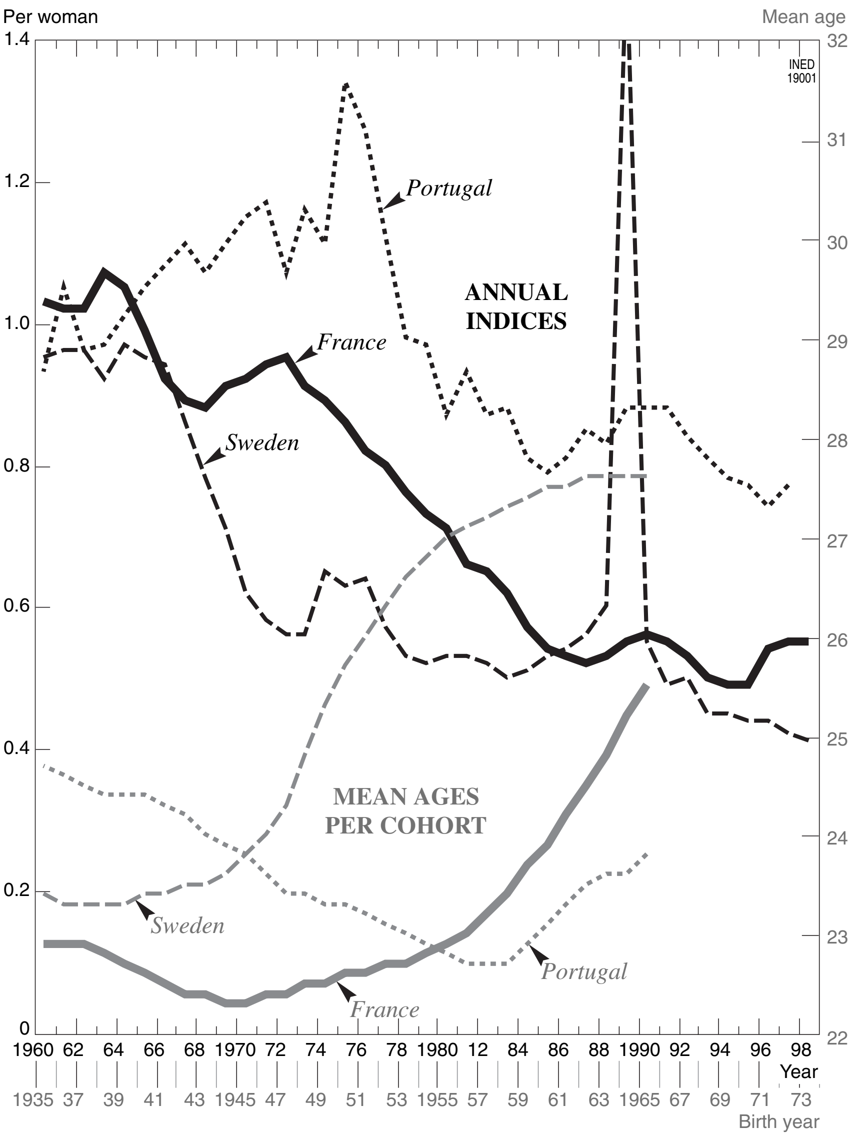


FIGURE 25–1 Annual nuptiality indices for never-married women and mean ages at first marriage by cohort.

or as a fashion has now become a long-term phenomenon. A sharply higher proportion of men and women will never marry—a rare occurrence in the past. The proportion of men and women born in the mid 1960s (i.e., still in their 30s at time of writing) who will not marry is estimated at about 40%, compared with less than 10% for men and women about 20 years older (Table 25–1 in Appendix).

There has been a parallel shift toward later marriage. Women born in the 1960s who marry do so on average at age 27.5 years, which is 4 years later than women born in the 1940s (Table 25–2 in Appendix).

One would have to go back more than 100 years to find such late ages at marriage in Swedish cohorts, when trans-Atlantic migration sharply disrupted the nuptiality of men and women who stayed behind. However, the proportions of never-married men and women have never been so high.

Marriage has declined across Western Europe. The pattern in the Nordic countries and France has been almost as spectacular as in Sweden. In France, the proportion of men and women who never marry has climbed to 30% in recent cohorts, and the age at marriage has risen to 26.5 years for women and 28.5 years for men. The trends in Finland and Norway are comparable. By contrast, the changes under way in southern Europe are distinctly mild: In Italy, for example, the proportion of never-married women born in the 1960s is still below 20%, and the rise in the age at marriage is modest; the extreme case—in total contrast to Sweden—is Portugal, where the very recent increase in age at first marriage has barely affected the cohorts already in their thirties (ODE, 1998).

In a shaded pattern running from the north of Europe to the south, the center (Germany, Austria, and Switzerland) holds a median position. Only France stands apart with a resolutely northern profile, which does not tally well with geography, but underscores the singularity of the marriage model developed in the past two decades.

V. DIVORCE TABLE

Like marriage formation, marriage termination can be analyzed by constructing a table. This enables demographers to identify—absent any other phenomenon that could conceal the risk of termination—the ultimate frequency of divorce and a summary index of its tempo. Some formal adjustments are required, however, before compiling a divorce table.

The impact of divorce is not so much on individuals as on couples. The unit taken into account is therefore the marriage, of which we monitor the duration or termination. The terminating events are of two types: separation of spouses or death of a spouse. The second event, in turn, can take two forms: widowhood of the husband or wife. For our purposes, we will use only the first distinction, without discriminating between terminations due to the death of the husband or wife.

Divorce is analyzed in terms of time elapsed since marriage. This method is consistent with a principle stated in prior sections: Divorce cannot occur except after marriage. Marriage is thus the origin-event that creates the population at risk.

The probability of divorce r_y is the ratio, at the duration $(y,y + \Delta y)$, of the number of divorces (D) $D(y,y + \Delta y)$ to the number of still-married couples Uy. Applying the same principle defined for nuptiality, we use the table to construct something that observation cannot measure—the pattern of extinction of a group of marriages, under the sole effect of divorce—whereas reality describes the combined effect of divorce (D) and widowhood (W). We thus replace the series Uy with the series u_y such that, with annual intervals between successive durations:

$$u_{y+1} = u_y(1 - {}_1r_y)$$

with:

$$_1r_y = D(y,y + 1)/U_y$$

and:

$$U_{y+1} = U_y - D(y,y + 1) - W(y,y + 1).$$

The hypothesis implicit in the combination of successive probabilities of not divorcing is that couples terminated early by widowhood would subsequently have been exposed to the same risks of divorce as the couples that have been preserved from mortality. It is the hypothesis of independence between the risks of widowhood and of divorce.

Using the table, we can determine a divorce quantum and measure it at 25 or 30 years of marriage: Beyond that duration, the number of new divorces becomes negligible. The table also enables us to measure the distribution of marriage durations up to divorce and to summarize them with a mean duration.

However, let us return to the initial choices that led us to group together newly formed couples and study the extinction of this group as it moves along the marriage-duration timeline. Why have we abandoned age as a time clock? First, we should discard the (genuine) objection that the marriage duration is common to both spouses, whereas age is not. Age is defined as that of the wife. As a result, we will refer to divorces of married women aged 20, 21, 22, and so forth. But between 20 and 21 years, for example, the group at risk is not only reduced by divorce and widowhood, it is also enlarged by newly married persons, who, at age 21 years, will not have the same behavior as the married persons who formed the initial group. This is known as the marriage-duration effect (i.e., same-aged women in marriages of different duration are not exposed to an identical risk of divorce). This violates the independence hypothesis, as an external event (marriage) modifies the change over time of the risk studied (divorce rate). The problem does not arise with a marriage cohort monitored from one duration to another, as the duration sequence does not introduce new unions.

Yet, the age issue is not eliminated. First, it would be wrong to pretend that age has no effect on the risk of divorce: At the same marriage durations, early-married women are more likely to divorce than later-married women. But the opposite is also true of the other risks of disunion, as these are linked to the death of one of the spouses (i.e., to mortality): At the same marriage durations, early-married women are less likely to die or lose their husbands than later-married women. Age at marriage therefore introduces a double heterogeneity: Between the risks of divorce *and* between the risks of other forms of termination. It also introduces a negative correlation between the two risks: when the divorce risk is high, the other risk is low, and vice versa. The combination of the heterogeneity and the correlation creates a dependence between the event studied (divorce) and the event that may prevent its occurrence (widowhood).

Let us assume, in an extreme simplification, that there are two ages at marriage of equal numeric importance; the divorce risk is high in the first and zero in the second, whereas the opposite is true for the widowhood risk. For all ages at marriage combined, we find medium-level risks of divorce and widowhood, and the death of a spouse seems to mask the divorce risk. In reality, widowhood does not prevent divorce at all because it occurs in a group where the probability of divorce is zero. The amalgamation of two different subpopulations creates a heterogeneity that contradicts the hypotheses needed for the analysis. Conversely, a good analysis requires a division into homogeneous groups, where the independence hypothesis is justified. For divorce and widowhood, the interaction is unlikely to be as strong as to bias conventional measurements, but we shall encounter a different situation later on.

Between divorce and widowhood, the interferences are so weak that we can often ignore them. As most marriages occur before age 35, and divorces are concentrated in the first 15 years of marriage, mortality is low at the time when divorces are most frequent. We can therefore safely omit widowhoods and simply calculate the ratio of divorces to initial marriages, at each marriage duration, in order to obtain a good estimate of divorces for the table and to deduce the table's other components.[5]

The statistics used to construct Table 3 in the Appendix have been analyzed using these methods. We have computed the ratio of divorces to the initial number of marriages in France at each marriage duration in the marriage cohort examined. *Disregarding the widowhood factor*, which we should have withdrawn from the denominator, the ratios are equal to the divorces of the divorce table. Their aggregation therefore measures the proportion of marriages terminated by divorce at the end of a marriage duration of over 40 years; the distribution of marriage durations at the time of divorce is summarized by a mean duration.

In 25 marriage cohorts (1946–1970), the increase in the ultimate proportion of terminated marriages has been spectacular, from 10% to almost 28%. The rise was moderate in the first 10 cohorts, then quickened: in the last 15, the ultimate proportion doubled. This trend is not confined to France, but is observed throughout Western Europe. It has been accompanied by changes in legislation. On January 1, 1976, France introduced a law liberalizing access to divorce by facilitating consensual procedures; it is doubtful, however, whether legislative reform is responsible for the rise in divorces.

In the cohorts studied, divorce remains a phenomenon very widely spread over time. The mean duration initially rose in cohorts affected later in life by the general change in behavior; it peaked at nearly 17 years in marriages contracted in around 1960, and then declined. The latter downtrend should gather speed in the post-1970 cohorts, in which there have been so many divorces in the first years of marriage that one can hardly believe they will be some day be as frequent as they used to be at the longer durations.

VI. WIDOWHOOD TABLE

The construction of a widowhood table obeys principles comparable to those used for divorce. Let us choose, as before, to study only terminations due to the death of *either* spouse. Divorce is the interfering event and marriage the origin-event that creates the population at risk. There are, however, some specific aspects of widowhood that prompt us to depart from the approach used for divorce when compiling the tables:

- The statistics by marriage duration of couples terminated by death are rarely produced. Yet they could be obtained simply by adding up the deceased married men and women at each duration. Moreover, it is surely easy to establish the marriage dates of the deceased, since one of the spouses almost always survives and is presumably the person who reports the death.
- Divorce prevents widowhood far more often than the opposite, as it occurs relatively early in marriage, whereas widowhood tends to occur late. We can therefore omit widowhood in a rough measure of divorce frequency, but the opposite is hardly possible.
- These constraints enhance the attractiveness of a classic, indirect, and convenient method of computation, which is to combine male and female life tables in order to produce a life table for couples. This is done by linking a cross-table of male and female ages at marriage x and y with the probabilities of survival of couples at successive marriage durations a:

$$(l^{M}_{x+a} \times l^{F}_{y+a} \div l^{M}_{x} \times l^{F}_{y})$$

- But this substitution method does have many imperfections:

[5] One must also consider the *interfering* role of migration, as it can be high at the ages when the risk of divorce is high as well. Emigration exerts an influence in the same direction as mortality and widowhood; immigration exerts an opposite influence. It should be noted, however, that the migration concerned is that of married persons, often far more modest than that of never-marrieds. We should also examine, on a case-by-case basis, whether divorces of migrant couples are pronounced in the couples' country of origin or in the country of destination.

- For a cohort computation, we need cohort life tables by sex, year of age, and year of birth. The available statistics are often far from offering such detail.
- We also need a measurement of the specific mortality of married men and women for each cohort—plus, of course, measurements differentiated by age at marriage, given the links between age at marriage and social group, which can influence mortality.
- Proxy measurements postulate a form of independence different from the hypothesis used until now in this chapter: the independence between the mortality of husbands and the mortality of wives. But, given the social homogamy of spouses, this independence is far from certain.

To these caveats we should add that widowhood plays a modest role in contemporary western societies, if nuptiality is viewed merely as an intermediate fertility variable, as it occurs in couples at ages and marriage durations where fertility has already been completed. But the situation is different if we take into account the sociological dimension of marriage. For example, it is commonly argued that the rise in divorces for marriages contracted after 1950 is linked to the extension of the potential duration of married life, consecutive to the increase in age at widowhood—although the latter trend cannot be adequately measured.

VII. REMARRIAGE

The principles of the remarriage analysis are modeled on those developed for divorce analysis. The duration to be measured is the time since marriage termination. However, we return to the individual—man or woman—as the unit of account. Death can prevent remarriage from occurring, and is thus a competing event. It is also very likely not be independent of remarriage, as early-widowed or early-divorced persons have a high probability of remarrying and a low probability of dying, while the opposite is true of their elders. But we simply need to introduce a distinction between widowers (or widows) and divorcees (of either sex) to restore the independence between death and remarriage, as the first group is, on average, far older than the second.

Moreover, the influence of mortality is all the more modest as remarriage generally occurs very soon after the termination of the earlier marriage, without leaving time for death to conceal the formation of new couples. Thus the ratio of the number of remarriages to the initial number of divorces or widowhoods at each marriage duration usually suffices to provide a good estimate of remarriages in a nuptiality table. This observation is even more valid if the data are gathered retrospectively through surveys of survivors. The figures obtained resemble the census proportions of never-marrieds mentioned earlier in our discussion of first marriages.

Thérèse Locoh and Marie-Paul Thiriat (1995) have used data of this type, compiled from surveys of women under 50 in Togo in 1988. The authors study the respondents' second marriages by introducing, as a time measure, the duration since divorce or the first husband's death. The 50-year age limit impairs the visibility of a phenomenon that occurs inevitably late, because it follows a first marriage and a termination. However, the age at first marriage is sufficiently young in Togo to give time for the termination to occur and to generate a population at risk of remarriage. It should be noted, however, that this population excludes terminations at older ages, whose number is surely not insignificant, especially for widowhood.

In Togo, second marriages are very frequent: 87% of divorced women and 76% of widows remarry. The difference between these two figures is probably due to the fact that divorces concern younger women than widowhood: the propensity to remarry generally declines with the age at marriage termination. In European societies, the frequencies are generally lower, most notably for widows, but the latter tend to be substantially older. New unions occur very soon in Togo, on average 1.8 years after divorce and 1.6 years after widowhood, which is much quicker that in France or its neighbors.

VIII. UNIONS AND MARRIAGE

For some years now in Western countries and for far longer in other societies, marriage has ceased to be the almost exclusive form of partner relationship. Different types of union are socially accepted and are just as likely to be adopted by couples who want to start living together. We shall not discuss here the potential diversity of these unions (with or without cohabitation, substitute for or prelude to marriage, etc.). We will simply examine the problems inherent in analyzing a situation in which marriage exists alongside other types of union. Solutions for these problems may be sought in two directions: aggregating all unions (marriages and other forms) and computing the total frequency of their formation or distinguishing between marriage and alternative forms and calculating the respective frequencies of either.

In the first case, as all forms of union are analyzed in the aggregate, their quantum is measured under the principles already developed for measuring the quantum of marriage alone. Accordingly, we classify unions by their rank in the life of male or female cohorts and we compile a first-union table similar to the first-marriage table, using probabilities by age (ratio of the number of first unions to the number of persons who have never lived in a partner relationship). To obtain a rough measure, we can calculate the following: (1) *first-union frequencies* (i.e., the ratio) at each age, of the number of first unions to the number of persons in the cohort, regardless of whether its members have already lived in a partner relationship or not (this is an equivalent of the unions in the first-union table); or (2) proportions of persons who have never lived in a partner relationship, (i.e., the ratio), at each age, of the number of these persons to the total size of the cohort, regardless of whether its members have already lived in a partner relationship or not (this is an equivalent of the frequency of life without spouse or companion in the first-union table).

The core difficulty lies not in the theoretical definition of these aggregates, but in collecting the statistics required to calculate them. By nature, unions without marriage go unrecorded in the administrative system and are not enumerated in the same way as events included in civil registration. Demographers usually resort to surveys of individuals to determine their marital life course and determine first-union formation retrospectively. The number of first unions at different ages and the population at risk are therefore calculated by aggregating individual data on survivors present at the time of the survey. As is always the case with retrospective observations, the statistics obtained are initially cleansed of the disruptive effects of mortality and migration; but we also need to make sure that the resulting selection in the composition of the respondent sample has not altered the representativeness of the sample relative to the population profile at the time the unions were formed.

A 1986 French survey was used to prepare a table of first-union formation for female cohorts born in 1938 and subsequent years. Table 25–4 in the Appendix shows the universality of partner relationships, whose frequency consistently approaches or exceeds 94%. By subtraction, the proportion of women never having lived with a husband or companion is thus under 6%. There is no perceptible change in the pattern over time. The first union occurs early, on average at age 22; the mean age declines slightly from one cohort to the next. These results deserve comparison with the frequency and tempo of first marriages in the same cohorts. In the oldest age group, the differences are modest: More than 92% of women born in 1938 to 1942 have married and more than 94% have lived in a partner relationship; the mean ages at first marriage and first union were 22.6 and 22.3 years, respectively. In an overwhelming majority of cases, therefore, first union and first marriage coincide. Fifteen cohorts later, the picture is different: Only 85% of women born in 1953 to 1957 have married and 97% have lived in a partner relationship; the mean age at first marriage is 22.7 years versus 21.5 years for the mean age at first union. Not all women who have lived in a partner relationship have married; first marriage does not coincide with first union, and often comes later. It is therefore well worth determining the form of the first union; we can always investigate later on the extent to which unions initially formed outside marriage were subsequently legalized.

We have already encountered a comparable problem in the analysis of terminations, where we distinguished between their two forms: divorce or death of a spouse. However, we eluded the issue, for example by stressing that divorce was rarely prevented by early death, as widowhood generally occurs at longer marriage durations, when the risk of divorce is low. We cannot apply this reasoning to the study of union forms, as the two alternative events (starting one's life in partnership with a marriage or a consensual union) can occur at the same ages. The difficulty is similar to the one encountered in measuring mortality by cause of death. Let us look at the options available:

1. We can always calculate the probabilities of a first partner relationship in marriage and the probabilities of a first consensual-union partnership for different ages. In a sufficiently short time interval, the probabilities will consist of the ratio of the number of events to the number of persons who have never yet lived in a partner relationship at the start of the interval. At the same age, the two probabilities are additive, given their common denominator. Their sum is a probability of first partner relationship, identical to the one computed earlier, for example in preparing Table 25–4 in the Appendix. We therefore obtain two series of probabilities and a sum-series; the difficulty is to combine the probabilities at successive ages.
2. The first solution consists in leaving the two series of probabilities together, preserving their additivity. We thus construct a double decrement table. From one age to the next, the number of persons never yet having lived in a partner relationship falls at the pace shown by the sum probability; in the interval, the number of new unions is distributed between marriage and consensual union in proportion to the

corresponding probabilities. The table therefore contains two branches—marriage and consensual union—and the size of each branch partly depends on the size of the other. For example, in the French female cohorts reported in Table 25–4 (Appendix), the distribution of first unions before age 36 between marriage and consensual union went from 87% to 13% for women born in 1938 to 1942 to 56.5% to 43.5% for women born in 1953 to 1957 (Thierry, 1993).

3. Can we move beyond that stage and supply a summary measure of the frequency of each form of union that is independent of the frequency of the other form? The answer is presumably negative, as it does not seem reasonable to postulate, for example, that the probability of a first consensual-union partnership, at age x, is not influenced by the number of marriages before age x. Absent independence, the combination at successive ages of the probabilities of a single series (first partner relationship in marriage or first partner relationship in consensual union) does not have the same significance as an intrinsic measure of the frequency of either type of union. It is preferable either to abandon this operation, or to try to comply with the independence condition.

Indeed, we saw that dependence resulted from the combination of two circumstances: in the group studied, the risk of either form of union differs between subgroups, and this heterogeneity is concurrent with a correlation between the two types of risk (either in subgroups where one is high and the other is low, and reciprocally; in subgroups where both are high; and in subgroups where both are low). We must therefore find a segmentation of the population into groups such that the heterogeneity or the correlation is eliminated in each. The same requirement exists when the aim is to obtain an intrinsic measure of surviving a given cause of death, *absent* the other causes (see Chapter 44). Such investigations are the focus of differential demography.

The same principles are at stake in the next step of the analysis when, once the union is formed, we examine its outcome. If the first union is not a marriage, it can be ended by two events: termination (generally through the partners' separation; far more rarely, under present circumstances in Europe, by the death of one of the partners) or legalization [i.e., the transformation of the consensual union into marriage]). The two outcomes are radically different. We cannot reasonably postulate, for example, that the probability of legalization computed for couples still united y years after their formation could have been equally valid for couples prematurely terminated through separation, *absent* such separations. This nonindependence leads us to calculate a double-decrement table for consensual-union terminations or seek to break up the cohort into more homogeneous subcohorts in an effort to create the statistical conditions for the independence of the separation and legalization risks.[6]

IX. NUPTIALITY AND FERTILITY

Nuptiality—in the broad sense, encompassing unions and marriage, formation and termination—has complex links with fertility. The simple notion that fertility is curbed by an abundance of never-marrieds and by late marriage applies only to populations with weak fertility control and strong social control over sex outside marriage. No Western country is in this situation any more. If the women who married earliest still have an above-average total number of children, it is no longer because of a greater *duration of exposure to risk* but because early marriage is associated with differences in sociocultural level, attitudes to the family, and contraceptive practices. Admittedly, some quasi-mechanical link may persist between age at marriage and age at childbearing. This is attested by the correlations between the two variables, measured between countries (early nuptiality and fertility in Central/Eastern Europe, late nuptiality and fertility in Western Europe) or across time (concomitant decrease in ages at marriage and at childbearing in Western Europe from 1950 to 1975, then reversal of the trend in the final quarter century). But the direction of the causalities is not established: Do people marry young *to have* children early or is childbearing delayed *because* marriage was postponed first?

This complexity is illustrated here with a few contemporary examples of the connections between unions, marriage, and fertility (Figure 25–2) and between divorce and the number of children.

1. The Place of Children in Marriage

In France, in the late 1960s, more than 60 of 100 children conceived by unmarried parents were born legitimate as a result of the fathers marrying the pregnant mothers. In addition, the other children, born illegitimate, did not all remain so: More than half of them were later legitimated by their parents' marriage. In all, more than 60% of the children conceived outside

[6] On all these issues, see the article by Xavier Thierry (1993) and the ensuing debate in *Population* (Courgeau and Lelièvre, 1994; Festy, 1994; Blayo 1995).

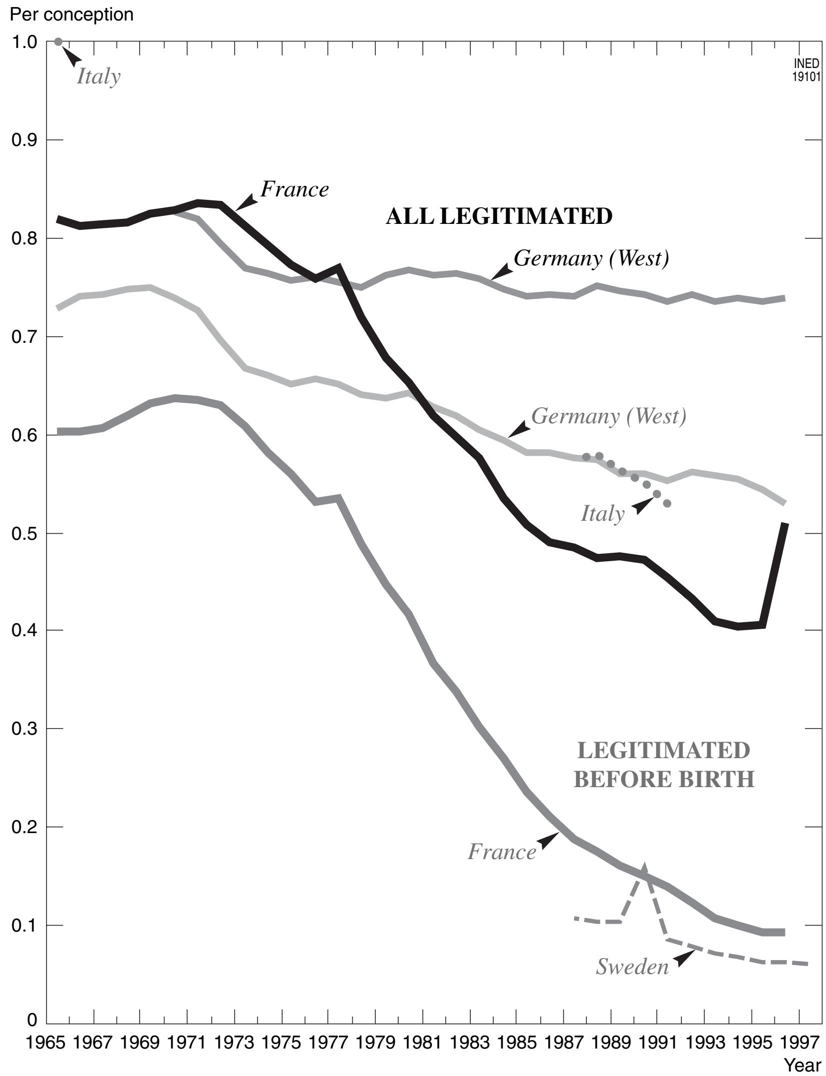


FIGURE 25–2 Proportion of legitimated conceptions outside marriage.

marriage were directly legitimated and 20% acquired legitimacy some time after birth through their parents' marriage—making a total of over 80%.

In the quarter century between the late 1960s and the mid 1990s, the number of children conceived outside marriage doubled, in France, from more than 150,000 to nearly 300,000 per year. But, in the same period, the number of unmarried women of child-bearing age doubled as well, from 3.8 million to 6.9 million. In other words, fertility outside marriage (measured at conception and not at childbirth) hardly varied in the interval. The increase in these births seems directly due to the fall in nuptiality, which swelled the never-married population without altering its fertility behavior.

By contrast, the fate of children conceived outside marriage has changed profoundly. Twenty-five years ago, nearly two thirds of them were legitimated by their parents' marriage even before childbirth; today, the proportion is less than one in 10. The desire to provide a robust legal framework to the birth of a child—a most eloquent indication of the importance assigned to marriage—has almost completely disappeared.

The future of children conceived *and born* outside marriage confirms this disaffection. Formerly, more than half of them would attend their parents' marriage and thus be legitimated (with a slight delay); today, the proportion has fallen to one third. In all, four in 10 children conceived outside marriage are raised in a marriage, one in 10 at childbirth and the three others a short while later (on average at 2 or 3 years old). The proportion in the late 1960s was more than eight in 10, of whom more than six at childbirth. The strong link that used to exist between the announcement of the birth of a child and the likelihood of the parents' marriage has been sharply weakened (Munoz-Pérez and Prioux, 2000).

This shift is, of course, largely due to the widespread acceptance of consensual union as an alternative partner relationship to marriage. An increasing number of children are conceived in these stable but nonformalized unions and, more and more often, the parents never marry. However,

> In unions of equal duration, unmarried couples are less fertile than married ones. From one cohort to the next, the relative fertility of cohabitants (in terms of conceptions) is [even] steadily declining, compared with married couples: if we set the conception rate for married couples to unity, the relative fertility of cohabitants, for a union of equal duration, is 0.75, 0.63, 0.50 and 0.39 for unions formed in ca. 1970, 1975, 1980, and 1985 respectively (Toulemon, 1994).

The proportion of couples who legitimate their pregnancies by marrying before the outcome has varied very widely from one country to another since the late 1960s, when it was high in almost every country, at around 60% or above. In northern Europe, including the United Kingdom, the proportion has fallen steeply—at least as much as in France—to 10% or even less; in Sweden, for example, only 4% of pregnancies outside marriage are legitimated by the parents before childbirth. In southern Europe, by contrast, the proportions have remained comparable to those of the 1960s; one example is Italy, where marriage still takes place in six pregnancies of 10. Between these two extremes, but close to the latter group, the countries of central Europe have recorded only a moderate decline in the nuptiality of pregnant women: in Germany (Western *Länder*) and Austria, more than half of all conceptions outside marriage are still legitimated before the completion of pregnancy. The case of Germany is all the more remarkable as, at the same time, the frequency of legitimations of born children has *increased* to well over 40%, versus barely 30% in the late 1970s. Today, three fourths of children conceived outside marriage in Germany are born in marriage or are legitimated shortly after birth. This has dramatically widened the gap with France in 25 years.

The dividing line in Europe separates the northwest from the southern and central parts, whereas the nuptiality trend isolated the south from the other countries. In fact, the differences in behavior toward children's status bear a greater resemblance to those that emerged during the decline in fertility in the past three decades than to the gaps that widened during the increase in age at marriage. Central Europe (in particular, Germany) and southern Europe (in particular, Italy) have recorded the sharpest falls in fertility, even as the decline in nuptiality was still moderate. In these countries, the attachment to marriage as the preferred framework for starting a family has blocked the rise in fertility outside of its traditional legitimate form. It has prevented the partial compensation of the decline in legitimate births by the rise in illegitimate ones—which has taken place more easily elsewhere, most notably in France.

2. On Divorce and Children

The relations between the presence of children in a couple and the risk of divorce are extremely complex and hard to unravel. The number of children is not the only determinant: the age of each, in particular the age of the youngest, also plays a role. Moreover, couples need time to have children and further time for the risk of divorce to show its effects. For France, the most eloquent results are the following (Toulemon, 1994):

- If we set the childless-couple divorce risk at unity for couples married in 1960 to 1979, the risk for a couple with child equals unity when the child reaches the age of four. From that age onward, the child's presence no longer makes a difference.
- Before the age of 4 years, the younger the child, the lesser the chances of divorce. The risk is 0.35 when the mother is still pregnant or when the child is in its first year, 0.55 when it is in its second year, and 0.75 thereafter.
- When there are several children, the risk remains below unity even when the youngest child reaches age 6, but the value is nearly 0.9 and the gap with childless couples is narrow. By contrast, the risk stays well below that level while the latest-born child is very young.
- After the birth of a child, the divorce risk therefore remains low, or even minimal, for a time. This *honeymoon* period is repeated with each pregnancy, but the number of children is not, in itself, a barrier to divorce: Once the youngest starts school, the difference with childless couples is very small. However, the childbearing period overlaps with the period of high separation risk consisting of the first years of marriage. The protection provided by the presence of very young children coincides with the period of highest divorce risk.

For all child ages combined, the presence of children lessens the risk of divorce in the first ten years of marriage by about 30%. The dissolution risk for unmarried couples with children is the same or slightly higher. By contrast, there is a large gap between the dissolution risks for the two types of union: it is five time as high for unmarried partners as for married couples, regardless of whether the couples in either category have children or not. The couple's status therefore seems to be a far more powerful differentiation factor than the presence and number of children.

We lack comparable studies for older periods that would confirm what more summary indicators suggest: the number of children no doubt provided more effective protection when the frequency of divorce was lower, i.e., in marriages contracted before 1960.

X. MALE AND FEMALE NUPTIALITY

We have measured first marriage and remarriage with reference to the male or female population. For example, we defined the proportion of men or women contracting a first marriage; the basic measure is a ratio between a number of marriages and a population at risk of marrying. But although the availability of men plays an essential role in the number of marriages of men of a given age, the availability of women is also an important factor. The *marriage market* is a convenient image to remind us that marriage results from an adjustment between male expectations and female expectations.

Initially, however, demographers' attention in this area focused not on nuptiality but on fertility. Births—like marriages—can be studied in relation to the male or female population (fathers or mothers): this yields a definition of male or female fertility (see Chapter 19). Now, it can be something of a shock to find that an identical number of births produces different rates, depending on whether it is measured against the male or female population of reproductive age. More specifically, it is astonishing to find population projections yielding different results depending on whether they are based on male fertility or female fertility rates, if both fertilities have been calculated from the same number of births. The initial priority is therefore to obtain fertility indices that avoid this inconsistency by taking both the male and female populations into account.

Transposed to nuptiality, the problem takes on a different form. It is not surprising that the male population marries more or less completely than the female population, as we can intuitively grasp that male nuptiality will be more intense than female nuptiality if there is a relative shortage of marriageable men, and less intense if there is a surplus. Nor is the frequency of these numeric imbalances surprising, given that men, on average, marry women from other cohorts—most likely of a different size than their own. In a population of, say, 100 men and 110 women, we expect that 90 first marriages between the two will leave 10 of 100 = 10% of never-married men and 20 of 110 = 18% of never-married women.

The surprise, therefore, lies not in the unequal frequency of female and male nuptiality, but in the size and nature of the inequality. Let us illustrate this with a concrete example borrowed from Louis Henry (1966). Human losses due to World War I created a deficit in a well-defined set of French male cohorts. This shortage predictably led to an intensification of male nuptiality in those cohorts and a decline in female nuptiality in the slightly more recent cohorts, given that men, on average, marry women who are younger than themselves. But:

1. In the cohorts most directly affected, the modulations in total marriage frequency remained well below what one would have expected using a simple computation as described above.
2. These modulations spread beyond the depleted cohorts (i.e., the phenomenon regained in scope part of what it lost in magnitude).
3. The repercussions are perceptible not only on the frequency of nuptiality, but also on age at marriage. In the depleted male cohorts, men married younger; in the female cohorts showing a *surplus*, women married later. The difference in age at marriage between men and women was therefore affected.

These three phenomena are closely linked. Men do not find their wives in a single, specific cohort, but in a range of cohorts. The range also varies with men's age at marriage. The reciprocal pattern applies to women. As a result, the deficit of one sex or the surplus of the other does not impact *one* cohort of the opposite sex, but a range of cohorts—hence a dilution of quantum effects. In addition, numeric imbalances initially alter the position and openness of the cohort range before causing an intensification or weakening of nuptiality: The excess women will wait longer to marry, and they will accept husbands with greater age differences.

From these observations, demographers have made various attempts to identify the male and female cohorts whose nuptiality would be influenced by numeric imbalances. The basic principle is to compare male and female cohort sizes allowing for the normal age gap between male and female marriage. The simplest measure is the ratio of the size of a male cohort to the size of a more recent female cohort, generally 2 to 5 years younger. This rough index is then improved in two ways. The first is to make allowance for the fact that matches occur not between two cohorts, but between two ranges of cohorts, whose definition may vary with successive ages at marriage. The second improvement reflects the fact that marriage is not only an adjustment between populations of marriageable males and females, but also a quest for a partner who displays (or does not display) certain characteristics (in particular, age characteristics); this may take time, as a fruitless search may lead an individual to stay in or leave the marriage market.[7]

These various attempts at measurement have one point in common. They seek to identify the cohorts whose nuptiality is liable to change because of the numeric inequality between men and women at risk of marriage. In the present state of research, demographers seem to have abandoned the notion of a single index that would take the male and female populations into account simultaneously and would therefore measure *pure two-sex nuptiality* (i.e., independent of any imbalances between men and women). We are not dealing here any longer (or we are not yet dealing) with the measurement of nuptiality; we are already involved in determining the factors that influence the quantum and tempo of marriage.

CONCLUSION

Demographers' interest in nuptiality, in all its forms, is due to the role played by the partner relationship in exposure to the risk of fertility. However, the measures described in this chapter do not aim to assess the quantitative incidence of a particular characteristic of marriage on childbearing. We have seen how to determine the quantum and tempo of nuptiality at its different stages (marriage or union formation; divorce, widowhood, or dissolution; remarriage or formation of a new union) without asking ourselves if the couples formed or dissolved were likely to have children.

Indeed, marriages and partner relationships are a major social phenomenon that deserves to be measured in itself—and is thus governed by the same principles of quantitative analysis as many other social phenomena. These principles are borrowed from the demographic methods for measuring mortality, fertility, and migration. The notion of risk (probabilities) and that of a table are shared with mortality; the breakdown of the marital sequence into stages has its equivalent, for fertility, in the succession of births in the same sibling group; the convergence between two populations at risk—which defines the marriage market—is closely related to the double influence of a population of origin and of a population of destination in the migration process. In fact, rather than an adaptation of tools specific to classic demographic phenomena, one should speak of mutual exchanges: the measurement of nuptiality has often produced innovations that have subsequently been taken up in the analysis of mortality, fertility, and migration.

This is because nuptiality, in fact, is not so much a marginal topic in a narrow demography concerned with the replacement of localized populations through natural and migration flows; rather, it lies at the heart of a broader demography, attentive to the characteristics of, and changes in, any population. The exchanges (transitions) between married and nonmarried populations are actually just a first example of this generalized approach, which also applies to transitions between economic activity and inactivity, employment and unemployment, good health and bad health, and so forth. This extension not only has the virtue of offering demography an infinite field of measurement and application. It also gives greater visibility to demography's ties with a diverse set of disciplines such as sociology, economics, and epidemiology.

[7] On all these points, see the article by McDonald (1995).

Acknowledgment

Translated from the French by Jonathan Mandelbaum.

References

Blayo Chantal, 1995 La condition d'homogénéité en analyse démographique et en analyse statistique des biographies, *Population*, vol. 50(6), p. 1501–1518.

Courgeau Daniel and Lelièvre Éva, 1994 Risques compétitifs et indépendance: cadre théorique d'une réflexion, *Population*, vol. 49(2), p. 482–489.

Festy Patrick, 1994. Quelques commentaires sur un problème fondamental de l'analyse démographique, *Population*, vol. 49(2), p. 490–498.

Henry Louis, 1966. Perturbations de la nuptialité résultant de la guerre 1914–1918, *Population*, vol. 21(2), p. 273–332.

Leridon Henri and Toulemon Laurent, 1997. *Démographie. Approche statistique et dynamique des populations*, Paris, Economica, 440 p.

Locoh Thérèse and Thiriat Marie-Paule, 1995. Divorce et remariage des femmes en Afrique de l'ouest, *Population*, vol. 50(1), p. 61–94.

McDonald Peter, 1995. L'équilibre numérique entre hommes et femmes et le marché matrimonial: le point sur la question, *Population*, vol. 50(6), p. 1579–1590.

Munoz-Pérez Francisco and Prioux France, 2000. Children born outside marriage and their parents: recognitions and legitimations since 1965, *Population: An English Selection*, vol. 12, p. 139–195.

ODE, 1998. Évolution récente de la nuptialité féminine en Europe occidentale, *L'Observatoire Démographique Européen vous informe*, no. 7, January 1998.

Sardon Jean-Paul, 1992. La primo-nuptialité féminine en Europe: éléments pour une typologie, *Population*, vol. 47(4), 855–892.

Sardon Jean-Paul, 1996. L'évolution du divorce en France, *Population*, vol. 51(3), p. 717–750.

Thierry Xavier, 1993. La nuptialité àl'épreuve de la cohabitation, *Population*, vol. 48(4), p. 939–974.

Toulemon Laurent, 1994. La place de l'enfant dans l'histoire des couples, *Population*, vol. 49(6), p. 1321–1346.

APPENDICES

TABLE 25–1 Proportion of First Marriages Contracted Before Age 50 In Female Cohorts

Cohort	Albania	Bulgaria	Hungary	Poland	East Germany	Romania	Czechoslovakia	Yugoslavia	Italy	Spain	France
1930	—	—	0.959	0.964	0.929	—	0.939	—	0.883	—	0.931
1931	0.928	—	0.962	0.953	0.933	—	0.951	—	0.892	—	0.920
1932	0.926	—	0.943	0.949	0.935	—	0.959	—	0.895	—	0.927
1933	0.918	—	0.948	0.954	0.939	—	0.961	—	0.903	—	0.923
1934	0.895	—	0.964	0.952	0.943	—	0.968	—	0.911	—	0.924
1935	0.868	—	0.950	0.946	0.944	—	0.965	—	0.912	—	0.920
1936	0.855	—	0.951	0.941	0.947	—	0.957	—	0.910	—	0.921
1937	0.849	—	0.954	0.937	0.946	—	0.958	0.884	0.912	—	0.924
1938	0.857	—	0.945	0.947	0.952	—	0.954	0.896	0.916	—	0.924
1939	0.874	—	0.946	0.966	0.934	—	0.955	0.913	0.913	—	0.923
1940	0.881	—	0.939	0.970	0.923	0.986	0.949	0.923	0.915	—	0.928
1941	0.901	—	0.938	0.953	0.928	0.979	0.942	0.921	0.913	—	0.924
1942	0.919	0.977	0.943	0.929	0.946	0.974	0.943	0.926	0.911	—	0.919
1943	0.883	0.964	0.955	0.912	0.957	0.976	0.949	0.925	0.920	—	0.926
1944	0.847	0.965	0.950	0.897	0.954	0.975	0.943	0.898	0.923	—	0.922
1945	0.845	0.971	0.944	0.895	0.891	0.964	0.941	0.882	0.934	—	0.920
1946	0.852	0.971	0.950	0.913	0.941	0.957	0.953	0.893	0.936	—	0.929
1947	0.857	0.975	0.949	0.931	0.971	0.959	0.945	0.901	0.923	—	0.916
1948	0.858	0.986	0.945	0.930	0.932	0.962	0.940	0.902	0.919	—	0.908
1949	0.854	0.987	0.947	0.925	0.953	0.962	0.945	0.901	0.917	—	0.907
1950	0.845	0.976	0.941	0.927	0.948	0.957	0.944	0.902	0.918	0.915	0.902
1951	0.841	0.965	0.952	0.928	0.930	0.953	0.945	0.899	0.910	0.898	0.892
1952	0.841	0.956	0.938	0.926	0.941	0.956	0.949	0.896	0.907	0.892	0.885
1953	0.852	0.952	0.938	0.922	0.939	0.965	0.947	0.893	0.935	0.867	0.877
1954	0.871	0.962	0.941	0.922	0.934	—	0.947	0.891	0.971	0.842	0.866
1955	0.900	0.972	0.943	0.928	0.936	—	0.944	0.883	0.933	0.831	0.854
1956	0.900	0.972	0.941	0.938	0.931	—	0.942	0.888	0.865	0.825	0.845
1957	0.867	0.968	0.943	0.944	0.929	—	0.940	0.900	0.853	0.827	0.833
1958	0.852	0.971	0.937	0.942	0.925	—	0.938	0.904	0.843	0.815	0.818
1959	0.855	0.958	0.934	0.934	0.927	—	0.937	0.903	0.824	0.798	0.800
1960	0.866	0.938	0.927	0.920	0.915	—	0.934	0.889	0.821	0.798	0.782
1961	0.872	0.934	0.927	0.904	0.910	—	0.934	0.883	—	—	0.757
1962	0.862	0.929	0.918	0.896	0.898	—	0.932	0.865	—	—	0.734
1963	0.855	0.924	0.898	0.892	0.885	—	0.928	0.840	—	—	—
1964	0.853	0.914	0.887	—	0.869	—	0.924	—	—	—	—
1965	—	0.898	0.872	—	—	—	0.922	—	—	—	—
1966	—	0.885	—	—	—	—	—	—	—	—	—

(continues)

TABLE 25–1 (*continued*) **Proportion of First Marriages Contracted Before Age 50 In Female Cohorts**

Cohort	Switzerland	The Netherlands	Belgium	West Germany	Austria	England-Wales	Finland	Sweden	Denmark	Norway
1930	0.861	0.942	—	0.930	0.893	0.958	—	0.919	0.942	0.905
1931	0.857	0.943	—	0.941	0.898	0.954	—	0.925	0.946	0.914
1932	0.855	0.943	—	0.935	0.897	0.961	—	0.919	0.946	0.915
1933	0.853	0.946	—	0.944	0.897	0.971	—	0.918	0.948	0.915
1934	0.847	0.955	0.945	0.942	0.893	0.974	—	0.913	0.950	0.918
1935	0.856	0.946	0.941	0.948	0.893	0.972	0.898	0.915	0.950	0.921
1936	0.861	0.950	0.934	0.945	0.890	0.970	0.891	0.911	0.950	0.919
1937	0.857	0.951	0.945	0.950	0.895	0.971	0.890	0.915	0.949	0.929
1938	0.854	0.952	0.942	0.945	0.894	0.972	0.903	0.909	0.955	0.934
1939	0.845	0.952	0.942	0.948	0.894	0.972	0.900	0.909	0.946	0.931
1940	0.838	0.950	0.949	0.942	0.877	0.970	0.907	0.906	0.950	0.905
1941	0.834	0.954	0.944	0.944	0.875	0.965	0.878	0.902	0.963	0.935
1942	0.825	0.947	0.939	0.936	0.871	0.973	0.894	0.899	0.941	0.909
1943	0.816	0.947	0.938	0.930	0.877	0.979	0.909	0.891	0.947	0917
1944	0.814	0.943	0.936	0.918	0.876	0.981	0.898	0.886	0.941	0.907
1945	0.811	0.943	0.933	0.929	0.871	0.971	0.908	0.874	0.938	0.911
1946	0.804	0.943	0.932	0.911	0.879	0.983	0.918	0.862	0.930	0.913
1947	0.805	0.940	0.925	0.914	0.884	0.980	0.904	0.850	0.924	0.920
1948	0.805	0.932	0.919	0.904	0.879	0.959	0.888	0.834	0.905	0.922
1949	0.803	0.926	0.918	0.902	0.868	0.954	0.880	0.820	0.906	0.919
1950	0.802	0.920	0.917	0.882	0.843	0.952	0.868	0.803	0.860	0.911
1951	0.794	0.911	0.915	0.879	0.837	0.953	0.848	0.791	0.852	0.907
1952	0.789	0.902	0.899	0.854	0.825	0.948	0.822	0.782	0.832	0.901
1953	0.786	0.894	0.898	0.850	0.826	0.936	0.813	0.774	0.824	0.884
1954	0.780	0.884	0.890	0.829	0.827	0.927	0.801	0.757	0.807	0.872
1955	0.776	0.873	0.880	0.838	0.824	0.922	0.801	0.746	0.789	0.862
1956	0.776	0.865	0.878	0.833	0.825	0.910	0.787	0.734	0.775	0.845
1957	0.771	0.850	0.869	0.826	0.820	0.895	0.794	0.721	0.756	0.824
1958	—	0.834	0.859	0.814	0.804	0.880	0.787	0.709	0.730	0.812
1959	—	0.820	0.852	0.802	0.798	0.857	0.772	—	0.710	0.796
1960	—	0.806	0.841	0.781	0.785	0.831	0.755	—	—	0.777
1961	—	0.795	0.823	—	0.773	0.812	—	—	—	0.748
1962	—	0.775	0.809	—	0.765	0.790	—	—	—	—
1963	—	—	0.782	—	0.746	—	—	—	—	—

Source: Sardon, Jean-Paul, 1992. La primo-nuptialité féminine en Europe: éléments pour une typologie, *Population*, vol. 47(4), 855–892.

TABLE 25–2 Mean Age at First Marriage in Female Cohorts, in Years

Cohort	Albania	Bulgaria	Hungary	Poland	East Germany	Romania	Czechoslovakia	Yugoslavia	Italy	Spain	France
1930	—	—	—	23.25	—	—	21.99	—	—	—	—
1931	20.62	—	—	23.16	—	—	21.82	—	25.07	—	23.01
1932	20.65	—	21.73	23.06	—	—	21.72	—	24.98	—	22.99
1933	20.66	—	21.59	22.92	—	—	21.63	—	24.89	—	22.98
1934	20.70	—	21.47	22.80	22.70	—	21.56	—	24.76	—	22.94
1935	20.80	—	21.42	22.70	22.61	—	21.58	—	24.62	—	22.93
1936	20.93	—	21.38	22.60	22.48	—	21.60	—	24.53	—	22.89
1937	21.04	—	21.29	22.52	22.43	—	21.55	22.93	24.43	—	22.83
1938	21.10	—	21.25	22.41	22.39	—	21.50	22.84	24.31	—	22.78
1939	21.09	—	21.22	22.26	22.20	—	21.43	22.73	24.24	—	22.71
1940	21.06	—	21.20	22.17	22.02	21.55	21.37	22.61	24.17	—	22.60
1941	21.00	—	21.26	22.17	21.95	21.51	21.41	22.53	24.09	—	22.53
1942	20.93	21.10	21.26	22.26	21.82	21.50	21.42	22.44	23.96	—	22.44
1943	21.00	21.17	21.29	22.42	21.70	21.50	21.44	22.38	23.83	—	22.37
1944	21.15	21.23	21.27	22.59	21.62	21.54	21.47	22.47	23.72	—	22.33
1945	21.19	21.27	21.27	22.70	21.65	21.62	21.55	22.55	23.63	—	22.32
1946	21.25	21.29	21.34	22.70	21.57	21.72	21.63	22.49	23.49	—	22.33
1947	21.31	21.30	21.25	22.65	21.48	21.82	21.69	22.44	23.45	—	22.41
1948	21.38	21.27	21.23	22.63	21.56	21.90	21.71	22.41	23.45	—	22.47
1949	21.46	21.23	21.16	22.64	21.59	21.99	21.73	22.40	23.44	—	22.49
1950	21.60	21.22	21.20	22.63	21.64	22.09	21.74	22.42	23.37	23.74	22.53
1951	21.74	21.25	21.24	22.64	21.77	22.16	21.78	22.43	23.34	23.66	22.54
1952	21.88	21.28	21.29	22.66	21.71	22.20	21.77	22.44	23.33	23.63	22.56
1953	21.97	21.28	21.33	22.67	21.76	22.19	21.75	22.47	23.33	23.58	22.59
1954	22.03	21.26	21.27	22.68	21.82	—	21.71	22.51	23.30	23.52	22.63
1955	22.07	21.23	21.24	22.66	21.84	—	21.67	22.58	23.31	23.48	22.68
1956	22.12	21.18	21.20	22.63	21.92	—	21.63	22.59	23.46	23.44	22.77
1957	22.26	21.15	21.07	22.59	21.98	—	21.61	22.53	23.55	23.38	22.90
1958	22.38	21.13	21.04	22.56	22.10	—	21.62	22.52	23.71	23.52	23.08
1959	22.42	21.11	21.01	22.55	22.22	—	21.63	22.56	23.92	23.61	23.26
1960	22.41	21.11	21.03	22.55	22.37	—	21.66	22.64	24.05	23.70	23.48
1961	22.40	21.13	21.09	22.56	22.54	—	21.68	22.69	—	—	23.69
1962	22.42	21.13	21.11	22.54	22.73	—	21.69	22.78	—	—	23.93
1963	22.45	21.12	21.16	22.51	22.88	—	21.71	22.94	—	—	—
1964	22.46	21.14	21.24	—	22.99	—	21.72	—	—	—	—
1965	—	21.18	21.29	—	—	—	21.73	—	—	—	—
1966	—	21.22	—	—	—	—	—	—	—	—	—

(continues)

TABLE 25–2 (*continued*) **Mean Age at First Marriage in Female Cohorts, in Years**

Cohort	Switzerland	The Netherlands	Belgium	West Germany	Austria	England-Wales	Finland	Sweden	Denmark	Norway
1930	25.11	24.61	—	23.97	24.26	23.19	—	23.77	—	23.88
1931	25.05	24.52	—	23.90	24.10	23.09	—	23.67	—	23.71
1932	24.95	24.33	—	23.82	24.02	23.01	—	23.58	—	23.49
1933	24.81	24.23	—	23.74	23.89	22.92	—	23.51	—	23.32
1934	24.65	24.09	22.82	23.65	23.78	22.82	—	23.45	—	23.14
1935	24.55	23.94	22.76	23.56	23.68	22.73	23.35	23.39	22.52	22.98
1936	24.43	23.88	22.68	23.47	23.54	22.63	23.30	23.38	22.52	22.92
1937	24.32	23.77	22.58	23.38	23.39	22.53	23.30	23.30	22.44	22.81
1938	24.23	23.69	22.51	23.29	23.26	22.45	23.31	23.26	22.47	22.78
1939	24.15	23.59	22.43	23.14	23.10	22.35	23.24	23.30	24.40	22.72
1940	24.06	23.43	22.36	22.97	23.00	22.29	23.20	23.42	22.46	22.63
1941	24.04	23.24	22.23	22.87	22.95	22.29	23.15	23.40	22.39	22.58
1942	24.03	23.11	22.20	22.78	22.84	22.28	23.15	23.51	22.46	22.66
1943	24.01	23.01	22.11	22.62	22.69	22.26	23.04	23.51	22.43	22.67
1944	23.99	22.88	22.12	22.38	22.52	22.23	23.00	23.58	22.41	22.75
1945	23.98	22.71	22.06	22.25	22.37	22.23	22.95	23.77	22.45	22.69
1946	23.94	22.72	22.06	22.14	22.36	22.23	22.96	23.92	22.52	22.75
1947	23.96	22.63	22.02	22.08	22.17	22.23	22.95	24.31	22.60	22.61
1948	24.01	22.61	22.03	22.13	22.18	22.28	23.03	24.70	22.79	22.61
1949	24.07	22.59	22.02	22.18	22.24	22.28	23.07	25.20	23.01	22.65
1950	24.16	22.60	22.03	22.27	22.31	22.26	23.20	25.64	23.39	22.66
1951	24.36	22.57	22.00	22.34	22.38	22.23	23.29	25.99	23.70	22.74
1952	24.56	22.62	22.04	22.50	22.57	22.25	23.47	26.29	24.10	22.87
1953	24.76	22.72	22.03	22.68	22.62	22.33	23.72	26.53	24.35	23.02
1954	24.98	22.84	22.07	22.90	22.75	22.43	23.88	26.76	24.66	23.12
1955	25.17	23.00	22.11	23.09	22.86	22.54	24.02	26.96	24.89	23.35
1956	25.32	23.30	22.19	23.28	23.02	22.68	24.17	27.11	25.10	23.52
1957	25.49	23.55	22.27	23.49	23.20	22.89	24.21	27.23	25.41	23.80
1958	—	23.83	22.41	23.72	23.36	23.10	24.27	27.36	25.70	24.03
1959	—	24.14	22.51	23.92	23.47	23.27	24.37	—	25.91	24.25
1960	—	24.40	22.64	24.12	23.58	23.42	24.51	—	—	24.50
1961	—	24.62	22.81	—	23.69	23.57	—	—	—	24.75
1962	—	24.83	22.96	—	23.76	23.75	—	—	—	—
1963	—	—	23.11	—	23.87	—	—	—	—	—

Source: Sardon, Jean-Paul, 1992. La primo-nuptialité féminine en Europe: éléments pour une typologie, *Population*, vol. 47(4), 855–892.

TABLE 25–3 Divorces per 10,000 Marriages by Duration in Marriage Cohorts (France)

Marriage duration, x	Marriage cohort												
	1946	**1947**	**1948**	**1949**	**1950**	**1951**	**1952**	**1953**	**1954**	**1955**	**1956**	**1957**	**1958**
0	*2.0*	*2.0*	*2.0*	*2.0*	*2.0*	*2.0*	2.1	2.0	1.9	2.7	2.9	2.0	2.2
1	*16.5*	*16.5*	*16.0*	*15.5*	*15.1*	15.1	14.2	13.5	14.1	13.7	15.2	15.4	14.4
2	*34.0*	*33.0*	*32.5*	*32.0*	31.7	32.5	31.4	30.9	29.2	31.0	35.1	34.0	34.1
3	*49.0*	*48.0*	*47.5*	47.0	44.6	44.3	46.8	45.1	43.8	46.3	44.6	44.5	45.4
4	*58.0*	*57.0*	56.2	51.8	50.4	53.3	52.5	49.1	50.7	47.8	53.9	52.3	51.9
5	*59.0*	58.5	55.1	53.4	56.4	57.3	54.2	54.7	52.3	52.2	55.8	56.0	52.6
6	54.7	55.8	55.2	56.1	59.3	56.0	58.0	53.3	52.3	55.2	53.2	54.1	59.3
7	50.0	52.9	57.4	56.4	56.8	54.3	54.0	50.9	51.3	51.6	52.2	56.1	57.7
8	48.6	54.0	54.7	55.5	55.9	52.2	51.5	52.0	50.2	49.0	56.3	55.4	55.2
9	47.5	51.1	51.7	52.4	48.0	48.7	49.9	47.4	48.5	51.8	53.8	52.6	54.4
10	44.9	48.2	50.4	48.3	47.0	50.0	47.5	46.3	47.6	52.0	52.7	51.4	48.6
11	40.8	44.0	45.7	47.2	44.0	45.0	44.9	47.3	48.2	50.2	51.4	44.2	46.9
12	37.9	39.2	42.6	41.8	40.0	40.9	45.8	44.7	45.8	46.4	42.0	45.4	45.4
13	33.4	37.2	39.4	38.2	39.3	40.8	42.5	44.3	42.9	41.9	40.3	42.3	53.4
14	32.1	35.1	37.9	35.3	38.5	40.9	41.5	39.6	38.4	38.7	41.4	45.3	48.2
15	30.0	31.5	30.9	35.8	35.0	38.2	40.1	36.3	35.5	37.2	46.4	43.3	43.0
16	27.5	28.5	34.3	34.9	38.0	34.3	34.3	34.8	35.6	40.0	41.6	41.0	49.0
17	27.4	28.7	32.8	34.3	35.4	31.9	31.4	35.3	39.3	37.7	41.0	43.0	44.0
18	25.3	27.4	30.1	31.5	28.8	28.6	30.9	34.4	34.3	35.0	42.0	39.0	41.0
19	25.6	28.5	30.9	28.7	28.1	30.2	33.1	32.1	34.0	39.0	41.0	39.0	42.0
20	22.7	25.5	25.2	26.5	28.5	31.1	29.6	31.0	34.0	37.0	37.0	42.0	46.0
21	21.4	21.5	24.1	25.8	28.3	26.9	30.0	33.0	34.0	33.0	38.0	40.0	*42.5*
22	19.1	20.9	23.6	27.0	25.2	25.0	29.0	31.0	30.0	35.0	37.0	*38.2*	*39.3*
23	17.9	19.3	22.9	23.0	23.0	26.0	27.0	27.0	31.0	33.0	*33.9*	*34.8*	35.7
24	15.8	19.4	21.2	19.0	25.0	25.0	24.0	27.0	30.0	*30.9*	*31.8*	32.7	36.6
25	17.4	17.5	18.0	21.0	22.0	19.0	27.0	26.0	*27.4*	*28.8*	30.2	33.2	35.4
26	13.9	14.0	16.0	18.0	19.0	22.0	25.0	*26.0*	*27.1*	28.1	30.5	33.2	34.2
27	12.0	14.0	15.0	16.0	21.0	20.0	*21.3*	*22.7*	24.0	25.5	28.0	29.7	31.5
28	12.0	13.0	14.0	17.0	20.0	*20.3*	*20.5*	20.8	22.4	25.7	26.0	28.2	29.2
29	9.0	11.0	15.0	16.0	*17.0*	*18.0*	19.0	19.7	21.2	23.2	24.3	24.0	24.5
30	10.0	12.0	14.0	*14.4*	*14.8*	*15.2*	17.2	19.3	20.4	20.7	22.9	22.3	22.3
31	11.0	12.0	*12.4*	*12.8*	*13.3*	15.0	15.8	17.2	16.8	18.2	19.3	20.2	21.1
32	9.0	*9.7*	*10.5*	*11.2*	12.7	12.9	14.6	15.6	15.9	17.2	17.6	17.2	17.4
33	*9.1*	*9.3*	*9.4*	10.7	12.6	12.1	13.5	14.4	13.3	14.3	14.2	16.8	16.4
34	*8.2*	*7.8*	*8.8*	9.1	10.4	11.1	12.0	11.7	13.0	13.6	14.1	14.7	15.5
35	*7.2*	*7.7*	*8.0*	*7.4*	*8.7*	*9.1*	*9.9*	*11.0*	*11.7*	*12.4*	12.1	13.7	*13.7*
36	*6.5*	*6.7*	*6.2*	*7.3*	*7.6*	*8.4*	*9.2*	*9.8*	*10.4*	9.8	11.0	*11.0*	*11.0*
37	*6.2*	*5.7*	*6.7*	*7.0*	*7.6*	*8.4*	*9.0*	*9.5*	9.5	9.3	*9.3*	*9.3*	*9.3*
38	*4.7*	*5.4*	*5.7*	*6.2*	*6.9*	*7.4*	*7.8*	*7.7*	8.1	*8.1*	*8.1*	*8.1*	*8.1*
39	*4.8*	*5.1*	*5.5*	*6.1*	*6.5*	*6.9*	6.0	*7.5*	*7.5*	*7.5*	*7.5*	*7.5*	*7.5*
40+	*23.9*	*26.5*	*21.6*	*22.4*	*26.1*	*32.3*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*
Total (per 10,000)	1005.9	1061.2	1107.2	1122.1	1150.5	1168.6	1210.2	1218.1	1239.5	1286.6	1351.6	1369.0	1422.0
Mean duration, x	14.08	14.33	14.53	14.84	15.27	15.51	15.87	16.21	16.40	16.49	16.47	16.65	16.70

(continues)

TABLE 25–3 (*continued*) **Divorces per 10,000 Marriages by Duration in Marriage Cohorts (France)**

Marriage duration, x	Marriage cohort											
	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
0	2.4	3.2	2.2	2.4	3.1	2.3	2.2	2.1	1.9	2.4	2.1	2.0
1	16.2	15.8	17.0	16.1	15.7	20.2	20.2	20.2	19.7	18.3	19.5	21.8
2	32.4	35.2	35.2	37.6	42.2	44.6	49.1	49.0	49.3	48.7	57.1	58.5
3	44.9	47.8	53.0	54.6	59.9	63.0	62.7	66.2	70.7	81.5	80.9	88.0
4	51.6	58.4	58.6	61.9	66.5	68.3	75.5	80.2	95.2	90.7	99.0	111.0
5	56.4	60.0	65.2	63.7	64.9	72.9	79.3	98.3	96.2	101.0	108.0	116.0
6	58.2	63.2	61.5	64.4	68.7	74.2	91.0	94.3	100.0	113.0	114.0	118.0
7	56.3	58.3	58.5	60.7	69.4	84.1	91.0	93.0	105.0	107.0	111.0	128.0
8	55.9	54.8	57.2	64.3	78.5	83.7	90.0	100.0	102.0	103.0	125.0	143.0
9	50.7	50.3	55.4	71.1	75.3	80.0	93.0	94.0	92.0	112.0	122.0	*122.5*
10	49.8	50.9	61.1	65.9	71.0	82.0	89.0	89.0	98.0	112.0	*113.9*	*115.9*
11	47.7	59.6	58.3	59.0	75.0	80.0	81.0	93.0	102.0	*104.9*	*107.8*	110.6
12	55.3	55.7	60.0	65.0	70.0	75.0	82.0	96.0	*98.4*	*100.8*	103.1	111.2
13	50.6	52.0	61.0	61.0	66.0	74.0	89.0	*91.6*	*94.2*	96.8	105.5	112.9
14	48.0	57.0	56.0	57.0	68.0	74.0	*77.4*	*80.8*	84.2	98.8	103.9	111.6
15	48.0	52.0	53.0	62.0	72.0	*74.8*	*77.6*	80.4	90.5	98.2	105.2	109.8
16	49.0	49.0	54.0	64.0	*68.2*	*72.3*	76.5	82.1	90.2	95.7	100.1	104.9
17	42.0	51.0	58.0	*61.5*	*65.0*	68.5	76.5	81.6	91.1	90.8	96.1	104.7
18	47.0	55.0	*57.2*	*59.5*	61.7	72.9	76.8	83.6	88.1	93.9	89.5	93.6
19	46.0	*48.5*	*51.0*	53.4	63.9	69.6	77.8	80.1	83.6	84.0	85.4	90.0
20	*47.6*	*49.2*	50.7	55.9	64.6	70.7	74.8	75.5	77.6	78.2	81.3	85.6
21	*45.0*	47.5	51.3	57.2	61.3	67.0	70.9	70.6	72.5	75.7	73.6	82.0
22	40.5	48.2	49.7	56.2	59.3	62.6	66.1	65.4	65.2	68.5	76.1	76.6
23	41.5	46.0	49.0	53.5	55.7	66.8	61.1	59.5	60.6	63.5	69.4	69.4
24	40.9	43.7	47.3	52.4	49.5	54.7	56.0	57.5	59.6	62.5	*62.5*	62.5
25	37.5	42.7	43.3	49.0	47.1	49.0	52.0	52.0	54.7	*54.7*	54.7	*54.7*
26	36.5	37.9	40.6	45.4	42.5	46.0	48.4	48.4	*48.4*	*48.4*	48.4	*48.4*
27	33.7	34.4	35.0	37.8	41.5	47.9	42.6	*42.6*	*42.6*	*42.6*	42.6	*42.6*
28	28.4	32.0	33.0	32.6	36.7	38.0	*38.0*	*38.0*	*38.0*	*38.0*	38.0	*38.0*
29	27.3	29.4	29.7	30.1	34.1	*34.1*	*34.1*	*34.1*	*34.1*	*34.1*	34.1	*34.1*
30	23.5	26.5	28.0	28.2	*28.2*	*28.2*	*28.2*	*28.2*	*28.2*	*28.2*	*28.2*	*28.2*
31	20.8	32.6	25.3	*25.3*	*25.3*	*25.3*	*25.3*	*25.3*	*25.3*	*25.3*	*25.3*	*25.3*
32	20.9	20.8	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*	*20.8*
33	17.4	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*	*17.4*
34	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*	*15.5*
35	*13.7*	*13.7*	*13.7*	*13.7*	*13.7*	13.7	*13.7*	*13.7*	*13.7*	*13.7*	*13.7*	*13.7*
36	*11.0*	*11.0*	*11.0*	*11.0*	*11.0*	11.0	*11.0*	*11.0*	*11.0*	*11.0*	*11.0*	*11.0*
37	*9.3*	*9.3*	*9.3*	*9.3*	*9.3*	9.3	*9.3*	*9.3*	*9.3*	*9.3*	*9.3*	*9.3*
38	*8.1*	*8.1*	*8.1*	*8.1*	*8.1*	8.1	*8.1*	*8.1*	*8.1*	*8.1*	*8.1*	*8.1*
39	*7.5*	*7.5*	*7.5*	*7.5*	*7.5*	7.5	*7.5*	*7.5*	*7.5*	*7.5*	*7.5*	*7.5*
40+	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*	*36.0*
Total (per 10,000)	1470.9	1587.2	1655.7	1767.9	1910.3	2066.2	2194.6	2292.0	2398.5	2512.6	2622.7	2760.7
Mean duration, x	16.85	16.89	16.73	16.67	16.38	16.22	15.91	15.67	15.51	15.35	15.15	14.93

Note: Italics indicate estimates.
Source: Sardon, Jean-Paul, 1996. L'évolution du divorce en France, *Population*, vol. 51(3), p. 717–750.

TABLE 25–4 Proportion (%) of Women Having Lived in at Least One Partner Relationship

Exact age, x	Cohort groups					
	1938–1942	1943–1947	1948–1952	1953–1957	1958–1962	1963–1967
18	4.7	6.9	4.8	11.0	9.3	6.3
20	23.3	25.4	26.8	36.2	33.4	27.0
22	47.9	57.0	55.8	61.7	57.5	—
24	73.0	75.9	74.2	80.2	75.7	—
26	85.3	85.0	83.0	85.8	—	—
28	89.6	90.8	86.8	88.8	—	—
30	90.7	93.8	89.7	91.9	—	—
32	92.1	94.4	92.0	—	—	—
34	93.3	95.4	92.4	—	—	—
36	94.4	96.3	93.6	97.0[a]	—	—
Mean age	22.3	22.0	22.1	21.5[b]	—	—
N	329	409	587	590	550	530

[a] Estimated by extrapolation.
[b] Mean value computed on assumption that tempo of first unions after age 30 years was identical to that of 1948–1952 cohorts.
Source: Thierry, Xavier, 1993. La nuptialité à l'épreuve de la cohabitation, *Population*, vol. 48(4), p. 939–974.

CHAPTER

26

Event-History Analysis of Nuptiality

An Application to Africa

PHILIPPE ANTOINE

Institut de récherche sur le développement (IRD), Dakar, Senegal

The study of nuptiality is relatively complex, as it concerns renewable events and events not always recorded statistically. Demographic studies on nuptiality rely either on cohort data, often from civil registration records (Chapter 25), or on cross-sectional observations, which allow only a partial and limited description of the phenomenon. Neither data source conveys the complexity of the marriage life cycle, a process that fits into an individual life course, itself shaped by the broad trends in the society to which the person belongs. The approach to nuptiality based on event-history methods (questionnaires and analytical procedures) yields an abundant harvest of results. Thanks to these methods, we can not only describe the evolution of a phenomenon, but also attempt to explain the change and isolate the demographic, economic, social, and even legislative factors involved (see Chapter 23). After addressing the inherent difficulties of tracking the marriage life cycle, we will discuss the collection and analysis of event-history data on nuptiality with the aid of African examples—in particular, a study carried out in Dakar (Senegal).

I. COMPLEXITY OF MARITAL STATUSES: FROM CONSENSUAL UNIONS TO POLYGAMY

The concepts and definitions of marriage used in the classical analyses of nuptiality focus on marriage recorded by administrative authorities (civil registration) or religious authorities (parish registers).[1] However, the data collected from these sources prove inadequate to describe the social reality of unions. Couple formation goes through several distinct phases: In Western countries, for example, an increasing number of couples begin with a consensual union; a proportion of them later officialize their partnerships, especially after the birth of children or for tax reasons. In France, prenuptial cohabitation rose sharply in a quarter century. Between 1960 and 1985, the proportion of unions starting with prenuptial cohabitation climbed from 10% to 43% for marriages before age 25 and from approximately 10% to 59% for those celebrated at age 25 or thereabouts[2] (Léridon and Villeneuve-Gokalp, 1994). These couples may eventually separate; later on, either spouse may enter into a new union, official or consensual, with another partner or remain single. Among the young, the variety of statuses is increasing—for example, cohabitation in one's parents' home, or intermittent cohabitation (Battagliola *et al.*, 1991). Only a survey of partner-relationship histories can describe the transitions between statuses. To illustrate the complexity of the marital life cycle, I will draw on African examples, which—in addition to the patterns mentioned above—frequently involve polygamy.

[1] In most European countries, religious marriage is on a steep downtrend (Dittgen, 1995); by contrast, consensual union is increasing sharply, and civil registration records now provide only a very incomplete picture of changes in nuptiality.

[2] These results were obtained from a specific cohort survey.

1. Stages in a Union

In Africa, marital life courses are made more complex by the fact that most marriages are not marked by a single event; they often require several compulsory stages before the spouses actually cohabit (van de Walle, 1968). In Bamako (Mali), for example, the first stage is typically engagement, which signifies that the parents of the wife-to-be accept the suitor as their daughter's future husband. We then need to distinguish between three types of marriage: religious, civil, and traditional. The three ceremonies are distinct and can be celebrated the same day or on different days, provided the religious marriage always precedes the traditional ceremony. Although the code stipulates that the civil marriage must be celebrated before the religious marriage, in practice the sequence matters little. The three marriages are rarely celebrated together, and one therefore observes lags between the ceremonies.[3] Cohabitation is sometimes allowed by the bride's family, which is empowered to lend its daughter as soon as the religious marriage has taken place. For a couple completing these stages at different dates, which will be regarded as the date of the marriage? Clearly, for the most religious, the most traditionalist or the most emancipated persons, the three celebrations will have different meanings (Djiré, 1993). Improvements in the distinction between legal, traditional, and religious marriage and in dating the start of cohabitation have been provided in some countries by the surveys executed under the World Fertility Survey (WFS) program. One example is the Côte d'Ivoire survey, in which the flexibility of questions on the type of union has made it possible to distinguish between the different forms of cohabitation (Brandon, 1991). In a recent survey on urban integration in Bamako, the authors have sought to pinpoint the marriage stages as precisely as possible (Ouédraogo and Piché, 1995).

The need to distinguish between union forms is, in itself, an indicator of the changes in marital behavior driven by the emergence of new practices that depart from traditional rules. These changes also require a better definition of the concept of divorce (formal dissolution of a formal marriage) or separation (informal dissolution of a formal or informal marriage). In Africa, the dissolution of unions is just as hard to track as their formation. Separation is not always final: In a customary marriage, families often attempt a reconciliation (Pison, 1986). Families may also intervene while divorce proceedings are under way. Typically, there is a rather long period between the end of cohabitation and the divorce ruling.[4] The question then arises: What should be regarded as the date of the divorce (Smith *et al.*, 1984)?

2. Changing Types of Union

A union changes over time: A consensual union may subsequently be confirmed by a civil or religious ceremony. In a recent study in Antananarivo, the capital of Madagascar, we have tabulated—with precise dates—the spectrum of union celebrations all the way from the absence of ceremony (consensual union) to marriage marked by a near-concomitant celebration of civil, customary, and religious marriage (Antoine *et al.*, 2000). Among women, we found a sharp increase in the proportion of nonformalized marriages (consensual union or *tokantrano maso*) or weakly formalized marriages (customary marriage or *vodiondry*) from one cohort to the next (Table 26–1). Purely religious marriage does not exist, and the proportion of unions starting with a religious ceremony has been falling steadily. In the aggregate, unions formalized at the town hall or (and) in a religious ceremony have fallen from 67.5% for the oldest cohort of our survey to 41.4% for the youngest. Thanks to event-history analysis, we can determine the outcome of unions at the end of a set period. We have specified an interval of ten years: by the end of the period about half of the informal unions had been formalized, while just under 20% had ended in separation (Table 26–2).

Marital mobility, in particular of men, poses another classification problem among ever-marrieds (Pison, 1986; Donadjé, 1991). In nuptiality studies, individuals

TABLE 26–1 Distribution (%) of Women by Type of Union and Cohort at Start of First Union in Antananarivo

Type of union	1943–1952 Cohort	1953–1962 Cohort	1963–1972 Cohort
Consensual union	10.9	14.3	22.5
Customary	21.6	27.1	36.1
Civil	4.4	3.3	3.2
Customary and civil	24.9	28.3	20.0
Civil and religious	3.5	2.8	0.0
Customary, civil, and religious	34.7	24.3	18.2

Source: Antoine Philippe, Bocquier Philippe, Razafindratsima Nicolas and Roubaud François (eds.), 2000. *Biographies de trois générations dans l'agglomération d'Antananarivo. Premiers résultats de l'enquête BIOMAD-98*. Paris, CEPED, 140 p. (Documents et manuels, no.11).

[3] In fact, the three celebrations are not always performed.

[4] In some cases, the separation is even harder to document because the spouses did not cohabit.

TABLE 26–2 Distribution (%) of Women by type of Union and Cohort at Start of First Union and Outcome of Union 10 Years Later for Two Cohorts in Antananarivo

Type of union	1943–1952 Cohort			1953–1962 Cohort		
	Same type	Other type	Separation	Same type	Other type	Separation
Consensual union	31	51	18	34	47	19
Customary	29	53	18	42	40	18
Customary and civil	64	26	10	62	21	17
Civil and religious	80	7	13	100	—	—
Customary, civil, and religious	92	—	8	94	—	6

Source: Antoine Philippe, Bocquier Philippe, Razafindratsima Nicolas and Roubaud François (eds.), 2000. *Biographies de trois générations dans l'agglomération d'Antananarivo. Premiers résultats de l'enquête BIOMAD-98*. Paris, CEPED, 140 p. (Documents et manuels, no.11).

TABLE 26–3A Period Marital Status by Previous Event (Men, Dakar)

	Period marital status			
Previous event	Monogamous	Polygamous	Divorced	Widower
Marriage	517	222	0	0
Divorce	50	7	76	0
Widowhood	7	4	0	8

Source: IFAN-ORSTOM 1989 survey.

are usually classified as never married, married (monogamous or polygamous), divorced, and widowed. In fact, however, many changes can occur in the individual's life cycle after a first marriage: A widower or divorcee at a given date may formerly have been polygamous. Polygamy is largely measured from cross-sectional indices such as the proportion of polygamous men among married men, the proportion of women in polygamous union, the mean number of wives per married man, or the mean number of women per polygamous man (Donadjé, 1992; Klissou, 1995).

However, the ever-marrieds in each of the standard categories display different past nuptial characteristics and have not necessarily contracted the same number of marriages before ending up in the same marital category: Some bigamists, for example, are in their second marriage, whereas others have celebrated more than two marriages but have kept only two wives at the study date. One should take greater account of the distinction between marital events and period marital status. Different kinds of events can lead to the same marital status. Donadjé (1992), who has collected marital event histories, distinguishes between stable monogamous men who have been in a single union, unstable monogamous men who have been in more than one union, stable polygamous men who, after the second wife, have never returned to monogamy, and unstable polygamous men who have become monogamous again at least once in their lives. However, this typology is far from exhausting all the possible combinations if we include divorcee status. A few simple cross-classifications (Tables 26–3A through 26–3C) based on the Dakar survey (see subsequent sections) reveal this complexity. For example, we can see that a large share of divorcees (43%) are men who have several wives and are therefore still married after divorcing (Table 26–3A). Likewise, 30% of marriages entail a transition to polygamy. Although the shift to monogamy is the only marital transition for a bachelor, monogamous men move to polygamous status more often than to divorced status, whereas polygamous men may return to monogamy (Table 26–3B). On average, monogamous men experienced 1.32 marriages, 0.25 divorces, and an infinitesimal proportion of widowhoods at the time of the survey. Logically enough, polygamous men experienced more marriages (2.5 on average) and relatively fewer divorces (0.2) (Table 26–3C).

Most demographic studies make no allowance for the complexity of the sequence of these marital events.[5]

[5] Demographic and Health Surveys (DHSs), in particular in Togo, occasionally fill part of the gap by gathering information on birth history and marriage history—without, however, the possibility of linking these to the person's occupational or residential histories. Some event-history analyses have been conducted on this type of data. Thiriat (1998) shows, in Togo, that urbanization and schooling are the main factors in the increase in age at marriage and that there is a widening dissociation between sexuality and nuptiality.

TABLE 26–3B Period Marital Status by Previous Marital Status (Married Men Older than 35 Years, Dakar)

Previous marital status	Period marital status			
	Monogamous	Polygamous	Divorced	Widower
Bachelor	451	0	0	0
Monogamous	0	180	76	8
Polygamous	57	53	0	0
Divorced	58	0	0	0
Widower	8	0	0	0

Source: IFAN-ORSTOM 1989 survey.

TABLE 26–3C Mean Number of Marital Events by Period Marital Status

Marital status at time of survey	Mean no. of marriages	Mean no. of divorces	Mean no. of widowhoods	Number of individuals
Monogamous	1.32	0.25	0.05	310
Polygamous	2.50	0.20	0.07	123
Divorced	1.22	1.22	—	18

Source: IFAN-ORSTOM 1989 survey.

Changes in marital life captured in a dynamic approach must be integrated into the complete life cycle to assess the interaction between marital events and the change in the individual's economic and social status. Event-history surveys yield new information on the marital life cycles of men and women and on their socio-occupational histories. These data can provide a better understanding of the evolution and determinants of nuptiality, in connection with other events in the individual's life.

II. NUPTIALITY ANALYSIS AND EVENT-HISTORY ANALYSIS

The principle of the retrospective event history is to record the main events known to an individual from birth concerning his or her activity (including education), marital life (formation and dissolution of households), and residential life (mobility, *decohabitation*, access to home ownership). This information, known as *tri-biographical* (Courgeau and Lelièvre, 1989), can also be supplemented by birth history or any other information that can offer additional insights for the analysis of individual histories. The complete event history sheds greater light on the interrelationships between working career and marital life or between family life and dwelling changes. As a result, demographers no longer separate the analysis of marital events from the changes in an individual's living conditions. Several studies have been performed using this approach with the aid of event-history methods. For example, some recent work in North America has focused on the interference between women's working careers and their marriage outcomes (Hiedemann *et al.*, 1998; Neill and Le Bourdais, 1998; Le Bourdais, Neill, and Vachon, 2000). The aim is a better understanding of the demographic, economic, and social determinants of the formation and dissolution of unions, by viewing marital events against other dimensions of personal life. The purpose is less to measure the scale of a phenomenon than to identify the main factors that can promote it or not. Three concrete examples from a study carried out in Dakar will be used to describe the collection and analysis methods and highlight the value of measuring the influence of male working careers on marital dynamics.

1. Description of Dakar Survey

The main purpose of the Dakar study is to show the impact of worsening living conditions on demographic and social behavior. The IFAN-ORSTOM team in Dakar selected three dimensions: access to work, access to housing, and the formation of the household (through marital history and children's life courses). How do city dwellers manage to satisfy various needs—particularly jobs and housing—when they may not possess the same strengths and do not have the same demands?

The collection of *quantitative event histories* relies on the respondent's supply of an accurate time line of events experienced (GRAB, 1999). The problem is that few people memorize the dates of all the events they have experienced; by contrast, the sequence of family events is easily stored in memory. The Dakar survey (Antoine and Bocquier, 1999) used the AGEVEN (age-event) form to help respondents date the main events of their lives. Before starting to fill out the questionnaire, the interviewer asks the respondent to locate in time the main events of his or her family life, migration and residential life, and working career. The data are not collected in a preset order. As the interview progresses, events are recorded on a form showing a time scale (in calendar years) and the time lapsed since the event. This allows the events to be classified without too much difficulty.

From a representative sample of Dakar and its suburbs comprising more than 2100 households and 17,900 individuals,[6] we drew a subsample for the event-history survey. The individuals were selected according to a stratification by sex and age group (25–34 years, 35–44 years, and 45–59 years), so that we were able to interview an equal number of persons in each cohort. This was needed to minimize problems due to inadequate numbers that could arise when comparing cohorts. In the end, we collected 1557 event histories in Dakar in the last quarter of 1989.

2. From Questionnaire to Statistical Processing

Event-history collection requires a fairly long and complex questionnaire. The most practical solution seemed to be to gather the information by broad topic: activities, nuptiality, migrations, and so forth. For example, interviewers posed an identical series of questions on each union. At the end of the operation, we had series of event-history information on specific individuals. The transition from the questionnaire to computer files was a delicate process. In the Dakar survey, we had parallel series of migration histories, occupational histories in Dakar, urban residential histories, marital histories, and descent histories. For each topic we created a file in which the events concerning a single individual were dated and classified on a time line. For marital data, the main information was the date of entry into union, characteristics of the union and of the spouse, and the date of the end of the union, either by separation or death of the spouse. We therefore needed to prepare the marital file in such a way as to identify the date of each marital event (union formation or break up), and the successive marital statuses that each individual may have experienced. We then had to merge the marital file with the other files containing the other histories to obtain a single file with all the events experienced by the individual, in chronologic order. This computer task required specific procedures for data merger in time. They were tested in the Dakar survey and developed by Bocquier (1996) with the aid of the Stata software program, which seems particularly well suited to this type of study.

For a single individual, the file contains as many lines of information as status changes. We can therefore determine, at each point of the individual's life, his or her occupation, place of residence, marital status, children born—along with information on each event—and establish connections between the items. The events are dated and, for each period of the individual's life, we know the time spent in a given status, as shown in our example (Table 26–4A and 3B). Indeed, it is essential to arrive at a consistent collection of dates of transitions between life stages. The novel feature of event-history analysis is the possibility of determining time relationships between life events. It is therefore indispensable, when collecting the data, to locate the events in each dimension relative to the others (for example, marriage relative to employment, employment relative to access to housing, etc.). For this purpose, the accurate recording of the event sequence matters more than the precise dating of the events (to the day or to the month). In particular, the date-recording system should be capable of locating the relative positions of same-year events, even when the exact months are not known (Bocquier, 1996; GRAB, 1999).[7]

Table 26–4A summarizes the history of a man who has experienced four unions. The date of each marital event (marriage or divorce) is converted into a number of months elapsed since January 1, 1900. The man was first married in January 1968 (817 months after January 1, 1900). Thirty-eight months later (855 − 817), he divorced his first wife. He remarried shortly after (November, 1971) and a year later took another wife and became polygamous (December, 1972). In July of 1973, he divorced his first wife (first wife in his household at that date, but second union in his life). He remained monogamous for nearly 8 years until

[6] Random two-stage sample drawn from districts in the 1988 census covering the entire Dakar conurbation.

[7] For example, in Dakar, we have used a coding system for events occurring in the same year, but in months that have not been precisely determined. If three events occur in the same year, the first is coded 33, the second 66, and the third 88. Later, during computer processing, a fictitious month will be assigned to each event to preserve the order of occurrence.

TABLE 26–4A Marital Statuses Experienced by a Specific Male Individual Since His First Union

Spouse rank	Number of spouses	Marital event	Marital status	Date of start of period	Date of end of period
1	1	Marriage	Monogamous	817	855
1	0	Divorce	Divorced	855	863
2	1	Marriage	Monogamous	863	876
3	2	Marriage	Polygamous	876	883
2	1	Divorce	Monogamous	883	982
3	0	Divorce	Divorced	982	992
4	1	Marriage	Monogamous	992	1080

Source: IFAN-ORSTOM survey, individual 803.

TABLE 26–4B Marital, Residential, and Occupational Careers of a Specific Male Individual Since His First Union

Marital status	Place of residence	Occupation	Children	Date of start of period
Monogamous	Plateau	Higher-level technician		817
Monogamous	Medina	Higher-level technician		835
Monogamous	Medina	Higher-level technician	1 with wife 1	840
Divorced	Medina	Higher-level technician		855
Monogamous	Medina	Higher-level technician		863
Monogamous	Medina	Higher-level technician		869
Monogamous	Medina	Higher-level technician	1 with wife 2	875
Polygamous	Medina	Higher-level technician		876
Monogamous	Medina	Higher-level technician		883
Monogamous	Medina	Higher-level technician	1 with wife 3	901
Monogamous	Medina	Higher-level technician	2 with wife 3	925
Monogamous	Medina	Large retailer		936
Monogamous	Sicap	Large retailer		938
Monogamous	Sicap	Large retailer		960
Monogamous	Grand Dakar	Large retailer		974
Divorced	Grand Dakar	Large retailer		982
Monogamous	Grand Dakar	Large retailer		992
Monogamous	Grand Dakar	Large retailer		1020
Monogamous	Grand Dakar	Sales representative		1022
Monogamous	Medina	Sales representative		1024
Monogamous	Plateau	Sales representative		1054

Source: IFAN-ORSTOM survey, individual 803.

October 1981. He then remained without wife for 10 months, and at the time of the survey (1080 = December 1989) he had been married again (for the fourth time) for more than 7 years (since August, 1982). The dating system is therefore very flexible and allows multiple analyses. For example, we can obtain the individual's age at each event. This man was born in 1947 (we arbitrarily assign him the month of July, or 571 months since 1900) and we can thus determine that he was 20 years old at his first marriage, 24 years old at his second, 25 years at his third, and 35 years at his fourth (ages in complete years).

But the example above does not exhaust the possibilities of this type of data file, which comprises a number of rows equal to the number of events. The principle is the same for each history and, in the case of the man we have just discussed, the file contains 21 rows for the 21 status changes since his first marriage. Table 26–4B recapitulates his marital status, neighborhoods of residence in Dakar, occupations, and number of children, as well as the mother's reference number. As we can see, a new row is created in the file with each status change. It includes the information from the previous row that has not changed and the updated information on the new event. The variables used in our example represent only a tiny fraction of the ones available in the file. Such a file makes it very easy to conduct complete event-history analyses covering an individual's entire life. Data and analysis of this kind make it possible to transcend cross-sectional analysis and take into account an individual's successive statuses. We can thus study marital history and occupational history, and factor in their reciprocal influences. Does marriage quicken women's entry into—or exit from—the labor market? Does divorce lead to new occupational statuses? These are just two of the many questions that can now be answered.

Preparing the data in this way may be fastidious, but it is a prerequisite for analysis. In particular, all the event histories must end on the same date (i.e., the survey date [1080]). Thanks to the time sequencing of events, we can move on to analyze the event histories themselves.

3. Fundamentals of Event-History Analysis: the Concepts of Population at Risk and Censoring

Once we have created the complete file, we can perform a number of analyses. Each specific analysis (entry into first union, end of union through divorce, remarriage after divorce, etc.) consists of measuring the time between a reference moment common to all the individuals and the date of the observed event or the date of exit from observation. This method chiefly demands a rigorous conceptualization of the issue studied. The population at risk, the event studied (the risk), and the competing risks that will cause the individual to exit from the observation must all be precisely defined. For example, if we study the transition from never-married status to first marriage for males in Dakar, the *population at risk* will comprise never-married men living in Dakar at age 18; the time interval will be measured from the 18-year anniversary date to the marriage date. A man arriving in Dakar after age 18 will not be included; a man leaving Dakar before his marriage will be included up to his departure date. This method of handling *censoring* makes it possible to take all event histories into account, even those of young cohorts observed up to the survey date (this is known as *right-censoring*). The procedure helps us to compute the regression coefficients. The diagram below shows a population at risk and the possible censoring events (Figure 26–1).

Using a similar approach, we can study male divorce (breakup of first union) or entry into polygamy. In both cases the population at risk is the same—men in first union—and the time examined runs from the date of the first union up to the date of the event defined as the risk. When the event studied is divorce, the censoring events are departure from

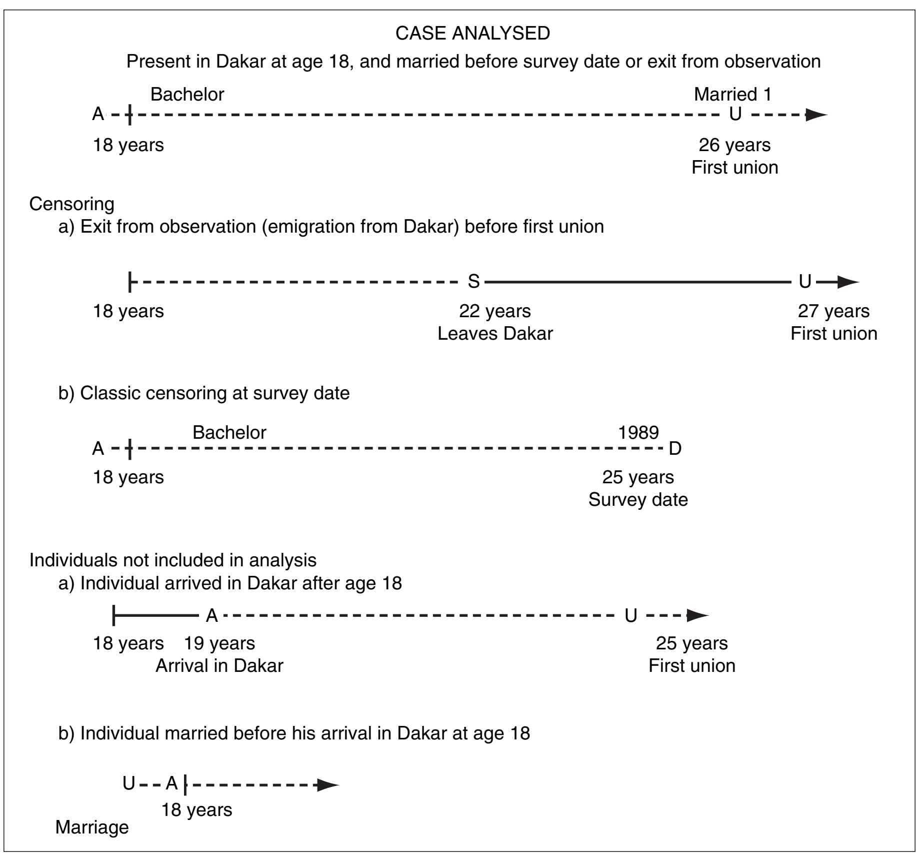


FIGURE 26–1 Population at risk: male bachelors living in Dakar at age 18. Risk: first union in Dakar.

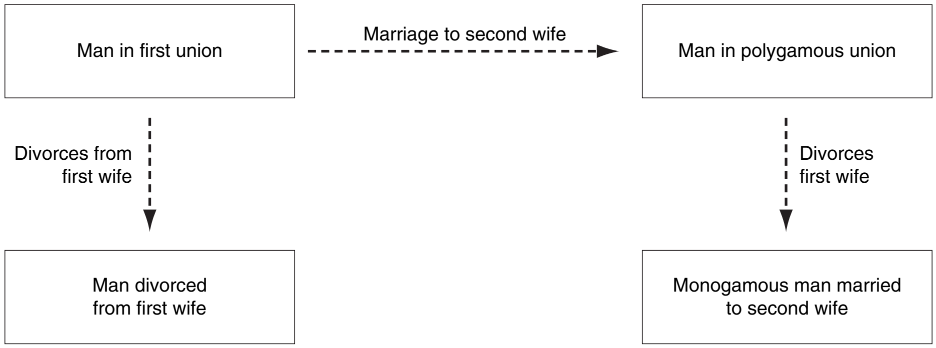


FIGURE 26–2 Diagram of possible statuses: transition from union with first wife to union with second wife (with or without divorce).

Dakar and widowhood. By contrast, entry into polygamous union is a variable that evolves over time and not a competing risk. Likewise, we can study entry into a polygamous union—but, in this case, divorce from the first wife constitutes censoring. The population at risk is that of men in first union, who may later be married to a second wife, either after divorcing the first or by becoming polygamous and later divorcing the first wife. The flowchart for this case is shown in Figure 26–2. Once the concepts used have been well defined, the causal analysis becomes easier.

4. The Semiparametric Cox Model: Introducing the Time Dimension into the Regression

Although event-history analysis emphasizes the time dimension, regression models omit it. We must therefore use methods that allow a combination of the time effect and the characteristics capable of explaining an event's occurrence. For example, we can study not only the time that elapses between the start of a union and its breakup in divorce, but also identify the factors that hasten the process.

Several methods and models are available (Courgeau and Lelièvre, 1989; Blossfeld, Hamerle, and Mayer, 1989; Trussell, Hankinson, and Tilton, 1992; Bocquier, 1996). Rather than give a catalogue of methods, it is more useful here to describe in detail some of the studies conducted with the aid of an event-history analysis method based on David Cox's semiparametric, proportional-hazards regression. Cox was the first to suggest combining regression analysis and the life-table principle (Cox, 1972; Cox, 1984).

The Cox model may be described as (1) taking account, by means of the regression, of the effect of explanatory variables in the life-table analysis or (2) the introduction of the time dimension into the regression. The advantages of one technique fill the gaps in the other. To solve the problem of duration and explanatory factors, the regression is performed not on the characteristic acquired by the individual at the end of his or her life, but on the characteristic acquired in each year of life up to the survey date. The reference status—required in the regression model—is not identical for the entire sample, but specific to each observation period. This series of probabilities makes it possible to define a baseline survival curve (for example, survival of never-married status if the analysis concerns first marriage), also called *baseline hazard function*: This is the model's nonparametric component. The regression model then computes the effect of explanatory variables on the annual risk of occurrence of the event. This is the model's parametric component, which is added to the nonparametric component to form a model known as semiparametric. A regression coefficient, measuring the variable's mean influence on the annual risk, is associated with each variable.

In a model of exit from never-married status, a variable that increases the risk of ending that status multiplies the basic risk by a quantity greater than unity; the converse is true of a variable that lessens the risk. For example, if the fact of being born between 1955 and 1964 rather than between 1930 and 1944 lowers the risk of marrying, the coefficient associated with the variable will be a number smaller than unity. If the value is 0.64, we will say that, all other things being equal, the modality concerned decreases the risk of marrying by 36% (or by just over one third). Alternatively, we can say that the modality divides the risk by 1.56, the reciprocal of the coefficient 0.64.

Time-dependent variables can be introduced into proportional-hazards models; in fact, that is one of the models' original features. The effect of independent variables whose value may vary during the observation is interpreted in the same manner as the effect of an independent variable with a constant value. A man who is successively an apprentice, a payroll employee in the informal sector, and unemployed contributes successively to the estimation of the risk of apprentices, of payroll employees in the informal sector, and of the unemployed. For interpretative purposes, this is the same as saying that the man is successively exposed to the risk change characteristic of apprentices, of informal-sector payroll employees, and of the unemployed.

III. FIRST MARRIAGE, DIVORCE, AND POLYGAMY IN DAKAR

The analyses presented in the following paragraphs are restricted to men aged 25 to 59 years (at the time of the survey) for the analysis of first marriage and to men aged 35 to 59 years for the analyses of divorce and polygamy. We have confined our sample to men as one of the problems in analyzing polygamy among women is the need for a complete record of husbands' marital histories. For each woman, we have information on all her unions and on the marital status of the husband at the time of the union, but we do not know the later changes in her husband's marital status (Does he have an additional wife? Has he divorced his previous wife?). This means we cannot monitor the dynamics of polygamy from data exclusively concerning women. With the latter data we can only study women's entry into union and try to establish, according to the husband's characteristic, if he was already in a union or not. We can also construct some simple indicators such as the proportion by age of women who begin their marital life in a polygamous union. In a more recent survey of the same type conducted by IFORD, some questions were introduced with the purpose of gathering the marital history of husband(s) from the wives (Kouamé *et al.*, 1999), but the results were not conclusive. Many women refused to give information on their co-wives (or rivals).

1. Explanatory Factors

Different explanatory factors, materialized by variables, are tested for each analysis. For the analysis of exit from never-married status, the following variables were selected.

- Birth cohorts are determined from dates of birth and reproduce the boundaries used in the sample stratification. The study focuses on the histories of three cohorts: men born between 1930 and 1944 (aged 45–59 years at the time of the survey), between 1945 and 1954 (aged 35–44 years at the time of the survey), and between 1955 and 1964 (aged 25–34 years at the time of the survey). The three cohorts experienced different economic conditions at the time of their entry into union, extending from Senegal's early independence years (1960s) to the slowdown of the 1970s and the crisis of the 1980s.
- The age at arrival in Dakar gives an idea of the migration history and allows a distinction between persons born in Dakar (the reference group) and the different types of migrants by age at arrival.
- The individual's ethnic group is that of the father, recorded by the survey. The largest and most urbanized ethnic group, the Wolof, is used as the reference.
- Islam is the dominant religion (95% of the population). We distinguish between the members of Muslim brotherhoods (reference), Muslims outside the brotherhoods, and Christians.
- The last grade attended serves as the indicator of education attainment. Five categories are used: no schooling; study at Koranic school for at least 7 years (without attending official school system); attendance in first grade (*cours préparatoire*) or second and third grades (*cours élémentaire*) (four years of schooling or less); fourth and fifth grades (*cours moyen*) or middle school (*collège*) (5–10 years of schooling); high school (*lycée*), technical education, or higher education. The reference group consists of men who have never attended school.
- Economic activity, residence status, and type of dwelling are independent variables that vary over time (i.e., the same person can experience different statuses during his or her lifetime).
- For *activity periods*, we distinguish between the following statuses: public-sector employee; private-sector employee; informal-sector employee; informal-sector self-employed; apprentice; student; unemployed.
- The *dwelling occupancy status* acts as an indicator of a certain degree of economic autonomy for individuals who are tenants or owners as against being housed by their parents or another relative.
- The *dwelling type* is an indicator of social differentiation, as a proxy for the social category of the household where the individual has lived in different stages of his life (Antoine *et al.*, 1995).

For the analysis of polygamy and divorce, we use only the two oldest cohorts. Some variables are identical to those used for the analysis of entry into first union: *age of arrival in Dakar, ethnic group* (of husband

and first wife), *religion*, *activity period*, and *dwelling occupancy status* with slightly different categories. A variable reflects the timing of the first union: *the man's age at first marriage*. These variables enable us to modulate cohort analysis by taking into account the variation in age at marriage, and to check whether polygamy or divorce vary as a function of later marriage.

Was the father polygamous (in the study on polygamy) or *was the father divorced* (in the study on divorce)? These two items tell us if family antecedents influence marital behavior. For polygamy, we can venture the hypothesis that the young man reproduces his father's marital behavior. *Kinship between spouses at first marriage* is also an indicator of marriages often arranged by the family. With this variable, we can test whether such unions are stabler. The questionnaire should have been more specific and should have been designed to establish whether the spouse was chosen or if the union was arranged. We can also advance the hypothesis that these unions, where the freedom of choice is low, can promote polygamy.

The last grade attended is used as a measure of *educational attainment*. Three levels are distinguished for the husband and for the first wife respectively: no schooling (reference category); elementary school; secondary school and higher. A hypothesis often put forward is that high educational attainment is the main inhibitor of polygamy.

In the analysis of divorce, the *arrival of the second wife* is taken into account from the date of the second marriage. The variable enables us to measure the effect of polygamy on a dissolution of the first union. In the case of divorce, we factor in the *number of children from the first marriage*.

2. First Marriage of Men in Dakar

In Dakar, we observe a sharp slowdown in the pace of entry into first union between the oldest and the youngest cohort (Antoine *et al*., 1995). Between the oldest cohort (1930–1944), with a median age at first marriage of 26 years, and the intermediate cohort (1945–1954), the gap in the median age at first union is two years. The gap widens further between the two youngest cohorts. The timing of first marriage has moved up five years to a foreseeable median age of 33 years for the 1955 to 1964 cohort. Despite such timing differences, marriage is by no means in jeopardy.

Table 26–5 reports the results of four analyses (Antoine *et al*., 1995; Antoine and Djiré, 1998). The first concerns all three cohorts combined; the other three analyses each concern one cohort, from the oldest to the youngest. The relative risk of entry into first marriage declines with cohort youth, but is not statistically significant (at the 5% limit) except for the youngest cohort. These youths are almost one half as likely as their elders to have entered into first union (0.64). Some individual characteristics promote a quicker entry into union, such as the fact of arriving in Dakar before the age of 12, or belonging to the Diola ethnic group. By contrast, males who stay in school beyond fifth grade, Muslims who do not belong to brotherhoods, and Christians marry later. But the strongest influence on the marriage rate is exerted by time-varying characteristics. For example, the fact of being in unemployment, in school, or in apprenticeship sharply slows the entry into first union, as is the fact of still living with one's parents or relatives. Crisis effects and cohort effect combine to make young cohorts marry later. But an analysis by cohort will provide a more detailed view.

The traditional marriage model still dominates the 1930 to 1944 cohort. Characteristics such as age of arrival in Dakar, ethnic group, religion or even educational attainment, activity status or residence status, and the dwelling category do not seem to influence first marriage (Table 26–5). Only apprenticeship seems to have some effect in reducing entry into first union. Marriage of men in this cohort seems virtually unaffected by economic factors. The marriages occurred between the end of French colonial rule and the late 1960s, which may be described as a period of relative expansion. There were fewer economic constraints, and household formation was a stage of the life cycle that did not depend significantly on economic well-being.

Some factors turn out to have a more discriminating effect on first marriage for the 1945–54 cohort. We observe sizable differences by religion. Muslims belonging to a brotherhood are the most prone to early marriage. They entered into first union twice as fast as Muslims outside the brotherhoods and eight times as fast as Christians. Those who have attended elementary or middle school are not significantly distinct from those with no schooling; only the young men with a high-school education or better marry at an older age. The latter display a rate of entry into union almost twice as slow as the members of their cohort who have never been to school. We find no meaningful difference between activity periods except for unemployment spells: The unemployed exhibit a marriage rate about five times as slow as that of the reference category (self-employed)—unlike their elders, who have almost never been out of work. It should be noted that unemployment has been a more powerful brake on marriage for this cohort than apprenticeship was for the preceding one. In the 1945 to 1954 cohort, employment and

TABLE 26–5 Coefficients of Cox Semiparametric Proportional-Hazards Model for Male Entry into First Union

Variable[a]	Category	Total cohorts	1930–1944 Cohort	1945–1954 Cohort	1955–1964 Cohort
Cohort (1930–1944)	1945–1954	1.06	—	—	—
	1955–1964	0.64[c]	—	—	—
Age on arrival in Dakar	Under 12 years	1.39[d]	1.20	1.43	1.21
Born in Dakar	Over 12 years	1.06	1.42	1.82	1.38
Father's ethnic group (Wolof)	Fulani	0.92	0.64	1.15	0.90
	Manding	0.98	1.13	0.71	0.51
	Sereer	0.88	1.53	0.52	0.82
	Diola	1.99[d]	3.13	3.39	1.41
	Other	1.41	1.60	1.71	1.60
Religion (Muslim brotherhood)	Muslim outside brotherhood	0.58[b]	1.12	0.46[b]	0.46[d]
	Christian	0.34[c]	0.39	0.12[c]	0.52
Educational attainment (no schooling)	Koranic school	0.98	1.81	1.10	0.40[d]
	Elementary school	0.99	1.49	1.58	0.62
	Middle school	0.64[c]	1.05	0.69	0.37[b]
	High school—higher education	0.64[d]	1.26	0.54[d]	0.45[d]
Activity status (self-employed in informal sector)	Payroll employee in informal sector	1.26	1.51	1.09	1.10
	Public-sector employee	1.32	1.33	1.66	0.52
	Modern formal private sector	1.31	1.85	1.05	1.61
	Unemployed	0.31[b]	1.02	0.18[b]	0.17[b]
	Student	0.50[c]	0.25	0.78	0.46
	Apprentice	0.39[b]	0.27[c]	0.53	0.40[d]
Residence status (owner or tenant)	Housed by father/mother	0.66[c]	0.68	0.73	0.74
	Housed by other relative	0.65[b]	0.61	0.74	0.87
Dwelling category (corrugated roof)	High- and medium-level quality	0.57[c]	0.45	0.76	0.24[c]
	Permanent roofing	1.18	1.57	1.17	1.05
	Shanty	0.94	1.14	0.47[c]	2.59[b]

[a] Reference category in parentheses.
[b] Significant at 1% limit.
[c] Significant at 5% limit.
[d] Significant at 10% limit.

housing constraints appear to emerge, already influencing entry into union.

More than for their elders, first marriage in the youngest cohort (1955–1964) has been more visibly subject to the effects of most of the characteristics specified, except for religion, which does not have the same hold as on the 1945 to 1954 cohort. Although the behavior of Christians does not differ significantly from that of Muslims belonging to brotherhoods, Muslims outside the brotherhoods have entered into union 2.5 times as slowly. Educational attainment is linked to lower entry into first union except for the men who have not gone beyond first grade or elementary school. The behavior of the latter category is largely explained by the activity period, and it is perhaps these young people who are most affected by unemployment.

Men in payroll employment show no perceptible difference in their rate of entry compared with self-employed men in the informal sector. For the 1955–1964 cohort, reaching majority and maturity in the 1980s was concomitant with greater tensions in the job market; as most of the men were seeking a first job, unemployment spells (as in the intermediate cohort group) and apprenticeship spells (as in the oldest cohorts) were the least conducive to early marriage. Apprentices were 2.5 times as slow to marry as self-employed men in the informal sector. But the differential was even wider in unemployment periods, where the risk of marrying was six times as small. Periods of residence in higher-quality dwellings (and hence of membership in better-off social categories) were not more propitious to an early first marriage, contrary to periods of residence in shanties, where the rate was 2.6 times the baseline. Middle-class marital behavior therefore seems to have been more affected by the economic crisis. The 1955 to 1964 cohort—for whom adolescence began with the 1970s recession that worsened in the 1980s—was the hardest hit in terms of family solidarity as well as on the individual level. Unemployment, the main consequence of the crisis, was a powerful brake to rapid entry into union.

3. Becoming Polygamous

In Dakar, the fact of having had a polygamous father quickens the entry into polygamous union, as if a section of Dakar society perpetuated a more traditional behavior (Table 26–6). Kinship with the first wife (synonymous with a family-arranged marriage) also fostered an earlier marriage to a second wife, confirming the hypothesis that when the husband has not chosen his wife himself, he will be quicker to marry another woman whom he may have met outside the family. By contrast, late marriage helps slow the entry into polygamous union. When the first wife's educational attainment reaches secondary school (or higher), the pace of entry into polygamy is halved: this corroborates the reluctance of educated women to accept the arrival of a second wife, or even their outright resistance to the idea. Christianity inhibits polygamy in Dakar. By contrast, the rate of entry into polygamous union seems unaffected by the age of arrival in the city, the husband's ethnic group, the first wife's ethnic group, the husband's educational attainment, or the husband's activity. As paradoxical as this may sound, Dakar is witnessing a diversification of polygamy patterns.

Urbanization does not spell the end of polygamy. It persists in Dakar, where it is vested with official, reli-

TABLE 26–6 Coefficients of Cox Semiparametric Proportional-Hazards Model for Male Entry into Polygamy and Divorce

Variable[a]	Category	Polygamy	Divorce
Cohort (1930–1944)	1945–1954	0.76	1.07
Age at first marriage (husband) (25–29 years)	Under 25 years	1.04	1.06
	30–34 years	0.72	1.07
	35 years and over	0.54	1.04
Age on arrival in Dakar (born in Dakar)	Under 12 years	1.40	0.94
	Between 12 and 17 years	1.29	0.87
	Between 18 and 24 years	0.88	1.27
	Over 25 years	0.88	0.43
Divorced father (no)	Yes	NC	0.98
Polygamous father (no)	Yes	1.45[c]	NC
Related to first wife (no)	Yes	1.56[d]	0.52[b]
Religion (Muslim)	Christian	0.31[c]	0.45
Husband's ethnic group (Wolof)	Fulani	0.76	1.31
	Sereer	0.97	1.48
	Other	1.49	0.85
First wife's ethnic group (Wolof)	Fulani	0.96	0.49[d]
	Sereer	1.72	0.52
	Other	0.88	0.77
Spouses' educational attainment (husband and first wife never went to school)	Husband: elementary	1.16	1.19
	Husband: secondary	1.29	2.16[c]
	Wife: elementary	1.16	0.75
	Wife: secondary	0.59[d]	0.43[c]
Activity status (skilled worker)	Informal sector	0.74	1.09
	Unemployed	0.45	3.26[b]
	Not in labor force	0.0003	0.0003
Residence status (owner or tenant)	Housed by father/mother	0.92	1.22
	Housed by other person	1.21	1.13
Number of children from first union	—	NC	0.67[b]
Entry into polygamous union (no)	Yes	NC	3.73[b]

NC, not concerned by this model.
[a] Reference category in parentheses.
[b] Significant at 1% limit.
[c] Significant at 5% limit.
[d] Significant at 10% limit.

gious, and social legitimacy, and acts as a social and demographic regulator. It enjoys a popularity and a normality status that causes it to be perceived as commonplace—most notably by men, for whom it remains a privilege that they can claim, an ambition enabling them to flaunt their social success, a means of controlling and dominating women (Antoine and Nanitelamio, 1995; Antoine *et al.*, 1998).

4. Joblessness and Polygamy: the Main Causes of Divorce

The statistical analysis of event histories concerning divorce yields more contrasting results (Table 26–6). Several factors determine the pace of divorce. Kinship and the number of children slow a separation from the first wife; a preferential marriage is harder to break up, whereas it promotes polygamy. In contrast, economic precariousness favors divorce. Unemployment accelerates the risk of divorcing by a factor of more than three. The husband is supposed to meet household expenses; when his economic status deteriorates, the union is undermined. Failure to provide for one's wife is often cited as one of the main causes of divorce. The issue of economic precariousness and activity is, however, complex. In a study using a similar methodology on female nuptiality in Dakar, Bocquier and Nanitelamio (1991) have shown a higher risk of divorce among the youngest cohorts by comparison with the periods before the 1980s. They note that women's economic activity after marriage is the chief factor that increases the risk of divorce. Work means potential financial independence for the wife, giving her the capability of supporting herself and her children in the event of a divorce (McDonald, 1985; Burnham, 1987). For women who have worked before marriage and have left their jobs after marriage, the risk of divorce is smaller, corroborating the hypothesis of economic precariousness as a deterrent to divorce for women in that situation.

Men with a secondary-school education divorce twice as fast as men with no schooling. By contrast, better-educated women divorce more slowly than lesser-educated ones. A separate analysis (Antoine and Nanitelamio, 1996), cross-classifying the education attainments of the husband and his first wife, shows that too wide an education gap between spouses increases the risk of divorce. A husband with a secondary-school education or higher and an illiterate wife are three times as likely to divorce as a couple in which both spouses are illiterate. In contrast, the study finds no differences between the other education-attainment inequalities.

Gage-Brandon (1992), using an identical method (but without time-varying characteristics), shows that polygamy is an important factor in the instability of unions in Nigeria. For the study of divorce in Dakar, the polygamy episodes recorded in the marital event history have been factored in to test the hypothesis that the arrival of a second wife quickens the departure of the first. Entry into polygamous union indeed sharply accelerates the risk of the man's divorcing his first wife. The model incorporates only the sequence of events declared by respondents. However, our analytic template is close to true-life situations, as in many cases the first wife is placed before a *fait accompli*. Sometimes, she is not informed until the second marriage has already been celebrated. After this transition to polygamy, divorce may occur after a phase of observation and conflict. But the size of the coefficient (the risk of divorce is accelerated 3.7 times) shows that the situation often leads to divorce. Women are increasingly initiating divorce.

CONCLUSION

Event-history analysis makes it possible to evaluate interactions between events. For example, in Dakar, the cohort effect persists for the first marriage. Secondary schooling (and higher) slows marriage and accelerates divorce. But the main lesson of the analyses performed in Dakar is the major role of the job-market crisis on union formation or dissolution. Although the type of activity does not, in itself, influence the pace of marriage or that of divorce or emancipation, unemployment spells have a decisive effect. Unemployment sharply slows entry into union and hastens divorce. Educated men divorce more quickly than illiterate men, but residence in a high- or medium-level dwelling halves the probability of divorcing. Job insecurity hastens divorce, but material well-being slows it (Antoine and Djiré, 1998). When his economic position worsens, the husband can no longer meet the household's economic and financial liabilities for which he is responsible, and the union is in many cases at risk. In sum, the crisis-induced economic precariousness promotes divorce. Our analysis identifies few socioeconomic factors capable of explaining a differentiated behavior of males in regard to polygamy. Only a limited number of demographic mechanisms influence the phenomenon. The analysis of divorce shows a high risk of break-up of the first union after the arrival of a second wife.

Another valuable aspect of event-history analysis is the opportunities it offers for studying reciprocity. For example, we can analyze young people's residential autonomy as a function of access to employment or of marriage events. Tokindang (1995) has shown that, for Dakar's young generations, marriage and having a child are no longer—paradoxically—an incentive for couples to find dwellings of their own. Lack of work and lack of housing combine to slow marriage for young men, but marriage does not hasten access to an independent dwelling. A distinction is drawn between the formation of the couple and the formation of a new household.

Acknowledgment

Translated from French by Jonathan Mandelbaum.

References

Antoine Philippe and Bocquier Philippe, 1999. L'enquête "Insertion urbaine à Dakar et Pikine," *in*: Groupe de réflexion sur l'approche biographique (ed.), *Biographies d'enquêtes. Bilan de 14 collectes biographiques*, p. 177–201. Paris, INED, IRD, Réseau Socio-Économie de l'habitat, 340 p. (Coll. Méthodes et Savoirs).

Antoine Philippe, Bocquier Philippe, Fall Abdou Salam, Guissé Youssouf Mbargane and Nanitelamio Jeanne, 1995. *Les familles dakaroises face à la crise*. Dakar, ORSTOM-IFAN-CEPED, 209 p.

Antoine Philippe, Bocquier Philippe, Razafindratsima Nicolas and Roubaud François (eds.), 2000. *Biographies de trois générations dans l'agglomération d'Antananarivo. Premiers résultats de l'enquête BIOMAD-98*. Paris, CEPED, 140 p. (Documents et manuels, no. 11).

Antoine Philippe and Djiré Mamadou, 1998. Crise et évolution des comportements matrimoniaux à Dakar, *in*: Francis Gendreau (dir.), *Crises, pauvreté et changements démographiques dans les pays du Sud*, p. 203–235. Paris; AUPELF-UREF and Éditions ESTEM, 415 p.

Antoine Philippe, Djiré Mamadou and Laplante Benoît, 1995. Les déterminants socio-économiques de la sortie du célibat à Dakar. *Population*, no. 1, p. 95–118.

Antoine Philippe, Djiré Mamadou and Nanitelamio Jeanne, 1998. Au cœur des relations hommes-femmes: polygamie et divorce, *in*: Philippe Antoine, Dieudonné Ouédraogo and Victor Piché (eds.), *Trois générations de citadins au Sahel*, p. 147–180. Paris, L'Harmattan, 281 p.

Antoine Philippe and Nanitelamio Jeanne, 1996. Can polygamy be avoided in Dakar? *in*: Kathleen Sheldon (ed.), *Courtyards, markets, city streets: urban women in Africa*, p. 129–152. Boulder, Westview Press, 342 p.

Battagliola Françoise, Bertaux-Wiame Isabelle, Ferrand Michèle and Imbert Françoise, 1991. *Dire sa vie: entre travail et famille. La construction sociale des trajectoires*. Paris, CSU, IRESCO, 280 p.

Blossfeld Hans Peter, Hamerle Alfred and Mayer Karl U., 1989. *Event history analysis: statistical theory and application in the social sciences*. Millsdale, Lawrence Erlbaum Associates Publishers, 294 p.

Bocquier Philippe and Nanitelamio Jeanne, 1991. Les déterminants socio-économiques des changements matrimoniaux des femmes de Dakar, *in*: UEPA (ed.), *Femme, famille et population*, vol. 1, Communications sollicitées, p. 369–386. Dakar, UEPA, 630 p. (Ouagadougou [Burkina Faso] Conference Proceedings 24–29 April 1991).

Bocquier Philippe, 1996. *L'analyse des enquêtes biographiques à l'aide du logiciel Stata*. Paris, CEPED, 208 p. (Coll. Documents et Manuels).

Brandon Anastasia, 1991. The Formation and Stability of Informal Unions in Côte d'Ivoire, *in*: UEPA (ed.), *Femme, famille et population*, vol. 1, Communications sollicitées, p. 330–349. Dakar, UEPA, 630 p. (Ouagadougou [Burkina Faso] Conference Proceedings 24–29 April 1991).

Burnham Philip, 1987. Changing Themes in the Analysis of African Marriage, *in*: David Parkin and David Nyamwaya (eds.), *Transformations of African Marriage*, p. 37–54. Manchester, Manchester University Press, 350 p. (International African Seminars, New Series, no. 3).

Courgeau Daniel and Lelièvre Eva, 1989. *Analyse démographique des biographies*. Paris, INED, 268 p.

Cox David Roxbee, 1972. Regression models and life tables, *Journal of royal statistical society*. B34, p. 187–220

Cox David Roxbee, 1984. *Analysis of survival data*. London, Chapman and Hall, 201 p.

Djiré Mamadou, 1993. *Un aspect de la dynamique de la nuptialité en milieu urbain: l'entrée des hommes en première union à Dakar (Sénégal)* . Montréal, Département de démographie, Université de Montréal, 118 p. + appendices. (Masters dissertation).

Dittgen Alfred, 1995. Un aspect méconnu de la transition démographique: l'évolution des mariages religieux, *in*: Dominique Tabutin, Thierry Eggerickx and Catherine Gourbin (dir.), *Transitions démographiques et sociétés*, p. 459–476. Louvain-la-Neuve, Academia L'Harmattan, 692 p. (Actes de la Chaire Quetelet, 1992, Louvain-la-Neuve, September 15–17, 1992).

Donadjé Florentin, 1991. Typologie des unions masculines au sud du Bénin, *in*: UEPA (ed.), *Rapport de l'UEPA sur la conférence Femme, famille et population. Ouagadougou, Burkina Faso 24–29 avril 1991*. vol 1, p. 387–406. Dakar, UEPA, 630 p.

Donadjé Florentin, 1992. *Nuptialité et fécondité des hommes au Sud-Bénin: pour une approche des stratégies de reproduction au Bénin*. Louvain-la-Neuve, Université Catholique de Louvain, Institut de Démographie, Académia, 222 p.

Gage-Brandon Anastasia J., 1992. The Polygyny-Divorce Relationsip: A Case Study of Nigeria, *Journal of Marriage and the Family*, no. 54 (May), p. 285–292.

GRAB (Groupe de réflexion sur l'approche biographique), 1999. *Biographies d'enquêtes. Bilan de 14 collectes biographiques*. Paris, INED, IRD, Réseau Socio-Economie de l'habitat, PUF/Diffusion Paris, 336 p. (Coll. Méthodes et Savoirs).

Hiedemann Bridget, Suhomlinova Olga and O'Rand Angela M., 1998. Economic independence, economic status and empty nest in midlife marital disruption, *Journal of Marriage and the Family*, 60 (February), p. 219–231.

Houèdokoho Thomas, 1995. *Approche biographique de la nuptialité des femmes béninoises*. Louvain-la-Neuve, Université Catholique de Louvain, Academia-Bruylant, l'Harmattan, 215 p.

Klissou Pierre, 1995. *La polygamie au Bénin. Une approche régionale des tendances et des déterminants*. Louvain la Neuve, Université Catholique de Louvain, L'Harmattan, Academia Bruylant, 257 p.

Kouamé Aka, Beining Eric, Gueye Abdoulaye, Kuepié Mathias and Kishimba Ngoy, 1999. L'enquête "Crise et insertion urbaine à Yaoundé," *in*: GRAB (ed.), *Biographies d'enquêtes. Bilan de 14 collectes biographiques*, p. 221–240. Paris, INED, IRD, Réseau Socio-Economie de l'habitat, PUF/Diffusion Paris, 336 p. (Coll. Méthodes et Savoirs).

Le Bourdais Céline, Neill Ghyslaine and Vachon Nathalie, 2000. Family disruption in Canada: Impact of the changing patterns of family formation and of female employment, *Canadian studies in population*, vol. 27, no. 1, p. 85- 106.

Léridon Henri, Villeneuve-Gokalp Catherine, 1994: *Constance et inconstances de la famille: biographies familiales des couples et des enfants*. Paris, INED, Travaux et documents, no. 134, 341 p.

Lesthaeghe Ron, Kaufmann Georgia and Meekers Dominique, 1989. The Nuptiality Regimes in Sub-Saharan Africa, *in*: Ron Lesthaeghe (ed). *Reproduction and social organisation in sub-Saharan Africa*. Los Angeles, University of California Press, p. 239–337.

McDonald Peter, 1985. Social organisation and nuptiality in developing countries, *in*: John Cleland and John Hobcraft (eds.), *Reproductive change in developing countries*, p. 87–114. Oxford, Oxford University Press, 301 p.

Neill Ghyslaine and Le Bourdais Céline, 1998. Dissolution des premières unions fécondes au Canada: une analyse de risques dans un contexte de changements. Louvain la Neuve, Institut de démographie, Louvain, 23 p. (Communication à la Chaire Quetelet 1998: "Ménages et familles face à la crise").

Ouedraogo Dieudonné and Piché Victor (dir.), 1995. *L'insertion urbaine à Bamako*. Paris, Karthala, 207 p.

Pison Gilles, 1986. La démographie de la polygamie, *Population*, vol. 41, no. 1, p. 93–122.

Smith David P., Carrasco Enrique and McDonald Peter, 1984. *Marriage Dissolution and Remarriage*. Voorburg, International Statistical Institute, 94 p. (World Fertility Survey Comparative Studies, no. 34).

Thiriat Marie-Paule, 1998. *Faire et défaire les liens du mariage. Évolution des pratiques matrimoniales au Togo*. Paris, CEPED, 295 p. (Les Études du CEPED, no. 16).

Tokindang Joël, 1995. *Mobilité et insertion résidentielle à Dakar*. Montreal, Département de démographie, Université de Montréal, 218 p. + appendices. (Demography Ph.D. thesis).

Trussell James, Hankinson Richard and Tilton Judith, 1992. *Demographic applications of event history analysis*. Oxford, Clarendon Press, 276 p.

van de Walle Étienne, 1968. Marriage in African censuses and inquiries, *in*: William Brass, Ansley J. Coale, Paul Demeny, Don F. Heisel, Frank Lorimer, Anatole Romaniuk and Étienne van de Walle (eds.), *The demography of tropical Africa*, p. 183–238. Princeton (New Jersey), Princeton University Press, 539 p.

C H A P T E R

27

The Complexities of Nuptiality

From Early Female Union to Male Polygamy in Africa

PHILIPPE ANTOINE

Institut de récherche sur le développement (IRD), Dakar, Senegal

Apart from its being less ostensible, in demographic terms, than mortality or fertility, nuptiality is a more delicate subject of study for two major reasons. The first is the ambiguity of the definition of marriage, which blends the legal and cultural concepts of the demographic event. The second is the renewability of marriage and especially the fact that it involves at least two statistical units (Tapinos, 1985) or even more in the case of polygamy. Despite its universality, demographers show little interest in the study of marriage and regard nuptiality as merely an intermediate fertility variable (Davis and Blake, 1956). However, nuptiality deserves more attention and is worthy of study in itself. The evolution of marital behavior reveals deep changes in society. The observation of marital change also sheds light on a society's preferred gender relationships (Hertrich and Locoh, 1999). A multifaceted phenomenon, nuptiality poses interesting analytical problems.

Sub-Saharan Africa offers a good example. Admittedly, in that part of the world, nuptiality-related issues have been partly masked by the need to produce relevant data for the study of trends in fertility, mortality, and spatial distribution of the population. However, experience demonstrates the necessity of paying special attention to marital change in order to understand the current transformations in African societies. In African cities, the present nuptiality transition is largely due to the enrollment of young girls in school but also to growing economic difficulties reflected in employment, housing, education, and other problem areas. The result is a weakening of the traditional social-management system in an urban setting.

The nuptiality transition concerns not just the age of entry into union, but the entire marital dynamics, including such factors as the change in the inter-spousal age gap, union terminations, and so on. A change in a single marriage-market parameter affects all behaviors. As we shall see, one of the preconditions for polygamy is a significant gap between men and women in the age at marriage. Later marriage among women, for example, will impact the overall marital dynamics that make polygamy possible.

I. NUPTIALITY IN AFRICA: DIVERSITY AND CHANGE

Marriage systems in Africa differ from one society to another. But some common features are found in these lineage-based societies: As a rule, unions concern lineage groups more than individuals. The asymmetry caused by the *loss of a woman* must be offset by what is known as *bridewealth*.[1] This compensation is usually more modest in matrilinear societies than in patrilin-

[1] Bertrand Lemennicier (1988) draws a bold parallel between bride wealth and pensions and alimony paid to wives in divorces in contemporary western societies. The author describes both kinds of payment as the price to pay for regaining freedom. "Paradoxically, it is the new wife who should indemnify the old one for the loss she causes by marrying the latter's ex-husband."

ear societies, where a larger set of belongings circulates between lineages (Dozon, 1986). Marriage is therefore treated primarily as a *family concern* and a *social concern* in which individuals play a passive role (Hertrich, 1996). Even if the system is relaxed, the marital processes are broadly complied with, both in the sub-Saharan rural area (Hertrich, 1996) and in cities with seemingly more liberal practices such as Douala, where marriage is still difficult without parental consent (Séraphin, 2000). Marriage thus leaves little room for the expression of feelings, which may even be viewed as a hindrance to family bargaining (Lallemand, 1985, quoted in Faizang and Journet, 1988). In Mali, for example, relations between men and women are built around two opposing and complementary poles: *furu* (marriage) and *kanu* (the love relationship) (Dumestre and Touré, 1998). Reconciling the two with the same spouse is often difficult: Studies show the development of strategies involving escape to avoid an imposed marriage, adulterous relationships, divorce followed by remarriage, and so on. Moreover, in an *arranged* marriage, incomprehension between spouses can be aggravated by their sometimes wide age gap (Tiemoko, 1999).

1. Marriage, Union, or Cohabitation? The Limits of Observation and Analysis

Before addressing nuptiality trends—and principally age at marriage—we should spell out the concept of union and give details of the measurement difficulties encountered. Most data-collection operations in Africa use marital categories that are identical to those of developed countries and are therefore barely relevant to African realities. Most African censuses include only a simple question on the marital status of respondents and, in some cases, a question on the number of wives (van de Walle, 1968). Absent specific data, the study of nuptiality is therefore dominated by the period approach. The task of defining the start of the union is often left to the respondent. Now, as seen earlier (Chapter 26), marriage or entry into union often involves several stages, and if we want to compare nuptiality data from different sources, we need to know how marriage is defined in each of the collection operations.

Thiriat (1999) clearly shows the difficulty of defining consensual unions and distinguishing them from premarital relationships. Changes in the type of union may accompany the rise in age at marriage. For example, in Antananarivo (Madagascar), the age of male exit from bachelorhood varies little from one generation to the next (approximately 25 years), but the first union increasingly begins as a consensual union: the percentage has risen from 32% of unions to 44%. Among women in the same city, not only has age at marriage risen from 20.8 years to 23.3 years in two decades, but the proportion of consensual unions has grown from 32% to 59%. The prevalence of consensual unions varies massively from one country to another. They comprise over half of total unions in Botswana (57%) and about 10% in Ghana, Nigeria, and Zambia. Their share is insignificant in Senegal (0.7%), and negligible in Kenya and Tanzania. There are sizable disparities even between countries that are geographically and culturally close such as Ghana (10%) and Togo (29%) (Thiriat, 1999). However, Étienne van de Walle (1996) observes that, apart from Togo, the surveys carried out in French-speaking countries mask the existence of consensual unions.

Beyond these definition problems there is also the issue of surveying practices in the field and the non-comparability of particular sources such as the Demographic and Health Surveys (DHSs) and censuses. Étienne van de Walle (1996) clearly demonstrates the limits of each of these sources: Censuses focus on formal unions, regardless of the type of cohabitation, whereas DHSs emphasize cohabitation regardless of the type of union.[2] Lastly, computation methods often differ from one operation to another. One of the methods most commonly used to calculate mean age at first union consists of determining a period index from a series of proportions of never-marrieds in different age groups, whether the data are derived from a census or a survey (Hajnal, 1953; Tabutin and Vallin, 1977). Another method relies on a direct computation from respondents' statements on their age at first marriage. The two methods yield significantly different results, owing to the calculation method, the memory effect, and the definitions used. Hajnal's method seems to overestimate age at first marriage, but some authors also believe the direct collection of age at first union encourages respondents to report a younger age at first marriage (Table 27–1). There is a wide discrepancy between the results of calculations based on the current status and those based on retrospective declarations.

Increasingly, demographers use as a nuptiality indicator the median age at first marriage, which gives the age at which half the women in a given cohort are married. This median age may or may not include women who will never marry. The difference is far from negligible in countries, such as Botswana, where a high proportion of women never marry (van de Walle, 1996). If we include all women, the median age

[2] DHSs are primarily concerned with the conception risk, rather than the form of the union.

TABLE 27–1 Mean Age at First Union According to Two Calculation Methods

Country	Mean age, years (Hajnal method)	Reported mean age, years
Benin (1989), men	28.0	25.5 (Men aged 35–44)
Togo (1988), women	20.3	18.4 (Women aged 30–39)

Sources: Donadjé, Florentin and Tabutin, Dominique, 1991. Nuptialité et fécondité des hommes au Sud-Bénin, *in*: IUSSP Committee on Comparative Analysis of Fertility and University of Zimbabwe (eds), *Seminar on the course of fertility transition in Sub-Saharan Africa*, p. 1–32. Liège, IUSSP, multiple p. numbers (proceedings of Seminar in Harare, Zimbabwe, November 19–22, 1991);

Thiriat, Marie-Paule, 1998. *Faire et défaire les liens du mariage. Évolution des pratiques matrimoniales au Togo*. Paris, CEPED, 295 p. (Les Études du CEPED, no. 16).

at first marriage is 26.2 years; if we include only ever-married women, the median age falls to 24.5 years. In countries where nearly all women marry, like Kenya, the two indicators converge (20.6 years and 20.5 years, respectively; van de Walle, 1996).

These methodologic issues call for prudence in drawing conclusions and commenting on changes in marital practices in Africa.

2. Increase in Age at First Marriage

In most parts of Africa, marriage is characterized by early first union among women and the near absence of never-marrieds among both men and women. The United Nations (1988) estimates that the proportion of never-married women at age 50 in Africa never exceeds 6%. In the early 1980s, the mean age at first marriage for women ranged between 17.7 years (Ethiopia) and 26.4 years (Botswana); only in a few countries did it exceed 21 years. If the proportion of never-married men never exceeded 10%, men's mean age at first marriage is far older than women's: It ranges from 25.5 years (Ethiopia, 1981) to 30.8 years (Botswana, 1981). Georgia Kaufmann *et al*. (1988) also find an increase in age at first marriage in most countries and for both sexes. First marriages occur at older ages in urban areas than in rural areas, and urban dwellers are entering into union increasingly late (Hertrich and Pilon, 1997).

The compilation of results of several recent DHSs confirms this pattern. We can compare the change in median age of entry into first union for two different cohorts: women aged 40 to 44 years and women aged 25 to 29 years at the time of the surveys (Figure 27–1). In most surveyed countries, the median age at first marriage in rural areas changes minimally from one cohort to the other (about half a year in 15 years, from approximately 17 years to 17.5 years), except in countries such as Liberia, Kenya, and Sudan, where the age at first union rises more sharply. In urban areas, the variation is greater (about 1.6 years): The age at marriage rises from just under 18 years to over 19.5 years. However, in some countries, the difference between cohorts remains minimal. This is the case in Burkina Faso, Togo, and Côte d'Ivoire. By contrast, the increase exceeds two years in Cameroon, and reaches almost 4 years in Rwanda, 5.6 years in Senegal, and 7.4 years in Sudan. In urban areas, the gap between countries is still modest for the 40- to 44-year-old cohort. The median age ranges from as low as 15.5 years in Niger (and similar values in neighboring sub-Saharan countries such as Chad) to 20.2 years in Kenya and 20.4 years in Madagascar. For the younger cohort aged 25 to 29 years, the gap is wider between countries like Chad (16.2 years) or neighboring sub-Saharan countries (Niger: 16.6 years) and countries such as Sudan (23.2 years), Senegal (23.3 years), and Rwanda (23.9 years). The shift in marital behavior is more pronounced in certain countries and the disparities between countries are growing.

This increase in the age at marriage in urban areas is confirmed in the countries for which multiple surveys are available (Figure 27–2). The comparison concerns the cohort of women aged 25 to 29 years at the time of the survey. In most countries, the median age at first marriage increases: This is the case in Cameroon between 1988 and 1994, in Mali between 1986 and 1996, in Niger between 1992 and 1998, and above all in Senegal between 1986 and 1997. Some countries (for example, Ghana) display virtually no measurable change; occasionally, as in Madagascar, the age of entry into first union decreases. Elsewhere, as in Senegal, the figures corroborate the sharp increase in age at first marriage of urban women[3] (from 18.8 years to 23.3 years in an 11-year period), in contrast with a minimal variation in rural areas (from 16.5 years to 17.7 in the same period). Young women's early entry into union—often presented as a major characteristic of nuptiality in Africa—seems to persist in rural areas, but is heavily challenged in urban ones.

Later first marriage of women is often attributed to a set of factors directly linked to urbanization: higher school enrollment for girls, participation of women in the labor force, particularly in the modern sector of the

[3] Interestingly, in a context of urbanization, increased school enrollment, and rising payroll employment, the age at marriage has moved in the opposite direction in France. Between 1821 and 1945, the mean age at marriage fell from 28.7 years to 25.2 years for men and from 26.1 years to 22.8 years for women (Segalen, 1981).

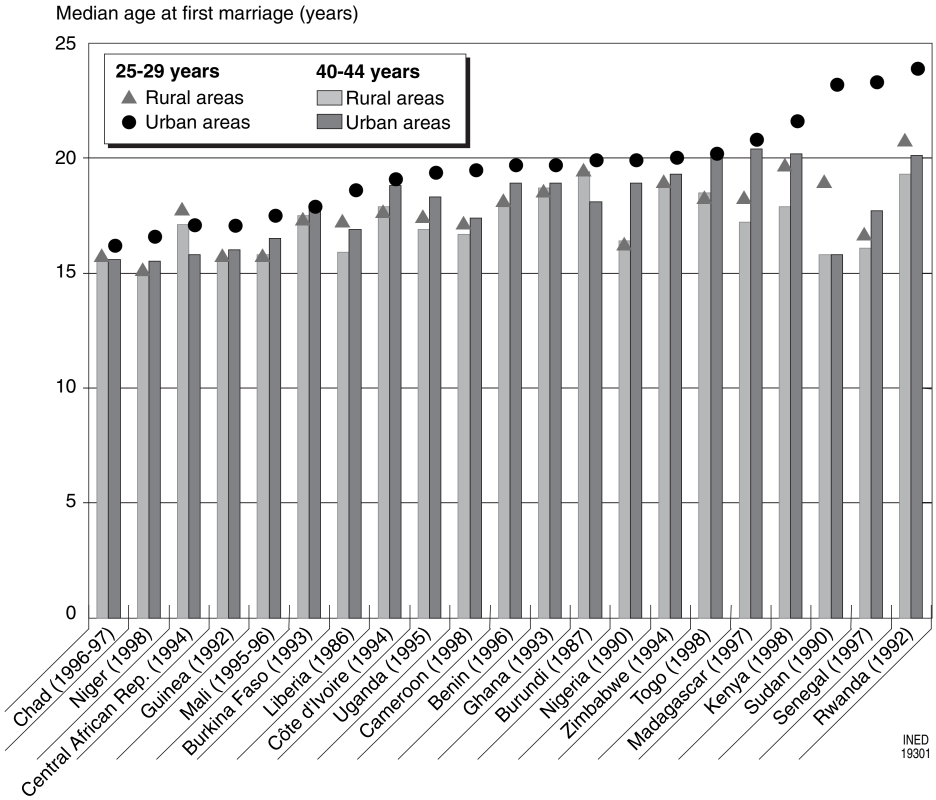


FIGURE 27–1 Change in median age at first marriage in urban area and rural area for two cohorts of women. The bars show median age at first marriage of women aged 40–44 at the time of the survey; the triangles and points are age at first marriage of women aged 25–29. Countries are ranked by median age at first marriage in rural areas for women aged 40–44. Source: Demographic and Health Survey data.

economy (Isiugo-Abanihe *et al.*, 1991), and the attendant difficulty of reconciling school or work outside the home with running a household. The adoption of new behaviors, new attitudes to intimate relationships before marriage, and new urban lifestyles are also major factors in women's postponement of first marriage (Lesthaeghe *et al.*, 1989). Nor should we overlook the existence of new legal rules on marriage and the family (in particular, the adoption of a family code), which offer women new rights.

There are very few similar results on the determinants of first marriage among men. The material difficulties facing most African urban males (i.e., problems of employment, housing, and bride wealth formation) are among the factors most often invoked to explain the increase in age at marriage among men (United Nations, 1988; Antoine and Nanitelamio, 1991; Donadjé, 1992; Djiré, 1993). This change in male nuptiality timing certainly accentuates the increase in age at first marriage among women.

The United Nations (1988) has developed a conceptual framework linking the timing and intensity of first marriage to socioeconomic and cultural factors via demographic factors, marriage standards, and individual factors. To perform their comparative analyses of nuptiality regimes in African ethnic groups, Ron Lesthaeghe *et al.* (1989) have used a theoretical anthropologic framework in which the main explanatory variables are degree of dependence of livelihood on agriculture and female labor, lineage organization, transfer of inheritance to women and endogamy, sociopolitical organization, castes and social stratification, and female education.

In the aggregate, the increase in age at first marriage is not confined to women. It also concerns men, and the wide age gap between spouses at first marriage—far from narrowing—tends to remain stable.[4] The age gap is particularly well correlated with the practice of

[4] One of the main social consequences of the age gap between spouses concerns widowhood. The age gap at marriage combined with a slightly longer life expectancy increases the duration of widowhood. The mean duration of widowhood estimated by Tiemoko (1999) varies from 7 years in Tunisia and Burundi to over 15 years in Côte d'Ivoire and Guinea, 17 years in Senegal, and 18.4 years in Mali.

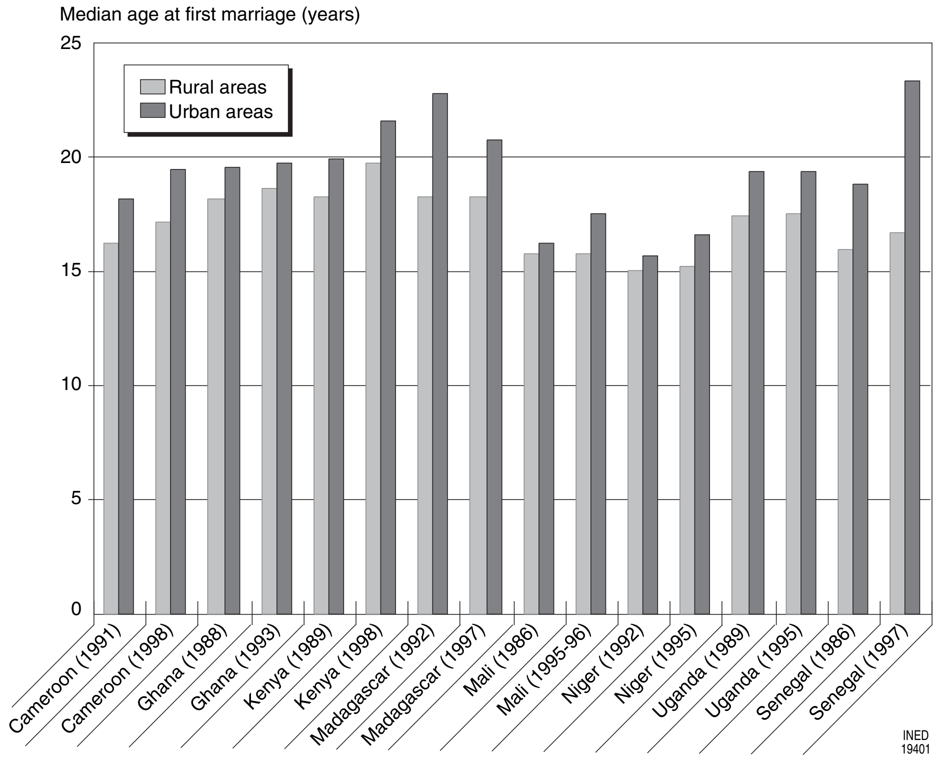


FIGURE 27–2 Comparison of change in median age at first marriage for women aged 25–29 between two surveys. Source: Based on data from Demographic and Health Survey.

polygamy and remarriage of women (Barbieri and Hertrich, 1999). The regions where women marry early and the age difference between spouses is greatest (over 9 years) coincide with those of high polygamy (chiefly West Africa). In Dakar, for example, women aged 20 to 24 years are on average 12 years younger than their spouses; the gap widens with the husband's age, particularly in the case of polygamy. Other studies confirm a progressive increase in age at first marriage and different marital behaviors in different neighborhoods of the same city (Antoine and Nanitelamio, 1991). In sub-Saharan urban areas, marriage, although delayed, remains the main concern of women in planning their future. Will women's later marriages alter the marriage market and undermine other marital practices such as polygamy? Is the African nuptiality transition driven by a tempo effect, with maximization of marriage by women still the rule? Thérèse Locoh (1988) notes that many factors combine to maximize marriage.

> Its frequency may or may not be rising, it may be inhibited for fairly long periods by the need to accumulate bride wealth (dowry), it may sometimes be hastened by the onset of pregnancy—but entry into union is a near-certain prospect for Africans. Never-married status is unacceptable and widow(er)hood or divorce without remarriage before a given age is condemned just as strongly in most countries.

The coexistence of these values with a divorce risk can surely point to favorable conditions for polygamy but doubtless cannot confirm the theory of a transition to monogamy.

II. DOES POLYGAMY STILL MATTER?

The term *polygamy* is often used as a synonym of *polygyny.* Polygyny is a form of marriage where the man has more than one wife at a time; the other form of polygamy, *polyandry*[5] (i.e., a woman with several husbands), is far rarer. Practiced in different periods in many societies, polygyny has never been the sole form of marriage in a society. As a rule, only rich and powerful men were or are able to support polygynous families; the majority of the population practiced

[5] Séraphin Ngondo A Pitshandenge (1996) describes an example of vanishing polyandry among the Bashilele of Zaire. Among Kotokoli (Togo) the practice of clan exogamy promotes relationships with official lovers in wives' native villages, all the more tolerated as the first husband is from another village and therefore resides elsewhere (Tchak, 1999).

monogamy.[6] This is not the case in several African countries where a large percentage of the population lives under a polygamous marital system. Although most married women may live in polygamous unions, polygamy invariably concerns only a minority of men in all the societies studied. Remember, however, that even in a polygamous society, a certain proportion of men and women will never enter into a polygamous union in their entire lives, but all members (more particularly, women) are exposed to the risk of living in this type of union.[7] In response to this threat, many women do not share their incomes with their husbands or may even incite their husbands to spend to thwart a new marriage. In a polygamous union, wariness toward husbands is compounded by wariness toward co-wives: The dependence and submissiveness of co-wives are aggravated by the rivalry and inequalities instituted between them (Hertrich and Locoh, 1999; Yana, 1997). The wife knows nothing about what goes on between her husband and his co-wife (or wives). Polygamy could be defined as a double (or even triple) monogamy: a serial monogamy for each wife (a few days a week) and concurrent monogamies for husbands (Fainzang and Journet, 1988).

Most attempts to explain polygamy are based on a ruralist perception of African societies, in the framework of a specific mode of production: a weakly mechanized subsistence economy, in which women play an important role as food producers. Polygamy in this context is viewed as cheap and profitable for the man (Boserup, 1970). This economic rationale is challenged by Jack Goody (1973), who points out that the highest polygamy rates are in West Africa whereas the highest proportion of women who farm is in East Africa. For Goody, the causes of polygamy are chiefly sexual and reproduction-related rather than economic and production-related. Polygamy allows the man to maximize his offspring. Polygamous union is a means of reconciling male sexuality with prohibitions such as *postpartum* abstinence. Polygamy is also a way to reconcile the group marital preferences and individual (and sentimental) preferences (Fainzang and Journet, 1988).

Other authors offer more political explanations; they emphasize the internal consistency of a marital system and a social organization in which power lies in the hands of elders (Meillassoux, 1975). Here, polygamy is regarded as a means of preserving the power of elders over juniors in societies where the access to women is controlled by elders. Abdoulaye Bara Diop (1985) sees several incentives to polygamy: It enables individuals to forge ties with several groups and confers a sociopolitical advantage. It provides an economic contribution, as the wife, by working or farming her personal fields, contributes to the upkeep of the household. Women sometimes seek this task-sharing, especially in rural areas: They may even propose a second wife to their husbands. Having children increases the household workforce and the prospects of being looked after by one's offspring in old age. Polygamy is also a factor of ostentation and prestige for privileged social categories.

Although Islam has consecrated polygamy in many African countries, it has never instituted it there. On the contrary, whereas traditional polygamy does not restrict the number of wives, Islam limits it to four. The Koran does not advocate polygamy, but restrains its practice: for example, the principle of strict equality between wives is a very difficult condition to meet.[8] Contrary to a widely held belief, polygamy is very rare in North African and Arab countries.[9] In Algeria, the proportion of polygamists among married men fell from 15% in 1886 to 1.5% in 1986 (Tabutin, 1974; Ajbilou, 1998). The percentage was about 3% in Egypt in the 1980s (Fargues, 1987), and 3.4% in Morocco in 1992 (Ajbilou, 1998). In these regions, the phenomenon is becoming less common if not disappearing altogether—for example in Tunisia, where polygamy has been abolished. When we look at the rather low polygamy rates in other Muslim countries such as Mauritania (where 18% of women live in polygynous unions[10]), and the high polygamy rates in some weakly Islamized countries such as Benin or Togo, one may question the true influence of Islam on polygamous behavior. In sub-Saharan Africa, Islam seems, if any-

[6] According to Jack Goody (1985), the Catholic Church had taken a stance on cohabitation as early as the fourth century, but he credits Pope Gregory VII (1073–1085) with the explicit condemnation of cohabitation and polygyny in Europe. The concubine was a member of the household and her children were entitled to the father's inheritance. The Church gradually abolished the traditional privileges of the concubine and her offspring.

[7] In many monogamous households the husband does not hesitate to threaten polygamy to ensure his wife's *docility* (Fainzang and Journet, 1988; Hertrich and Locoh, 1999).

[8] According to Ayatollah Mortadhâ Motahhary (n.d.), both Shiite and Sunni sources quote this *hadith* (narrative, utterance) of the prophet, "*Whoever has a two wives and does not treat them equally, by displaying an inclination [or preference] for one of them, will be treated on the Day of the Resurrection in such a manner that one side of his body will be dragged along the ground when he is taken to Hell*."

[9] As various authors point out (Fargues, 1986; Goody, 1985), polygamy in the Arab world seems to have primarily struck Western observers. In reality, however, polygamy was and still is practiced by only a small minority of the population in Middle Eastern countries.

[10] Fargues (1986) notes that polygamy in Mauritania concerns only a small proportion of Arab women: About 3% of married women there live in polygamous unions.

thing, to have channeled older traditional practices by legitimizing them.

Whatever its justifications, polygamy often enjoys legal status and recognition as a marital regime on a par with monogamy. Legislation in several countries allows spouses to choose between different marital regimes (Mali, Senegal, Togo). The abolition of polygamy elsewhere—as in Guinea (1962) or Côte d'Ivoire (1964)—has not eliminated it: 24% of married men in Côte d'Ivoire were polygamous in 1988 (Klissou 1995).

1. Demographic Conditions of Polygamy

The demographic mechanisms that make polygamy possible are relatively well identified (Pison, 1986; Pilon, 1991). Polygamy implies more married women than married men. But, contrary to a view still widely held—especially in Africa—polygamy is not due to a natural excess of women in the population, nor to an intake of women from neighboring societies, nor to a high percentage of never-married men. In fact the *excess of women* is the consequence of marital practices fostered by a specific demographic context: the two prerequisites for massive polygamy are a wide age gap at first marriage (with women marrying far earlier than men) and a young population pyramid. As men marry women 5 to 15 years younger than themselves, they have the impression of being in an unbalanced marriage market where women are more numerous. The age gap between spouses is far wider in a polygamous union than in a monogamous union (respectively, 14.5 years and 10.4 years in Mali and 11 and 7.6 years in Benin, for example) (Barbieri and Hertrich, 1999). Other factors abet the system: rapid and frequent remarriage of divorced or widowed women; a near-zero percentage of never-married women; a male mortality rate at older ages inducing widowhood for several women at once[11]; and a demographic growth that accentuates the imbalances between successive cohorts (Pison, 1986; Goldman and Pebley, 1989; Pilon, 1991).

Combined with specific kinship and marriage systems, these causes and demographic conditions of polygamy make it a key institution in the societies concerned—with multiple psychological, social, and economic implications. In societies where marriage remains urgent, competition between women is intensified by their relative excess in the marriage market. Some women who delay marriage too long run the risk of spinsterhood or of having to accept a first marriage with an already married man (Antoine and Nanitelamio, 1996). However, polygamy also implies marital competition between men of different cohorts (Clignet, 1997), as elders will lure women away from their juniors.

2. From East Africa to West Africa: Increasing Number of Polygamous Men

The practice of polygamy in sub-Saharan Africa displays major variations from one region or lineage system to another. The phenomenon is most prevalent in West Africa (the ratio M of married women to married men, adjusted for numeric imbalances due to migrations, generally exceeds 1.3), except among the nomadic populations of the Sahel (sub-Saharan savanna) and the Sahara (Lesthaeghe *et al*., 1989). The ratio is lower in Central and East Africa—where M exceeds 1.3 in few regions—and especially in Southern Africa, where M barely reaches 1.2. Moreover, polygamy is most widespread in patrilinear or bilinear societies.

For the preindependence period, van de Walle (1968) finds an incidence of polygamy among married men ranging from 17% in Congo (1956–1957) to 38% in Guinea (1954–1955). For the following period, according to Georgia Kaufmann *et al*. (1988), the proportion of married women living in polygamous union varied between 16.1% in Zaire[12] in 1975 and 51.8% in Senegal in 1976. Where data for different periods are available, the trend indicates a relative stability (Kaufmann *et al*., 1988). The high-polygamy zone is centered on West Africa, both in the coastal areas and the Sahel, where nearly one woman married in two lives in a polygamous union, even in urban areas. Polygamy is most frequent in Togo, Benin, and Burkina Faso, ahead of Senegal and Mali. In the countries bordering the Gulf of Guinea, in the same cultural area, polygamy is distinctly less common in the two English-speaking countries (Ghana and Nigeria) than in the neighboring French-speaking countries (Côte d'Ivoire, Togo, and Benin).

3. Polygamy Persists in Urban Areas

The DHS results show a disparity in polygamy levels between urban areas and rural areas (Figure 27–3). The low-polygamy zones lie in East and Southern Africa, where 11% to 31% of married women in rural areas live in polygamous unions, versus 7% to

[11] Levirate deserves special mention: This practice calls for a married man's widow(s) to be married by a younger brother or another relative of the deceased, who may already be married.

[12] Since renamed Democratic Republic of the Congo.

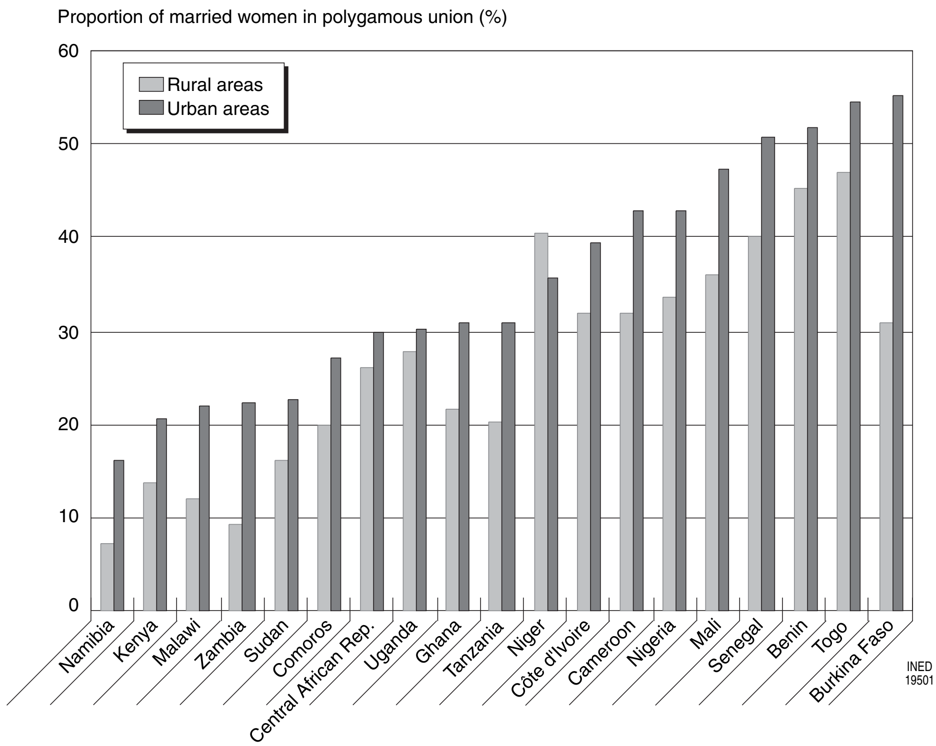


FIGURE 27–3 Proportion of married women in polygamous union in rural areas and urban areas. Source: Based on data from Demographic and Health Survey.

20% in cities and towns. Throughout sub-Saharan Africa, polygamy is far more prevalent in rural areas, most notably in southern Africa. The higher the frequency of polygamy in a country, the narrower the gaps between urban and rural areas. Only Niger exhibits a different profile: There, polygamy is more common in urban areas, where it is practiced in affluent social categories.

This overall analysis must, however, be complemented by examining the behavior of the younger generations, who—in principle—are the vectors of current change. Let us focus on the proportions of women in polygamous union among married women aged 20 to 24 years at the time of the DHS. In rural areas, we find the same evolution as described earlier for all women aged 15 to 49 in the aggregate. Polygamy increases from East to West, peaking at 45% of young women married to polygamous men in rural Burkina Faso and comparable proportions in rural Togo and Benin. By and large, the proportion of polygamous women is 25% lower in urban areas than in the countryside. A few countries, such as Burkina Faso and Zimbabwe, form an exception to the rule: There, the proportion of young rural women in polygamous union is one half that of the corresponding percentage in urban areas. Elsewhere—in Kenya, Togo, Ghana, Sudan, and Uganda—the proportions of young women living in polygamy in urban areas and rural areas are similar; in Niger and Burundi, the proportion in urban areas is actually higher.

In some countries, we have several observations from successive DHSs (Figure 27–4). These series show changes over time in a few cases only. The proportion of young women in polygamy decreases in Kenya, in rural Ghana, Mali, and to a substantial degree in Uganda, but the situation is unchanged (or displays a mild upward trend in urban polygamy) in Cameroon and Senegal.

Because of its structure and economic production system, the urban area—as well as the new ideals and lifestyles prevailing in it—should have had a constraining effect on polygamy (Gendreau and Gubry, 1988; Chojnaka, 1980). The media-propagated aspiration to Western lifestyles and family values could have entailed the gradual disappearance of polygamy in urban areas (Burnham, 1987; Clignet, 1987; Lesthaeghe *et al.*, 1989). As we have just seen, however, that has not happened: Urbanization does not appear to be an obstacle to polygamy. Polygamy can actually promote new distributions of activities in urban households. In Mali, for example, according to Marcoux (1997), women in polygamous households are more likely to engage in gainful activities, as they are freed from some domestic chores.

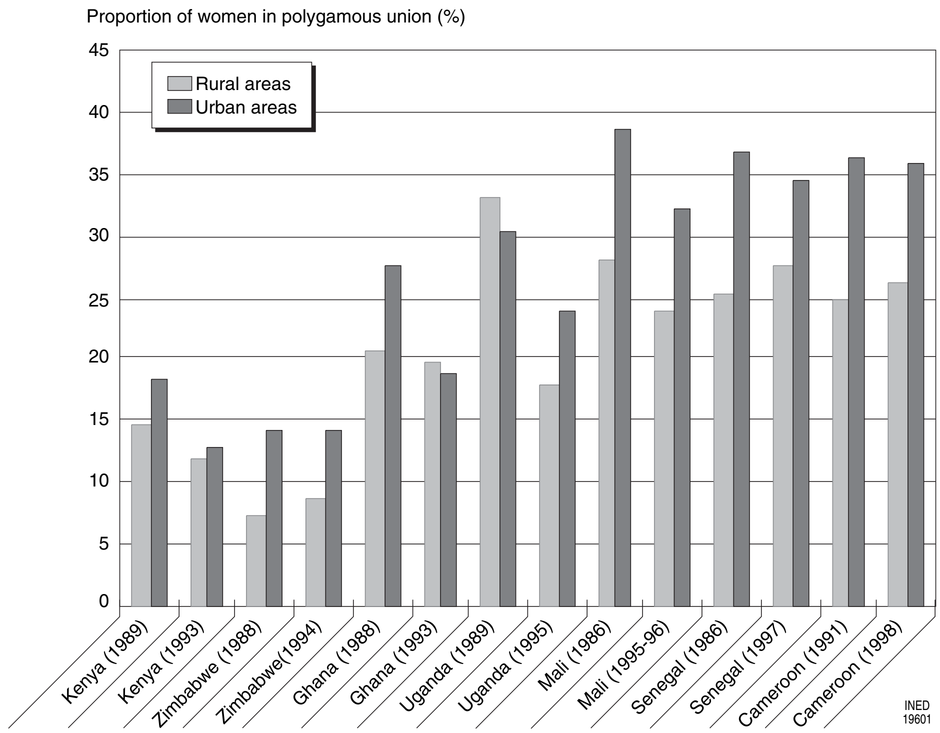


FIGURE 27–4 Proportion of women living in polygamous union among women aged 20–24 currently living in union. Source: Based on data from Demographic and Health Survey.

Back in the 1950s, Paul Mercier (1960) had observed the same phenomenon in Senegalese urban centers such as Dakar and Thiès: Urban life did not entail a rapid decrease in the polygamy rate. Mercier identified two phases in the urban evolution of the phenomenon:

> In a first phase, the influence of the rural tradition favorable to polygamy apparently remains intact, and would seem to be fully effective once individuals reach a certain income level.

In a second phase, aspirations to another lifestyle would appear to flourish and promote monogamy. The phases conducive or unfavorable to polygamy in an urban area overlap (Mercier, 1960). Luc Thore (1964), who is also very nuanced in his analyses of polygamy, stresses the complexity of the phenomenon, noting the coexistence of ambivalent attitudes: an opposition to polygamy on grounds of principle and an endorsement of the practice in actual behavior. Examining Zaire, Séraphin Ngondo A Pitshandenge (1992) notes that polygamy, hitherto characteristic of rural areas, has spread to several major cities, "The intrusion of polygamy in the urban area may be regarded as a revolutionary development" as it has occurred in spite of the structural hostility of the urban area (housing shortage, legal discrimination, respectability of monogamy, ban on polygamy by Judeo-Christian religions, etc.). The author suggests Zaire may be experiencing a ruralization of urban behaviors. He also ventures the hypothesis that (at least in central Africa) polygamy is more widespread among young generations less influenced by the colonial model.

At the individual level, research has found few discriminant socioeconomic characteristics of urban males in regard to polygamy. All social categories are concerned (Clignet, 1975 and 1987; Timæus and Reynar, 1998). Neither the husband's educational attainment, nor the fact of working in the modern sector of the economy, nor even ethnic-group identity seems to determine the practice of polygamy in African cities. In Dakar and Bamako, for example, most men are potentially polygamous (Antoine and Nanitelamio, 1996). Not all men become polygamous—about one in two will at some point in their lives—but no socioeconomic factor can enable us to predict who will become polygamous. However, the spells of nonemployment and economic precariousness sharply reduce entry into polygamous union in Dakar: The odds of becoming polygamous are divided by three (Antoine *et al.*, 1998). In Bamako, two socio-occupational categories are more likely to practice polygamy: small-business owners and self-employed men employing payroll workers, for example, "Two groups whose methods of accumulation and wealth creation are informed by the principles of a family-based economy" (Marcoux and Piché, 1998).

In sum, polygamy persists both in the countryside and in towns, and its dynamics deserve explanation.

4. The Demographic Perception of Polygamy

Most measures of polygamy are flawed to a certain extent as they basically rely on period data and therefore provide only a static, frozen image of the practice.[13] The demographic perception of polygamy is still strongly influenced by the male perspective. The intensity of the phenomenon, measured by the *average number of wives per polygamous husband*, generally varies between 2.0 and 2.5. The most prevalent form is bigamy, which concerns over 80% of polygamous men. The average number of wives per married man, usually ranging between 1.02 and 1.50, increases with age (Gendreau and Gubry, 1988). The *polygamy rate* commonly used gives the proportion of married men in polygamous union. This index is most often below 30%. On the women's side, however, the numbers tell a very different story: For example, whereas slightly over one third (35%) of married men in the Moba-Gurma ethnic group of northern Togo were polygamous in 1985, more than half (57%) of married women were living in polygamous unions (Pilon, 1991). An age breakdown of the data on the same population shows that, by age 25, more than half the women were living in polygamous unions; by age 50, the proportion had reached about 70%. Among men, the proportion of polygamists does not rise above half until age 50, reaching 60% or so. Obviously these figures may vary from one population to another, but without invalidating these gender differences, which reflect very different dynamics and life experiences.

Three indicators summarize the intensity of polygamy: the proportions of married women and men in polygamous union and the mean number of wives per polygamist. As an example, Table 27–2 reports these indicators for two cities: Bamako and Dakar.

These indicators vary over time, as illustrated in selected data on Dakar. Women aged 50 to 54 years had experienced, on average, 1.71 unions in 1955 (Service de la Statistique, 1962) and slightly under 1.63 in 1989

[13] For example, Pilon (1991) has determined that period indicators understate the intensity of polygamy by nearly 25% for the Lobi-Dagara of Burkina Faso and the Moba-Gourma of northern Togo—at all ages. Donadjé (1992) makes the same observation about Benin. According to him, the underestimation varies from 18% for men aged 20 to 54% for men aged 55, especially owing to the greater frequency of changes in marital status at older ages.

TABLE 27–2 Polygamy Indicators for Dakar and Bamako

Indicators	Dakar	Bamako
Proportion of married women in polygamous union (%)	66.3	64.2
Proportion of polygamists among married men (%)	28.1	25.1
Mean number of women per polygamist (at age 50)	2.2	2.4

Source: Antoine, Philippe, Djiré, Mamadou, and Nanitelamio, Jeanne, 1998. Au cœur des relations hommes-femmes: polygamie et divorce, in Philippe Antoine, Dieudonné Ouédraogo, and Victor Piché (eds), *Trois générations de citadins au Sahel*, p. 147–180. Paris, L'Harmattan, 281 p.

(Antoine, Djiré, and Nanitelamio, 1998). In 1989, 47% of Dakar women aged 50 to 54 were living in polygamous union. The proportion is highest in that age group, decreasing thereafter because of widowhood. Among men aged 55 to 59 years, the proportion of polygamists rose from about 30% in the 1955 census to 40% in the 1976 census and 45% in 1989. In 1955, men aged 60 years had experienced an average 2.6 marriages; by 1989, the number had fallen to 2.2; by contrast, at the same age, men in 1955 had an average of 1.45 wives, whereas in 1989 they had 1.70. This variation denotes a lesser marital mobility and perhaps a greater prevalence of polygamy in Dakar in recent years.

Another way to assess polygamy dynamics is to compare the mean number of marriages and the mean number of spouses for men and women. Within the same male and female cohorts aged over 45 at the time of the survey (i.e., the 1930–44 birth cohorts), we have examined two subpopulations: persons who have experienced at least one period of polygamous union in their lives and those who have stayed monogamous. The curves in Figure 27–5 show the mean number of marriages and the mean number of spouses at different ages. At age 15, no male is married whereas females have already experienced an average 0.34 marriages (about one young woman in three is already married). Through divorces and remarriages the number-of-marriages curve diverges from the number-of-spouses curve: At each age, individuals have experienced more marriages than they have had spouses. The gap is particularly wide for men who have been polygamous at some point in their lives. At age 50, they have experienced an average of 2.44 marriages, but have had an average of 1.96 wives—the reason being that some of these polygamists have reverted to monogamy. As the chart shows, male polygamy comes late (after 40) and is most often

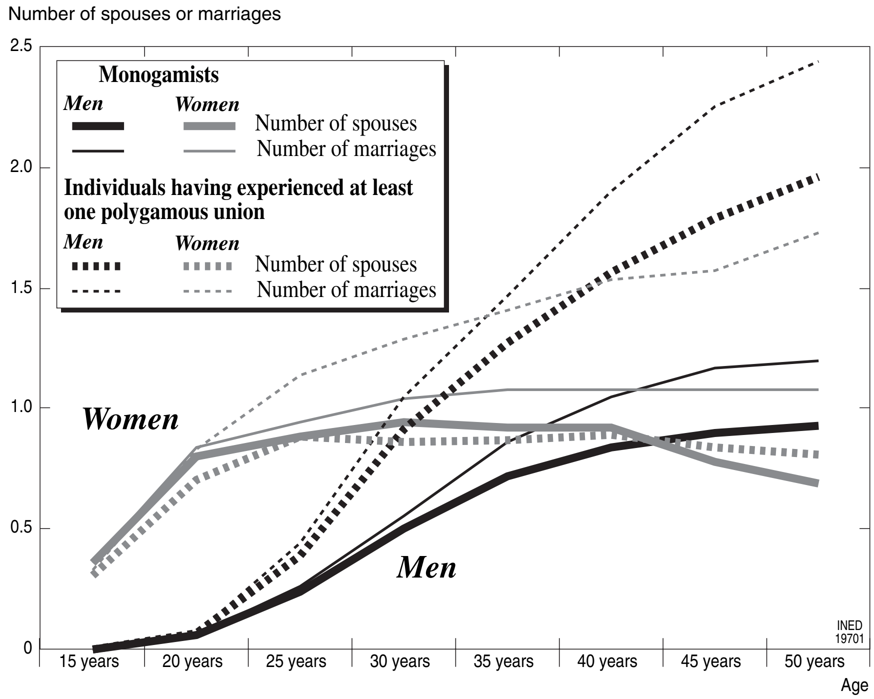


FIGURE 27–5 Comparison of number of unions and number of spouses by age for male and female 1930–1944 birth cohorts (Dakar, 1989). Source: Based on data from Dakar survey (IFAN-ORSTOM).

confined to bigamy. In the end, owing to divorce, men who are polygamous in the course of their lives fail to maximize the number of their wives. Polygamists, on average, have only 1.96 wives. Polygamy mainly accelerates the turnover of spouses. Among monogamists, the gap between the number of marriages and the number of spouses is smaller, suggesting a greater stability in monogamous unions than in polygamous ones.

We find a similar pattern on the female side. Women who have lived in a polygamous union at some time or other in their lives have had far more spouses than women who have remained monogamous (at 50 years, the number of marriages is 1.73 and 1.08, respectively). Monogamous women divorce far less often than women in polygamous unions.[14] By contrast, the number of spouses between ages 25 and 40 holds steady at around 0.9, indicating that at no age are all women married.

Marital status thus evolves throughout individual life courses; it is most notably subject to the random factor of mortality and dependent on the relationships between spouses. There is a considerable difference in the individual marital life courses of men and women: Although every polygamist was initially monogamous, a woman may shift directly from never-married status to polygamy by marrying an already married man. A man (or woman) in a monogamous union, widowed or divorced at time *t*, may have lived in a polygamous union beforehand. A fairer measure of polygamy should take this time dimension into account. Whereas a married, widowed, or divorced man will easily be able to remarry a never-married woman, the likelihood of a widowed or divorced woman entering into a new marriage with an already married man increases with her age. It is not uncommon for a woman to divorce her husband who has become polygamous and subsequently marry an already married man. These examples suggest the great complexity of these marital event histories and their specificity depending on whether we view them from the male or female perspective.

Figure 27–6 illustrates the example of two cohorts—one female, one male—of Dakar residents aged 45 to 59 years. The *marital dilemmas*, as Rémi Clignet (1997) puts it, differ for men and women. However, the situation seems more complex than he describes. According to him, the alternative for men is restricted to *an additional spouse or divorce*, whereas women have more options. Men, as well, have several possible choices, but men have greater control over the outcome than women, especially in the case of polygamy.

[14] Polygamy fosters divorce (see below); also, however, a divorced woman has hardly any choice other than to marry a polygamist.

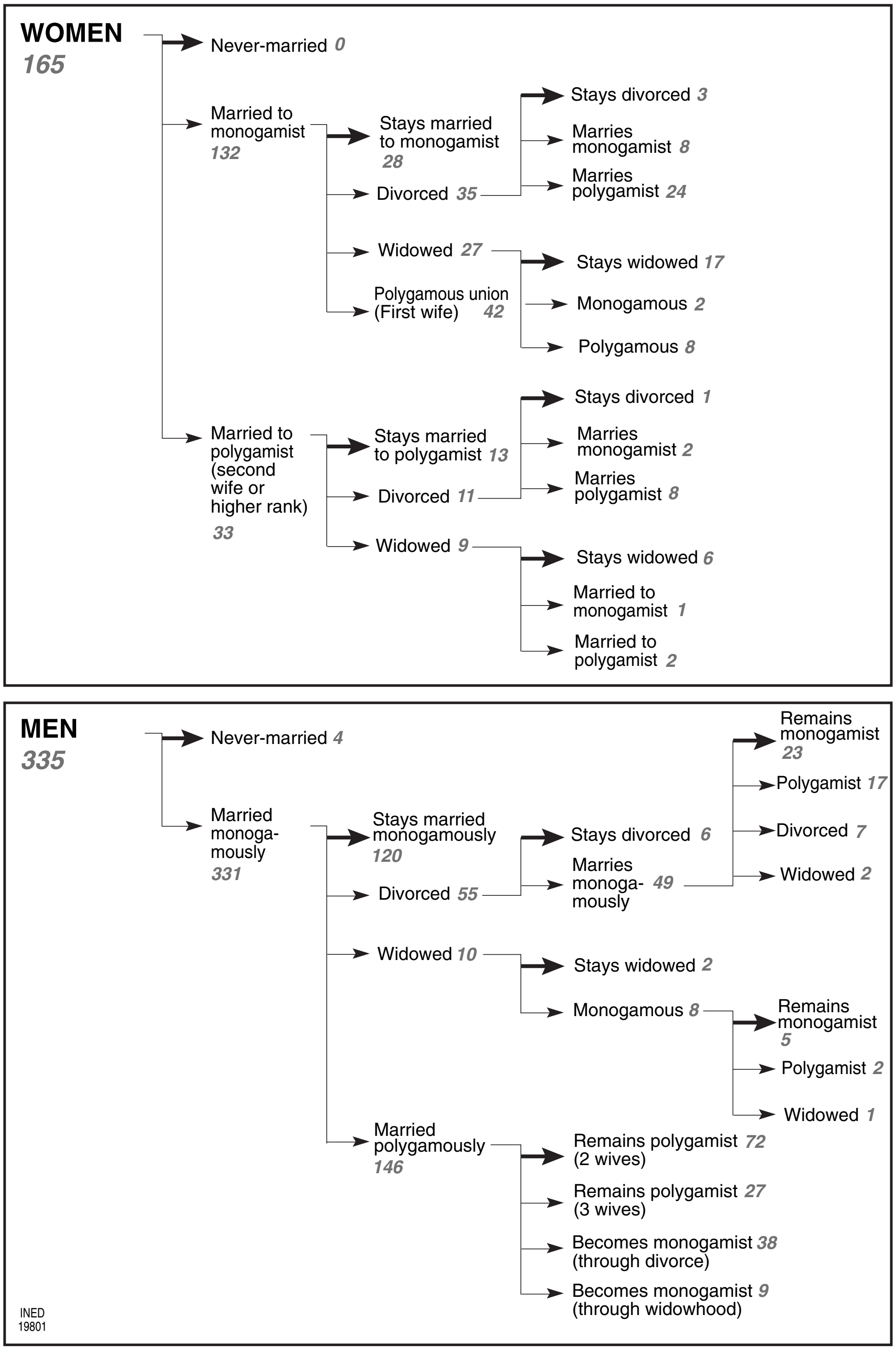


FIGURE 27–6 Female and male marital dilemmas (individuals aged 45–59). Source: Based on data from 1989 IFAN-ORSTOM event-history survey.

In Dakar, for men and women alike, divorce status is often only transitional. Among the 331 married men in the study (who are therefore monogamous at the start of their marital life courses), 165 have been polygamous at one time or other in their lives, and 166 have never been. Among the 165 women, at least 107 have been in a polygamous union at some point in their lives. However, that figure is understated: 35 divorcees and 27 widows began their marital life courses with monogamous unions, but we do not know if these subsequently became polygamous.[15] Women—partic-

[15] It is difficult to reconstruct the marital life courses of women's spouses from women's own accounts (see Chapter 26).

ularly divorcees—frequently remarry. Of the 46 divorcees in our sample, only four remained so; 42 remarried, of whom 32 joined a polygamous household. Divorce seems to be a form of resistance to forced polygamy, whereas remarriage leads to freely chosen polygamy.

5. Polygamy and Divorce

Regions with high polygamy are also characterized by more frequent dissolutions of unions, followed by rapid remarriages. However, the variations in union instability are still fairly underdocumented in Africa. The proportion of women who stay divorced is low in most countries. For example, in Burkina Faso—a country with a high polygamy rate—it varies between 1.8% at 20 to 24 years and 1% at 45 to 49 years. But despite this tiny proportion of women who stay divorced, nearly one in four women older than 35 years has already experienced several unions (DHS, Burkina Faso), a sign of rapid remarriage. As polygamy allows more frequent remarriage, it probably leads to a lower proportion of divorced women. In Ghana, by contrast, the proportion of female divorcees rises with age, from 5.2% at 20 to 24 years to 12.8% at 45 to 49 years (DHS, Ghana). The situation thus varies according to country, culture, legislation, and women's status.

We know nearly nothing about the changes in the frequency of divorce or about its determinants (Kaufmann *et al*., 1988). It is estimated that about 40% of women who marry before age 20 become widowed or divorced before age 50 (Lesthaeghe *et al*., 1989). Among women, the frequency of divorce and separation is higher in the first 10 years of marriage and especially at the youngest ages (Smith *et al*., 1984).[16] Union terminations are often concealed by rapid remarriages. As divorce undermines the alliances between families formed by the marriage, it is perceived as disruptive and hence treated as an event not to be talked about (Locoh and Thiriat, 1995). Remarriage becomes more difficult among women past childbearing age. Gainful employment is an important factor in divorce. It creates an opportunity for the woman to be financially independent. She can thus ensure her own livelihood and that of her children in the event of divorce (McDonald, 1985; Burnham, 1987). For women who have held a job before marriage and have ceased to do so after marriage, the risk of divorce is lower: Economic insecurity probably deters these women from divorcing. Divorce is sometimes a social-promotion strategy for women thanks to remarriage with a more affluent man. Independent of these material considerations, a woman who remarries gains more autonomy vis-à-vis her new husband (Dial, 1999).

The arrival of a second wife emerges as the main cause of divorce by the first wife in Dakar (see Chapter 26). Yet, according to some anthropologic studies, several factors may keep the first wife from divorcing. The first marriage often seals an alliance between families (Fainzang and Journet, 1988). Even if the arrival of a second wife is not a legitimate cause of divorce, first wives do seem to divorce more than second wives in urban polygamous unions. The first marriage is often a prescriptive marriage, which the woman sometimes experiences as an obligation from which she tries to free herself through divorce (Locoh and Thiriat, 1995).

A risk of divorcing or separating from the first or second wife may be deemed to exist from the very outset of polygamous union. There is a suitable method of analysis for measuring competing risks of this type: the Aalen estimator[17] (Trussell *et al*., 1992; Courgeau and Lelièvre, 1989). In the three cities studied (Bamako, Dakar, and Yaoundé) where event-history data is available for such analysis, the trend is always the same: divorce from a first wife occurs sooner than a divorce from a second wife (Figure 27–7). The survey does not make it possible to determine who initiates the divorce, but it is highly likely to be the wife: Faced with the *fait accompli* of the arrival of a second wife, the first will decide to leave her husband. Polygamy is not necessarily the main cause. Rather, divorce is triggered by unequal treatment or rivalry with the second wife.

6. Trends in the Phenomenon

The theory of modernization called for a fertility transition and a convergence toward the nuclear-family model. Within this rationale, many authors predicted a gradual disappearance of polygamy by the 1960s—in particular owing to urbanization, higher school enrollment, an opening to Western values, and so on. But the data available show that, overall, the institution continues to thrive. The links with urbanization, educational attainment, and socioeconomic status sometimes prove ambivalent—an ambivalence that would surely be accentuated by the inclusion of informal unions (mistresses, *deuxièmes bureaux*,[18] and

[16] In Senegal, the percentage of unions terminated in the five first years of marriage is as high as 17% (Smith *et al*. 1984).

[17] The interpretation chart for this estimator is constructed by comparing the slopes of the curves of instantaneous cumulative probabilities, which gives an idea of the intensity of each risk involved at any given moment.

[18] Literally, *second offices*, a euphemism for *outside wives* in French-speaking Africa.

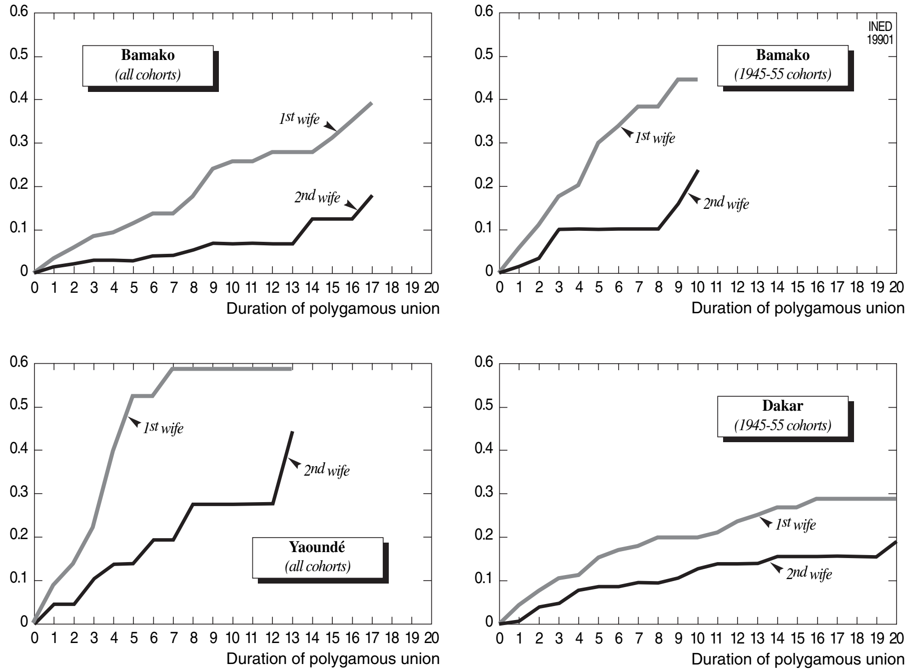


FIGURE 27–7 Divorce or separation from first or second wife by duration of polygamous union (in years). Source: Based on data from event-history surveys in Bamako, Dakar, and Yaoundé.

other arrangements) (Lacombe, 1983; Wa Karanja, 1987).

The World Fertility Survey (WFS) and DHSs give us several fairly reliable statistics for a number of countries. Figure 27–8 illustrates five countries where we can track the changing proportion of married women in polygamous unions. The sharpest fall is recorded in Kenya, the country with the lowest proportion of polygamists in the 1970s. The rate there fell by more than a third, from 29.5% in 1977 to 19.5% in 1993. In Ghana as well, polygamous unions have declined, albeit more slowly. In Côte d'Ivoire during the same period (1975–1994), the rate decreased by only 5 points, from 41.4% to 36.6%. In Mali and Senegal, no significant variation was measured (Antoine and Pilon, 1998).

The English-speaking countries are thus the ones where polygamy seems to be falling most steeply. However, this measure of polygamy masks certain changes. The indicator chosen concerns only married women. If we take into account all women aged 15 to 49 years regardless of their marital status (married or not) and if we calculate the proportion of polygamous women in that aggregate, the changes are more conspicuous. In Kenya, between 1977 and 1993, the ratio of women in polygamous unions to the total number of women aged 15 to 49 years fell from 21% to 11%. Even in a country such as Benin, where the ratio of women married to polygamists to all married women has varied very little since the 1960s (52% in 1961, 51% in 1992, and 49.6% in 1996), the proportion of women in polygamous unions fell from 43% to 34% of all women aged 15 to 49. Later marriage and divorce explain this decline in the frequency of polygamy when the ratio denominator consists of all women regardless of their marital status.

Polygamy plays an important role in the family strategies for controlling the married couple. When the spouses choose monogamy, they are subject to pressure from their family circle. Women—who have more reservations toward polygamy—often exhibit ambiguous attitudes and behaviors that reinforce this institution, particularly in the sub-Saharan countries. The deep change in population structure that will soon result from the decline in fertility will constitute the first major inhibitor of polygamy by curbing the present imbalance in the marriage market.

CONCLUSION

The decline in polygamy has been predicted by several anthropologic theories. By increasing the female age at marriage, education was supposed to contribute to the increase in the male population in the marriage market, entailing a shift in the numerical balance with adverse effects on polygamy.

In recent years, a number of changes have been taking place in African societies. In particular, worsen-

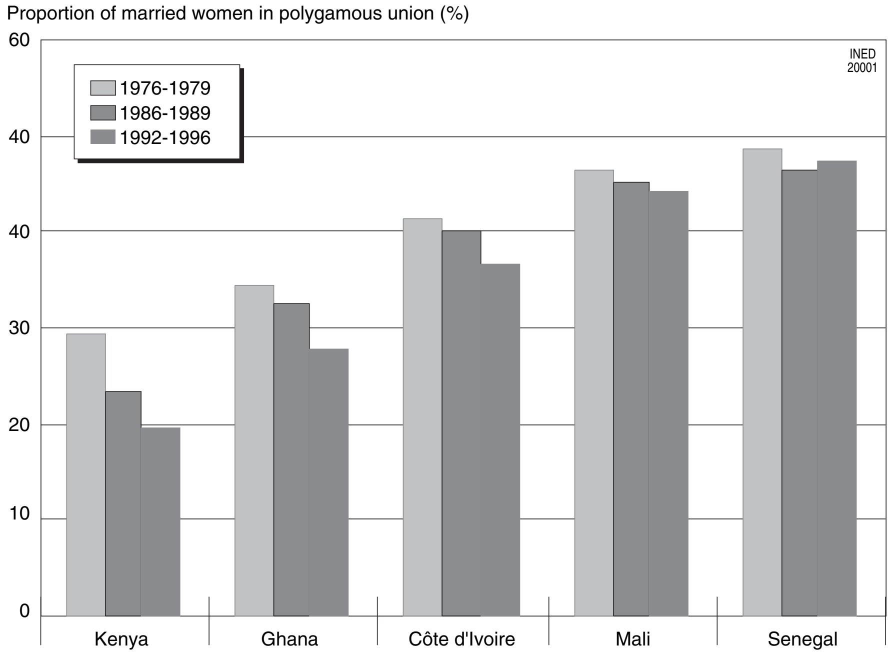


FIGURE 27–8 Proportion of married women in polygamous unions.

ing economic conditions have led to a sizable increase in age at first marriage and have made unions more vulnerable. The age gap between men and women at first marriage is shrinking in nearly all countries. The aspiration to large families has waned, and there are growing signs of a fertility transition in countries such as Kenya, Botswana, Senegal, and Côte d'Ivoire.

These significant changes in the levels of several nuptiality determinants may therefore entail either a shift in the prevalence of polygamy or a modification of other factors. If the age gap at marriage narrows, polygamy can survive only through an offsetting increase in the male never-married rate. On the other hand, an abandonment of polygamy would have tremendous implications for the organization and functioning of the societies concerned—starting with a challenge to the power of men over women, of elders over their juniors. As we can see, the present changes are bound to affect polygamy, but we still know little about the ways in which they will do so, in what time frame, and through what mechanisms.

Acknowledgment

Translated from the French by Jonathan Mandelbaum.

References

Ajbilou Aziz, 1998. *Analyse de la variabilité spatio-temporelle de la primo-nuptialité au Maghreb (1970–1980).* Louvain-la-Neuve, Academia Bruylant, 343 p.

Antoine Philippe, Djiré Mamadou, and Nanitelamio Jeanne, 1998. Au cœur des relations hommes-femmes: polygamie et divorce, in Philippe Antoine, Dieudonné Ouédraogo, and Victor Piché (eds), *Trois générations de citadins au Sahel,* p. 147–180. Paris, L'Harmattan, 281 p.

Antoine Philippe and Nanitelamio Jeanne, 1991. "More single women in African cities: Pikine, Abidjan and Brazzaville," *Population, English Selection,* no. 3, p. 149–169.

Antoine Philippe and Nanitelamio Jeanne, 1996. Can polygamy be avoided in Dakar? *in:* Kathleen Sheldon (ed.), *Courtyards, markets, city streets: urban women in Africa*, p. 129–152. Boulder, Westview Press, 342 p.

Antoine Philippe and Pilon Marc, 1998. La polygamie en Afrique: quoi de neuf? Paris, *La Chronique du CEPED*, no. 28, Jan.–March 1998, p. 1–4.

Barbieri Magali and Hertrich Véronique, 1999. Écart d'âges entre conjoints et transition de la fécondité en Afrique sub-Saharienne, *in:* UEPA and NPU (eds), *La population africaine au 21e siècle*, vol. 2, p. 163–211. Dakar, 650 p. (Third African Conference on Population, Durban, South Africa, December 6–10, 1999).

Boserup Ester, 1970. *Women's role in economic development*. Londres, Allen and Unwin, 283 p.

Burnham Philip, 1987. Changing themes in the analysis of African marriage, *in:* David Parkin and David Nyamwaya (eds), *Transformations of African marriage*, p. 37–54. Manchester, Manchester University Press, 350 p. (International African Seminars, New Series, no. 3).

Chojnacka Helen, 1980. Polygyny and the rate of population growth, *Population Studies*, no. 1, p. 91–107.

Clignet Rémy, 1975. Distribution et fonctions de la polygamie en milieu africain: ses effets sur les rôles familiaux, *Psychopathologie Africaine*, vol. 11(2), p. 157–177.

Clignet Rémy, 1987. On sait que la polygamie est morte: vive la polygamie, *in:* David Parkin and David Nyamwaya (eds), *Transformations of African marriage*, p. 199–209. Manchester, Manchester University Press, 350 p. (International African Seminars, New Series, no. 3).

Clignet Rémy, 1997. *Un sociologue entre Afrique et États-Unis. Trente ans de terrains comparés.* Paris, Karthala-ORSTOM, 227 p.

Courgeau Daniel and Lelièvre Eva, 1989. *Analyse démographique des biographies*. Paris, INED, 268 p.

Davis Kingsley and Blake Judith, 1956. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol. 4, p. 211–235.

Dial Fatou Binetou, 1999. *Divorce et promotion sociale et économique de la femme*. Saint-Louis (Senegal), Université Gaston Berger, 96 p. + 36 p.

Diop Abdoulaye Bara, 1985. *La famille wolof: tradition et changement*. Paris, Karthala, 262 p.

Djiré Mamadou, 1993. *Un aspect de la dynamique de la nuptialité en milieu urbain: l'entrée des hommes en première union à Dakar (Sénégal).* Montreal, Université de Montréal, Département de Démographie, 118 p. + appendices (master's dissertation).

Donadjé Florentin, 1992. *Nuptialité et fécondité des hommes au Sud-Bénin: pour une approche des stratégies de reproduction au Bénin*. Louvain-la-Neuve, Université Catholique de Louvain, Institut de Démographie, Académia, 222 p.

Donadjé Florentin and Tabutin Dominique, 1991. Nuptialité et fécondité des hommes au Sud-Bénin,*in:* IUSSP Committee on Comparative Analysis of Fertility and University of Zimbabwe (eds), *Seminar on the course of fertility transition in Sub-Saharan Africa*, p. 1–32. Liège, IUSSP, multiple p. numbers (proceedings of Seminar in Harare, Zimbabwe, November 19–22, 1991).

Dozon Jean-Pierre, 1986. En Afrique, la famille à la croisée des chemins, *in:* André Burguière, Christiane Klapisch-Zuber, Martine Segalen, and Françoise Zonabend (eds), *Histoire de la famille. 3. Le choc des modernités*, p. 393–442. Paris, Armand Colin, 736 p.

Dumestre Gérard and Touré Seydou, 1998. *Chroniques amoureuses au Mali*. Paris, Karthala, 312 p.

Fainzang Sylvie and Journet Odile, 1988. *La femme de mon mari. Anthropologie du mariage polygamique en Afrique et en France.* Paris, L'Harmattan, 172 p.

Fargues Philippe, 1986. Le monde arabe: la citadelle domestique, *in:* André Burguière, Christiane Klapisch-Zuber, Martine Segalen, and Françoise Zonabend (eds), *Histoire de la famille. 3. Le choc des modernités*, p. 443–484. Paris, Armand Colin, 736 p.

Fargues Philippe, 1987. La démographie du mariage arabo-musulman: tradition et changement, *Machreb-Machrek*, no. 116, Apr.—June 1987, p. 59–73.

Gendreau Francis and Gubry Françoise, 1988. La nuptialité en Afrique: niveaux, tendances et caractéristiques socio-économiques, *in:* UIESP (ed.), *Congrès africain de population*, *Dakar* p. 5.1.1–5.1.18. Liège, UIESP, multiple p. numbers.

Goldman Noreen and Pebley Anne, 1989. The demography of polygyny in sub-Saharan Africa, *in:* Ron Lesthaeghe (ed.), *Reproduction and social organization in sub-Saharan Africa*, p. 213–237. Berkeley, University of California Press, 556 p.

Goody Jack, 1973. Polygamy, economy and the role of women, *in:* Jack Goody (ed.), *The character of kinship*, p. 175–190. London and Cambridge, Cambridge University Press, 251 p.

Goody Jack, 1983. The development of the family and marriage in Europe. Cambridge, Cambridge University Press, 308 p.

Hajnal John, 1953. Age at marriage and proportions marrying, *Population Studies*, no. 2, p. 111–36.

Hertrich Véronique, 1996. *Permanences et changements de l'Afrique rurale: dynamiques familiales chez les Bwa du Mali.* Paris, CEPED, 570 p. (Les Études du CEPED, no. 14).

Hertrich Véronique and Locoh Thérèse, 1999. *Rapports de genre, formation et dissolution des unions dans les pays en développement.* Liège, UIESP, 46 p. (Gender in Population series).

Hertrich Véronique and Pilon Marc, 1997. *Transitions de la nuptialité en Afrique*. Paris, CEPED, 27 p. (research report no. 15).

Isiugo-Abanihe Uchec, Akin Ebigbola J., and Adewuyi Alfred A., 1991. Urban nuptiality patterns and fertility in Nigeria, *in:* UEPA (ed.), *Femme, famille et population. Actes de la conférence de Ouagadougou, Burkina Faso 24–29 avril 1991*, vol. 1, Communications sollicitées, p. 205–228. Dakar, UEPA, 647 p.

Kaufmann Georgia, Lesthaeghe Ron, and Meekers Dominique, 1988. Les caractéristiques et tendances du mariage, *in:* Dominique Tabutin (ed.), *Population et sociétés en Afrique au sud du Sahara*, p. 217–247. Paris, L'Harmattan, 551 p.

Klissou Pierre, 1995. *La polygamie au Bénin. Une approche régionale des tendances et des déterminants*. Louvain-la-Neuve, Université Catholique de Louvain, L'Harmattan, Académia Bruylant, 257 p.

Lacombe Bernard, 1983. Le deuxième bureau. Secteur informel de la nuptialité en milieu urbain congolais, *Stateco*, no. 35, p. 37–57. Paris, INSEE.

Lallemand Suzanne, 1985. *L'apprentissage de la sexualité dans les contes d'Afrique de l'Ouest*. Paris, L'Harmattan, 294 p.

Lemennicier Bertrand, 1988. *Le marché du mariage et de la famille*. Paris, PUF, 226 p.

Lesthaeghe Ron, Kaufmann Georgia, and Meekers Dominique, 1989. The nuptiality regimes in sub-Saharan Africa, *in:* Ron Lesthaeghe (ed.), *Reproduction and social organization in sub-Saharan Africa*, p. 213–237. Berkeley, University of California Press, 556 p.

Locoh Thérèse, 1988. Structures familiales et changements sociaux, *in:* Dominique Tabutin (ed.), *Population et sociétés en Afrique au sud du Sahara*, p. 441–478. Paris, L'Harmattan, 551 p.

Locoh Thérèse and Thiriat Marie-Paule, 1995. Divorce et remariage des femmes en Afrique de l'Ouest. Le cas du Togo, *Population*, vol. 50(1), p. 61–94.

Marcoux Richard, 1997. Nuptialité et maintien de la polygamie en milieu urbain au Mali, *Cahiers québécois de démographie*, vol. 26(2), p. 191–214.

Marcoux Richard and Piché Victor, 1998.—Crise, pauvreté et nuptialité à Bamako, *in:* Francis Gendreau (ed.), *Crises, pauvreté et changements démographiques dans les pays du Sud*, p. 237–255. Paris, AUPELF-UREF and éditions ESTEM, 415 p.

McDonald Peter, 1985. Social organisation and nuptiality in developing countries, *in:* John Cleland and John Hobcraft (eds), *Reproductive change in developing countries*, p. 87–114. Oxford, Oxford University Press, 301 p.

Meillassoux Claude, 1975. *Femmes, greniers et capitaux.* Paris, Maspero, 251 p. (Fondations series, no. 29).

Mercier Paul, 1960. Étude du mariage et enquête urbaine, *Cahiers d'études africaines*, no. 1, p. 28–43.

Motahhary Mortadhâ, n. d. *Les droits de la femme en Islam* (transl. Abbas Ahmad al-Bostani). Montréal, La Cité du savoir, 315 p.

Ngondo a Pitshandenge, Emam, 1992. Nuptialité et structures familiales en Afrique au Sud du Sahara, *in:* UEPA (ed.), *The structure and dynamics of family formation in Africa*, p. 28–45. Dakar, UEPA and NPU.

Ngondo a Pitshandenge, Séraphin, 1996. *La polyandrie chez les Bashilele du Kasaï Occidental (Zaïre). Fonctionnement et rôles*. Paris, CEPED, 22 p. (Les Dossiers du CEPED, no. 42).

Pilon Marc, 1991. Contribution à l'analyse de la polygamie, *Étude de la Population Africaine*, no. 5, p. 1–17.

Pison Gilles, 1986. La démographie de la polygamie, *Population*, vol. 41(1), p. 93–122.

Segalen Martine, 1981. *Sociologie de la famille.* Paris, Armand Colin, 335 p.

Séraphin Gilles, 2000. *Vivre à Douala. L'imaginaire et l'action dans une ville africaine en crise* Paris, L'Harmattan, 302 p.

Service de la Statistique, 1962. *Recensement démographique de Dakar (1955). Résultats définitifs*. Dakar, Ministère du Plan, 143 p. + appendix.

Smith David P, Carrasco Enrique and McDonald Peter, 1984. *Marriage dissolution and remarriage*. Voorburg, International Statistical Institute, 94 p. (World Fertility Survey Comparative Studies, no. 34).

Tabutin Dominique, 1974. La polygamie en Algérie, *Population*, no. 2, p. 313–325.

Tabutin Dominique and Vallin Jacques, 1977. La nuptialité, *in:* Sources et analyse des données démographiques, part 3, vol. 3, p. 5–56. Paris, INED, INSEE, ORSTOM, 120 p.

Tapinos Georges, 1985. *Éléments de démographie: analyse, déterminants socio-économiques et histoire des populations*. Paris, Armand Colin, 367 p.

Tchak Sami, 1999. *La sexualité féminine en Afrique. Domination masculine et libération féminine.* Paris, L'Harmattan, 240 p.

Thiriat Marie-Paule, 1998. *Faire et défaire les liens du mariage. Évolution des pratiques matrimoniales au Togo.* Paris, CEPED, 295 p. (Les Études du CEPED, no. 16).

Thiriat Marie-Paule, 1999. Les unions libres en Afrique subsaharienne, *Cahiers québécois de démographie*, vol. 28(1–2), p. 81–116.

Thore Luc, 1964. Polygamie et monogamie en Afrique Noire, *Revue de l'action populaire*, no. 180, p. 807–822.

Tiemoko Richmond, 1999.—Vieillir ensemble: le rôle de l'écart d'âges entre époux dans le bien être familial, *in:* UEPA and NPU (eds), *La population africaine au 21e siècle*, vol. 3, p. 383–394. Dakar, UEPA and NPU, 712 p. (Third African Conference on Population, Durban, South Africa, December 6–10, 1999).

Timæus Ian and Reynar Angela, 1998. Polygynists and their wives in sub-Saharan Africa: an analysis of five Demographic and Health Surveys, *Population Studies*, vol. 52(2), p. 145–162.

Trussell James, Hankinson Richard, and Tilton Judith, 1992. *Demographic applications of event history analysis*. Oxford, Clarendon Press, 276 p.

United Nations, 1988. *First marriage: patterns and determinants*. New York, United Nations, Department of International Economic and Social Affairs, 110 p. (Doc. ST/ESA/SER.R/76).

van de Walle Étienne, 1968. Marriage in African censuses and inquiries, *in:* William Brass, Ansley J. Coale, Paul Demeny, Don F. Heisel, Frank Lorimer, Anatole Romaniuk, and Étienne van de Walle (eds), *The demography of tropical Africa*, p. 183–238. Princeton, Princeton University Press, 539 p.

van de Walle Étienne, 1996. L'âge au mariage: tendances récentes, *in:* Karen Foote, Kenneth H. Hill, and Linda G. Martin (eds), *Changements démographiques en Afrique subsaharienne*, p. 119–154 Paris, PUF/INED, 371 p. (Travaux et documents, Cahier 135).

Wa Karanja Wambui, 1994. The phenomenon of "outside wives": some reflections on its possible influence on fertility, in Caroline Bledsoe and Gilles Pison (eds), *Nuptiality in sub-Saharan Africa: contemporary anthropological and demographic perspectives*, p. 194–214. Oxford, Clarendon Press, 326 p.

Yana Simon, 1997. Statuts et rôles féminins au Cameroun, *Politique africaine*, no. 65, p. 35–47.

C H A P T E R

28

Factors in Couple Formation

THÉRÈSE LOCOH

Institut national d'études démographiques (INED), Paris, France

In demographic terms, couple formation could be a very simple process. Barring accidental circumstances, all human populations comprise roughly the same number of males and females. Young boys and girls reach sexual maturity within few years of one another and—even if attraction to individuals of the same sex is a phenomenon encountered everywhere—most sexually mature young people are strongly attracted to their contemporaries of the opposite sex. Biology and demography combined could lead to the formation of couples most of whom would be fertile and would ensure the reproduction of the population. But this *plain tale* does not exist in any society. Quite to the contrary, couple formation is a process particularly subject to the formulation of social, religious, and customary norms that multiply the varieties of union to infinity. Bans and obligations restrict individual choices, creating distances between candidates to union and—depending on the circumstances—shortages and excesses of marriageable persons.

The central concern of all cultures is to control the formation of unions with the aim of demographic reproduction. Marriage is never treated as a straightforward sexual pact; it includes other major dimensions of social reality, most notably economic and religious. It establishes alliances between groups and defines the integration of the individual into the group by the recognition of his or her descent.

Whom can one marry? When can one marry? How can one marry? Who decides the answers to these questions? By the solutions it offers for them, each social group defines the marriage market, determines the stock of marriageable individuals (i.e., of those who have the most direct access to reproduction). The rules that will be adopted, valued, and handed down across the generations or challenged serve to control individual behavior and the attraction between the sexes—in sum, to manage alliances between groups and the formation of descents, the two basic activities of the social factory.

The conditions for access to marriage are defined by each society's culture, consensually accepted rules, and the degree of individual compliance with these rules. Religious beliefs, sex and marital education, the designation of the persons vested with authority over individuals' marriage choices, and the conditions defined for accessing a partner (sex, age, previous marital status, social or economic characteristics) are all ways of structuring the *marriage market*, which, in turn, will shape each society's specific marriage behavior.

I. NUPTIALITY, ALLIANCE, AND REPRODUCTION

In demography, the study of nuptiality is often essentially regarded as a facet of the study of fertility, because the conditions of couple formation in a society have obvious consequences on reproduction. Women's delayed entry into union, or even the exclusion of a portion of the adult population from the marriage market, may constitute the first form of birth control. This has been termed the *Malthusian transition* in a reference to Malthus, who invoked the *moral constraint* as a means of curbing population growth (Coale, 1974; Chesnais, 1986).

In Africa, where almost everyone marries, and where women marry extremely young, the increase in the age at marriage is beginning to have a tangible effect on the decline in urban fertility (Lesthaeghe, 1989; Pison *et al.*, 1995). The same pattern is found to some extent in Asia, where the increase in the age at marriage is playing a part in the fertility downtrend (Casterline, 1994; Smith, 1980). In Latin America, the recent fall in fertility appears to have little connection to a change in patterns of entry into union (Rosero-Brixby, 1990).

In all societies, however, the issue of couple formation goes beyond the concern over making sure that its members reproduce. In societies bearing witness to the oldest types of social organization, anthropologists have found that unions—whatever the form in which they are promoted—are not guided solely by the concern to ensure demographic reproduction; they are also designed to provide a group with opportunities for expansion through alliances with other groups, by *exchanging women* (Lévi-Strauss, 1958).

In postmodern societies, in a totally different context, ensuring offspring is now only one of several goals of partner relationships, whether involving marriage or not. There is no longer any need for the rigid set of rules formerly designed to ensure numerous offspring, and the freedom of each individual to enter into a union or not is considerable. The main priorities are personal attraction and satisfaction of the partners. In these societies, families have lost most of their control over unions, and social norms themselves may be ignored without incurring punishment. Tolerance grows as the stakes become more narrowly confined to the choices of individuals themselves. The sphere of social control tends to be limited to protecting the interests of children and to settling conflicts in this area. New modes of couple formation emerge, determining ever more diversified individual statuses and residential and family arrangements. Changes in nuptiality may therefore be determined by cultural changes more than by demographic rationales concerning reproduction.

Will the analysis of nuptiality become obsolete as nuptiality becomes distinct from fertility behavior? Demographic and Health Surveys (DHSs), for example, which are one of the major sources of demographic data for the developing countries, devote little attention to unions, except for current marital status. Yet the study of unions is a crucial area of demography, for at least three reasons. The first is that marriage arrangements are a decisive factor in fertility in many societies. The second is that unions play an important role in structuring exchanges of goods and services in a society, and therefore have unquestionable implications for adult residential arrangements, the care of children and dependent old people, and even the transmission of property. The third concerns the measurement of social change. The age of entry into union and the new forms of couples (consensual unions, unions without co-residence, homosexual unions, etc.) are a highly sensitive indicator of that change. In particular, the age of entry into union is strongly correlated with educational attainment, female autonomy, and access to payroll employment. The study of nuptiality is therefore a valuable means of assessing social change.[1]

1. Alliance by Marriage: a Principle at the Core of Social Life

In many societies, couple formation depends very little on individual decisions: The sexual and emotional attraction that draws young people to one another is strictly regulated and controlled by those who hold power in a society. Couple formation has always been under tight social control.

At the root of all social life lies the ban on incest, which, as Lévi-Strauss (1958) demonstrated, expresses the transition from the natural fact of consanguinity to the cultural fact of alliance: "The incest ban is not so much a rule that forbids marrying a mother, sister or daughter as a rule that obliges a mother, sister or daughter to be given to someone else."

It is merely the correlative of the obligation to exchange. As the archetype of exchange, marriage is central to kinship and social life.

Lévi-Strauss explains this starting from a simple observation: the contingencies of birth are such that each individual was very unlikely to find a suitable sexual partner in the small groups that initially constituted human societies. People therefore had to seek partners in another group. There were only two ways of doing this: war or exchange.

> People had to choose between isolated biological families juxtaposed like self-perpetuating closed units, overwhelmed by their fears, hatreds, and ignorance or—thanks to the incest ban—the systematic establishment of intermarriage chains, allowing the construction of an authentic human society on the artificial basis of affinity links, despite the isolating influence of consanguinity and even in opposition to that influence. (Lévi-Strauss, 1971b).

[1] For example, it is by analyzing the change in ages at first marriage, age gaps between spouses, and births out of wedlock that David Shorter has charted the emergence of the modern family in Europe in the 18th century (Shorter, 1975).

This explanation is perfectly illustrated by the reply of an Arapesh to Margaret Mead (1935), who asked him if men could marry their sisters:

> What, you would like to marry your sister? What is the matter with you anyway? Don't you want a brother-in-law? Don't you realize that if you marry another man's sister and another man marries your sister, you will have at least two brothers-in-law, while if you marry your own sister you will have none? With whom will you hunt, with whom will you garden, whom will you go to visit? M. Mead (1950, p. 68).

2. Prescribed Marriages and Forbidden Marriages

In the marriage legislation of industrial societies, the only bans concern parents and brothers and sisters. Marriages between cousins are legally possible. But, while Western societies have laid down only negative rules, others, on the contrary, in addition to banning marriage with certain groups, set positive rules for prescribing spouses. The most remarkable example is that of preferential marriage with cross cousins (an individual cannot marry parallel cousins—the children of the father's brothers or the mother's sisters—and must therefore marry, insofar as possible, a cross cousin, i.e., the children of the mother's brothers or the father's sisters).

In these societies with elementary structures studied by Lévi-Strauss (1971a), the principles of exchanging women apply between easily identified groups. The same is not the case in societies with complex structures. The length of reciprocity cycles is unassignable. A person renounces a close relative and thus acquires the right to take someone else's close relatives, but nothing guarantees that this right will be followed by effects. Reciprocity becomes a general principle here, not an immediately enforceable rule (Cuisenier, 1985). Many societies accordingly promote an endogamy that can be based on family (marriage between crossed or parallel cousins, depending on the society), lineage, or ethnicity. This approach is justified by the notion that a certain closeness between spouses makes it easier for the young bride to adapt to her new environment and facilitates the preservation of land and other possessions in the family group. At the opposite end, other societies, like those of northern India, prefer strict exogamy rules. Endogamy and exogamy create different statuses for the wife in the family she has married into. In societies practicing strict exogamy, the wife is more isolated in a family that is unrelated to her and treats her as an outsider. In a society where marriages are managed on endogamous principles, young wives remain in familiar territory.

In Africa, even absent strict endogamy or exogamy rules, marriages are typically determined by a complex game of gifts and counter-gifts; the elder kin know which families are likely to give one of their daughters and can be approached for the purpose. Preferential marriages between cousins are still common in Africa and the Arab countries. In Murdock's census (1969),[2] slightly over one third of the societies studied practice preferential marriage between cousins, 34% forbid it only between first cousins, and 28%, on the contrary, strictly forbid it between any cousins. The first type of marriage is more common in the societies of the Mediterranean rim, East Asia, and South America (Table 28–1).

3. Couples, Unions, Marriages: a Reality Hard to Measure Statistically

Unions are formed in a very complex range of ways that resist statistical observation and require localized anthropological studies: The mechanisms include prescriptions and bans, provisions and counter-provisions under rules recognized by a social group, alliances between families or decisions of two individuals. The demographic study of nuptiality has long relied on marriage statistics of the kind that exist in the developed countries, which is why statistical studies on couple formation have concentrated on formal marriage, registered by a religious or civil authority. However, the latter is only one of many partnership arrangements. Where the recording of vital statistics is not the rule, most unions, although fully recognized by society, are not registered. In such cases, the statistical analysis of marriages will draw on surveys where all unions described as marriages by the interviewees will be regarded as such.

Moreover, in industrial societies, a growing number of couples that do not formalize their living arrangements define their unions—on the contrary—as informal or *de facto*. These are sometimes called *unregistered unions*: They can be long-lasting and lead to the founding of a family. Civil marriage statistics cover only a portion of couples actually formed; new survey methods are also needed to track consensual unions, which are increasingly subsumed under the

[2] Murdock's atlas is a standardized compilation of information on 863 societies studied by anthropologists and ethnologists. The census offers a statistical picture of the diversity of cultural choices in regard to marriage arrangements. Of course the distributions shown apply to the populations of societies surveyed, not to the populations exhibiting each characteristic described. It is not a representative sample of current living arrangements but rather a description of the diversity of *life rules* conceived by human societies.

TABLE 28–1 Preferential Marriage between Cousins

Forms of marriage	Ethnographic regions[a]						
	HAS	C	E	I	N	S	Total
Preferential between cousins	34.3	58.9	64.4	29.3	17.3	63.0	37.5
Banned between all cousins	24.5	20.5	8.0	27.6	53.0	5.5	28.5
Banned between first cousins	41.2	20.5	27.6	43.1	29.7	31.5	34.0
TOTAL[b]	100	100	100	100	100	100	100
	(204)	(73)	(87)	(123)	(202)	(73)	(762)

HAS, Sub-Saharan Africa (except Madagascar); C, Mediterranean rim (Europe, North Africa, Turkey, Caucasus, Near East); E, East Asia (except Formosa, Philippines, Indonesia, and including Madagascar and other Indian Ocean islands); I, Pacific islands (Oceania, Australia, Indonesia, Formosa, Philippines); N, North America (including native societies, and extending south to Tehuantepec isthmus); S, South America (including West Indies and Yucatan).

[a] As defined by George P. Murdock.

[b] Data unavailable for 101 of the 863 societies listed by the author.

Source: Murdock George Peter, 1969. *Ethnographic atlas*. Pittsburg, University of Pittsburg Press, 128 p.

confidential domain of private life without undergoing social approval or a legal registration, except in order to qualify for social benefits hitherto reserved for married couples. So great is the diversity of partner relationships that it is becoming ever more difficult to define a union.

The multilingual dictionary of demography (van de Walle and Henry, 1982) defines marriages as "unions between persons of opposite sexes which involve rights and obligations fixed by law or custom."

This definition is broad enough to include unions that are not recorded in writing, as still occur in many developing countries where unions are guaranteed solely through custom, but it excludes homosexual unions, for which some activist movements demand and sometimes obtain quasi-marriage ceremonies. Lastly, the definition does not include *de facto* unions, increasingly common in the developed countries, whose rights and obligations are not in any way set by law or custom; it does not even include unions assigned an intermediate status such as France's *civil solidarity pact* (*pacte civil de solidarité* [PACS]).

Nevertheless, all societies seek to control couple formation and promulgate norms for entry into union—even the societies that accept the fact that an ever greater number of individuals do not comply with the norms. The latter concern the conditions required for entering into marriage or a partner relationship, the partners that an individual may (or must) marry, their characteristics (age, earlier marital status), and their number (in polygamous societies). Marriages also differ according to the authorities through which they are recognized (religious bodies, families, representative of a community authority, the state, etc.).

4. Legal, Religious, Customary, and Consensual Unions

Control over sexuality and the formation of families, and obligations toward children, have always been major areas of expression of rules, often based on religious beliefs. In nearly all societies, unions—whether determined by families or individuals—are marked by ceremonies that attest to their recognition by the community. The recognizing entity varies from one group to another. It may require religious rites whose symbolism reflects the principles that must govern the spouses' behavior; these rites are intended to signal the union's recognition by a designated representative of a religious community. In Muslim countries, marriage is pronounced by an imam except when the prevailing legislation also allows the civil authority to pronounce marriages (in Tunisia, for example). With the secularization of social life, civil authorities are increasingly in charge of registering and recognizing unions. But in many African societies, recognition may simply consist in an agreement between the representatives of two lineage groups with or without the presence of the traditional chief. The union is no less recognized, however, if the groom's family has presented the required bride wealth to the bride's family.

In Europe, some countries recognize only civil marriage; religious marriage, being a matter of private choice, can be celebrated only after a civil marriage (France, Belgium). Other countries, such as those of Northern Europe and Italy, recognize either civil marriages or religious marriages. The choice of ceremony will depend on the beliefs of each or the desired resonance of the ceremony (Dittgen 1994). With the secularization of societies, the civil authorities have gradually taken over from religious authorities as marriage registrars. Religious precepts nevertheless remain crucial determinants of marriage behavior, and sometimes the religious ceremonies are socially more significant than the civil marriages.

Civil marriage includes the registration of the marriage decision by an official of the State (as a rule, spousal consent is the mandatory expression of the decision) and a public reminder of the principles accepted by the community, principles that the spouses undertake to respect in their life together. Civil marriage assumes certain prior conditions have been met, most notably the minimum age requirement and the earlier marital status of candidates to marriage (not already married, except in the event of a polygynous union). The minimum age for young girls is almost always lower than the one required for young men. These principles mostly apply to the couple's residential arrangements and to child-raising, but they can also define other issues. In African civil marriages, the groom commonly states whether he intends to live in monogamy or polygamy.

The registration of a union in the civil records makes it a marriage recognized by society and gives access to potential family entitlements such as State allowances and tax benefits. It defines certain rights and duties of spouses in the marriage. Noncompliance with the norms of married life may constitute a motive for dissolution of the union; this too will be recorded, enabling either spouse to remarry.

Religious marriages are governed by specific norms that can strengthen or dilute the norms spelled out by civil society.[3] They are based on a set of prescriptions concerning not only residence and child-raising but also sex before and in marriage, as well as the proper attitudes of the spouses toward each other. They thus play an important role in determining the prescribed gender roles in married life.

Catholicism, for example, defines a strict framework for sexual activity and starting a family. Intercourse is allowed only in marriage. Marriage is based on the principle of indissolubility, except through the death of one of the spouses. As a result, divorced persons cannot remarry. Marriage is regarded as the bond of a stable and autonomous family. Spouses must consent freely to marriage; "this is why a man leaves his father and mother," as stated in the Bible (Matthew 19-5). Marriage transmits values of strict control over sexuality and is predicated on a stable union in a family cell built around the conjugal core. Last, a high value is attached to celibacy vows (priests still cannot marry). This set of rules significantly narrows the population on the marriage market at a given moment. The Catholic model of marriage has been of major importance throughout Europe. It strongly influenced the behavior of candidates to marriage in all countries of Catholic tradition until the late 1950s. However, current trends show the lessening sway of the Church. In France, 90% of marriages registered in the civil records were also celebrated in church in 1950; today, formal marriages are declining and disaffection toward religious marriage is gaining, since only 65% of civil marriages[4] are now also celebrated in church (Dittgen, 1991).

In parts of the developing world, most marriages are *customary* (or *common-law*) *marriages* since formal registration is not widely practiced and the use of civil records is very limited. Customary marriages differ from one social, religious, and ethnic group to another. They can be validated by the presence of an authority recognized by the community but are often simply a family matter. When two families have struck an agreement and the exchanges of goods and services (symbolic or not) have taken place, the marriage is celebrated with varying degrees of publicity and recognized by the community. The payment of bride wealth or a dowry is often a key step in the process leading to the union. In the event of a separation, a customary village authority may be asked to determine the compensations due to the family that deems to have suffered from the breach of what the community regards as a contract.

In the transition countries, where the practice of civil marriage is gradually spreading, common-law marriage nevertheless endures, and family consent is sought if the initiative for the union comes from the future spouses themselves.

When no civil, religious, or customary marriage is involved, the unions are known as *consensual* or *visiting unions*. They are very common in Latin America, accounting for nearly 50% of unions in certain countries of Central America and the Caribbean. They are also fairly unstable, which suggests that many of them may be trial marriages (Rosero-Bixby, 1990). They differ from the noncohabiting unions found in Africa, which are unions recognized by society under the rules of polygyny (Locoh, 1994; Oppong, 1992).

The recent spread of consensual unions in Europe has been very rapid. In the early 1970s, when the

[3] Christian marriage has, for example, insisted since the Middle Ages on the need for the free consent of spouses. This represents a break with the patriarchal principle, then prevailing, that required children to obey parental decisions (Leridon, 1991).

[4] Estimated percentage of marriages of never-divorced individuals, the only marriages allowed by the Catholic Church.

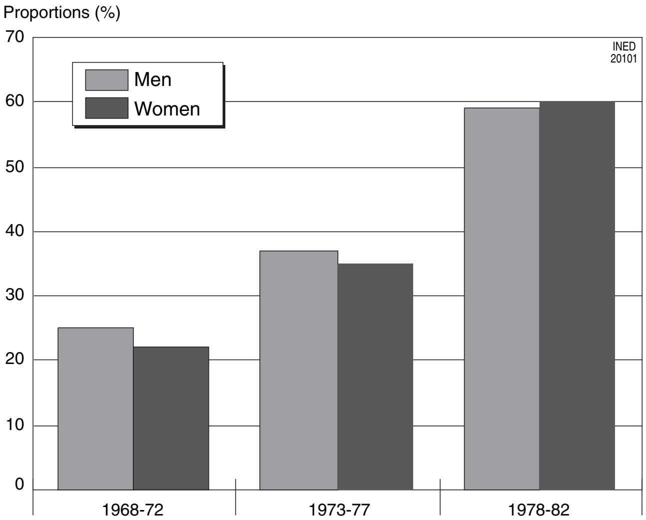


FIGURE 28–1 Proportions of men and women whose unions have begun outside marriage. Source: LERIDON Henri and VILLENEUVE-GOKALP Catherine, 1994. *Constance et inconstances de la famille: biographies familiales des couples et des enfants*. Paris, INED, PUF, 341 p. (Travaux et documents, Cahier no. 134).

phenomenon emerged, it was described as juvenile cohabitation. We now know that it was not a temporary living arrangement for young couples but a radical change of behavior that leads a growing number of adults to form partner relationships without necessarily marrying (Leridon and Gokalp, 1994) (Figure 28–1).

5. Marriages between Women

Anthropologists have noted marriages between women in several societies, for example in northern Benin and among the Nuer of Sudan. These marriages are entered into by infertile women to ensure that they will have descendants notwithstanding. A woman can marry a young woman and pay the bride wealth that seals the union. The children fathered by the husband with *his young wife* will be recognized as the descendants of the woman who has paid the bride wealth, not of the husband.

6. Homosexual Unions

Reported in several cultures as episodes of the youth life cycle (Badinter, 1992), homosexual relations are an abiding category of human behavior. In a recent survey on sexual behavior in France, for example, 4% of men reported having had a homosexual relationship at least once in their life. Similar figures (between 5% and 7%) are advanced for England and the United States (Messiah and Mouret-Fourme, 1993). With the concern over reproduction having given way to the satisfaction of partners as the main goal of couple formation, homosexual couples have acquired legitimacy, and the ideal of respect for individual rights has inevitably led to the recognition of their rights. The dramatic experience of the AIDS epidemic in homosexual communities has focused attention on the specific problems of gay couples and on the possibility of institutionalizing solidarity bonds comparable to those that exist for heterosexual couples in which a partner is mourned. In some countries, homosexual couples are entitled to official certificates of cohabiting partnership, which give them reciprocal rights. Religious ceremonies are also organized by some churches that recognize these unions (in the United States, for example; Collins, 1988). However, this recent tolerance must not hide the fact that, in many countries, homosexual unions are punished by harsh legislation, sometimes extending to deprivation of freedom.

Whatever the authority that sets the rules of marriage, or more generally of union, the rules mainly concern the age of access to sex and marriage, the control over spousal choice, the choice between monogamy and polygamy, partnership arrangements, and access to divorce. These five areas are of direct interest to the demographer as they determine at any given point in time the number of *marriageable* persons, and therefore whether the marriage market will function in harmony or in conflict. More broadly, they also define the balance of power between men and women, between young and old, and hence the decision-making capabilities of each player in regard to fertility.

II. AGE OF ACCESS TO SEX AND AGE AT MARRIAGE

1. Age at Puberty and Age at Marriage

The conditions of access to sex, to couple formation, and to parental responsibility shape an individual's entire adult life. Young girls in particular, when they enter this area before reaching physiologic maturity, run risks that can compromise their health, or even their later fertility. The same is true of their psychological maturity, which, if inadequate, will make the young woman dependent on her spouse or family-by-marriage. Indeed, most societies endorse the notion that men should demonstrate some maturity before marrying, even though the same societies extol early marriage for young girls.

When unions are subject to legal requirements, one of the conditions is a legal minimum age for marriage. As a rule, it reflects a concern of a physiological nature regarding the future spouses' capacity to engage in sexual activity; in principle, it is higher than the age at puberty. A United Nations survey (United Nations, 1988) showed that the statutory minimum age ranges from 14 years in some Latin American countries to 22 years in China for men and from 12 to 20 years for women in the same countries. The minimum ages are highest in Europe and lowest in Latin America. In the developing countries, the setting of a legal minimum age is a fairly recent concern. For girls, in particular, puberty was often the criterion that determined marriageable status. For young men, initiation rites were and still are arranged. They endow young males with the status of adult, a prerequisite for marrying. The notion that a young person needs to reach sexual or psychological maturity is a fairly widespread but not universal rule. In India, where child marriages are practiced, a little girl may be socially married very young but she does not go to live with her husband until she reaches puberty, after a new ceremony (United Nations, 1988). In Africa, customary marriage still prevails, and the legislation newly promulgated by central governments is not always enforced. Some social groups allow pre-pubescent girls to marry because, it is argued, they will thus adjust to their family-by-marriage as early as childhood, or because they have been promised as a consequence of matrimonial reciprocity relationships between two lineage or clan groups. The marriage is not consummated until puberty. This was also the case in China before Mao, where marriages arranged in childhood were a means of ensuring a low-cost marriage for a son.

> Since males often needed to pay a bride price and most of the wedding expenses, a son often needed for his family to accumulate enough resources for him to get married. The waiting period tended to be especially long for males in poor families and ones with many brothers. The widely documented practice of "little sister-in-law" (i.e., adopting and raising a future daughter-in-law to avoid the high costs of a son's marriage) is just one of the social practices arising from these economic realities. (Feng and Tuma, 1993)

Access to marital life is undeniably far more dependent on social factors than on physiological-maturity factors, as shown in Table 28–2, which reports the results available on the link between young girls' age at puberty and age at marriage (Udry and Cliquet, 1982). The link is obvious in some social environments and very weak in others. The connection is fairly strong among U.S. African Americans, in Pakistan, and in the Malaysian population; by contrast, in the Belgian sample and among young girls of Chinese descent living in Malaysia, it is weak or even nonexistent. Even more interestingly, the variations in the age at marriage are wider, for a given age at puberty, between the societies sampled than they are by age at puberty in the same cultural group. Among young girls who have had their first period at age 12, the age at marriage varies from 15.2 in Pakistan to 21.1 among Malaysian Chinese; in contrast, the maximum difference in age at marriage by age at puberty observed in Pakistan is smaller (from 15.2 to 19.7 on average). In individual terms, early puberty is correlated with early marriage. But the relationship is reversed at the collective level. The developed societies are the ones with the youngest mean age at puberty and the oldest mean age at marriage. The influence of age at sexual maturity on the age at first marriage or the age at first intercourse differs according to whether the society encourages early marriage or not. If early marriage is encouraged, the age at puberty is an important determinant of age at marriage. In the opposite case, social control over sexuality and reproduction is exercised and plays the leading role.

Some developing countries are in a transition phase where the improvement in living conditions leads to

TABLE 28–2 Mean Age at Puberty and Age at Marriage for Selected Populations

Age at puberty, years	United States (African Americans)		Belgium		Pakistan		Malaysia		Chinese (Malaysia)	
	m	N	m	N	m	N	m	N	m	N
≤11	17.9	269	21.30	452	—	1	16.19	47	20.04	21
12	18.85	385	21.54	713	15.24	17	15.60	129	21.06	65
13	18.82	395	21.91	831	16.92	52	16.57	110	20.38	81
14	18.69	189	22.15	650	18.14	72	17.15	124	20.86	88
≥15	19.55	196	22.19	502	19.65	43	18.10	139	20.59	212

m, mean age at marriage; N, number of observations.

Source: Udry Richard J. and Cliquet Robert L., 1982. A cross-cultural examination of the relationship between ages at menarche, marriage and first birth, *Demography*, vol. 19(1), p. 53–64.

earlier puberty and to an increase in the age at marriage—to allow schooling or work experience before entry into union. In Thailand, for example, a survey has shown that 60% of urban-dwelling young girls enrolled in school have had their periods before age 13 versus only 11% among those living in a rural area (Muangman, 1979, cited in Udry and Cliquet, 1982). For the first group, there follows a length of time—not experienced by earlier generations—in which they have reached sexual maturity but are not in a partner relationship. If they have not had access to contraception information and resources, they will be at increased risk of unwanted pregnancy, births out of wedlock, and possibly abortions. This is often the case in Africa and in many disadvantaged communities in the developing countries.

2. Tolerance of Premarital and Extramarital Sex

Some societies are fairly permissive about sex for unmarried young people others are repressive and strictly control access to sexuality. In China, marriage is the ceremony that gives access to sexual relations. The same is true in Christian populations, for which the religious ceremony of marriage is normally a prerequisite to intercourse and designates the only authorized sexual partner. In some societies, marriage ceremonies include the exposure of blood-stained bed linen as proof of the bride's virginity. In India, family honor is at stake in the preservation of the virginity of marriageable daughters. For this purpose, the inhabitants resort to *purdah* (reclusion of young girls) and a high segregation of female and male activities.

In African societies, family and social controls are totally focused on marriage. Young people marry partners designated by the elders, not chosen by themselves. This rigid control of marriage can go hand in hand with a moderate degree of sexual freedom for youth. Many African societies regard prenuptial births as normal, proving the young woman's fertility. In Kenya in 1989, for example, 41% of females aged 15–49 had had a child or been pregnant before marriage. In a rural Islamized population of Senegal, thought to disapprove of births outside marriage, these accounted for 20% of registered births between 1983 and 1991 (Delaunay, 1994).

Judging from Murdock's atlas (Table 28–3), the premarital-sex ban is not the most widespread norm. Premarital sex is heavily punished in one-fourth of the societies observed; 10% of the societies punish it when it results in pregnancy; 21% ban it in theory but punish it lightly; 7% of societies have adopted the practice of early marriage for women to avoid their having sex outside marriage, whereas one third of the societies are permissive in this respect. The greatest tolerance toward premarital sex is found in Africa (45%) and the Pacific Islands (49%); the heaviest disapproval is shown in the societies of the Europe and Mediterranean rim group (50%).

The most recent observations from the DHSs show that African societies remain the most tolerant of premarital sex. The highest frequency of premarital sex is found in sub-Saharan Africa (except Mali and Burundi). By contrast, the two Asian countries for which data are available, Indonesia and Thailand, show a broad coincidence between first intercourse and marriage. Latin America and the Caribbean are

TABLE 28–3 Permissiveness in regard to premarital sex

Premarital sex	Ethnographic regions[a]						Total
	HAS	C	E	I	N	S	
Not allowed	44.7	17.3	41.3	49.3	22.2	39.6	35.2
Strongly punished	21.3	50.0	27.0	18.7	24.6	14.6	25.1
Forbidden, but lightly punished	12.8	21.2	17.5	13.3	35.7	20.8	21.6
Punished if pregnancy occurs	12.8	3.8	7.9	12.0	12.7	10.4	10.7
Early marriage for women	8.5	7.7	6.3	6.7	4.8	14.6	7.4
TOTAL[b]	100.0 (94)	100.0 (52)	100.0 (63)	100.0 (75)	100.0 (126)	100.0 (48)	100.0 (458)

HAS, Sub-Saharan Africa (except Madagascar); C, Mediterranean rim (Europe, North Africa, Turkey, Caucasus, Near East); E, East Asia (except Formosa, Philippines, Indonesia, and including Madagascar and other Indian Ocean islands); I, Pacific islands (Oceania, Australia, Indonesia, Formosa, Philippines); N, North America (including native societies, and extending south to Tehuantepec isthmus); S, South America (including West Indies and Yucatan).

[a] As defined by G. P. Murdock (see Table 28–1).

[b] Data unavailable for 101 of the societies enumerated by the author.

Source: Murdock George Peter, 1969. *Ethnographic atlas*. Pittsburg, University of Pittsburg Press, 128 p.

TABLE 28–4 Distribution per 100 women by Interval between Age at First Intercourse and Age at First Marriage (Demographic and Health Survey, 1986–1989)

Countries	Relative to first union, first sexual intercourse took place:				Total
	More than a year earlier	At the same time	More than a year later	Data unavailable	
Sub-Saharan Africa					
Botswana	67.2	14.6	11.7	7.6	100.0
Burundi	19.4	60.3	19.1	1.1	100.0
Ghana	59.2	26.1	12.8	1.9	100.0
Liberia	58.1	28.3	10.3	3.4	100.0
Mali	12.0	66.0	11.5	10.5	100.0
Uganda	53.1	32.8	13.9	0.2	100.0
Togo	64.2	24.2	9.5	2.1	100.0
Zimbabwe	46.9	40.3	10.4	2.4	100.0
Asia					
Indonesia	2.7	88.6	7.7	1.0	100.0
Thailand	3.7	86.7	9.0	0.5	100.0
Latin America/Caribbean					
Brazil	30.6	57.8	5.2	6.6	100.0
Colombia	40.4	49.9	9.7	0.1	100.0
Dominican Republic	18.1	77.1	4.7	0.1	100.0
Ecuador	28.4	68.9	1.3	1.3	100.0
Guatemala	25.1	57.4	11.4	6.1	100.0
Mexico	20.2	54.6	24.0	1.2	100.0
Peru	44.8	52.1	3.0	0.2	100.0
Trinidad and Tobago	24.1	57.5	16.3	1.9	100.0

Source: IRD, 1990. *DHS, Methodological reports 1, An assessment of DHS-1 data quality*. Washington, Institute for Resource Development Inc., 143 p.

also relatively permissive societies, most notably Peru and Colombia (Table 28–4).

Shorter (1977) has shown that, in post-Renaissance Europe, premarital births have signaled the decline in social control of sexual activity and the emergence of more independent behavior among youth. In Africa, where the concept of marriage is very different, the tolerance of premarital sex must be interpreted otherwise.

3. Access to Sex and Access to Contraception

When the gap widens between age at puberty and age at first childbirth, one generally finds greater social control over access of young people to sex by means of bans and moral or religious rules. The population policies adopted in high-fertility countries typically add the increase in age at marriage to the arsenal of fertility-curbing measures. However, in the first stages of the fertility transition, one often observes both the increase in the age at marriage and a rise in extramarital births. For young people, the access to contraception—seldom facilitated in these societies—and the easing of social control on sex that usually goes with schooling, urban life, and access to nonfamily jobs initially translates into a greater number of out-of-marriage pregnancies.

By contrast, in the developed countries, such control is gradually replaced by access to contraception—but this does not mean that young people do not begin their sexual activity as soon as they reach sexual maturity. Along with greater sexual freedom, there has been a downward trend in the age at first intercourse and a higher frequency of its occurrence outside marriage.

The 1991 to 1992 survey on the analysis of French sexual behavior, for example, indicated a fall in the age at first intercourse—in moderate proportions for men, far more sharply for women, in the previous forty years. The age fell from 18.4 years for men born between 1922 and 1941 to 17.2 years for males born in 1972 to 1973. For women, the age fell from 21.3 to 18.1 years. Increasingly early sexual experience has coincided with a convergence between female and male ages at first intercourse. The gap has narrowed from 3 years in the older generations to less than 1 year today. General access to contraception is one of the important factors behind this change, but so is the changing status of women, in particular their growing autonomy in the workplace (Bozon, 1993).

4. Factors in Age at Marriage

While parental or social-group control of access to sex is exercised differently from one culture to another, the difference is even greater as regards access to marriage—a quintessential social act—and especially as regards age at marriage. Control often has divergent effects for the two sexes. Many societies encourage early marriage for girls so they can have children as soon as they are physically capable of doing so. By contrast, it is not uncommon for young men's access to marriage to be delayed, either to make them undergo initiation rites, or to accumulate bride wealth. This is the case in Africa, where gaps in the age at marriage are, as a result, particularly wide. Appendix 28–1 reports a set of data on age at marriage[5] by sex in different regions of the world and for two successive dates. The figures reveal three basic phenomena:

1. All over the world, *men enter into union on average later than women*. Their mean age at first marriage varies between about 24 years (Guatemala, China, Iran, and Turkey, for example) and 30 years (Sweden, Norway, and Switzerland). The earliest marriages take place in West Asia, China, and South Asia; the latest, in Europe and in West Africa. For women, the mean age at first marriage exhibits a wider spread of values, from about 18 years (Burkina Faso, in particular) to 30 years (Sweden). By and large, women marry earliest in sub-Saharan Africa, where the mean age rarely exceeds 20 years. Next comes West Asia (the Near Eastern countries) where women marry at a mean age of 21. Women marry latest in Northern and Western Europe. Very recently, however, the age at first marriage has risen steeply in North Africa.[6]
2. The smallest *age gaps* are found in Europe and North America (one year in Ireland, two years in several European countries, Canada, and the United States) but also in some Asian countries, notably China (1.9 years in 1987) and Thailand. The other countries of Asia, Latin America, North Africa, and Eastern Europe display differences in mean age at marriage ranging between 2.5 and 5 years. The widest gaps are in West Africa—for example, 10 years in Mali and Senegal. Their role in polygyny is well known. In societies where individual choice ranks second to family choice, the age gap between spouses is often wide, increasing women's dependence.
3. Age at marriage is trending up in most countries. The spread of education and nonfamily employment are the essential factors in the easing of controls on unions. When the young partners have personal sources of income, they are freer to take the decision to marry, but as a rule they will take it later.

In Europe, the loosening of the family grip formerly translated into a fall in the age at first marriage (Shorter, 1975): for women, it used to be about 25 years, but in developing countries where people marry far earlier, youth emancipation through schooling or access to nonfamily jobs almost invariably drives up the age at marriage. The decrease in male age at first marriage in China—owing to laws banning arranged marriages and dowries—is an exception in the contemporary world. But this trend has gone with a decline in early marriage. The percentage of men married before age 20 has fallen from 35% in the 1900 to 1925 cohorts to 6% in the cohorts born in 1960 to 1964. For women, the proportions have shrunk from 70% to 20%, respectively.

Most of today's industrial societies no longer encourage early marriage. Advances in female education, the very high integration of women into the labor market, and the widespread use of contraception have all combined to delay first marriage. Moreover, pressures on the young to marry legally are now weak in some societies where the custom of consensual unions is deep-rooted or in those where the free decision of individuals is such a strong value that parents no longer feel they have a right to control their children's choices.

At the individual level, all the developing countries exhibit a rise in the age at marriage (or at union, in countries where marriages are not officially recorded) as a function of the number of years of education. Delayed entry into union for educated women is particularly in evidence among the most highly educated women of sub-Saharan Africa. Between illiterate women and those who have at least 10 years of schooling, the difference in median age at first marriage is 4 years in Kenya, Togo, Liberia and Zimbabwe, and 6 years in Senegal and Uganda (United Nations, 1995).

[5] The data have been calculated using Hajnal's method (1953) from the proportions of single persons observed in censuses or surveys. In some societies, consensual unions may be reported as marriages in surveys.

[6] It now stands at 27 years in Tunisia (Bensalem and Locoh, 2001) and 28 years in Algeria (Ouadah and Vallin, 2000a and 2000b).

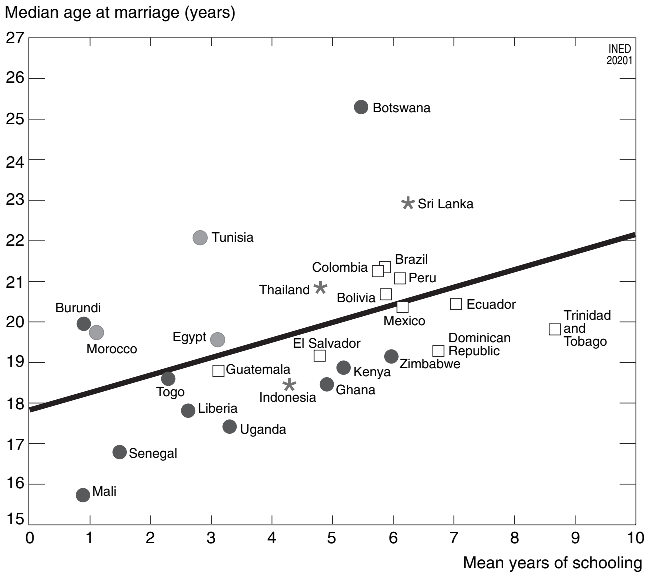


FIGURE 28–2 Median age of women at marriage, by mean years of schooling. Source: United Nations, 1995. *Women's education and fertility behaviour: recent evidence from the DHS*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, 113 p.

Figure 28–2 plots the correlation at the collective level. Here, the link is less obvious, as countries with comparable educational levels may have fairly different median ages at marriage due to the influence of other cultural factors concerning marriage behavior.

Some countries have made the increase in age at marriage one of the components of their fertility-control policy. In particular, China introduced coercive measures in the 1960s. Coale (1989) estimates at 60 million the number of births avoided in the 1970s by the increase in age at marriage. Since 1980, controls on that age have eased again, and the age at marriage for both sexes has fallen by about two years. This reflects young people's desire to marry earlier than would have been allowed during the harshest period of enforcement of the late-marriage policy (Feng and Tuma, 1993). One of the most visible trends in the change in marriage patterns in China is the convergence of male and female behavior. As men have married younger and women later, the age gaps have narrowed. The age gap between spouses or partners tends to fall with the autonomization of women and their growing access to education and employment—and, more generally, when they attain the same empowerment over their own lives as men.

5. Can One Escape Marriage?

The proportions of men and women married (or in union) at age 50 reported in Appendix 28–1 show that marriage is the common lot of the overwhelming majority of adults. In a great many societies, only physical incapacities are an obstacle to marriage. However, some religions attach value to consecrated celibacy—for example, Catholicism, which makes celibacy mandatory for priests and nuns.

There are also situations of imbalance between the number of marriageable men and women that can entail unwanted adult celibacy. This has occurred in the French countryside, where the excess of female migration flows over male migration flows toward urban communities has sharply driven up male celibacy rates in the past thirty years (Courgeau and Lelièvre, 1986). More generally, the phenomenon arises in all social environments where sex-differentiated migrations create imbalances in the marriageable population. Such imbalances are also observed after devastating wars in which most casualties are male. This can lead the excess female population to adjust its behavior in order to escape celibacy (Henry, 1966). Last, economic crises can exclude certain population categories from access to marriage owing to a lack of resources for starting a household. Farm laborers of the Crofting Counties in Scotland between 1850 and 1900 had never-married rates of up to 30% (Anderson and Morse, 1993).

The probability of never marrying depends on cultural factors—social pressure in favor of marriage, values respectively assigned to celibacy and marriage, financial requirements for the future spouses, and so forth—but also demographic factors that determine the size of the marriageable male and female populations. These numbers, as well, are influenced by cultural factors. A temporary or permanent ban on access to the marriage market for specific categories (determined by age, marital status, or social condition) increases the risk of a numerical imbalance. If, instead, ever-married men and divorcees are allowed to become candidates to marriage again, the marriage market will regain flexibility.

In Figure 28–3, we have plotted the mean age at first marriage against the proportion of ever-married 50-year-olds for a selection of countries. We can clearly identify two distinct nuptiality regimes: Europe and North America differ from the other continents by a higher celibacy rate and older ages at marriage, for men and women alike. Again, the data concern cohorts aged 50 in 1980 or thereabouts. Indeed, if the marriage behaviors currently observed in Europe and North America endure, the proportion of never-marrieds is likely to rise substantially in the years ahead. The nuptiality rates in industrialized countries show a clear disaffection for marriage since 1970 (see Table 28–5). If the present behavior continues, we might see birth

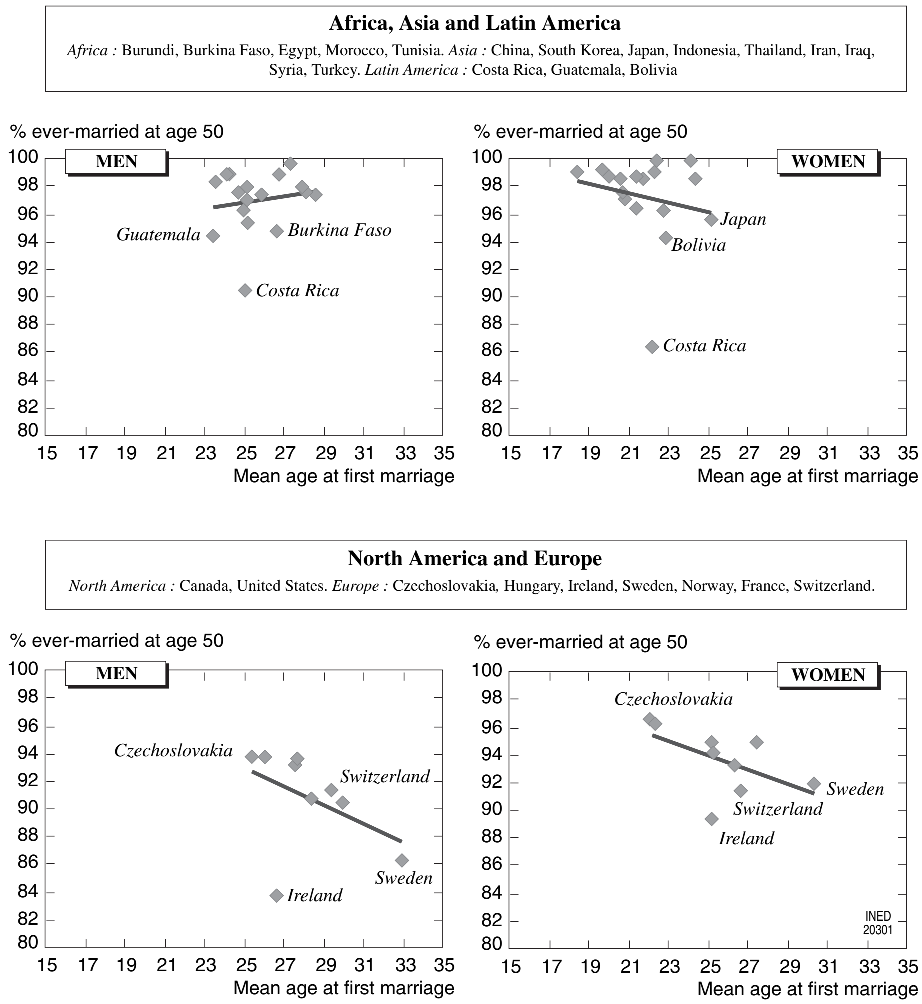


FIGURE 28–3 Correlations between mean age at first marriage and proportion ever-married at age 50 (Demographic and Health Survey, 1980–1990), by sex. Source: see table in appendix.

cohorts where one-third or even one-half of adults would never marry. However, they would have experienced one or more periods of partner relationships. What has changed is the type of relationship and its duration. Statistics will need to be based on a new definition of celibacy.

III. COUPLE FORMATION AND THE CHOICE OF PARTNER: FROM SOCIAL CONTROL TO PERSONAL DECISION

1. Dowry and Bride Wealth: How to Create Shortages to Control a Marriage Market

The need to accumulate goods in order to secure a partner is one of the crucial mechanisms for controlling access to marriage and the choice of partner. By requiring candidates to marriage to provide resources, the community decision-makers wield a powerful mechanism for regulating access to the marriage market. There are two contrasting methods: dowry, which is transmitted by the young girl's family to that of the future husband, and bride wealth, offered by takers of women to givers of women. Curiously, they have parallel effects on the marriage market by creating—via the socially accepted norm—shortages of either sex that do not exist in purely demographic terms. It is interesting to compare two matrimonial exchange systems: dowry in India and bride wealth in sub-Saharan societies. Both rely on the age difference between spouses.

In India, dowry has been subject to inflation for the past forty years or so (Rao, 1993). Some scholars have attributed this to a phenomenon known as Sanskritization (i.e., the poor imitating the behavior of the ruling classes, the Brahmans). Caldwell *et al.* (1983) have advanced the alternative hypothesis of a

TABLE 28–5 Total Nuptiality Rate (First Marriages per 1000 Men or Women), in Selected Industrialized Countries

Countries	Men			Women		
	1970	1980	1990	1970	1980	1990
Belgium	966	746	676	981	775	727
Denmark	752	491	652	799	532	596
England-Wales	1009	754	—	1040	761	—
France	915	689	551	919	706	563
Germany	920	680	593	980	690	639
Greece	1062	848	697	1044	851	743
Italy	1017	764	660	1007	765	670
The Netherlands	1012	660	619	1063	683	657
Norway	922	615	521	956	648	551
Portugal	1187	840	806	1102	878	838
Spain	1030	749	662	1003	735	680
Sweden	584	486	524	624	525	557
Canada	977	698	631[a]	921	695	674[a]
United States	1068	806	—	970	808	—
Australia	913	683	638	827	683	662

[a] Year 1991.

Source: Monnier Alain and Guilbert-Lantoine Catherine de, 1994. La conjoncture démographique: l'Europe et les pays développés d'outre-mer, *Population*, vol. 49(4–5), p. 1107–1128.

dual demographic constraint. First, rapid population growth and the sharply increasing frequency of a wide age gap between spouses creates a relative shortage of marriageable young men. Second, Indian society advocates early marriage of young girls; for a family, housing an unmarried woman aged over 20 is a particularly heavy burden given the few work opportunities for women and the disapproval of what is seen as late female celibacy.

These mechanisms create a *husband hunt*, with dowry as one of the stakes. The excess of marriageable women incites families to make ever higher offers to young men's families. The numerical-imbalance constraint (Caldwell *et al.*, 1983) inevitably leads to market adjustments, as the social pressure is so strong that all girls must eventually marry and succeed in doing so.

In a study of the states of Maharashtra and Andra Pradesh, Rao (1993) has observed a significant increase in dowries along with a generalization of marriage. The 1981 census found that the vast majority of women over 20 in these Indian states were married. However, adjustments have taken place. First, the female age at marriage has risen sharply (from 15.57 years in 1931 to 19.86 years in 1981), and faster than that of men (from 23.6 years to 25.5 years in the same interval): This has narrowed the age gap between spouses from 8 years to 5.6 years. One could have imagined other adjustments via divorce and the *recycling* of men in new marriages or polygyny. But this region of southern India practices neither polygyny nor divorce (only 1% of Indians of either sex are divorced). In the end, the adjustments were confined to the gap in ages at marriage, but the social pressures accompanying the shift have driven up dowries.

This example clearly shows the extent to which culture—by playing a role in setting norms for age at marriage and by institutionalizing differences between young women and young men—perpetuates numerical imbalances in situations where a balance could be achieved. It is interesting to compare India's situation with that of the countries of sub-Saharan Africa. There as well, age-at-marriage norms create numeric imbalances but through a different mechanism and with entirely different matrimonial consequences than those observed in southern India.

In Africa, the age gap between marriageable males and females ranges from 3 to 10 years depending on the society (see table in appendix). In these societies, marriageable young men belong to older age groups and are thus fewer in number than the young girls entering the marriage market. Despite the resulting deficit, African societies practice bride wealth (i.e., a transfer of wealth from the men's families to the women's families), contrary to the Indian custom. The reason is that the goal of men in African societies is to have access to several women, as polygyny is a positive value. The *surplus* of young marriageable women allows this type of marriage. *Bride wealth* also acts as a regulator in polygynous marriage systems. By requiring the accumulation of livestock, money or time working in the fields belonging to the lineage whose daughter is being courted, the system fixes a waiting period for marriageable young men and puts them in competition against older men, who are candidates to a new union and are probably more likely to assemble the bride wealth. The accumulation of bride wealth is often handled by the family members rather than by the young man himself: this also introduces a form of control by elders on the sexuality of the young. By and large, the elders are competing with young bachelors for the favors of marriageable young girls.

2. Family Systems and Marriage Decisions

The importance of family systems in understanding the patterns of couple formation stems from the fact that marriage forms the context in which families are established and grow. If the family is the birthplace of society, marriage is the birthplace of the family (United Nations, 1988). To analyze the different modes of family intervention in marriage decisions, we can distinguish—for simplicity's sake—between extended families and nuclear families.

In societies with a strong seniority-based hierarchy where the family authority is the structuring principle and the *extended family* is the most common living arrangement—as in many African societies—couple formation is an integral part of family, lineage, or clan strategies. In these societies, couple formation is driven by two basic motives: the desire to preserve and develop alliances and the need to manage each family group's reproduction as efficiently as possible. Alliance and descent are the core concerns of these societies. Marriage candidates' personal choices count for little, and the satisfaction of spouses in a quality conjugal relationship—beyond the necessary peace that must reign in a good marriage—is not the prime concern of match-makers. What matters is a family group's decision in conformity with consensual rules. These include control over sexual activity, the choice of partner, the type of bride wealth and bride wealth compensation to be exchanged by the families, and age at marriage. The stricter the rules, the greater the intervention of families in spousal choice. For example, in the Bwa country (Mali):

> Because of the inclusion of marriages in alliance strategies determined by community issues, and the recognized authority of family members in the formation of unions, couples neither form nor define themselves in terms of a plan developed and shared by the two partners. On the contrary, husband and wife each preserve their specific family identity, and very often they exhibit individual characteristics that do not promote any convergence (for example, a significant age gap). (Hertrich, 1994, p. 5).

The kinship group must preserve control of its members and thus maintain its capacity to manage its survival. Control by the extended family covers three aspects of the marriage: timeliness of the marriage, age at marriage, and choice of partner. The timeliness of the marriage is usually very high in these societies, where extended families are the most widespread system, and individuals are under pressure of different kinds not to remain single. The organization of domestic life in extended-family societies favors early marriage for girls, most notably in order to achieve reproduction goals, which require high fertility. When new couples join a family group of residence and contribute to a family production unit, there is no reason to delay marriages. The domestic group that hosts the new family cell will benefit from the young woman's labor, as it is almost always the woman who joins the husband's group. The greater the family involvement in the marriage processes, the younger the age at marriage, at least for girls, and the lesser the say of the future spouses in decisions concerning them. Because of the production and social-exchange objectives prevailing in extended families, spousal choice is heavily constrained by the family group. Under the authority of its elders, the group will enforce compliance with norms concerning exchange, endogamy, and exogamy. Until the 1950s, marriages in China were still governed by similar norms:

> Early and universal marriage has been a central feature of the Chinese cultural tradition. This tradition stipulates not only that both males and females should marry in order to have a complete life, but also that they should marry at early ages, normally between 16 and 18 for females and slightly older for males. The function of marriage was primarily to fulfill the collective goal of continuation and expansion of the family system by way of forming new reproductive unions, rather than to satisfy the individuals involved in the process. Therefore, until very recently, marriages were usually arranged by elder members of the clan, and not by the individuals themselves. (Feng and Tuma, 1993).

Advances in education and the extension of employment outside the family substantially loosened the grip of the traditional Chinese marriage model in the twentieth century. The State has also banned arranged marriages, required the free consent of spouses for their marriage, and introduced late-marriage policies.

In other transition societies, the monetarization of the economy brings a greater freedom of spousal choice by reducing the share of domestic production. Sometimes, parents are merely consulted about the choice made by their children—a modified form of the old patriarchal controls (Domingo and King, 1992). Table 28–6 compares the influence of parents in their daughters' spousal choices in two south Asian

TABLE 28–6 Mean Age at Marriage (Years) and Role of Parents in Daughter's Spousal Choice

Indicator	Pakistan			Thailand		
	Rural	Urban poor	Urban middle-class	Rural	Urban poor	Urban middle-class
Mean age at first marriage	17.2	18.2	20.1	21.9	20.1	22.5
Proportion (%) of women married as a result of the following						
Decision by parents	98.7	90.5	78.7	11.1	11.5	13.5
Decision by parents with daughter's approval	1.3	7.8	17.4	17.9	7.9	11.8
Decision by daughter with parents' approval	0.0	1.0	1.6	14.6	11.4	17.4
Decision by daughter	0.0	0.8	2.2	56.4	69.2	57.4

Source: Domingo Lita J. and King Elisabeth, 1992. The role of the Family in the Process of Entry to Marriage in Asia, *in*: Elza Berquo and Peter Xenos (eds.), *Family systems and cultural change*, p. 87–108. Oxford, Clarendon Press, 222 p. (International Studies in Demography).

countries: Pakistan and Thailand. In Pakistan, parental decisions still play an important role in young girls' marriages; in Thailand, as in many countries in transition toward monetarized economies, marriage behavior is evolving. In the most affluent strata of Thai society, where women marry at 22.5 years (a fairly high age), slightly over half the women (57.4%) have chosen their spouses freely, 17% have submitted their choice to parental approval, and one fourth have accepted the husbands chosen by their parents. By contrast, women in rural Pakistan—who marry, on average, at 17.2 years—never choose their husbands.

In nuclear families, whose rise has been facilitated by industrialization and urban living arrangements, parental intervention in spousal choice is less frequent. The interests of *nuclear* parents differ from those of extended families. True, they want their children to enter into prestigious marriages that will procure gains in status or property; they also seek a degree of geographic, religious or social endogamy in the marriage. But the young betrothed will establish their own cell. Their labor will therefore not be monopolized by their family group but, conversely, they will have to ensure the livelihood of their new family cell. In these societies, early marriage loses its justification and family pressures on marriage are weaker than in patriarchal societies. Moreover, the young usually acquire some autonomy, and parents have fewer means to force them to make choices against their will.

3. Endogamy and Homogamy: Individual Choice and Influence of the Social Environment

The closer one moves toward societies that value individual satisfaction, the more the couple-formation criteria reflect the partners' aspirations and the less the decision is shaped by social norms and pressures from the family circle.

> We are no longer dealing with groups that exchange their members, nor with families forming alliances, but with free and equal individuals deciding to unite in order to build their happiness together—and just as free to terminate their union later on, if it does not bring them the expected satisfactions. (Girard, 1981).

Yet, even in societies where individuals freely choose their spouses or partners, that liberty is conditional. Beyond their romantic quality, spontaneous attractions—and even *love at first sight*—are largely influenced by social conditioning, which shapes adolescent and adult preferences as early as childhood. Studies on social homogamy have shown that, even in industrialized countries, people continue to choose their spouses in *marriage markets* bounded by specific economic or social criteria. Doctors' sons have never really stopped marrying doctors' daughters, even though unions are slightly more imaginative than in the past and society is more tolerant toward what used to be called *mismatches* (Girard, 1981).

Endogamy is defined by the fact of choosing a partner in a group. Unions are called endogamous on the basis of precise characteristics. We can speak of ethnic, cultural, religious, geographic, and family endogamy. The most commonly studied family endogamy consists in marrying one's relative or a member of one's lineage group. This practice was inexorably eroded by the rise of migration (Fargues, 1986) and geographic mobility in the twentieth century and is certainly a factor that has hastened the transformation of marriage in many societies with endogamous preference. But society always locks the liberty of candidates to marriage in a wide variety of determinisms.

Geographic, social, and cultural endogamy is manifest in all societies. While losing some ground to more unpredictable unions because of geographic, social, and cultural mobility, it is clearly visible in most surveys.

Vicinity continues to play a major role in marriage choices. Despite migration flows, marriages are often still determined by ethnic, cultural, and regional preferences. Community and residential networks foster encounters between candidates to marriage and make others unlikely. Spousal choice is more open in today's urban societies than in older rural societies or in the present-day rural societies of developing countries, but traces of formerly prevailing conditions persist. In a survey of France of the 1970s, Alain Girard has observed (1981) that there was still

> a one-in-five chance that husband and wife came from the same village, a one-in-two chance of their being born in the same urban district, and a three-in-four chance that they came from the same département.

In societies whose members are free to choose their spouses, endogamy has given way to homogamy, which consists in marrying someone who resembles you. Homogamy can involve the family's social origin but also, increasingly often, common personal characteristics. In industrialized countries, cultural homogamy measured by educational attainment tends to prevail over social and economic homogamy. In France, the analysis of a 1989 labor-force survey showed education to be a more influential criterion than social origin in spousal choice. For a growing number of couples, being united through cultural homogamy is a means of accepting a degree of social heterogamy (Forste and Chauvel, 1995).

Cultural or social homogamy goes beyond the mere fact of marrying one's kindred spirit because he or she is closer, because one can meet him or her more easily, and because one is being encouraged by one's contact circle to form a union with a partner who is, in some sense, *familiar*. François de Singly (1987) ventures the hypothesis that the marriage market, like the labor market, is a space where individuals are recognized for their social value. He analyzes homogamy in the following terms:

> Homogamous unions are formed not out of respect for a social norm but on the basis of each partner's well-controlled interests [. . .] Heterogamous unions signal a strategic success for one of the two partners, since one of the two has apparently or actually succeeded in exchanging a value greater than his or her own. (p. 189)

Homogamy can also concern age. In advanced societies—including in China, where the break with traditional marriage has been strongly encouraged by the Communist regime—the age difference between spouses is narrowing. Traditional societies have hitherto been characterized by strong age heterogamy. Grooms are older than brides, sometimes much older. This type of heterogamy introduces a distance between spouses that has a powerful impact on their married life. It has a specific function in the workings of marriage systems where heavy female subordination is the rule. Dowry and bride wealth help perpetuate this wide age gap between husbands and wives in some societies. The gap has not vanished in advanced societies, where it is a survival from the days when people thought it normal that age gives the man additional authority over his wife.

IV. LIFE AND DEATH OF UNIONS

1. Monogamy, Polygamy, Polyandry: One Partner or More?

Monogamy, the union of one woman and one man, is the governing principle for all legal unions in the industrialized countries. Under the historical influence of the Christian concept of marriage, the industrialized countries have developed legislation consistent with the founding principle of the nuclear family, which—at least in theory—promotes an egalitarian relationship between spouses and makes reciprocal fidelity the cornerstone of exchanges between spouses. Yet monogamy can very well go hand in hand with a blatant inequality between spouses. The situation of European women until a recent period is proof of this. Bourgeois women in the nineteenth century had a lifelong status of minors, and, in certain countries, the financial independence of married women did not become reality until the 1950s.[7]

While monogamy is currently the most widespread form of union, this apparently has not always been the case in societies observed by ethnographers. Murdock's atlas (1969) indicates that the number of societies accepting polygamy far exceeds the number of societies advocating monogamy. Polygamy, in the strict sense, denotes the marriage of an individual with several persons of opposite sex. It includes polygyny (one man married to several women) and polyandry (one woman married to several men). Polygyny is by far the more common variant. Polyandry is extremely rare. It is observed in only a handful of societies (the Toda in India, the Bahini in East Africa, a few Eskimo tribes, and most conspicuously in Tibet, in the form of

[7] Conversely, polygyny, which is radically inegalitarian, is sometimes more conducive to the autonomy of both spouses.

fraternal polyandry). A closer analysis of polyandric practices shows that they typically result from a demographic imbalance due to girl infanticide. Societies incapable of ensuring their offspring's livelihood would appear to have practiced female infanticide as a means of controlling population growth—hence an imbalance between sexes leading to polyandry (Eskimos, Toda in India).

Polygyny is most common in Africa, particularly south of the Sahara (see Chapter 27). Contrary to what is sometimes believed, it predates the establishment of Islam. Everywhere else it is either heavily restricted or outlawed. In Muslim countries it is allowed but under strict conditions of equal treatment of wives, whose number is limited to four. In fact, polygyny is declining (between 1% and 11% in the countries of the Near and Middle East). In China, it was practiced by allowing concubines alongside the legitimate wife, but was abolished after the establishment of the Communist regime (Goode, 1963). In the developed countries, it survives informally and residually in a few religious groups, such as the Mormons.

Polygyny in sub-Saharan Africa is relatively stable. So far, it has also survived in the transitional urban environments, despite the introduction of modern, less favorable legislation and, in certain countries, the influence of Christian religions (Donadjè, 1992). Polygyny is a source of economic enrichment for the husband, either directly, through the appropriation of the wife's economic activity, or indirectly by taking advantage of the services of the children to whom the women will give birth. Polygyny is most common in the sub-Saharan African countries where women are most involved in agricultural production (Lesthaeghe, 1989). It also exists in Muslim countries, but Islam limits it to four co-wives. Most of the civil codes recently adopted in Africa restrict polygyny to two co-wives (Ngondo a Pitshandenge, 1994).

A large proportion of African women live in polygynous households: between one woman in three and one woman in two, depending on the country (Table 28–7). However, the new male cohorts hit by the economic crisis enter the marriage market at an older age and may find it difficult to access polygyny (Antoine *et al.*, 1995).

Marriage norms in these societies promote an age gap between spouses with the aim of creating a numerical imbalance between marriageable men and women. The age difference also increases the likelihood of widowhood among women and hence a contingent of newly *marriageable* women, sometimes systematically remarried in accordance with the practice of the *levirate* (see subsequent section on remarriage).

TABLE 28–7 Percentage of Women Living in Polygamy in Africa (1976–1987)

Countries	Percentage of women in polygamous union
Benin (1981–1982)	36.5
Cameroon (1978)	42.9
Côte d'Ivoire (1980–1981)	41.4
Ghana (1979)	34.4
Kenya (1977)	29.5
Liberia (1986)	38.0
Mali (1976)	46.3
Morocco (1987)	5.1
Mauritania (1981)	14.8
Rwanda (1983)	18.4
Senegal (1978)	48.5
Sudan (1979)	16.8
Zaire (1982–1984)	32.1

Source: United Nations, 1990. *Patterns of first marriage: timing and prevalence*. New York, United Nations, 327 p.

As long as the marriageable young agree to play by the rules—early marriage for women, late marriage for men, and the need to accumulate bride wealth—there is room for polygyny, despite the demographic equilibrium between the sexes! And, apparently, these rules are still well entrenched in African societies: the gap in the age at marriage remains stable, even in urban and schooled populations, and so does polygyny.

Polygyny is often justified by the desire to have many children in a context of high mortality. One of the reasons cited for polygyny is *postpartum* abstinence to preserve the health of the newborn child. In reproductive terms, polygyny has two effects that need to be separated. At the individual level, it has sometimes been argued that this type of marriage had a negative effect on the fertility of wives (Bean and Mineau, 1986). The effect is negligible when the wives' marriage-order effects can be controlled (Pison, 1986). At the collective level, polygyny has an undoubtedly positive effect on the fertility of the population by promoting a highly fluid marriage market. Every married man theoretically remains a candidate to marriage; as a result, all women can easily find themselves in a state of union and exposed to the risk of conception throughout their childbearing period—outside of *postpartum* abstinence periods—provided they accept life in a polygynous household.

2. Residential Arrangements and Partner Relationships

In couple formation, societies regulate not only access to union and the choice of partner. They often intervene in residential arrangements and in the spouses' married life.

The cohabitation of couples of different birth cohorts is still fairly frequent in traditional societies. This leaves room for controlling the life of couples, curbs their intimacy, and restricts their latitude to take decisions as an independent family unit. Typically, residence is patrilocal or virilocal. The wife will thus have the status of guest at best and of stranger at worst.

With modernization, residence is no longer necessarily determined by family considerations, but rather by economic constraints. In some cases, cohorts are forced to live together by the difficulty of finding accommodation for a new family cell. In Algeria, the present housing crisis is leading to a very pronounced delay in marriage and to unwanted situations where young couples are living with their parents.

Residential problem never before encountered are beginning to appear in the industrialized countries, where both spouses work but in different locales. This has prompted the invention of part-time conjugal living arrangements and, in some cases, dual-residence couples.

Polygyny, as well, gives rise to multiresidence households. In cities, *polygyny without co-residence* often replaces traditional polygyny (Locoh, 1994). Each woman has her own residence, and the husband goes from one to another. In traditional polygyny, one also finds forms of residential independence, with each wife assigned a hut of her own. The family authorities in the compound can thus exercise effective control over *postpartum* abstinence:

> Some sexual freedom among the unmarried youth is countenanced in many societies, but sexual relations of married couples tend to be carefully managed and scrutinized by senior kin, with the result that for many years of married life the majority of women are sexually inactive. (Oppong, 1992, p. 79).

3. Divorce and Remarriage

In the industrialized countries, the secularization of marriage, rising life expectancy, the reduction in the number of children desired, and the changing aspirations of men and women have combined to make divorce one of the possible and now accepted exits from marriage (Roussel, 1989). Legislation has gradually been aligned with social practice by making divorce more consensual and less disruptive for the ex-spouses.

Let us take the case of France. The divorce rate has risen from only 12% in the 1955 marriage cohort to 24% in the 1968 to 1970 cohorts, and will very probably reach 30% in the 1980–82 cohorts. The frequency of divorce has thus already doubled. From 1968 to 1980, with the swift spread of cohabitation, the frequency of divorce continued to rise, but at a slower pace. In fact, we are dealing here with two aspects of the same phenomenon: the decline in *marriage as institution*: "People are ever less convinced of the need for marriage (hence the fall in nuptiality) or of the need to preserve it when difficulties arise (hence the frequency of divorce)" (Leridon and Gokalp, 1994).

To different degrees and at different speeds, the phenomenon is affecting all European countries (Table 28–8). Italy seems to endure as a bastion of indissoluble marriage, but divorce was not legalized there until 1971. Likewise, Greece and Poland still report modest divorce rates. But in all the other countries of Europe and North America for which data are available, the period indicators show that, on present trends, between one-third and one-half of marriages will end in divorce. The highest proportion is found in the United States (55% in 1985).

Today, couples can form and break up several times in a life cycle. An increasing number of adults will experiment with multiple conjugal careers. Divorces and terminations of informal unions put candidates to marriage back in the running. This makes the marriage market—and, more generally, the partner-relationship market—more fluid in countries where the only conceivable form of couple is monogamous. A greater frequency of divorce increases the number of candidates to a union (official or not) at any given time, especially as fertility is no longer the cornerstone of marriage.

TABLE 28–8 Total Divorce Rate (Divorces per 100 Marriages)

Countries	1970	1980	1990
Austria	18.2	26.2	32.8
Belgium	9.6	20.8	31.0
Denmark	25.1	39.3	42.8
France	12.0	22.2	32.1
Norway	13.4	25.1	42.9
Netherlands	11.0	25.7	28.1
England-Wales	16.2	39.3	42.5
Sweden	23.4	42.2	44.1
Greece	5.0	10.0	12.0
Italy	5.0	3.2	8.0
Bulgaria	14.8	18.5	17.0
Poland	14.6	13.6	15.0
Canada	18.6	32.8	38.3
United States[a]	42.3	58.9	54.8

[a] For United States, last column shows 1985 data.

Source: Monnier Alain and Guilbert-Lantoine Catherine de, 1994. La conjoncture démographique: l'Europe et les pays développés d'outre-mer, *Population*, vol. 49(4–5), p. 1107–1128.

Divorce allows remarriage and therefore a conjugal mobility that some have described as *serial polygamy*. However, these successive monogamies have virtually nothing in common with African-style polygynous unions. The nature of relationships between partners, the impact on the fertility of couples, and the effects on the *marriage market* are radically different.

In the developing countries, many marriages are customary and not recorded by a civil or religious authority. Divorces, *a fortiori*, are even less often decided in court and recorded, so the scale of the phenomenon is harder to assess. In India, divorce seems practically nonexistent (Rao, 1993). In Africa, on the other hand, it is frequent but in the vast majority of cases it results in a *de facto* separation without judicial proceedings. In Togo, for example, by the age of 50, one woman in three has seen her first union terminated by a separation—quickly followed, in most cases, by remarriage (Locoh, 1985). In African societies where fertility remains the chief motive for couple formation, infertility is a major cause of divorce or remarriage for the husband. And after a separation, divorce or widowhood, women almost always remarry or are remarried, especially if they are still capable of bearing children. The practice of *levirate*—the remarriage of a widow with a brother or close relative of the deceased husband—illustrates this concern with fertility.

4. Widowhood

With the rise in life expectancy, the probability of a termination of the union through widowhood has decreased, leaving room for an increase in divorce. Widowhood obviously depends primarily on mortality but it is also determined by two factors of a more cultural nature: the difference in mortality between the sexes and the age gap between spouses. Where the age gap is wide, women are more likely to be widowed. In many societies, widowhood is an unenviable status. The extreme example is *suttee*, the cremation of a widow on her husband's funeral pyre, a practice now banned but formerly honored in India. In Africa, widowhood ceremonies are the occasion for rites involving violence toward the widows, who may also be remarried under levirate rules. While the practice is supposed to protect widows by placing them under the responsibility of their new husbands, in reality it only aggravates their dependency.

OVERVIEW AND CONCLUSION

The widening separation between marriage and maternity is a clearcut trend in most developed countries, although Southern Europe still exhibits fairly low rates of births outside marriage by comparison with Northern Europe. In North America, the proportion of out-of-wedlock births rose from about one in 10 in 1970 to one in four in 1990. In France, births outside marriage were six times as frequent in 1990 as in 1970 (Monnier and Guibert-Lantoine, 1994). Youth cohabitation is no longer a prelude to marriage; if a birth does take place it is usually desired and does not necessarily lead the partners to the mayor or priest. The *traditional* purpose of couple formation—that of begetting children and raising them together—has ceased to monopolize the conjugal universe in societies where falling mortality has made the task of reproduction less imperative. Behavioral norms are less constraining, and not all adults regard family and children as indispensable. Some claim the right to have children without necessarily forming a couple to raise them; others form couples without fertility plans. The satisfaction of partners can become the sole aim of couple formation. In the developing countries, by contrast, prenuptial births are increasing even though nuptiality remains intense and the proportion of never-marrieds is still very small (Antoine and Nanitélamio, 1991).

Couplehood and the Status of Men and Women

Couple formation plays a decisive role in determining the statuses of men and women. Age gaps, the origins of each partner, the involvement of their families and social groups, the values defining their couple (stability, precariousness, indissolubility), and the goals it embodies—all these factors reflect the status assigned by society to an individual as an adult member and potential parent; they can also influence that status, first within the couple, but also in society. These factors accordingly shape not only the partners' fertility decisions, but their entire family life and social life.

Under the influence of the human-rights ideal and feminist movements, modern legislation in the industrialized countries is increasingly aimed at ensuring equal rights and duties for men and women in marriage. The structure of the couple—formerly based on a hierarchic principle that subjected the wife to her husband's authority—has been rebuilt on a more egalitarian foundation. The implementation of sexual equality in marriage may take one of two forms: (1) recognizing for each partner, and more specifically for the woman, independence and freedom in the conduct of certain activities (choice of occupation, etc.); and (2) making both spouses jointly responsible and providing for the common exercise of their powers (joint liability for household debts, joint management of the family dwelling, and so on). In certain countries, estate law has gradually been adapted to the new concept of inter-spousal relationships.

This radical revision trend is producing different effects in countries that have recently introduced modern legislation. Much remains to be done in many countries as regards adapting the existing laws, but even more so as regards daily practice. All cultures, in varying degrees, rely on traditions of strong male dominance. The Cairo Conference on Population (1994) and the Beijing Conference on Women (1995) powerfully underscored the importance of striving toward equality between men and women and particularly toward the respect of women's rights in public life and private life.

The facets of marriage that societies seek to influence—age norms, partner-choice norms, and remarriage norms—govern couple formation in each period of the societies' history. Amid the maze of demographic, economic, and social determinants, what is the place of the individual, of his or her choices and inclinations? It is a subtle alchemy that succeeds in extracting the nectar of passion from so many combined determinants. And yet it exists: All societies have their tales of passionate love, their *Thousand and One Nights*, their *Romeo and Juliet*—which stand out precisely as an antidote to the constraints imposed by society, as an assertion of liberty, often for a fleeting but intense moment in the lives of individuals.

Acknowledgment

Translated from the French by Jonathan Mandelbaum.

References

ANDERSON Michael and MORSE Donald J., 1993. High emigration, low nuptiality: adjustment processes in Scotland's demographic experience, 1861–1914, *Population Studies*, vol. 47(2), p. 319–344.

ANTOINE Philippe, BOCQUIER Philippe, FALL Abdou Salam, GUISSE Youssouf Mbargane and NANITELAMIO Jeanne, 1995. *Les familles dakaroises face à la crise.* Dakar, ORSTOM-IFAN-CEPED, 209 p.

ANTOINE Philippe, DJIRÉ Mamadou and LAPLANTE Benoît, 1995. Les déterminants socio-économiques de la sortie du célibat à Dakar, *Population*, vol. 50(1), p. 95–117.

BADINTER Elisabeth, 1992. *XY, de l'identité masculine*. Paris, Odile Jacob, 314 p.

BEAN Lee L. and MINEAU Geraldine P., 1986. The polygyny-fertility hypothesis: a re-evaluation, *Population Studies*, vol. 40(1), p. 67–81.

BERQUO Elza and XENOS Peter (eds.), 1992. *Family systems and cultural change*. Oxford, Clarendon Press, 222 p. (International Studies in Demography).

BLEDSOE Caroline and PISON Gilles (eds.), 1994. *Nuptiality in sub-Saharan Africa, contemporary anthropological and demographic perspectives*. Oxford, Clarendon Press, 326 p. (International Studies in Demography).

BOZON Michel, 1990. Les femmes et l'écart d'âge entre les conjoints. Une domination consentie. II. Modes d'entrée dans la vie adulte et représentations du conjoint, *Population*, vol. 45(3), p. 565–602.

BOZON Michel, 1993. L'entrée dans la sexualité adulte. Le premier rapport et ses suites: du calendrier aux attitudes, *Population*, vol. 48(5), p. 1317–1352. (Special issue edited by Michel Bozon and Henri Léridon, *Sexualité et sciences sociales*).

BOZON Michel and HÉRAN François, 1987. L'aire de recrutement du conjoint, *in*: *Données sociales 1987*, p. 338–347. Paris, INSEE, 625 pp.

CALDWELL John C., REDDY Palli H. and CALDWELL Pat, 1983. The causes of marriage change in South India, *Population Studies*, vol. 37(3), p. 343–362.

COALE Ansley J., 1974. The demographic transition, *in*: *IUSSP Congrès international de la population, Liège, 1973*, vol. 1, p. 53–72. Liège, IUSSP, 494 p.

COALE Ansley J., 1989. Marriage and childbearing in China since 1940, *Social Forces*, no. 67, June, p. 833–850.

COLLINS Randall, 1988. *Sociology of marriage and the family: gender, love and property*, 2nd ed. Chicago, Nelson-Hall, 536 p.

COPPET Danielde, 1985. Systèmes de parenté, *in*: *Encyclopedia Universalis*, vol. 13, p. 496–501. Paris, Encyclopedia Universalis, 1102 pp.

COURGEAU Daniel and LELIÈVRE Éva, 1986. Nuptialité et agriculture, *Population*, vol. 41(2), p. 303–326.

CUISENIER Jean, 1985. Inceste, *in*: *Encyclopedia Universalis*, vol. 12, p. 5–7. Paris, Encyclopedia Universalis, 1051 p.

DELAUNAY Valérie, 1994. *L'entrée en vie féconde. Expression démographique des mutations socio-économiques d'un milieu rural sénégalais*. Paris, CEPED, 326 p. (Les Études du CEPED series, no. 7).

DITTGEN Alfred, 1991. Les mariages religieux en France. Comparaison avec les mariages civils, *in*: Thérèse Hibert and Louis Roussel (eds.), *La nuptialité: évolution récente en France et dans les pays développés, Actes du IX^e Colloque national de démographie, décembre 1991*, p. 7–18. Paris, INED, 279 p. (Congrès et colloques series, no. 7).

DITTGEN Alfred, 1994. La forme du mariage en Europe. Cérémonie civile, cérémonie religieuse. Panorama et évolution, *Population*, vol. 49(2), p. 339–368.

DOMINGO Lita J. and KING Elisabeth, 1992. The role of the Family in the Process of Entry to Marriage in Asia, *in*: Elza Berquo and Peter Xenos (eds.), *Family systems and cultural change*, p. 87–108. Oxford, Clarendon Press, 222 p. (International Studies in Demography).

DONADJÈ Florentin, 1992. *Nuptialité et fécondité des hommes au Sud-Bénin: pour une approche des stratégies de reproduction au Bénin*. Louvain-La-Neuve, Université catholique de Louvain, Institut de démographie, Academia, 222 p.

DYSON Tim and MOORE Mick, 1983. On kinship structure, female autonomy, and demographic behavior in India, *Population and Development Review*, vol. 9(1), p. 35–60.

FARGUES Philippe, 1986. Le monde arabe: la citadelle domestique, in André Burgière, Christiane Klapisch-Zuber, Martine Segalen and Françoise Zonabend (eds.), *Histoire de la famille. 3. Le choc des modernités*, p. 443–484. Paris, Armand Colin, 736 p.

FARGUES Philippe, 1988. La baisse de la fécondité arabe, *Population*, vol. 43(6), p. 975–1004.

Feng Wonget Tuma Nang B., 1993. Changes in Chinese marriage patterns during the twentieth century, *in*: UIESP (ed.), *Congrès international de la population, Montréal, 1993*, vol. 3, p. 337–352. Liège, UIESP, 483 p.

Forste Michel et Chauvel Louis, 1995. L'évolution de l'homogamie en France. Une méthode pour comparer les diagonalités de plusieurs tables, *Revue française de sociologie*, vol. 36(1), p. 123–142.

Girard Alain, 1981. *Le choix du conjoint. Une enquête psycho-sociologique en France*. Paris, INED, PUF, 216 p. (Travaux et documents, Cahier no. 70).

Godelier Maurice, 1995. Sexualité et société. Propos d'un anthropologue *in*: Nathalie Bajos, Michel Bozon *et al*. (eds.), *Sexualité et sida, recherches en sciences sociales*, p. 117–123. Paris, ANRS, 382 p.

Gokalp Catherine, 1981. *Quand vient l'âge des choix. Enquête auprès des jeunes de 18 à 25 ans: emploi, résidence, mariage*. Paris, INED, PUF, 180 p. (Travaux et documents, Cahier no. 95).

Goode William J., 1963.—*World revolution and family patterns*. New York, Free Press of Glencoe, 432 p.

Goody Jack, 1983. *The development of the family and marriage in Europe*. Cambridge, New York, Cambridge University Press, 308 p.

Hajnal John, 1953. Age at marriage and proportions marrying, *Population studies*, no. 2, p. 111–136.

Henry Louis, 1966. Perturbations de la nuptialité résultant de la guerre 1914–1918, *Population*, vol. 21(2), p. 273–332.

Hertrich Véronique, 1996. *Permanences et changements de l'Afrique rurale: dynamiques familiales chez les Bwa du Mali*. Paris, CEPED, 570 p. (Les Études du CEPED, no. 14).

IRD, 1990. *DHS, Methodological reports 1, An assessment of DHS-1 data quality*. Washington, Institute for Resource Development Inc., 143 p.

Kono Shigemi and Hayase Yasuko (eds.), 1994. *Fertility in the developing countries: a comparative study of the demographic and health surveys*. Tokyo, Institute of Developing Economies, 310 p.

Lathrop Mark and Pison Gilles, 1982. Méthode statistique d'étude de l'endogamie. Application à l'étude du choix du conjoint chez les peulhs bandé, *Population*, vol. 37(3), p. 513–542.

Lefranc Christophe, 1995. Le mariage en déclin, la vie en couple aussi, *Insee Première*, no. 392, juillet 1995, 4 p.

Leridon Henri, 1991. Pourquoi le démographe s'intéresse-t-il à la nuptialité? *in*: Thérèse Hibert and Louis Roussel (eds.), *La nuptialité: évolution récente en France et dans les pays développés, Actes du IX[e] Colloque national de démographie, déc*, 1991, p. 7–18. Paris, INED, PUF, 279 p. (Congrès et colloques, no. 7).

Leridon Henriand Villeneuve-Gokalp Catherine, 1994. *Constance et inconstances de la famille: biographies familiales des couples et des enfants*. Paris, INED, PUF, 341 p. (Travaux et documents, Cahier no. 134).

Lévi-Strauss Claude, 1958. *Anthropologie structurale*. Paris, Plon, 453 p.

Lévi-Strauss Claude, 1971a. *Les structures élémentaires de la parenté*, reprint. Paris, Mouton, 591 p.

Lévi-Strauss Claude, 1971b. La famille, *in*: *Le regard éloigné*, p. 5–29. Paris, Plon, 398 p. (Annales de l'Université d'Abidjan, series F, III).

Lévy Michel-Louis, 1982. Mariage et chômage des jeunes générations, *Revue d'économie politique*, no. 5–6, p. 595–611.

Locoh Thérèse, 1994. Social change and marriage arrangements: new types of union in Lomé, Togo, *in*: Caroline Bledsoe and Gilles Pison, (ed.), *Nuptiality in sub-Saharan Africa, contemporary anthropological and demographic perspectives*, p. 215–231. Oxford, Clarendon Press, 326 p. (International Studies in Demography).

Locoh Thérèse and Hertrich Véronique (eds.), 1994. *The onset of fertility transition in sub-Saharan Africa*. Liège, IUSSP, Derouaux Ordina Éditions, 308 p.

Mair Lucy, 1974. *Le mariage, étude anthropologique*. Paris, Payot, 235 p.

Martin Claude, 1994. Diversité des trajectoires post-désunion. Entre le risque et la solitude, la défense de son autonomie et la recomposition familiale, *Population*, vol. 49(6), p. 1503–1525.

Mead Margaret, 1935. *Sex and temperament in three primitive societies*. New York, W. Horrow, XXII+ 355 p. (reprinted by Mentor Books, 1950).

Messiah Antoine and Mouret-Fourme Emmanuelle, 1993. Homosexualité, bisexualité, éléments de socio-biographie sexuelle, *Population*, vol. 48(5), p. 1353–1380. (Special issue edited by Michel Bozon and Henri Léridon, *Sexualité et sciences sociales*).

Monnier Alain and Guilbert-Lantoine Catherine de, 1994. La conjoncture démographique: l'Europe et les pays développés d'outre-mer, *Population*, vol. 49(4–5), p. 1107–1128.

Muhsam Helmut, 1990. Social distance and asymmetry in intermarriage patterns, *Journal of Comparative Family Studies*, no. 3, p. 307–324.

Murdock George Peter, 1969. *Ethnographic atlas*. Pittsburg, University of Pittsburg Press, 128 p.

Ngondo a Pitshandenge Emam, 1994. Marriage law in sub-Saharan Africa, *in*: Bledsoe Caroline and Pison Gilles (eds.), *Nuptiality in sub-Saharan Africa: contemporary anthropological and demographic perspectives,* p. 117–130. Oxford, Clarendon Press, 326 p. (International Studies in Demography, IUSSP).

Oppong Christine, 1992. Traditional family systems in Africa, *in*: Elza Berquo and Peter Xenos, *Family systems and cultural change*, p. 69–86. Oxford, Clarendon Press, 222 p. (International Studies in Demography).

Pinelli Antonella, 1995. Women's condition, low fertility and emerging union patterns in Europe, *in*: Karen Oppenheimer Mason and An-Magritt Jensen (eds.), *Gender and family change in industrialized countries*, p. 82–104. Oxford, Clarendon Press, 329 p. (International Studies in Demography).

Pison Gilles, 1986. La démographie de la polygamie, *Population*, vol. 41(1), p. 93–122.

Pison Gilles, Hill Kenneth H., Cohen Barney and Foote Karen A. (eds.), 1995. *Population dynamics of Senegal*. Washington, D.C., National Academy Press, 254 p. (Committee on Population, Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy of Sciences).

Rao Vijayendra, 1993. Dowry "inflation" in rural India: a statistical investigation, *Population Studies*, vol. 47(2), p. 283–295.

Rosero-Bixby Luis, 1990. *Nuptiality trends and fertility transition in Latin America*. Liège, UIESP, 22 p. (Paper given at IUSSP seminar on "Fertility transition in Latin America" Buenos Aires, April 3–6, 1990, organized by Committee on Comparative Analysis of Fertility and Family Planning).

Roussel Louis, 1989. Types of marriage and frequency of divorce, *in*: Eugene Grebenik, Charlotte Höhn and Rainer Mackensen (eds.), *Later phases of the family life cycle: demographic aspects*, p. 20–36. Oxford, Clarendon Press, 249 p. (International Studies in Demography).

Shorter Edward, 1975. *The making of the modern family*, New York, Basic Books, 369 p.

SINGLY François de, 1987. Théorie critique de l'homogamie, *L'année sociologique*, vol. 37, p. 181–205.

SMITH Peter C., 1980. Asian marriage patterns in transition, *Journal of Family History*, vol. 5(1), p. 58–97.

UDRY Richard J. and CLIQUET Robert L., 1982. A cross-cultural examination of the relationship between ages at menarche, marriage and first birth, *Demography*, vol. 19(1), p. 53–64.

UNITED NATIONS, 1987. *Fertility behaviour in the context of development: evidence from the World Fertility Survey*, New York, United Nations, 150 p.

UNITED NATIONS, 1988. *First marriage: patterns and determinants*. New York, United Nations, 110 p.

UNITED NATIONS, 1990. *Patterns of first marriage: timing and prevalence*. New York, United Nations, 327 p.

UNITED NATIONS, 1991. *United Nations nuptiality chart*. New York, United Nations, 4 p.

UNITED NATIONS, 1995. *Women's education and fertility behaviour: recent evidence from the DHS*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, 113 p.

VAN DE Walle Étienne and HENRY Louis (eds.), 1982. *Multilingual demographic dictionary*, 2nd ed. Liège, IUSSP, 163 p.

WEISS Yoram, 1994. Les économistes et la formation des couples. Le fonctionnement du mariage et du marché matrimonial, *Population*, vol. 49(4–5), p. 1015–1039.

APPENDIX TABLE Mean Ages at First Marriage (MAFM), Percentage of Ever-Married Persons at age 50, and Difference in Mean Age at First Marriage

	Date	Men		Women		Difference (1)–(3)
		MAFM (1)	% Ever-married (2)	MAFM (3)	% Ever-married (4)	
East Africa						
Burundi	1979	24.4	98.8	20.8	97.2	3.6
	1988	25.2	97.9	21.7	98.5	3.5
Kenya	1979	25.8	95.0	20.4	97.9	5.4
	1988/89	—	—	19.1	—	—
Tanzania	1973	—	—	18.6	98.4	—
	1978	24.9	95.9	19.1	98.5	5.8
Zimbabwe	1982	25.4	95.7	20.4	97.4	5.0
	1988/89	—	—	19.6	—	—
West Africa						
Burkina Faso	1975	27.0	93.7	17.4	97.5	9.6
	1985	26.7	94.8	18.4	99.0	8.3
Ghana	1971	27.1	96.2	19.4	99.4	7.7
	1988	—	—	19.7	—	—
Liberia	1974	26.6	93.2	19.3	97.2	7.3
	1986	—	—	19.4	—	—
Mali	1976	28.2	96.9	18.1	98.3	10.1
	1987	—	—	19.9	—	—
Senegal	1978	28.3	97.5	18.3	99.5	10.0
	1986	—	—	19.4	—	—
Togo	1971	26.5	96.9	18.5	99.9	8.0
	1988	—	—	19.6	—	—
North Africa						
Egypt	1980	26.8	98.9	21.4	98.7	5.4
	1988/89	—	—	21.3	—	—
Morocco	1982	27.9	97.9	22.3	99.1	5.6
	1987	—	—	21.3	—	—
Tunisia	1984	28.1	97.5	24.3	98.5	3.8
	1988	—	—	24.0	—	—
Latin America						
Costa Rica	1973	25.4	89.4	21.7	84.7	3.7
	1984	25.1	90.4	22.2	86.4	2.9
Guatemala	1981	23.5	94.3	20.5	93.8	3.1
	1990	23.9	95.9	21.4	96.5	2.5
Bolivia	1976	24.5	94.4	22.1	92.3	2.4
	1988	25.2	96.9	22.8	94.3	2.4
North America						
Canada	1981	25.2	92.4	23.1	94.1	2.1
	1989	27.6	93.1	25.2	94.2	2.4
United States	1980	25.2	94.0	23.3	95.4	1.9
	1990	27.7	93.6	25.2	94.9	2.5

(continues)

APPENDIX TABLE *(continued)*

Date		Men MAFM (1)	Men % Ever-married (2)	Women MAFM (3)	Women % Ever-married (4)	Difference (1)–(3)
East Asia						
China	1982	25.1	96.3	22.4	99.8	2.7
	1987	23.9	—	22.0	—	1.9
Japan	1980	28.6	97.4	25.1	95.6	3.5
	1985	29.5	—	25.8	—	3.7
South Korea	1980	27.3	99.6	24.1	99.8	3.2
	1985	27.8	—	24.7	—	3.1
South Asia						
Indonesia	1980	24.1	98.9	20.0	98.7	4.1
	1985	24.8	—	21.1	—	3.7
Thailand	1980	24.7	97.6	22.7	96.2	2.0
	1987	—	—	21.9	—	—
West Asia						
Iran	1976	24.2	98.8	19.7	99.2	4.5
	1986	23.8	—	20.2	—	3.6
Iraq	1977	25.2	95.3	20.8	97.1	4.4
	1987	26.3	—	22.3	—	4.0
Syria	1970	25.9	97.3	20.7	97.5	5.2
	1981	25.7	—	21.5	—	4.2
Turkey	1980	23.6	98.3	20.6	98.6	3.0
	1985	24.6	—	21.5	—	3.1
Eastern Europe						
Czechoslovakia	1980	24.7	94.3	21.6	96.3	3.1
	1989	25.4	93.8	22.2	96.5	3.2
Hungary	1980	24.8	95.3	21.0	95.9	3.8
	1989	26.1	93.8	22.4	96.3	3.7
Northern Europe						
Ireland	1981	24.4	76.1	23.4	85.5	1.0
	1988	26.7	83.8	25.3	89.4	1.4
Norway	1980	26.3	88.9	24.0	94.3	2.3
	1990	30.0	90.4	27.5	94.8	2.5
Sweden	1980	30.0	87.1	27.6	93.1	2.4
	1988	32.9	86.2	30.4	91.9	2.5
Western Europe						
France	1982	26.4	89.5	24.3	92.9	2.1
	1989	28.4	90.8	26.4	93.3	2.0
Switzerland	1980	27.9	91.4	25.0	90.3	2.9
	1988	29.4	91.4	26.7	91.4	2.7

Sources: United Nations, 1988. *First marriage: patterns and determinants*. New York, United Nations, 110 p.; United Nations, 1991. *United Nations nuptiality chart*. New York, United Nations, 4 p.; Shigemi Yasuko, 1994.

C H A P T E R

29

Marital and Nonmarital Fertility

SILVANA SALVINI AND ANTONIO SANTINI

Dipartimento di Statistiche, Università degli Studi di Firenze, Florence, Italy

At both the phenomenologic and methodologic levels, demography evolves in step with the behaviors of the populations it studies. When a particular phenomenon or some of the variant forms of a given process become an important factor in population change, they enter demography's field of study, not by virtue of their intrinsic characteristics, but because of the historic emergence of new behaviors. To corroborate this obvious fact, we need not even refer to the extreme situations described by Bourgeois-Pichat (1987), who contrasts the populations of the Paleolithic and those of the 21st century. Marriage and marital fertility are two emblematic examples.

Marriage is in no way a natural prerequisite for reproduction[1]; strictly speaking, a marital union, legally authorized or recognized, is not a demographic event. Nevertheless, from the outset, demography has been interested in the characteristics of marriage and has paid special attention to the distinction between procreation in marriage and outside marriage. The reasons for this are well known: Marriage—by nature a rite of passage, a legal contract, and an instrument of social organization—enters the field of demography because of the reproductive function assigned to it by nearly all societies. Since time immemorial, all the European cultures (or cultures of European origin) have been dominated by a *Malinowskian legitimacy principle*, whereby reproductive functions were entrusted by society only to married couples and denied to the never-married. Being vested with such functions, marriage plays a key role in fertility and thus becomes a fundamental demographic event for population dynamics. But if these functions change or cease in the course of social evolution—which is what has been happening in the Western world for a quarter century—the phenomenologic (and methodologic) framework defined in a previous age may become obsolete. The new behaviors designate new objects of study and require new methodologies.

We therefore cannot generalize the role of marriage and the effects of nuptiality—its *quantum* or *intensity*, the proportion of persons who marry, and its *tempo*, the age at which people marry—on fertility. We cannot analyze marital fertility without considering the historic context in which the populations studied live. Nor can we understand, out of context, how marriage can effectively become the focus of the observation and measurement of the reproductive process.

In historic terms, we must accordingly distinguish at least three different demographic regimes. Within each one, we need to examine the specific relationships between marriage and fertility and determine the most suitable approaches for analyzing them. The three regimes are (1) the *ancien regime*, (2) the phase long referred to as the *demographic transition* and today renamed *first demographic transition* in contrast to (3) the following phase, still in progress, which many already call the *second transition* (see also, on this topic, Chapters 37, 38, and 39).

[1] The debate on the demographic nature of the marriage event has been intense, especially in the late 1960s. This can be seen from the report on session 6.4 of the 1969 IUSSP Congress in London (Galan Vivas, 1969).

I. THE ANCIEN REGIME

We need not dwell at length here on a historic period that has been comprehensively examined by historic demographers and has made a powerful contribution to the analysis of population dynamics (Livi Bacci, 1989). Our discussion will be confined to the aspects directly connected to the topics of this chapter.

The demographic *ancien regime* was essentially characterized by high mortality and, in consequence, a high frequency of widowhood: In fact, (early) death played a comparable role with respect to marriage as divorce does today. However, even more important, it was a regime where fertility was not controlled and where illegitimate fertility was a demographically marginal phenomenon. Absent voluntary action to curb fertility according to the number of children already born (an absence that Louis Henry regards as the fundamental characteristic of *natural fertility*), the role of marriage is central. Marriage is not only the mandatory locus of reproduction, but a restraining factor with decisive direct effects on fertility. It is thus one of the determinants—along with mortality—of population dynamics: The larger the proportion of never-married women and the older the mean age at marriage of women, the lower the mean number of children per woman. In other words, the reproductive process is regulated by the proportions of never-marrieds, the frequency of remarriages, and age at marriage. As Dupâquier emphasizes (1972), marriage was the basic self-regulatory mechanism of traditional populations.

If there is no voluntary control of fertility, we can say that the reproductive events in a couple are independent of its will: What matters is the woman's *marital status*, her *age*, and her *age at marriage*. A reading of the reproductive process as the result of events willed by the couple is consequently of little use in analyzing and understanding the phenomenon. For an identical age at marriage and an identical mortality, the union duration is the same and the completed fertility tends to be the same: In marriage, the age-specific variations in fertility depend solely on biologic factors. Figure 29–1, reproduced from one of Henry's most famous historic studies (1958), is a well-known confirmation of this.

For the reasons just described, the sequence of births in the married couple is not the prime focus of the analysis of fertility of traditional populations. Also, marital fertility is generally difficult to measure and describe—even more so in successive marriages, unless we use simulations. Yet the historic information on marital-status–specific fertility is, curiously, more common and significant than the information on the

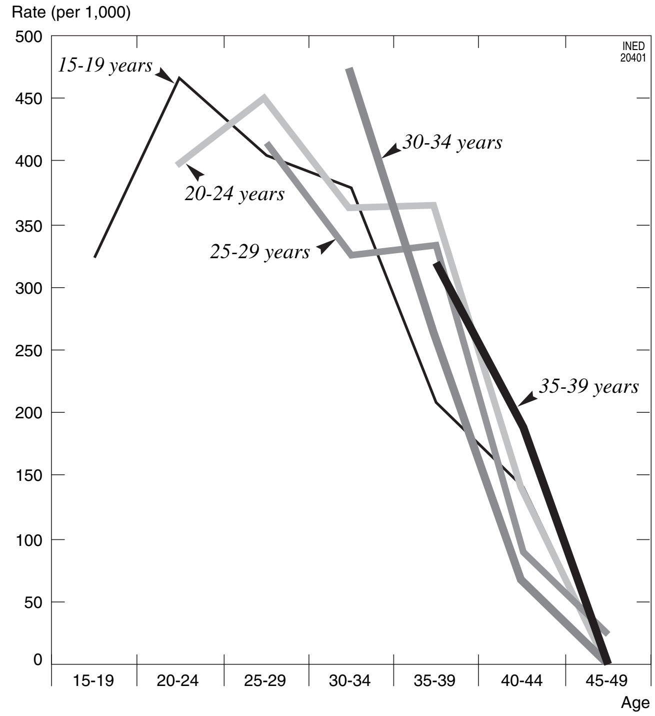


FIGURE 29–1 Legitimate fertility rates by age at marriage in a traditional population (Crulai, a parish in Normandy) (source: Louis Henry, 1958. *La population de Crulai paroisse normande*. Paris, INED and PUF, 270 p. Travaux et Documents series, *Cahier* no. 33).

population, even if it concerns very small demographic groups. This is strictly due to information availability: If the data on demographic characteristics and events are not systematically and periodically collected by official agencies, we will know very little indeed about a population's reproduction level and, in any event, the heuristic potential of official data is always constrained by their having been gathered for *nonscientific purposes*. Yet, thanks to the reconstitution of families (Henry, 1985) from the records of marriages, baptisms, and burials kept by the parishes—which are small communities—demographers have been able to document the reproductive history of couples with great precision. For *cohorts of marriages* celebrated in the same time interval (typically a decade or so), and distinguishing subcohorts by wife's age group at marriage, Henry (1972) has suggested using indicators of the following type:

$$f_x = \frac{\sum_y f_x^y \cdot \bar{F}_x^y}{\sum_y \bar{F}_x^y}$$

where

$$f_x^y = \frac{LB_x^y}{\bar{F}_x^y}$$

and where, for each age group at marriage ($y,y+a$), legitimate births (LB) are divided by the number of women-years ($\bar{F}$) effectively exposed to the maternity risk in the age group ($x,x+a$) (Del Panta, 1997).[2]

II. DEMOGRAPHIC TRANSITION

The change known as *demographic transition* in populations living in the area of European civilization was characterized by a sharp and fairly rapid fall in mortality, which became a general phenomenon by the mid 19th century. Its initial, direct effect was on fertility: By reducing widowhood, it promoted stable unions, made second marriages less frequent, and hence increased the duration of exposure to the fertility risk.

However, whereas marriage remained the cornerstone of family formation, birth control gradually spread across social strata and marital fertility decreased in the second half of the century—within the framework of the *European model* (Hajnal, 1965) and in step with its dissemination. In a cultural context where marriage remains the sole means of accessing reproduction, where fertility outside marriage remains socially stigmatized, and where the proportion of *illegitimate* births therefore persists at very low levels (with exceptions such as Portugal), marital status endures as a crucial factor in procreation. The spread of a Malthusian mentality and the concomitant rise of a new behavior in the population thus give the marriage-reproduction relationship a different content from that specific to the *ancien regime*. The study of the relationship therefore demands new analytic approaches.

The central link between completed fertility and age at marriage loses much of its significance. A growing number of couples want to form a family with one or two children—a goal that can be easily reached in the first years of marriage. This makes age at marriage secondary: Even the couples formed at a relatively old age will be able to satisfy their desire for children. Fairly soon after marriage, most couples behave, voluntarily, as if they were sterile. Figure 29–2 clearly shows this new reality.

[2] Even in the case of aggregate data (i.e., not obtained from a family reconstitution with personalized data), population fertility is measured more accurately by (1) exploiting the data on legitimate births through indicators derived from an adaptation to historic data of the method for analyzing legitimate fertility by year of marriage of the spouses, introduced in the 1930s by Corrado Gini (1932), or (2) establishing the ratio of this type of event to denominators that incorporate the change in marriage frequency over a given number of years prior to the period studied. However, as we can see, both of these are *period* (i.e., cross-sectional) measurement techniques.

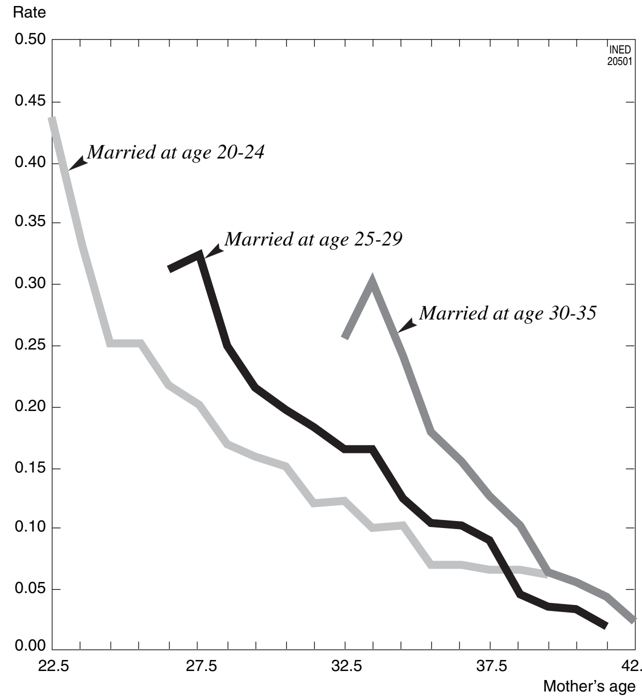


FIGURE 29–2 Age-group-specific legitimate fertility rates in a post-transition population (source: David V. Glass and Eugene Grebenik, 1954. *The trend and pattern of fertility in Great Britain, Part II*. London, HMSO, 692 p.).

In a controlled-fertility regime, reproduction is chiefly linked to marriage duration for all couples. Duration is merely the difference between the two effective determinants of legitimate fertility: age at marriage and the wife's current age. Duration has become a fundamental factor in behavior and obliges demographers (or gives them the opportunity) to replace the wife by the union itself as the reference unit for studying the reproductive process. In fact, one of the basic principles of demographic analysis is to link each demographic event experienced by an individual to the event that necessarily preceded it and determines its occurrence. The *origin-event* of a legitimate birth is marriage and, within marriage, time is effectively expressed by the duration elapsed since the event. As we saw, this measure of time was already implicitly present in traditional populations. However, in the new context shaped by the desire and ability of couples to control their fertility, it is destined to play a major role.

For each subcohort defined by the combination of year of couple formation and age of wife at marriage, we can assess the influence of marriage duration on fertility from three distinct angles. First, as the marriage duration increases, the proportion of couples who, with age, become physiologically sterile rises,

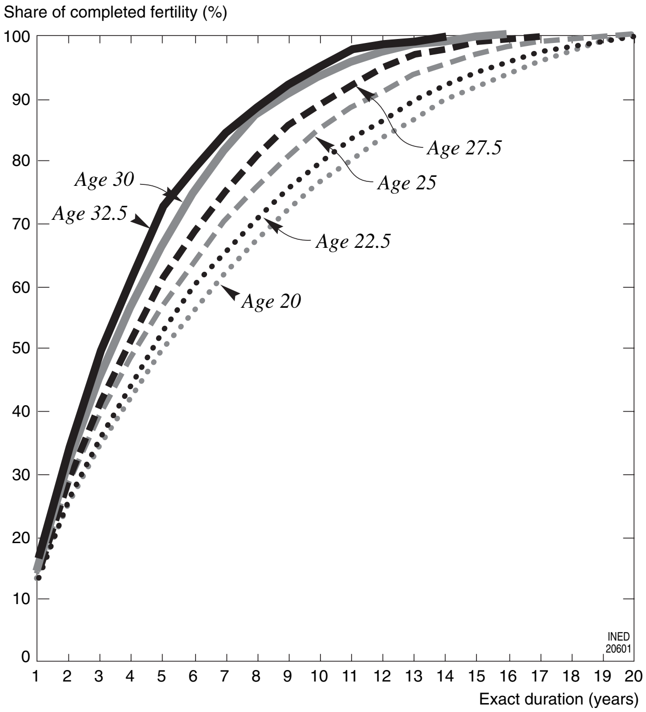


FIGURE 29–3 Family formation by age-group at marriage (source: G David V. Glass and Eugene Grebenik, 1954. *The trend and pattern of fertility in Great Britain,Part II*. London, HMSO, 692 p.).

because the wife's age increases with marriage duration. Second, and more important, marriage duration also determines the proportion of voluntarily infertile couples. Third, marriage duration, in itself, can also be a fertility-reducing factor if, over time, the couple tires of intercourse and thus reduces its frequency (Henry, 1953).

Henry (1953) argues that the marriage-duration factor, while operating with other factors, can be examined separately. This allows a relatively simple but, in fact, comprehensive analysis of fertility: The link between aging and marriage duration is a functional link that nothing can sever. The relationship between marriage duration and voluntary infertility is typical of post-transition populations where couples control their births; apparently, it prevails over the three effects describe (as shown in Figure 29–3).[3] We should bear in mind, however, that this relationship merely reflects the strict correlation between marriage duration and the number of children already born. In other words, in modern populations where couples curtail their fertility, an increase in marriage duration produces a rise in the proportion of couples who have reached their family-size goal and then remain voluntarily infertile or minimally fertile. Thus, in reality, these proportions increase not so much because of the marriage duration but because of the number of children already born. As the desired number is reached in each household, it is ultimately that number that becomes the main factor in procreative behavior.

Moreover, absent events that would force couples to interrupt their family formation or their partner relationship in a lasting manner, there is a close relationship between completed fertility and marriage duration, and it seems useless to explicitly introduce the number of children already born. In these conditions, the age-at-marriage/marriage-duration combination seems largely adequate to resolve a number of issues (Henry, 1953). However, the influence of the number of children already born on procreative behavior or, more specifically, on the definition of family strategies to reach the desired family size is sufficiently strong to require the explicit introduction of this factor into the analysis.[4]

III. FOUNDATIONS OF MARITAL-FERTILITY ANALYSIS

The spread and consolidation of birth control in Western societies has spawned a new methodology and stimulated a growing number of studies that describe the procreative process as a function of the woman's marital status and of her marital-event history.[5] However, the cohort approach to *legitimate fertility* (which would be more aptly described today as *marital fertility*)—based on the adoption of the marriage cohort as the unit of study—did not emerge and gain acceptance in demography until the 1950s. In that same period, national statistical institutes became increasingly involved in collecting demographic data extending beyond purely administrative needs,

[3] In fact, we can add a fourth: the time elapsed since the previous birth. This plays a role both for Malthusian couples wishing to space their births and for other couples because of the duration of pregnancy and postpartum amenorrhea.

[4] In this connection, it must be recalled that one of the most classic fertility measures in which birth order is taken into account—parity progression ratios, which measure the probability of an increase in family size—was introduced by Louis Henry and Norman Ryder in their study of marriage fertility, although it was later extended to that of overall fertility.

[5] Besides the studies published between the two World Wars in which demographers tried to take account of marital status to measure reproductivity (then a topical issue), we should also recall the studies by Jean Bourgeois-Pichat (see, for example, 1950) aimed at producing total fertility indicators based on marriage duration and age at marriage.

whereas theoretical inquiries led to the first attempts at cohort analysis[6]: One of the leading figures here was John Hajnal, who worked on the results of the large-scale fertility survey carried out with the 1946 British census (Hajnal, 1950); the two other protagonists were Louis Henry (1953) and Norman B. Ryder (1956), who, without coordinating their efforts, simultaneously set forth the principles of methods that have become well known and remain in widespread use.

As Henry explicitly argued, the analysis of marriage fertility is—for the reasons recalled earlier—the most suitable means for studying the reproductive behavior of populations that largely practice contraception but where a vast majority of births are *legitimate*. Although this is true, there are many practical difficulties involved in such an approach. The application of rigorous analytic techniques is chiefly dependent on—and often restricted by—the kind of information available. The resulting set of problems depends, in turn, on the type of observations available. In particular, they depend on whether the data are gathered by means of continuous observation (i.e., of the kind used to compile vital statistics) or by means of retrospective observation (i.e., of the sort used to compile most of the results from censuses, initially, and from *ad hoc* surveys subsequently).

What classical analysis is invariably asked to provide is a correct solution to the following basic problem: "Absent an intrusive event such as death, divorce or migration, what would be the mean number of children by marriage and how, in these same conditions, would the children's births be distributed by marriage duration?"

1. Continuous Observation

The general rule (Henry, 1972) according to which the description of a process should be tied to its origin-event implies that the reference cohort should be a couple cohort (i.e., the set of couples formed during a given period).[7] Within the cohort, the process will unfold as a function of marriage duration. Completed fertility is primarily determined by this duration. However, the ratio of fertility to marriage duration is far more rigid that the ratio of the wife's age to fertility. In consequence, the latter ratio cannot be overlooked, as it represents a heterogeneity factor for the structure of the observation unit. We can take it into account by creating subgroups (subcohorts) based on the wife's age at marriage. This requirement has obvious consequences on the event-classification criteria (for which we need to know three time variables simultaneously: year of marriage, marriage duration, and mother's age at marriage) and on the determination of person-years effectively lived in each duration interval by the cohort members. Birth classifications that meet these needs exist for a few European countries—France, for example, from 1946—at least for limited or recent periods of time.[8] In general, however, official statistics do not provide an update on current marriage duration after, for example, a divorce, separation, widowhood, or migration. For each marriage cohort formed in T with wives aged y, the measure of fertility at marriage duration x—called *legitimate-birth frequency*, is generally obtained from the ratio of events (LB) not to marriage-years ($\overline{M}^y_{T,x}$) of the interval of duration x, expressed by:

$$f^y_{T,x} = \frac{LB^y_{T,x}}{\overline{M}^y_{T,x}}$$

but to the initial number of marriages ($M^y_{T,0}$), whatever the duration examined:

$$f^y_{T,x} = \frac{LB^y_{T,x}}{M^y_{T,0}}$$

This has obvious consequences on the process quantum: Mean fertility—obtained, of course, by summing the rates—is routinely underestimated, unless the entrances of married persons exceed deaths and exits. There are also consequences, albeit less visible, on the tempo of marriage fertility, as measured by the mean index of the distribution of duration-specific rates. Naturally, beyond the approximations due to the denominator of duration-specific rates, the overall quantum and tempo indices will not express the marriage cohort's propensities unless the behaviors of arriving or departing couples are identical to those of the initial population.

[6] Actually, the first studies on marriage fertility date from the early 1930s, when some national statistical institutes, such as Italy's (Gini, 1932), began compiling and publishing statistics on legitimate births by parents year of marriage (or by marriage duration). These were, however, period analyses rather than cohort analyses.

[7] According to Ryder, studying fertility in marriage cohorts is also a way to skirt the issue of the difference between male and female fertility functions. By choosing marriage as the origin-event, since the duration is the same for both partners, we arrive at a reconciliation within one marriage cohort of two fertilities—male and female—with very different quantum and tempo characteristics.

[8] Very often, legitimate births are classified according to a single time criterion, such as year of marriage or marriage duration. Sometimes, a second criterion is used: the mother's age at marriage. In these cases, current statistics do not allow the computation of rates assignable to specific cohorts or covering correct durations (Santini, 1974; Calot, 1984; Prioux, 1991).

2. Retrospective Observations

As a rule, the analyses that rely on data gathered in retrospective surveys or from responses to census-schedule questions on reproductive history do not allow a reconstitution of marriage cohorts. However, to study female fertility, we usually have the wives' dates of birth and marriage, as well as the dates of birth of their children—or, at least, of their live-born children. We can thus construct female cohorts by year of birth and year of marriage, classify them by age at marriage, and locate the births of children on this double time scale.

Even with this type of observation, demographers studying legitimate fertility more commonly use marriage-duration–specific rates than age-specific rates, at least for contemporary populations. Hence the terminologic analogy: The term *marriage fertility*—obviously more suitable to continuous observation—still prevails in retrospective observation.

In principle we should subdivide the cohorts of married women by age at marriage and calculate separate marriage-duration–specific rates of legitimate fertility for married women in the successive 5-year age groups. Since the units examined are those still living and present at the survey or census date, consistently with the formulation given by Wunsch in Chapter 12, the rates (or *legitimate-birth frequencies*) should take the form:

$$f_{x,x+a}^{y,y+a} = \frac{LB_{x,x+a}^{y,y+a}}{Fm_u^y} \quad \text{or} \quad f_u^{y,y+a} = \sum_{x=0}^{u-a} \frac{LB_{x,x+a}^{y,y+a}}{Fm_u^y}$$

where, as before, y is the age at marriage, x the marriage duration to date of birth, u the marriage duration at the survey date, and *Fm* the number of married women.

The mean number sought for is the sum of births in marriage at each duration, in the absence of death, divorce, and migration, again subject to compliance with the *independence* hypothesis (and, implicitly, the *continuity* hypothesis). Since all the unions survive at time of survey, the age-at-marriage distribution stays the same: At any given moment, it is identical to the one observed at the time of the survey and—in keeping with the basic hypotheses—to the distribution prevailing at the time of cohort formation.

In determining the marriage-fertility quantum, we cannot disregard the composition of subcohorts by age at marriage. Unions with older wives are more rapidly terminated by death. After, say, 10 years, they are proportionally less numerous than the others, barring fortuitous compensations through divorce or migration; however, these unions are also the least fertile, so that the fertility of unions surviving after 10 years is higher than the fertility of all unions in the cohort, absent mortality.

To calculate the overall fertility, we therefore need to know the marriage-duration–specific fertility rates for different ages or age groups at marriage and combine them as they would be if the proportions of ages at marriage had remained unchanged. The error committed by performing a lump-sum computation—without distinguishing between the subcohorts of ages at marriage—is tolerable. But if we start excluding the marriages in which the wife is over 50 years, if the mortality at reproductive ages is fairly low and if the breaks in married life due to separation and migration are rare, then the study of the influence of the age at marriage on fertility is sufficiently relevant to make it worth renouncing, in practice, a purely methodologic simplification.

IV. POSTTRANSITION SOCIETIES

Since the late 1960s, the Western world has experienced major changes in collective behavior, concerning not only reproduction (end of the baby boom and lasting fall in total fertility rate well below the level needed for cohort replacement; van de Kaa, 1994) but the entire system of family formation and development. Of all these changes in the past 30 years, none has captured the attention of demographers and alarmed officials more than the drastic slump in fertility. This is not only because of the incredibly low levels reached, but also because of the differences in the pace of the trend across Europe. In particular, the Mediterranean countries, which initially lagged, have seen a violent plunge in their fertility after 1975—to well below the rates in central and northern Europe. But the transformations in marriage systems are no less important and decisive for the process we are seeking to estimate here; more generally, the crucial factors are the demographic behaviors governing the formation and stability of partner relationships and, consequently, family formation.

The changes have been sufficiently important to be taken as evidence, by some specialists, of a *second demographic transition* (van de Kaa, 1994). This new evolutionary phase, they claim, signals a radical break with the previous regime, even if, in some national contexts, the succession of birth cohorts does not display a genuine discontinuity (Santini, 1995). From 1970 to 1985, nuptiality has declined substantially. According to projections for relatively recent cohorts (born in 1960 and after), the proportion never married may reach nearly 50% in some countries, particularly of northern Europe (Roussel, 1987). Overall, the

Western world is heading toward a situation where one third of each cohort may remain outside marriage, at least in legal terms.

These fewer and later marriages are also more fragile. Around the 1960s, divorces occasionally peaked at 10% of marriages. More recently (2001), the period indicators have very often reached values near 40%; in northern Europe (Austria, Belgium, Finland, The Netherlands, England and Wales, Sweden, Denmark), that level has been far exceeded. In central–northern Europe, the ultimate proportion of divorced couples in the cohorts of marriages celebrated in recent decades will probably not be under 30%[9]—although that will in no way imply a recovery via second marriages. Moreover, although the phenomenon is hard to quantify, it is certain that a growing number of marriages are effectively terminated even if the break is not legally recognized: in other words, the evidence suggests more *de facto* separations than official divorces.

However, the difference with what we have called the *first transition* is even greater. The trend in *de jure* unions has been matched by profound changes in *de facto* unions. The decline in *legitimate nuptiality* has been broadly offset by the increase in consensual unions. This phenomenon is most visible among the under-30s. It has taken on considerable importance in some areas such as Scandinavia—most notably Sweden and Denmark, where, before anywhere else in Europe, nearly all marriages have been preceded by a fairly long cohabitation. But the model has spread widely across western Europe, except in the southern countries. There, couple formation is still mostly governed by traditional rules, although it will likely follow the path taken by the rest of the continent—with some local adjustments. In several countries, including France and Switzerland, half the cohorts of married people have cohabited before marriage.

Not all cohabitations, however, can be described as prenuptial union. Among young people and divorcees, there has been a gradual increase in unions that do not seem aimed at marriage. We are dealing no longer with *engagement* periods or trial marriages, but with true *marriages without certificates*. These unions—like some prenuptial cohabitations, in fact—no longer exclude the birth of children. As a result, fertility outside marriage has grown at an unprecedented pace since the 1980s. Once upon a time, fewer than 10% of births were *illegitimate*; the percentage was often much smaller, as in Italy and Spain. Today, the proportion now easily reaches or exceeds 20%—as in, for example, France, England and Wales, and the United States.[10] In Scandinavia, it ranges as high as 50%. Since the mid 1980s, many of these characteristics have consolidated and we are witnessing new changes, such as the upturn in fertility above age 30 and, in some countries, a halt in the decline in fertility among the young, either because the levels reached are extremely low, or because of a rise in teenage fertility. In certain countries, the upswing in fertility rates among the over-30s has led to a mild increase in total period fertility.

True, this summary may give an oversimplified image of a complex evolution, which is highly diversified in space and time, and extraordinarily important. As with the first demographic transition, some heterogeneity factors are emerging that may outweigh the common aspects, as different populations are following different paths. However, there are some basic shared characteristics. The first fertility transition was matched by a rise in individual autonomy, which has led to the rejection of religious bans on contraception; but the manifestation of this disagreement has remained totally private and the first transition was, as a result, very silent. By contrast, the second transition is driven by a far more vocal demand for even greater individual autonomy—in radical opposition to all forms of external institutional authority (Lesthaeghe, 1992; Lesthaeghe and Moors, 1991; Simons, 1994). Such is the context in which the form of unions is changing: fewer and fewer marriages, more and more consensual unions, a near-rejection of any official link that could institutionalize a sentimental attachment. The generalization of these new family structures is still confined to specific countries and large urban areas, and to the demographically most modern social strata (De Sandre *et al*., 1999). Apart from the recent plunge in their fertility, the Mediterranean countries, for example (particularly Italy and Spain), have not (yet) experienced a full-scale inversion of the trends in the complex phenomena that shape family formation (Lesthaeghe, 1992; Santini, 1995; De Sandre *et al*., 1997).

V. NEW FAMILY STRUCTURES AND THE MEASUREMENT OF FERTILITY: THE PROBLEM OF HETEROGENEITY

These changes in demographic behavior have revolutionized the analysis of marriage fertility. A few

[9] Nor should we overlook the fact that the tempo has changed: Today, couples divorce much sooner after marriage than in the past.

[10] In 1994, half of all pregnancies in the United States took place outside marriage. Half of these out-of-wedlock pregnancies ended in birth (54%) (Darroch *et al*., 1999). In 2001, the proportion in France, Norway, Denmark and Sweden exceeded 40% of all births (UNECE, see the Web site www.unece.org/stat/scriptsdb/variables.asp).

features need to be highlighted in order to properly understand the obstacles presented by the new demographic regime to measuring the phenomenon, whether from continuous observations or, for different and more complex reasons, from retrospective information.

First, we must stress that the disinstitutionalization of marriage has led to a proliferation of *de facto* situations at the expense of *de jure* situations; this applies both to the formation of couples living maritally and to the stability of unions, legitimate or not.

Second, despite variations in time and space in the dissemination of the new cultural and moral principles (variations already mentioned repeatedly), it seems hard to argue that the Western world still enforces the legitimacy principle (in the Malinowskian sense referred to at the start of the chapter) whereby the social right to have children is confined to marriage. Legal marriage is no longer, in practice, the exclusive locus of reproduction.[11] The measurement of marital fertility therefore loses its significance—not only in the countries where the proportion of births outside marriage is already very high, but also in more traditional societies where, in any event, the social stigma of illegitimate fertility has been drastically attenuated, if not totally dispelled.

Third, if the demographic reference unit for the reproductive process is no longer legal marriage alone, but encompasses all forms of heterosexual union, we must reckon with the fact that each of these unions seldom constitutes a reproductive history in its own right any longer; ever more often, it is merely one of several partnership sequences in an individual's fertility history.

The consequences are immediate and very clear for the analysis of continuous-observation data. These are extracted from declarations made to administrative agencies empowered to keep track of a *legal obligation* for individuals to report acts formally recognized by prevailing law. We cannot, therefore, use such data as the basis for defining a cohort of unions or monitoring it over time. Let us assume we wanted to restrict our attention to a conventional marriage cohort—despite the high probability of selecting different fertility behaviors among individuals who choose to marry versus all couples in the aggregate. We would still need to measure the fertility rate on the basis of the cohort's initial size in the denominator at each duration, despite the obstacles that are augmented and aggravated by the growing fragility of marriages—in *de jure* and *de facto* terms.

[11] This, of course, would also imply a reappraisal of the demographic role of nuptiality.

Because *de facto* unions can be determined only by questioning the couples directly, the reconstitution of the reproductive path—now linked to the woman's partnerships rather than to the woman alone—calls for survey-based observation. The spread of the new couple-formation patterns has given rise, since the late 1970s, to large-scale retrospective surveys. These are the only data sources enabling demographers to compile information on unions not regulated by legislative provisions and on *reconstituted* families or stepfamilies (Villeneuve-Gokalp, 1990). However, even in a retrospective observation framework, there is no shortage of problems involved in measuring a phenomenon as complex as the fertility of these undocumented unions.

The alternative cohabitation forms can be described according to the event that follows entry into union—marriage, birth, separation—and the union duration up to the event (Villeneuve-Gokalp, 1990). First, cohabitation can be regarded as a transitional arrangement that precedes but does not replace marriage. In this case, children will not be born until after the union is legalized. If the union is short-lived, it will be merely a prelude to marriage; if it has lasted longer, it may constitute a trial marriage. Alternatively, cohabitation may begin without any plan to continue it in marriage and without any commitment. It may end swiftly in separation, in which case it is only an ephemeral situation. But, equally well, it may endure long enough to form a lasting (stable) union, even without a special event such as marriage or birth (Munoz-Perez and Prioux, 1999a, 1999b). Last, the partnership may take the form of what is generally known as a consensual union (in French, *union libre*): A lasting cohabitation in which the partners live as husband and wife. Not only is their union stable but it acts as a locus for motherhood and fatherhood, without the legal ties of marriage. This is a new type of family, an additional kind of marital-event history.

However, the classification gets complicated when we consider the possibilities of entering into successive unions, legal or not, with different partners. The propensity to have children then becomes a composite function of the entry into first union, the precariousness or stability of the cohabitation (does it turn into marriage or not?), its duration, the propensity to form a new union after a separation, and so on.

In sum, these new family structures require—before any fertility measurement—an analysis of the heterogeneity of subgroups resulting from union terminations, the variety of union forms (marriage, cohabitation, etc.), and sequences of successive unions.

We are now measuring family formation not in a single marriage, but in a succession of partnerships with different legal statuses. Accordingly, what issues do we need to address in the retrospective survey to achieve a satisfactory measure?

For the sake of rigor and completeness, the analysis must obviously concern a cohort of individuals who have already formed a union at least once and have thus been able to express their fertility. The variability factors—the source of the group's heterogeneity—are numerous. In measuring the fertility of a consensual union, we must therefore simultaneously take into account exits due to termination of the union by marriage or by death, migration, and so forth. This requires a highly detailed classification into subgroups. For example, we will need to calculate fertility rates for specific union durations and marriage durations (as the case may be)—allowing for union terminations and competing events—by duration until the event; alternatively, we must use methods based on a dynamic analysis of individual data similar to the *life cycle* approach. We will briefly discuss event-history analysis in the subsequent sections. To simplify the presentation of measurements based on aggregate data, we can describe several typical situations, each reflecting a single aspect of an increasingly complex reality (Festy and Prioux, 1989).

1. Fertility in the Event of a Marriage Termination

For simplicity's sake, let us consider the measurement of marital fertility with frequent divorces. As seen earlier, continuous observation gives us virtually no data on marriages surviving at different durations. When we compute legitimate-birth frequencies, the denominator therefore consists of the initial number of marriages, which—if divorces and separations are deemed negligible—offer a correct proxy for the fertility quantum. The rising instability of marriages, however, makes this estimation somewhat irrelevant, obliging us to turn to survey data. Thanks to the retrospective data on ever-marrieds, we can make allowance, within a marriage cohort, for the duration elapsed since the marriage by aggregating the surviving marriages and unions terminated by divorce or separation. In such conditions, given the high propensity to separation and divorce, should we calculate the marriage-duration–specific rate by (a) setting the denominator to the decreasing numbers due to the exit of some couples from the cohort of ever-marrieds surviving at the time of the survey (regarded as the initial population) or (b) using a constant number of married persons, in keeping with the standard method described in Post-transition Societies, disregarding marriage terminations occurring since the start of the process?

Approach A assumes the truth of the independence hypothesis (and, of course, that of the continuity hypothesis). This holds that terminated couples—who have left the observation field—would have behaved, had they stayed together, in the same way as the couples still together: the implication is that fertility is not selective with respect to exits from the cohort. Approach B requires the additional hypothesis that marital fertility does not differ according to whether the couples break up before the potential end of the process or—on the contrary—stay together until the end of the reproductive period. The first hypothesis, of course, is totally unverifiable. But, if the survey data are sufficiently detailed, we can observe possible differences in behavior.

The importance of these issues is naturally increasing with the rise in divorces, which entails an ever greater number of premature marriage terminations. Special care should be taken to classify the respondents who, in the survey, have described themselves as separated or divorced by year of marriage and by marriage duration at the union-termination date. For each marriage cohort, we will thus construct subcohorts by duration at the time of separation and calculate (retrospectively) marriage-duration–specific fertility rates. By comparing the duration-specific rates for the two subgroups, we will be able to see whether a difference exists between stable marriages and interrupted marriages: this hypothesis turns out to be plausible, for example, for French cohorts in the 1960 to 1969 period (Festy and Prioux, 1989).

2. Marital Fertility versus Consensual-Union Fertility

Earlier, we stressed the growing numeric and social importance of cohabitation. Our main purpose was to remind the reader that in measuring total fertility we need to abandon the classical distinction between *legitimate* or marital fertility and *illegitimate* or nonmarital fertility when the spread of alternative forms of unions ceases to be negligible. We could, however, restore some meaning to the concept of illegitimate fertility (with no moral connotation, of course, but only as a label denoting an exceptional situation) by confining it to births occurring in these unions, which we previously described as *ephemeral* in contrast to stable unions. The term *illegitimate* would therefore be restricted to specific population groups such as teenagers (incidentally, we will see in the following paragraphs that the fertility of very young girls poses other measurement problems). By contrast, under these definitions, births occurring in stable unions where partners live as if they were married are in no way illegitimate and could usefully serve to compute *union fertility* as distinct from *marriage fertility*.

At this stage, the correct measurement of fertility in and outside legal marriage raises two types of questions: For a given union duration, does the propensity to have a child differ according to the type of union and des the transition between the two statuses entail a change in reproductive behavior? Both questions are relevant to the issue of measuring the fertility quantum of unions, taking into consideration the fact that some informal unions turn into legal ones. Would cohabiting couples who decide to marry have behaved, for the same duration, in the same manner as couples who do not? One could, in a sense, regard marriage as an additional intrusive event, but it would certainly be wrong to apply the independence hypothesis to marriage, because—by definition—*de facto* union and marriage are competing and dependent events (Thierry, 1993). Conversely, we may well ask whether a synthetic indicator of the marital fertility quantum is still correct when the group of marriages to which it applies is determined, at least in part, by the legalization of prenuptial cohabitations. Here as well, we are dealing with a heterogeneous group of marriages, and our analysis must rely on the classification of subgroups by union duration at the time of legalization. Fertility should therefore be measured for different subgroups, with a distinction between the period before the celebration of marriage and the period after; the measurement should thus be based on the rates by union duration at the legalization date and by marriage duration. This will enable us to determine whether the marriage event produces a behavioral discontinuity.

The French survey of female cohorts born between 1917 and 1964 yields two further observations: (a) There seem to be no differences in legalized-union fertility by union duration at the legitimation date but (b) the fertility of nonlegitimated unions is distinctly lower.

The study on French consensual unions that have turned into marriages after 5 years shows (on a common time axis) a bimodal tempo, with a first peak at the start of cohabitation and a second after legalization. We can thus conclude that fertility is not the consequence of union formation but—on the contrary—a cause; likewise, some of the legalizations are prompted by a pregnancy. The distinction between consensual union and marriage loses its relevance as regards the propensity for childbearing, and we can study duration-specific union fertility ignoring marital status: The quantum index obtained (by summing consensualunion periods and marriage periods) plays the same role as the overall fertility rate with respect to the legitimate-fertility and illegitimate-fertility rates.

However, we should continue to give separate treatment to the group of unions that do not turn into marriage. For this category, the behavioral heterogeneity reflected in the diversity of union forms is in itself a heterogeneity factor in the propensity for childbearing.

3. Fertility of Successive Unions

The increasing instability of marriage and the high probability of consensual-union termination have produced a growing number of stepfamilies or *reconstituted families* (i.e., unions above order one, whose origin-event is the end of a previous union). As a rule, later unions have a lower fertility than first unions because the new couples have already had children from the earlier union.

One way to simplify measurement problems would be to regard the total duration in union (regardless of union order) as a time axis on which to compute the fertility rate. This would be tantamount to assuming a continuum between the end of the previous union and the start of the following one, as if the partners in the new couple tended to complete the family formation begun in the preceding union. Obviously, this hypothesis obliges us to ignore the periods of nonunion (i.e., of nonexposure to risk).

We have already emphasized, in our discussion of retrospective surveys, that there is no basic difference in reproductive behavior, for an identical union duration, between the subgroups of couples who are still together at the end of the process and those who separate instead. Once again, however, we need to check the heterogeneity of the subgroups by union duration at the termination of the first union: For a given total union duration, there must be no significant differences in behavior between these subgroups, or between them and the unions that have never broken up. Absent such differences, the new couple's fertility would take up—so to speak—where the fertility of the previous union had left off. We could thus measure fertility by complete duration of successive unions as if they formed a single union. But, if we rely once again on the retrospective analysis of the fertility of French cohorts, the facts do not seem to fit that assumption (Festy and Prioux, 1989). When we compare the duration-specific rate for stable couples and the rate for separated couples, and then reformat the rate for the hypothetical durations of the unions if they had not been terminated, we find a sharp discontinuity: the new unions are far more fertile than the first ones. In sum, to obtain a measurement of the phenomenon that can be correctly interpreted, it is better to control for the heterogeneity by keeping the two subgroups apart in the analysis.

VI. EVENT-HISTORY ANALYSIS AND FERTILITY IN MULTIPLE UNIONS

The analysis of aggregated survey data leads us to examine special groups of unions but, as we have been able to observe, this forces us to make simplifying hypotheses. In the reconstruction of a cohort's typical event history through a retrospective survey, the interviewees are the set of *present and surviving* members of the original cohort. The population thus remains constant throughout the observation; it never deviates, numerically, from the initial population due to losses caused by negative events or gains caused by positive events. The hypothesis of *independence* between competing risks of events that do not alter the population size guarantees the relevance of quantum measurements. The respondents, however, must not constitute a biased sample of the initial set. In other words, the persons who have left the cohort before the survey must have the same propensity to experience the event studied as the *present and surviving* respondents. To ensure that the indicators calculated on data obtained from retrospective observations correspond to the raw figures, as in the continuous-observation method, the independence between the competing phenomena and the event studied must be documented before the occurrence of the event itself and must persist after it (Thierry, 1993; Salvini and Santini, 1999). The problem becomes complicated if the event studied, which changes the status of cohort members, is in competition not only with the intrusive events but also with other events leading to the same status change (Thierry, 1993). When we want to keep the different outcomes separate, the independence hypothesis is rarely tenable. A preferable alternative is to replace it—where the phenomenon studied allows—by a *complementarity* hypothesis. This holds that the events examined, rather than *interfering* with one another, flow from a prior decision. That is precisely the case with cohabitation and marriage: The complementarity hypothesis assumes that

> Each individual made a final and deliberate choice between the two forms of union and thus joined one of two distinct groups: those who have opted for cohabitation or those who have preferred marriage (Festy, 1986; p. 343).

To measure the fertility of these different forms of union, it therefore becomes essential to divide the cohorts into subgroups. We can thus check for heterogeneity according to distinct duration criteria for each type of union. This can quickly lead to an over-fragmentation of the cohort studied, without necessarily ensuring the explicit inclusion of all the heterogeneity components.

The availability of individual biographical data thanks to retrospective observation and the techniques of *event-history analysis* allow a treatment of initial or acquired heterogeneity (Salvini and Santini, 1999). In other words, individual fertility paths can be analyzed in parallel with marital event histories in order to measure the conditional risks of each status. The assessment of the risk of experiencing a given event rests on an examination of dependencies between events. For simplicity's sake, let us confine the discussion to two processes (union and fertility). The dependencies will be measured via the observed effect on the distribution of the first event when the second occurs and, subsequently, the changes in the distribution of the second event when the first occurs.

Using risk functions, we can compute summary quantum and tempo indicators with no need to assume independence and continuity between phenomena (Courgeau and Lelièvre, 1994 and 1996; Festy, 1994; Blayo, 1995). To measure the quantum of the phenomenon, event-history analysis relies on the computation of *cumulative risks* (i.e., the sum, over time, of the instantaneous risks of experiencing the different events in interaction). Unlike survival functions, this indicator allows us to regard the processes as independent even when the risks are dependent (Andersen *et al*., 1992; Courgeau and Lelièvre, 1996).

VII. FERTILITY OUTSIDE MARRIAGE AND ADOLESCENT FEMALE FERTILITY

It is generally admitted that births outside marriage have long been a minor phenomenon in population dynamics and for the understanding of the reproductive process.

Let us go back to the *ancien regime* societies described earlier. We can distinguish three long periods in the evolution of fertility outside marriage (or illegitimate fertility, even though the two concepts should not be strictly equated without a rigorous definition or cultural precautions[12]: Laslett *et al*., 1980). As

[12] The historic literature on illegitimacy is considerable, but the tone of the studies on the subject generally depends on the prevailing behavioral ethics. Before laicization—in the 18th, 19th, and early 20th centuries—the frequency of illegitimate births was actually regarded as a symbol of the moral condition of society: the increase in births outside marriage, as a sign of decadent morals; its reduction, as a sign of victory over moral decline by authorities and the clergy (Shorter *et al*., 1980). A more objective vision was introduced with the birth of the social sciences: Kingsley Davis (1939a and 1939b) authored the works that, in a sense, inaugurated the specifically sociologic era of illegitimacy studies. The adoption of ethical impartiality or neutrality toward the illegitimate child status sig-

far back as the data available on this issue go (i.e., from the 16th century to the first half of the 18th century), illegitimate births represented a tiny minority of total births, from 1% to 4% at most. It is in the second half of the 18th century that we find a gradual increase in the phenomenon, across most of Europe, up to a peak in the 1850s to 1880s. The following period saw a sharp down swing: By the eve of World War II, illegitimate births in most European countries had returned to very low levels, well below the 19th century values (Shorter *et al*., 1971).

In the 1950s, however, the number started trending up again, but the phenomenon did not become significant until the 1980s. This period witnessed the gradual increase in the number of nonlegalized unions among the young and divorcees. In their aims and their stability, such unions assumed the characteristics of full-fledged *marriages without certificate*. The fact that these unions do not exclude the birth of children accounts for the sharp rise in fertility outside marriage. Before this radical shift in family structures in European societies, fewer than 10% of births were *illegitimate*.[13]

Today, births outside marriage have risen to and even exceed 30% in a number of places including France, England and Wales, and the United States; in Scandinavia, they account for nearly half of total births. In Italy, the proportion is far lower, albeit constantly expanding: Births outside marriage, which represented under 7% of the total until the 1990s, rose to 9.7% by 2001 (Council of Europe, 2002). In sum, fertility outside marriage has become, in some cases, the rule rather than the exception, making *illegitimacy* a totally improper designation for these children living in families whose characteristics are comparable to those based on a legal marriage. Indeed, even in the past, other historic vicissitudes of the social, legal, or political environment make it necessary to provide a precise description of the context and an adequate definition, more relevant to the period and the culture—under pain of producing meaningless results (Laslett *et al*., 1980).

We have already discussed the problems of documenting and measuring fertility created by these new forms of union, but we thought it would be useful to elaborate here on the very specific issue of teenage pregnancies. These generally occur outside of a stable union and often cause trauma. Not until the 1960s to 1970s did teenagers become interested in fertility outside marriage—a phenomenon that, in some groups, is threatening to become a social ill.

From this standpoint, the reproductive behavior of female adolescents is of special interest. There has been much research on the social effects of teenage fertility, and the political sensitivity to the issue has prompted *ad hoc* surveys, since motherhood before age 19 is seen—in demographically advanced societies—as a behavior liable to disturb teenage girls maturing process, their future occupational performances, and, ultimately, their well-being and quality of life (Montgomery and Casterline, 1996).

This growing interest is closely linked to the fact that, in recent years, the phenomenon has been dramatically amplified in many countries (Foster and Hoffman, 1996; Hollander, 1996). According to the latest United Nations data, the fertility rate at age 15 to 19 varies from a low of 3.7% in Japan (but we also find fairly low values in Mediterranean Europe) to a peak of 59.2% in the United States (Singh and Darroch, 1999; Darroch *et al*., 1999).[14] The other Anglo-Saxon countries display rates ranging from 28.6% in the United Kingdom to 33.6% in New Zealand.

In the Western world, the problem is rather new. Before the recent changes in the *second demographic transition*, only a few Western countries had high teenage fertility rates linked to high levels of illegitimate fertility. As we have seen, during this process, the concept of illegitimacy seems to have lost its meaning and substance because of the rising number of births outside marriage, occurring in stable unions. Illegitimacy is therefore concentrated in the youngest age groups, and often has a disruptive effect on personal and family relationships. This gives a greater relevance to the study of illegitimacy, including its measurement.

Like the rates at other fertile ages, the data cited above are computed in terms of birth frequency (Darroch *et al*., 1999) or by dividing the number of births to girls aged 15 to 19 years by the number of

naled a break with the past—although reproduction outside marriage long continued to be treated as a deviant behavior (Shorter *et al*., 1980). The evolution of public morality on this issue reflects changing attitudes toward marriage: the change in the frequency of illegitimate births seems to reveal the erosion of marriage and other family institutions.

[13] Except in the entirely atypical case of Portugal, where, for different reasons, the proportion has been much higher in the past: From 12% in 1928, it rose to 17% in 1937, before declining steadily since (Livi Bacci, 1971).

[14] In 1992, 112 pregnancies per 1000 girls aged 15–19 years were observed; 61% ended in birth, 36% in intentional abortion. The fertility rate of this age group rose by nearly 12 points between 1987 and 1991 (Henshaw, 1997). The teenage birth rate dropped 5% in 2002, to 43.0 births per 1000 females aged 15–19 years, another historic low. This rate has fallen 30% since 1991 (61.8). Rates declined for teenagers of all ages; the rate for the youngest teenagers, aged 10–14 years, declined to 0.7 per 1000, exactly half the rate reported for 1994 (http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_10.pdf).

young women in that age group. We disregard the conceptual and methodologic problems specific to the measurement of teenage fertility—problems that, once again, are chiefly due to the heterogeneity of situations. From a biologic (sexual maturity) and behavioral (sexual intercourse) standpoint, teenage girls form an extremely diverse aggregate. The process of change and the high variability within the group raise technical issues that are not simple to resolve. This provides an at least partial explanation of why sometimes, despite the wealth of social analyses, the literature on the statistical tracking and processing of female-adolescent fertility data is truly scarce. Similarly to the problems mentioned earlier concerning the other ages of reproductive life, the distinction between legitimate and illegitimate teenage fertility is even more fraught with imprecision and misleading appearances, particularly because of those who—not so long ago—spoke of *reparatory marriage*, namely, marriage not as the locus of reproduction, but as its effect. In dealing with this issue, it is crucial to remember the tremendous importance not only of the changing legislation on minimum age at marriage but also of the regulations on access of female minors to effective contraception methods and legal abortion (Livi Bacci and Ventisette, 1980). More than at other ages, heterogeneity concerns both individual factors (biologic, for example) and contextual factors (legal norms, among others).

Consequently, we should stress that—despite their widespread use—the conventional measures (legitimate and illegitimate fertility rates) are not entirely satisfactory for teenagers, because we cannot easily identify the population at risk of conception and childbearing. Unlike with other age groups, it is essential to the analysis that two preconditions be met: The young girl must be sexually mature, and she must have intercourse. Naturally, the retrospective surveys that allow an analysis of the variability of risk exposure for this age group are rarely available. Without such information, no meaningful comparison between populations is possible: the differences observed using classic age-specific rates may largely reflect the differences in the proportions of teenage girls actually at risk of pregnancy.

Acknowledgment

Translated from the French by Jonathan Mandelbaum.

References

Andersen Per Kragh, Borgan Ørnulf, Gill Richard, and Keiding Niels 1993. *Statistical Models Based on Counting Processes*. New York, Springer Verlag, 768 p.

Blayo Chantal, 1995. La condition d'homogénéité en analyse démographique et en analyse statistique des biographies, *Population*, vol. 50(6), p. 1501–1518.

Bourgeois-Pichat Jean, 1950. *Mesure de la fécondité des populations*. Paris, INED, 152 p. (Travaux et Documents series, *Cahier* no. 12).

Bourgeois-Pichat Jean, 1987. La demografia e le altre scienze: interazione di metodo e contenuto, *in:* Eugenio Sonnino *et al.* (eds), *Demografia: scienza, insegnamento, professione*, p. 53–57. Milan, F. Angeli, 571 p.

Calot Gérard, 1984. La mesure des taux en démographie. Paris, INED, PUF, 322 p. (Travaux et Documents series, *Cahier* no. 104).

Conseil de l'Europe, 1999. *Évolution démographique récente en Europe*. Strasbourg, Éditions du Conseil de l'Europe, 643 p.

Conseil de l' Europe, 2002. *Évolution démographique récente en Europe*. Strasbourg, Éditions du Conseil de l'Europe.

Courgeau Daniel and Lelièvre Éva, 1994. Risques compétitifs et indépendance: cadre théorique d'une réflexion, *Population*, vol. 49(2), p. 482–489.

Courgeau Daniel and Lelièvre Éva, 1996. Les motivations des comportements migratoires, *in:* Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Demografia: analisi e sintesi. Cause e conseguenze dei processi demografici. Seminario Internazionale, Siena, 22–24 aprile*, vol. 1, p. 143–164. Rome, DSD, Università di Roma La Sapienza, 252 p.

Darroch Jacqueline E., Landry David J., and Oslock Selene, 1999. Pregnancy rates among U.S. women and their partners in 1994, *Family Planning Perspectives*, vol. 31(3), p. 122–126.

Davis Kingsley, 1939a. The forms of illegitimacy, *Social Forces*, vol. 18(1), p. 77–89.

Davis Kingsley, 1939b. Illegitimacy and the social structure, *American Journal of Sociology*, vol. 45(2), p. 215–266.

Del Panta Lorenzo and Rettaroli R. S., 1997. Metodi della demografia storica, *in:* Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Démographie: analyse et synthèse. Causes et conséquences des évolutions démographiques*. Proceedings of the San Miniato (Pisa) Seminar, December 17–19, 1997, vol. 2, p. 205–230. Rome: DSD; Paris: INED, 236 p. (Materiale di studi e di ricerche series, special issue).

De Sandre Paolo, Ongaro Fausta, Rettaroli Rosella, and Salvini Silvana, 1997. *Matrimonio e figli, tra rinvio e rinuncia*. Bologna, Il Mulino, 218 p.

De Sandre Paolo, Pinnelli Antonella and Santini Antonio (eds.), 1999. *Nuzialitàe fecondità in trasformazione: percorsi e fattori del cambiamento*. Bologna, Il Mulino, 837 p.

Donovan Patricia, 1999. The "illegitimacy bonus" and state efforts to reduce out-of-wedlock births, *Family Planning Perspectives*, vol. 31(2), p. 94–97.

DupÂquier Jacques, 1972. De l'animal à l'homme. Le mecanisme autorégulateur des populations traditionnelles, *Revue de l'Institut de Sociologie*, no. 2, p. 177–211.

Festy Patrick, 1986. Cohabitations et cohabitants: enquête et indices, *in:* AIDELF (ed.) *Les familles d'aujourd'hui. Démographie et évolution récente des comportements familiaux*, p. 343–355. Paris: Association Internationale des Démographes de Langue Française, 600 p. (International colloquium, Geneva, September 17–20, 1984).

Festy Patrick, 1994. Quelques commentaires sur un problème fondamental de l'analyse démographique, *Population*, vol. 49(2), p. 490–498.

Festy Patrick and Prioux France, 1989. L'analyse de la fécondité en pays developpés: nouveau défis, *in:* IUSSP (ed.), *International population conference, New Delhi*, vol. 1, p. 285–302. Liège: IUSSP, 448 p.

Foster E. Michael and Hoffman Saul D., 1996. Non-marital childbearing in the 1980s: assessing the importance of women 25 and older, *Family Planning Perspectives*, vol. 28(3), p. 117–119.

Galan Vivas M. C., 1969. Report of rapporteur: problème de la nuptialité, *in:* IUSSP, *International population conference, London 1969*, vol. 3, p. 2182–2185. Liège: IUSSP, 2296 p.

Gini Corrado, 1932. Di un procedimento per la determinazione del numero medio di figli per matrimonio, *Metron*, vol. 10(1–2), p. 1–31.

Glass David V. and Grebenik Eugene, 1954. *The trend and pattern of fertility in Great Britain. Part II*. London, HMSO, 692 p.

Hajnal John, 1965. European marriage patterns in perspective, *in:* David V. Glass and David E. C. Eversley (eds.), *Population in history: essays in historical demography*, p. 101–143. Chicago and London, Aldine Publishing Company, 682 p.

Henry Louis, 1953. *Fécondité des mariages. Nouvelle méthode de mesure*. Paris, INED, PUF, 180 p. (Coll. Travaux et documents, *Cahier* no. 16).

Henry Louis, 1958. *La population de Crulai paroisse normande*. Paris, INED and PUF, 270 p. (Travaux et Documents series, *Cahier* no. 33).

Henry Louis, 1972. *Démographie, analyse et modèles*. Paris, Larousse, 381 p.

Henry Louis, 1985. La reconstitution des familles. Origines et perspectives, *European Journal of Population*, vol. 1(4), p. 305–307.

Henshaw Stanley K., 1997. Teenage abortion and pregnancy statistics, 1992, *Family Planning Perspectives*, vol. 29(3), p. 115–122.

Hollander Dore, 1996. Non-marital childbearing in the United States: a government report, *Family Planning Perspectives*, vol. 28(1), p. 29–32.

Laslett Peter, 1980. Introduction: comparing illegitimacy over time and between cultures, *in:* Peter Laslett, Richard M. Smith, and Karla Oosterveen (eds.), *Bastardy and its comparative history: studies in the history of illegitimacy and marital nonconformism in Britain, France, Germany, Sweden, and the United States*, p. 1–68. Boston, Harvard University Press, 431 p.

Lesthaeghe Ron, 1995. La deuxième transition démographique dans les pays occidentaux: une interprétation, *in:* Dominique Tabutin, Thierry Eggericks, and Catherine Gourbin (eds.), *Transitions démographiques et sociétés. Chaire Quetelet 1992*, p. 133–180. Louvain-la-Neuve, Institut de démographie, Université catholique de Louvain, Académia, L'Harmattan, 691 p.

Lesthaeghe Ron and Moors Guy, 1991. Razionalità, coorti e riproduzione, *in:* Antonio Golini *et al*. (eds), *Famiglia, figli e società in Europa*, p. 235–266. Turin, Edizioni della Fondazione G. Agnelli, 359 p.

Livi Bacci Massimo, 1971. *A century of Portuguese fertility*. Princeton: Princeton University Press, 149 p.

Livi Bacci Massimo, 1998. *Storia minima della popolazione del mondo*. Bologna, Il Mulino, 300 p.

Livi Bacci Massimo, and Ventisette Moreno, 1980. Pregnancy and childbearing in adolescence in countries with low fertility, *in:* National Institutes of Health (ed.) *Social, economic, and health aspects of low fertility*, p. 235–257. Washington, U.S. Department of Health, Education and Welfare, Public Health Service, 320 p.

Montgomery Mark R. and Casterline John B., 1996. Social influence, social learning and new models of fertility, *in:* John B. Casterline, Ronald L. Lee and Karen A. Foote (eds.), *Fertility in the United States: new patterns, new theories* (supplement to *Population and Development Review*, no. 22), p. 151–175. New York, The Population Council, 337 p.

Munoz-Perez Francisco and Prioux France, 1999a. Une enquête dans les registres d'état civil. Filiation et devenir des enfants nés hors mariage, *Population*, vol. 54(3), p. 251–270.

Munoz-Perez Francisco and Prioux France, 1999b. Les enfants nés hors mariage et leurs partens. Reconnaissances et légitimations depuis 1965, *Population*, vol. 54(3), p. 481–508.

Prioux France, 1991. La fécondité en début de mariage. Réflexions sur le calcul des taux, *Population*, *Hommage à Roland Pressat,* vol. 46(6), p. 1491–1512.

Roussel Louis, 1987. Deux décennies de mutation démographique (1965–1985) dans les pays industrialisés, *Population*, vol. 42(3), p. 429–448.

Ryder Norman B., 1956. La mesure des variations de la fécondité au cours du temps, *Population*, vol. 11(1), p. 29–46.

Santini Antonio, 1974. *La fecondità delle coorti. Studio Longitudinale della fecondità italiana dall'inizio del secolo XX*. Florence, Dipartimento Statistico-Matematico, 272 p.

Santini Antonio, 1995. *Continuitàe discontinuitànel comportamento riproduttivo delle donne italiane nel dopoguerra: tendenze generali della fecondità delle coorti nelle ripartizioni tra il 1952 e il 1991*. Florence: Dipartimento Statistico, 60 p. (Working Paper 53).

Salvini Silvana and Santini Antonio, 1999. Dalle biografie alle coorti, dalle coorti alle biografie, in *Quadri concettuali per la ricerca in demografia: quattro saggi*, p. 40–54. Florence, Dipartimento di Statistica G. Parenti, 101 p. (Ricerche teoriche series no. 15).

Shorter Edward, Knodel John, and van de Walle Étienne, 1971. The decline of non-marital fertility in Europe, 1880–1940, *Population Studies*, vol. 25(3), p. 375–593.

Simons John, 1994. The cultural significance of western fertility: trends in the 1980s. Liège: IUSSP, 17 p. (paper given at Values and Fertility Change Seminar, Sion, Switzerland, February 16–19, 1994).

Singh Susheela and Darroch Jacqueline E., 1999. Trends in sexual activity among adolescent American women: 1982–1995, *Family Planning Perspectives*, vol. 31(5), p. 212–219.

Thierry Xavier, 1993. La nuptialité à l'épreuve de la cohabitation, *Population*, vol. 48(4), p. 939–974.

van de Kaa Dirk J., 1994. The second demographic transition revisited, *in:* Gijs Beets *et al*. (eds.), *Population and family in the low countries 1993: late fertility and other current issues*, p. 81–126. NIDI/CBGS Publications nr. 30, Amsterdam/Lisse: Swets & Zeitlinger, 246 p.

Villeneuve-Gokalp Catherine, 1990. Du mariage aux unions sans papiers: histoire récente des transformations conjugales, *Population*, vol. 45(2), p. 265–298; in English as: From marriage to informal union: recent changes in the behaviour of French couples, *Population: An English Selection*, vol. 3, 1991, p. 81–112.

P A R T

II

BIOLOGIC AND SOCIAL FACTORS OF FERTILITY IN INDIVIDUALS AND COUPLES

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

The formation of some kind of union is a near prerequisite to fertility, but not the only one. The union so formed must also be fertile. There are biologic conditions to this related to fecundity and its converse, infecundity, as well as behavioral requirements: sexual intercourse and ways of influencing the outcomes—contraception and abortion.

Following the overview given by Henri Leridon of these biologic and social factors of fertility as they relate to the individual and the couple in Chapter 30, five further chapters will consider in turn each factor of the various elements that determine human fertility.

In Chapter 31, Henri Leridon considers the factors of fecundability—the probability of conception during a menstrual cycle—and those of the nonsusceptible period (i.e., the period in which conception is impossible because the woman is either pregnant or subject to postpartum infecundability). The former is the result of positive fertility behaviour in a fecund period, the latter sets the schedule of fertility in line with the tempo dictated by both biology (period of gestation) and behaviors (duration of breast-feeding).

Clearly, however, the process of pregnancy can be interrupted before a live birth is produced, thereby reducing effective fertility. In Chapter 32, Catherine Gourbin gives a descriptive survey of the factors of noninduced-abortion–related fetal mortality. Likewise, in Chapter 33, Henri Leridon addresses the causes and solutions to the other biologic fertility-limiting factor: infecundity.

Summarizing the effects of these proximate determinants of fertility not expressly controlled by individuals or couples, the same author reviews the levels and models of so-called *natural* fertility in Chapter 34.

Finally, in Chapter 35, Gianpiero Dalla Zuanna considers the factors of induced abortion, whereas in Chapter 36, Michel Bozon assesses the role of context-specific sexual behaviors.

CHAPTER

30

Biologic and Social Factors of Fertility

An Overview

HENRI LERIDON

Institut national d'études démographiques (INED), Paris, France

I. FERTILITY AND FECUNDITY: THE VOCABULARY

The terms *fertility*, *fecundity*, and *sterility* are widely used not just by demographers but also by health professionals and the general public. Yet they are often a source of confusion, because except within the small world of professional demographers, this vocabulary is far from being precisely defined. Indeed, the same terms are used with completely opposite meanings by different people and linguistic conventions often change between different languages.

The most straightforward task for the demographer when observing a group of individuals and, on occasion, a particular individual (or a couple), is to *count the events that occur in a given period of time*—live births, still births, spontaneous and induced abortions. Thus we say that a couple *has been fertile* in the first 5 years of its marriage if at least one live birth is recorded during this period; in the opposite case (no birth) we say that the couple has *remained infertile* over the same period, which, however, has no implications for its future fertility. For a group of individuals, the *fertility rate* over a fixed interval of time expresses the ratio of the number of live births to the total (mean) membership of the group exposed to the risk. What we are concerned with here, therefore, are measures of *results* that can be, and actually are, observed.

The term *fecund*, on the other hand, attaches to a concept of *capacity.* It would apply, for example, to a woman whose physiologic condition was such that she could conceive if exposed to unprotected intercourse. The opposite of fecund is *infecund* or *sterile.* This concept appears to have obvious relevance for individuals and couples but it is of mainly theoretical value, because although we can sometimes know for certain that a particular man or woman is incapable of procreating (because of a clinically diagnosed pathology), in most cases the fecundity or sterility of individuals or couples is impossible to determine at the individual level. At the aggregate level, however, the concept remains of interest. If we first measure the fecundity of a group of people and then analyze the differences in this fecundity by various sociodemographic characteristics (beginning with age), a probable fecundity can be deduced that is valid at the individual level for each person with these characteristics.

Another important distinction that must be made is between capacity *for fertilization* (for a man, to fertilize; for a woman, to be fertilized) and capacity *for procreation*, meaning (for the woman) being able to carry a pregnancy to term and produce a live-born child. Civil registration statistics usually record only live births (i.e., the products of procreation and [at most] still births). Demographic surveys, by contrast, often contain questions about pregnancies that have terminated with either spontaneous abortion (miscarriage) or induced abortion.

Table 30–1 summarizes the terminology that follows from the preceding comments. The table incorporates the terms whose usage has already been codified in

TABLE 30–1 Conventional Terms and Definitions for Fertility, Fecundity, and Sterility

Event	Concept	
	Capacity (at a point in time)[a]	Result (over a given time period)[b]
Fertilization (conception, whatever the ultimate outcome)	**Fecundable**: capable of being fertilized, of conceiving (M, capable of fertilizing) Ant.: **Infecundable**	Who has conceived, who has been fertilized (M, who has fertilized) ***Ant.:*** who has not conceived (e.g., nulligravida: who has never conceived)
Fertilization followed by intrauterine death	Fecundable, but infecund (or pseudo sterile)	Partially fertile
Procreation (live birth)	**Fecund**: capable of procreating (a live-born child) ***Ant.*: infecund, sterile**	**Fertile**: who has actually procreated (e.g., primiparous, multiparous: who already has one or more children) ***Ant.*: infertile** (e.g., nulliparous: who has never procreated)

M, equivalent for men (the concepts are defined in the table primarily for women); ***Ant.***, antonym; e.g., example (particular case).

[a] More precisely: in the course of one menstrual cycle, which forms the indivisible unit of time (because the risk in fact varies from day to day in the cycle).

[b] The period can be defined between two dates, two birthdays, or two marriage durations. It is also possible to use the whole of the reproductive period before a particular age (e.g., *previously, or already, fertile*) or after this age (e.g., *subsequently, or permanently, infertile*).

specialist publications (Pressat, 1979; IUSSP, 1981), gives indications on their usage, and introduces some additional terms to maintain overall consistency.

Observed infertility, whether temporary or permanent, can be voluntary or involuntary. *Involuntary infertility*, assuming exposure to the risk of conception, can result from sterility of one or other member of the couple (inability to conceive or to carry a pregnancy to term); *voluntary infertility* can result from use of contraceptive methods or a recourse to induced abortion.

Infertility (retrospective) is said to be *primary* when there has been no live birth since the start of the reproductive period and *secondary* when there has been at least one birth.

It can be noted that among demographers the convention is that *fecundability* measures the capacity to conceive for *nonsterile* couples. This is the probability of conceiving in the course of a normal menstrual cycle, in relation to the frequency and distribution of coitus over the cycle. We shall return to this concept later, but it is relevant to note here that it can be applied to all conceptions or only to those that result in a live birth. A distinction is made between the following:

- *Effective* fecundability, which is limited to conceptions that terminate in a live birth;
- *Apparent* fecundability, which also includes any nonproductive pregnancies the woman has been aware of and that she can report in a medical interview or a survey;
- *Recognizable* fecundability, which includes all the conceptions that are detectable using existing techniques (in practice, a pregnancy has to have lasted at least two weeks from conception);
- *Total* (or physiologic) fecundability, which includes all successful fertilizations.

The Limits of the Reproductive Period

The limits of the reproductive period in females are clearly defined by puberty and menopause, which in physiologic terms correspond to the onset and the end of ovulation. This in turn prompts the question of how to identify these stages satisfactorily.

In the case of *puberty*, the appearance of the first menses is a fairly reliable indicator, even though the early menstrual cycles are frequently irregular and sometimes anovular. Table 30–2 gathers data from tables published by Wood (1994), many of them in fact taken from Eveleth and Tanner (1976) and Becker (1993). These make clear the diversity of past and present situations, with just a few significant results being selected in each case.

For European or European-origin populations and even for those of Latin America, the results indicate a median of 13 years. The age range broadens (at its upper end) for Asian and African populations: In rural regions the medians can rise to 15 and 16 years and even to 17 or 18 years for groups living in more specific ecosystems (highland regions, in particular). This amount of variation is not surprising because the relatively low ages at puberty observed in developed countries are known to be the product of a fairly recent trend. The median age in the Scandinavian countries

TABLE 30–2 Median Age at First Menses (Menarche)

Population	Date	Median age	Source
Europe			
Italy, Carrara	1968	12.6	Eveleth and Tanner, 1976
England, London	1966	13.0	Eveleth and Tanner, 1976
Hungary	1965	13.1	Eveleth and Tanner, 1976
Norway, Oslo	1970	13.2	Eveleth and Tanner, 1976
Romania (rural)	1963–1966	14.2	Eveleth and Tanner, 1976
Other Western populations			
United States (whites)	1960–1970	12.8	Eveleth and Tanner, 1976
Australia, Sydney	1970	13.0	Eveleth and Tanner, 1976
New Zealand	1970	13.0	Eveleth and Tanner, 1976
Canada, Montreal	1969–1970	13.1	Eveleth and Tanner, 1976
Latin America			
Chile, Santiago (middle ESC)	1971	12.3	Eveleth and Tanner,1976
Mexico, Xochimilco	1966	12.8	Eveleth and Tanner, 1976
Guatemala, Guatemala City	1965	13.3	Eveleth and Tanner, 1976
Asia			
Singapore (middle ESC)	1968	12.7	Eveleth and Tanner, 1976
India, Madras (urban)	1960	12.8	Eveleth and Tanner, 1976
India (urban)	1956–1965	13.7	Eveleth and Tanner, 1976
India (rural)	1960	14.4	Eveleth and Tanner, 1976
Tibet	1970	16.1	Pawson, 1976 (in: Wood)
Nepal (Sherpas)	~1970	18.1	Pawson, 1976 (in: Wood)
North Africa, Middle East			
Turkey, Istanbul (high ESC)	1965	12.4	Eveleth and Tanner, 1976
Iran (urban)	1963	13.3	Eveleth and Tanner, 1976
Tunisia, Tunis (low ESC)	1970	14.0	Eveleth and Tanner, 1976
Iraq, Baghdad (low ESC)	1969	14.0	Eveleth and Tanner, 1976
Sub-Saharan Africa			
Uganda, Baganda (high ESC)	1959–1962	13.4	Eveleth and Tanner, 1976
Ethiopia (rural)	~1970	14.7	Pawson, 1976
Egypt (Nubians)	1966	15.2	Eveleth and Tanner, 1976
Rwanda (Hutus)	1957–1958	17.0	Eveleth and Tanner, 1976

ESC, economic and social category.

was approximately 17 years in the mid 19th century compared with 13 years at present (Eveleth and Tanner, 1976). The general change in living conditions, and in particular the sustained improvement in nutrition, is responsible for this rapid decline, which has coincided with the demographic transition and industrialization.

The data concerning *menopause* are less reliable for at least three reasons. First, the process is more continuous—cessation of menses is often gradual, and a minimum duration of amenorrhea must be defined to conclude that menopause has occurred (usually 1 year, although if spotting is taken into account the estimation can become difficult). Second, in the countries not operating an official registration system, adults approaching age 50 years do not always know their exact age. Third, in the industrialized countries the menopause is increasingly influenced by hormone replacement therapies and surgery to the reproductive organs, with the result that for many women, going through menopause is no longer a natural process. Even more than with puberty, therefore, caution is in order when interpreting the data in Table 30–3.

The median age of menopause appears to lie between 50 and 51 years in industrialized countries. Elsewhere the range is wider, with some values close to those of the first group (49–50 years) whereas others are significantly lower, falling to 43 and 44 years according to some sources, notably in Asia; values of around 46 years seem to be confirmed by the Demographic and Health Surveys (DHSs) recently conducted in the various states of India (International Institute for Population Science, 1995). This time there is no evidence of any large change in age at menopause over time, as could have caused the values for the populations of the developed countries to rise from below 45 years to over 50 years. In view of the difficulties described above in this section, a *cautious* conclusion is that the median ages observed in some developing countries may well be appreciably lower (by several years) than those in the most developed countries, perhaps, again, related to nutritional conditions. Age at natural menopause seems not to vary significantly in the latter countries, the only known differentiating factor being smoking, which can reduce age at spontaneous menopause by 1 or 2 years (Willet *et al.*, 1983).

From the data available for the industrialized countries, we can estimate that 10% to 15% of women (not receiving treatment for perimenopausal conditions) experience menopause before 45 years, around 50% before 50 years, and 90% to 95% before 55 years; 75% to 80% reach the menopause between 45 and 55 years. Figure 30–1 summarizes this progression with age.

Turning now to men, the physical signs of puberty are less directly linked to onset of reproductive capacity (start of spermatogenesis), making this stage much harder to observe. There are, however, grounds for

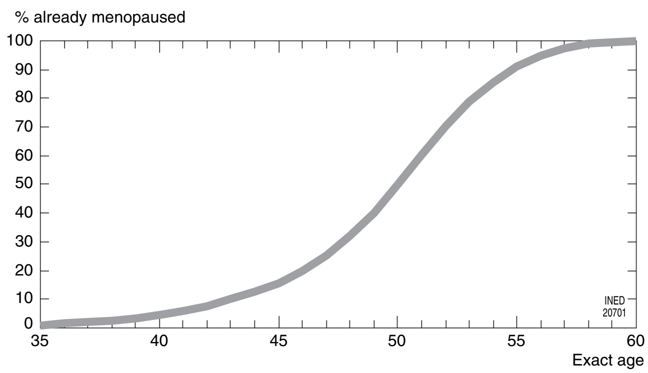


FIGURE 30–1 Estimated percentage of women who have reached (natural) menopause, by age (industrialized countries). Source: Adapted from Stanford *et al.*, 1987.

TABLE 30–3 Median Age at Menopause

Population	Date	Median age	Source
Europe			
United States (whites)	1981–1982	51.4	Wood, 1994
England	1965	50.8	Wood, 1994
New Zealand	1967	50.7	Wood, 1994
France, Paris (high ESC)	1957–1962	49.8[a]	Bourlière (in: Leridon, 1977)
France, Plozévet (rural)	1963	49.0[a]	Bourlière (in: Leridon, 1977)
Asia			
Nepal (Tibetans)	1981	46.8	Wood, 1994
India, Punjab	1966	44.0	Wood, 1994
Bangladesh	1975–1978	43.6	Wood, 1994
Africa			
Botswana (Dobe Kung)	1963–1973	49.1	Wood, 1994
South Africa (blacks)	1960	48.1	Wood, 1994
Central Africa	~1970	43.0	Wood, 1994

ESC, economic and social category.
[a] Mean age.

thinking that it occurs at a similar age to female puberty (La Rochebrochard, 1999). The main difference concerns the cessation of reproductive capacity, which is much more gradual in men than in women—fecundity declines slowly and surely from 30 or 40 years but we do not know at what age it gets close to zero. This upper age limit is clearly much higher than 50 years, but the small number of observations at high ages (of men still seeking to fertilize nonsterile women), related notably to loss of sex-drive, makes it hard to describe accurately the end of the period of male fecundity (La Rochebrochard, 2000). Analysis thus has to be based on the length of the *female* reproductive period, while bearing in mind that at all ages a couple's fecundity depends on both the man and the woman.

II. FROM ONE BIRTH TO THE NEXT: THE COMPONENTS OF BIRTH INTERVALS

Between two successive births, a woman passes through a number of clearly distinct physiologic stages. Ovulation does not resume immediately at the end of a pregnancy. The delay is approximately 2 months without breast-feeding and considerably longer in the presence of breast-feeding, especially when this is the infant's main source of food. It will be seen later that the duration of postpartum amenorrhea is as follows (approximate estimates):

4 months for a woman who breast-feeds for 6 months
6 months if the woman breast-feeds for 12 months
10 to 12 months if the woman breast-feeds for 18 months
14 to 16 months if the woman breast-feeds for 24 months

After resumption of ovulation, the woman can again conceive although the risk of doing so is far from 100% each month. This probability, termed *fecundability*, varies between couples but is on average situated between 20% and 30% in the high fertility ages, for an ordinary monthly frequency and distribution of coitus. This gives a mean waiting time to conception of approximately 5 to 8 months, in addition to the duration of amenorrhea.

Two live births can be separated by one or several conceptions that lead to miscarriages, as a result of *intrauterine mortality*, thus lengthening somewhat the apparent time to conception.

Finally, the interval ends with the 9 months of *gestation*, the average duration for a pregnancy terminating in a live birth (Figure 30–2).

In the context of premodern Europe, for example, the typical duration of an interval between live births would have ranged from 10 + 5 + 9 = 24 months to 16 + 8 + 9 = 33 months, depending on the assumptions made about fecundability and breast-feeding, findings that are fully consistent with the available observations on interval durations in this type of traditional society. It can also be noted here that stopping breast-feeding completely has the immediate effect of shortening considerably the interval between two births, unless it is compensated for by contraceptive (or abortive) practices: without breast-feeding, the mean completed family size from a marriage made at age 20 and still intact at age 50 would be over 12 children!

The diagram in Figure 30–2 summarizes the process, which is also illustrated by Figure 30–3. On the latter have been plotted the mean interval between two births in a population of rural Senegal (Sine Saloum) around 1965, by age at death or age at weaning of the infant born at the start of the interval: weaning or, *a fortiori*, death have the consequence of marking the end of the anovulatory period associated with breast-feeding (we will see in subsequent sections of this chapter that it ends, in fact, significantly earlier). When the infant dies or is weaned (because of inadequate maternal milk supply) in the first month, the

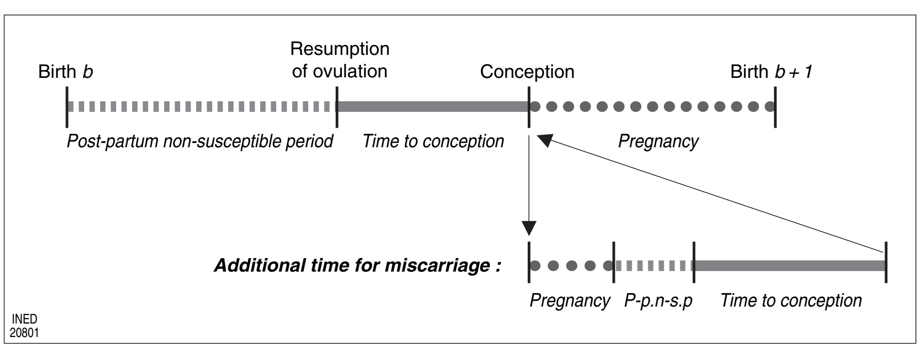


FIGURE 30–2 Decomposition of interval between two live births.

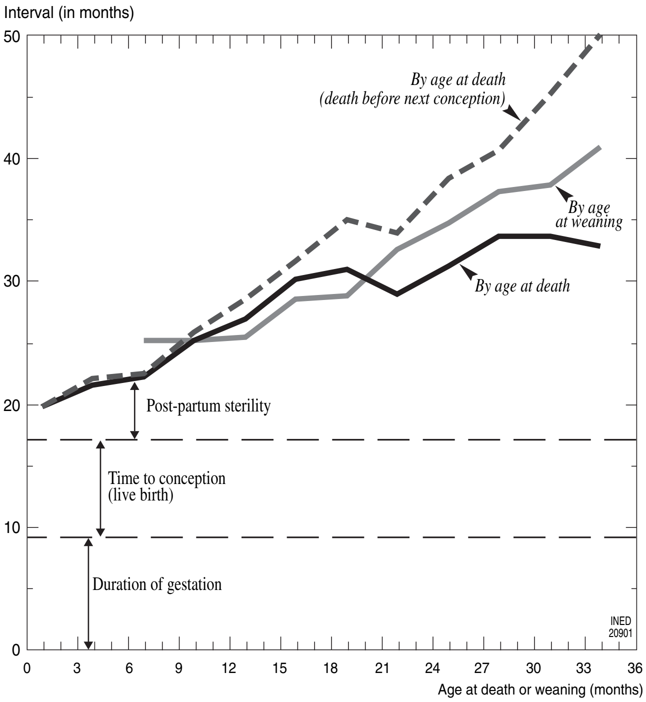


FIGURE 30–3 Mean birth intervals, by age at weaning or age at death of the first of two children (Senegal). Source: CANTRELLE PIERRE and LERIDON HENRI, 1971. Breast-feeding, child mortality and fertility in a rural zone of Senegal, *Population Studies*, vol. 25(3), p. 505–533.

interval until the next birth is around 19 months, representing 9 months' gestation, 8 months time to conception (including the time lost due to a possible miscarriage), and 2 months of postpartum sterility. There is no reason for the first two components to change with age at weaning, so it is the duration of breast-feeding (which in this population can last up to 3 years) that explains the lengthening of the interval, up to 40 months for a weaning that takes place at 3 years. The effect is slightly smaller when we consider age at death, if this takes place after 2 years: In this case, weaning may have preceded death by several months.

III. THE MENSTRUAL CYCLE AS A UNIT FOR MEASUREMENT OF TIME

Research by Ogino and Knaus in 1925 to 1930 established that fertilization is possible only during a short period in the middle of the menstrual cycle, around the point of ovulation. The precise length of this fecund period cannot be known because of the difficulty of determining the exact time of ovulation. Empirical studies such as those by Barrett and Marshall (1969) have shown that the probability of conception was significantly different from zero for 6 days, mostly before the date of ovulation *J* and with a maximum at day *J*-2. But this fecund period can be lengthened artificially by uncertainties in identification of the date of ovulation, which is detected here by variations in body temperature. However, more recent research based on biologic tests for the detection of ovulation confirms that the fecund period lasts 6 to 7 days (Wilcox *et al.*, 1998).

This heterogeneity of fecundity within the cycle means that this cycle must be taken as the minimum unit of time for the study of fertility, and therefore we need to know its duration. This seems to be between 26 and 30 days (Matsumoto, 1964) but the existence of appreciably longer cycles increases the mean duration to approximately 30 days. Consequently for the purpose of fertility studies, it is legitimate to equate menstrual cycle with calendar month and thus count 12 cycles per year. The duration of the cycles varies between women (ranging from 10 to 45 days) as well as between cycles and by age for the same woman. It also has to be noted that use of hormonal contraception establishes a strict 28-day cycle: In this case it is essential to count 13 cycles in the year.

A point to remember is that not all cycles are ovulatory. Anovulation is common at the beginning of the reproductive period (and may be responsible for the phenomenon of *adolescent sterility* described by Montagu in 1946), and again in the second half of this period; we will return to this point in our discussion of age-related change in sterility. The frequency of anovular cycles in women aged 20 to 35 years varies between 5% and 15% depending on the author.

IV. FECUNDABILITY

In practice, the concept of fecundability is only meaningful in statistical terms. For any given couple, the time required to conceive is measured by the number of cycles during which they have had unprotected intercourse. A satisfactory estimation of individual fecundability would require the test to be repeated many times, yet each couple wants to have only a small number of children. It is by combining the experiences of a certain number of couples that we are able to calculate the (monthly) rates of conception, at least for the conceptions that result in recognizable (and recognized) pregnancies. However, to do this we have to be certain that all the couples are indeed exposed to the risk of conception at the start of the period, a situation that is actually not very common. In traditional, noncontracepting populations, at any given time a large proportion of women will be in the nonsusceptible period (i.e., unable to conceive) because of a pregnancy already under way or the subsequent anovular period (that breast-feeding practices may extend to 1 year or more, as we have seen). In this case, the only favorable circumstance for the risk of conception is marriage, if it is accepted that this coincides approximately with the start of regular sexual relations for the couple.

That last assumption is unlikely to hold true for the low-fertility populations of the developed countries. Here, specific observations have to be made to determine the dates at which contraception was discontinued and the length of time that was then necessary for conception to occur.

The two types of observations lead to the following results:

a. Effective fecundability (rate of conception measured in live births) usually has an average value between 0.2 and 0.3 (20% to 30%);
b. The probability of conception decreases progressively, month on month, after the ninth month, revealing the heterogeneity of the cohorts being studied with respect to the risk of conception (Table 30–4).

A beta distribution function has proved extremely practical for representing the distribution of fecundability between couples. As well as enabling us to define a large variety of distributions with just two parameters, the beta law provides a simple yet effective tool for analyzing the progressive selection process of the less fecund couples. For if the parameters of the initial distribution are a and b, the fecundability of the couples who have not conceived after n months of exposure to the risk is also distributed according to a beta distribution with parameters a and $b + n$: Thus we immediately know all the characteristics of this new distribution and, in particular, the mean fecundability, which is equal to $a/(a + b + n)$.

So, for example, if the mean fecundability of the group is 0.25, with $a = 2.75$ and $b = 8.25$, after 1 year without conception (12 months) the fecundability of the residual group is only $0.12 = a/(a + b + 12)$ (Figure 30–4); but 71% of them will manage to conceive in the following 12 months. Even after 2 years without conception, when the residual fecundability has fallen to 0.08, the probability of conceiving in the next year exceeds 57%.

This approach can be used to clarify certain clinical practice problems. The first concerns the detection of so-called sterile couples, a category that is generally taken to include all couples who have still not conceived after trying for 12 months. This definition has the effect of grouping together—unfortunately in unknown proportions—couples who are pathologically sterile and couples who are merely subfecund, although for the latter, the probability of conception in the absence of any treatment is far from nil.

This clearly needs to be taken into account when estimating the effectiveness of an infertility treatment. If treatment is given uniquely to couples who are 100% sterile, any conception that results can be attributed to the treatment; if not, we must try to ascertain whether

TABLE 30–4 Distribution of 1000 First (Legitimate) Births, by Marriage Duration

Completed months of marriage	Number of first births (p. 1000)	Monthly fertility probability (p. 1000)
0	15	15
1	25	25
2	34	35
3	34	37
4	41	46
5	45	53
6	46	57
7	32	42
8	68	93
9	146	220
10	112	217
11	77	192
12	59	181
13	44	166
14	34	153
15	26	136
16+	162	—
Total	1000	1000

Source: Paul Vincent, 1961. *Recherches sur la fécondité biologique*. Paris, INED, PUF, 272 p. (Travaux et Documents, *Cahier* no. 37), Tables 21 and 27.

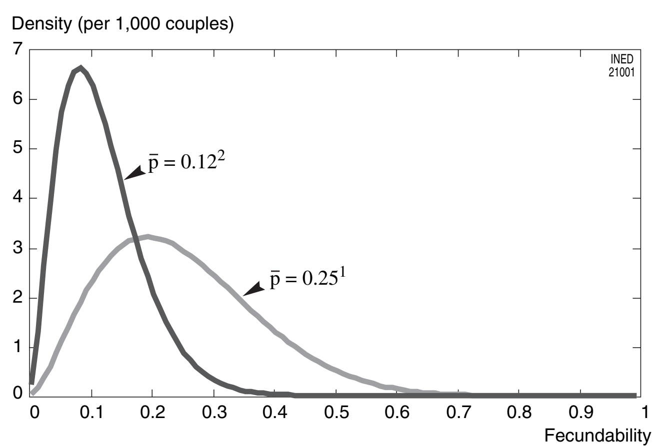


FIGURE 30–4 Distribution of fecundability by a beta function ($a = 2.75$ and $b = 8.25$).

[1]For all couples: mean fecundability $\bar{p} = 0.25$

[2]For hypofertile couples (no conception during the first 12 months): $\bar{p} = 0.12$

the result is better than would have been obtained without any treatment and this comparison can produce unexpected results.

In addition, the new techniques of *assisted reproduction* require us to rethink some aspects of the problem. For example, IVF (*in vitro* fertilization) specialists have adopted the concept of a *natural fecundability* in the order of 0.25, and on this basis conclude that when their *success rate from oocyte transfer* reaches 25% they are doing as well as nature. However, this reasoning neglects the fact that transfer involves the following stages:

- Hormonal stimulation of ovulation, intended to ensure that conditions for ovulation are normal, or even better than normal, because several eggs are usually collected;
- Attempted fertilization at the optimal point in the cycle and of several eggs at once;
- Transfer of several fertilized eggs to increase the chances of at least one developing.

These conditions are far removed from the average situation of ordinary couples, who experience merely one ovulation per cycle and who do not necessarily have intercourse at the *best* moment in that cycle. Furthermore, Barrett and Marshall (1969) and Schwartz, MacDonald and Heuchel (1980) determined that if intercourse takes place once a day throughout the fecund period, the chance of obtaining a *recognizable conception* can be as high as (or exceed) 50% per cycle. This is the type of figure that should be used when comparing the success rate per treatment cycle of AID (artificial insemination by donor) or of IVF, with allowance made for the pregnancies that do not run to term.

V. DURATION OF PREGNANCY

For live births, the duration of pregnancy traditionally seemed to vary little (on average) between different populations. A precise distribution of gestation durations was produced by Hammes and Treolar (1970) for almost 200,000 single births registered in California in 1966. Table 30–5 shows this distribution after conversion of the *conventional durations* (counted from the first day of the last menstrual period) to *actual durations* (counted from the day of ovulation) by the simple procedure of subtracting 2 weeks. The right-hand section of the table gives the distribution in calendar months (of 30.4 days), which is useful for some applications.

The mean duration of gestation is 265 days (38 weeks) when counted from the probable day of ovulation (or 279 days from the start of the last menstrual cycle). The median is also close to 38 weeks, whereas

TABLE 30–5 Distribution of Pregnancy Durations (Live Births) (Duration Calculated from Theoretical Day of Ovulation)

Completed weeks x	%	Cumulative % from 0 to x (inclusive)	Completed days	Completed months: y	%	Cumulative % from 0 to y (inclusive)
<26	0.5	0.5	<182	<6 (<182 *d*)	0.5	0.5
26–28	0.4	0.9	182–202	6 (182–212 *d*)	1.0	—
29	0.4	1.3	203–209			1.5
30	0.4	1.7	210–216			—
31	0.6	2.3	217–223	7 (213–242 *d*)	6.0	7.5
32	1.2	3.5	224–230			—
33	1.9	5.4	231–237			—
34	3.0	8.4	238–244	8		—
35	5.8	14.2	245–251			—
36	12.8	27.0	252–258	8 (243–273 *d*)	66.0	73.5
37	21.6	48.6	259–265			—
38	23.2	71.8	266–272			—
39	14.7	86.5	273–279			—
40	7.1	93.6	280–286	9 (274–303 *d*)	25.5	99.0
41	4.1	97.7	287–293			—
42	0.5	98.2	294–300			—
43	0.8	99.0	301–307			—
44	0.1	99.1	308–314	10 (304 *d* and +)	1.0	100.0
>44	0.9	100.0	>314			—

Source: Laurel M. Hammes and Alan E. Treolar, 1970. Gestational interval from vital records, *American Journal of Public Health*, vol. 60(8), p. 1496–1505.

the mode is between 38 and 39, thus giving proof of skewness in the distribution.

This distribution has certainly been modified by improved medical care for premature babies in the countries with the most advanced health facilities. Because the scale of changes is closely dependent on the introduction and effectiveness of these new practices, we prefer to retain the Hammes and Treolar (1970) distribution as reference standard.

For pregnancies that do not end in live births it is hard to give a standard distribution. Because declaration of these pregnancies is not usually compulsory, we have to rely on information obtained in surveys. The result is then dependent on the proportion of early miscarriages that women report (see Chapter 32 for a comprehensive discussion of intrauterine mortality). For practical purposes, a mean of approximately 3 months can be used.

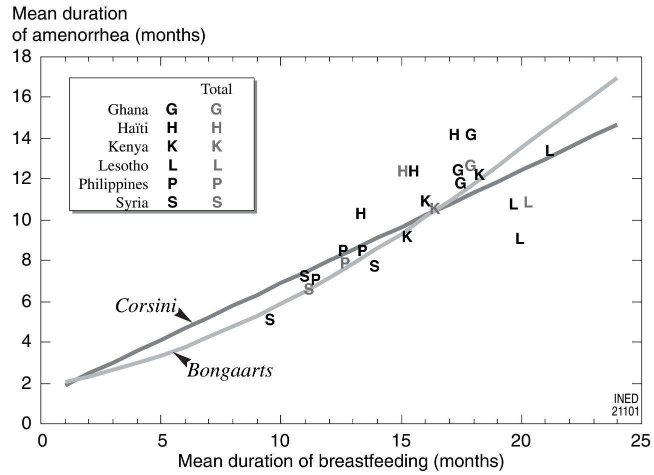


FIGURE 30–5 Relationship between mean durations of breast-feeding and amenorrhea (World Fertility Surveys and standard distributions). Source: Leridon and Ferry, 1985.

VI. THE POSTPARTUM NONSUSCEPTIBLE PERIOD

Birth is followed by a period in which the risk of conception is zero, a situation that in traditional demographic regimes contributes to a significant spacing of births, even when this outcome is not consciously sought. Although breast-feeding, by temporarily inhibiting ovulation, has long been known to be the main factor responsible for this phenomenon, the debate has suffered from two confusions. First, the relationship between the duration of breast-feeding and the duration of postpartum sterility is purely *statistical*—when working at the individual level, the second duration cannot be predicted from the first, and in most cases ovulation resumes before the infant has been weaned. Second, anthropologic data have established the existence in some populations of taboos on sexual intercourse for a period following birth. Only at the end of the 1970s did analysis of data from the World Fertility Survey (WFS) program make it possible to demonstrate the usually secondary role of sexual prohibitions compared with breast-feeding. We return to this point below.

The effect of breast-feeding has been proved in several ways. The indirect data provided by studies of historic demography (which have related the length of the interval between two births to the age at death of the first child, as represented by one of the curves on Figure 30–3), have been confirmed by data from surveys containing direct questions on the duration of breast-feeding. Furthermore, the workings of the biologic mechanism involved are now understood. Secretion of a specific hormone, prolactin, which begins during, and continues after, pregnancy, is responsible for maternal milk production and (perhaps not exclusively) for suppressing release of other hormones (luteinizing hormone [LH] and follicle-stimulating hormone [FSH]), thereby preventing the resumption of ovulation.

The most obvious sign of the resumption of ovulation is the return of menstruation (the end of amenorrhea, or the *return of menses*). In point of fact, this is merely an approximate indicator, since the cycle preceding the first menses is sometimes though not always ovulatory. On average, however, the approximation is adequate. Figure 30–5 shows the possible relationships between the mean duration of breast-feeding and postpartum amenorrhea: These data provided the basis for the plausible estimates given earlier (6 months of amenorrhea for 1 year of breast-feeding, 14–16 months for 2 years). The larger data output from the DHS surveys will enable further refinement of this relationship. We are concerned here with *total* durations of breast-feeding: It is important to appreciate that during much (indeed usually most) of this period, the infant is not nourished exclusively on mother's milk. But the effect of breast-feeding on postpartum sterility is diminished when feeds are less frequent and of shorter duration. This is the reason, at least in part, for the fairly weak correlation between total duration of breast-feeding and duration of amenorrhea.

The duration of postpartum sexual *abstinence*, as noted, seldom exceeds the duration of amenorrhea and thus *a fortiori* the duration of breast-feeding. But the impact on the effective duration of sterility is not necessarily negligible, since we cannot here reason on mean figures—each individual case where the first

TABLE 30–6 Mean Durations of Breast-feeding, Amenorrhea, Sexual Abstinence and Postpartum Sterility (Months), in Selected Countries (World Fertility Survey)

Country	Breast-feeding	Amenorrhea	Abstinence	Nonsusceptible period
Ghana	17.9	12.5	10.2	14.2
Kenya	16.3	10.6	3.2	11.3
Haiti	15.2	12.3	6.5	14.0
Lesotho	20.2	10.8	16.3	17.7

Source: Leridon and Ferry, 1985.

duration exceeds the second contributes a little to increasing the overall mean duration of sterility. Table 30–6 shows this clearly for various countries, based on the WFS. In this case too, the more plentiful data from the DHS will allow us to confirm this analysis.

Acknowledgment

Translated from French by Godfrey I. Rogers.

References

Barrett John C. and Marshall John, 1969. The risk of conception on different days of the menstrual cycle, *Population Studies*, vol. 23(3), p. 455–461.

Becker Stan, 1993. The determinants of adolescent fertility with special reference to biological variables, *in:* Ronald Gray, Henri Leridon, and Alfred Spira (eds.), *Biomedical and demographic determinants of reproduction*, p. 21–49. Oxford, Clarendon Press, 482 p. (International Study in Demography)

Bourlière François, Cendron M., and Clément F., 1966. Le vieillissement individuel dans une population rurale française: étude de la commune de Plozévet, *Bulletins et Mémoires de la Société d'anthropologie de Paris*, *série 11*, no. 10, p. 41–101.

Bourlière François, Clément F. and Parot S, 1966. Normes de vieillissement morphologique et physiologique dans une population de niveau sociologique élevéde la région parisienne, *Bulletins et Mémoires de la Société d'anthropologie de Paris*, *série 11*, no. 10, p. 11–39.

Cantrelle Pierre and Leridon Henri, 1971. Breast-feeding, child mortality and fertility in a rural zone of Senegal, *Population Studies*, vol. 25(3), p. 505–533.

Eveleth Phyllis B. and Tanner James M., 1976. *Worldwide variation in human growth*. Cambridge, Cambridge University Press, 497 p.

Gray Ronald, 1976. The menopause: epidemiological and demographic considerations, *in:* Robert J. Beard (ed.), *The menopause, a guide to current research and practice*, p. 25–40. Lancaster, MTP Press, 269 p.

Gray Ronald, Leridon Henri, and Spira Alfred (eds.), 1993. *Biomedical and demographic determinants of reproduction*. Oxford, Clarendon Press, 482 p.

Hammes Laurel M. and Treolar Alan E., 1970. Gestational interval from vital records, *American Journal of Public Health*, vol. 60(8), p. 1496–1505.

IIPS, 1995. – *National family health survey, India 1992–1993*. – Bombay, IIPS (International Institute for Population Sciences), 402 p.

IUSSP, 1981. *Multilingual demographic dictionary* (2nd ed). Liège, Ordina, 180 p.

La Rochebrochard Elise (de), 1999. Le sâges de la puberté des filles et des garçons en France. Mesures à partir d'une enquête sur la sexualité des adolescents, *Population*, vol. 54(6), p. 933–962 (in English as: Age at puberty of girls and boys in France: measurements from a survey on adolescent sexuality, *Population: An English Selection*, vol. 12, 2000, p. 51–80.

La Rochebrochard Elise (de), 2000.—*Modélisation de la baisse de la fertilité avec l'âge de la femme et l'âge de l'homme*. Paris, Université Pierre et Marie Curie, 423 p. (doctoral thesis).

Leridon Henri and Ferry Benoît, 1985. Biological and traditional restraints on fertility, *in:* John Cleland and John Hobcraft (eds.), *Reproductive change in developing countries. Insights from the World fertility survey*, p. 139–164. Oxford, Oxford University Press, XVI+ 301 p.

Leridon Henri, 1977. *Human fertility: the basic components*. Chicago, The University of Chicago Press, 202 p.

Matsumoto Seiichi, 1964. Statistical survey of the length of the menstrual cycle and the duration of menstruation in 3000 ovulatory cycles based upon BBT findings, *in:* Changing patterns in fertility. *Proceedings of the 7th Conference of the International Planned Parenthood Federation, February 1963 (Singapore)*, p. 202–225. London, Exerpta Medica, 748 p. (International Congress Series no.72).

Montagu Ashley, 1946. *Adolescent sterility: a study in the comparative physiology of the infecondity of the adolecent organism in mammal and man*. Springfield, Charles C. Thomas, 148 p.

Pawson Ivan G., 1976. Growth and development in high altitude populations: a review of Ethiopian, Peruvian and Nepalese studies, *Proceedings of the Royal Society of London*, *series B*, no. 194, p. 83–98.

Pressat Roland, 1979. *Dictionnaire de démographie*. Paris, PUF, 295 p.

Schwartz Daniel, Mac Donald Peter, and Heuchel V., 1980. Fecundability, coital frequency and the viability of ova, *Population Studies*, vol. 34(2), p. 397–400.

Stanford Janet L., Hartge Patricia, Brinton Louise A., Hoover Robert N., and Brookmeyer Ronald, 1987. Factors influencing the age at natural menopause, *Journal of Chronicle Diseases*, vol. 40(11), p. 995–1002.

Vincent Paul, 1961. *Recherches sur la fécondité biologique*. Paris, INED, PUF, 272 p. (Travaux et Documents, *Cahier* no. 37).

Wilcox Allen J., Weinberg Clarice R., and Baird Donna Day, 1988. Post-ovulatory ageing of the human ovocyte and embryo, *Human Reproduction*, vol. 13(2), p. 394–397.

Willett Walter C., Stamper Meir J., Bain Christopher, Lipnick R., Speizer Frank E., Rosner Bernard, Cramer D. W., and Hennekens Charles H. 1983. Cigarette smoking, relative weight, and menopause, *American Journal of Epidemiology*, vol. 117(6), p. 652–658.

Wood James W., 1994. *Dynamics of human reproduction: biology, biometry, demography*. New York, Aldine de Gruyter, XVI +653 p.

C H A P T E R

31

The Factors of Fecundability and the Nonsusceptible Period

HENRI LERIDON

Institut national d'études démographiques (INED), Paris, France

The various biologic parameters of reproduction were presented in the last chapter, along with some components of their spatial and temporal variability. In this chapter we examine the possible causes of these variations, including those resulting from sexual behavior and contraceptive practices.

I. AGES AT MENARCHE AND AT MENOPAUSE

In most of the developed countries, the median age *at menarche* (for girls) is between 13 and 14 years. Although the data published by Tanner in 1962 describing the rapid decline in age at first menstruation during the 20th century have been criticized (Tanner inferred a reduction of 4 months per decade), the general trend is not disputed. Most available data point to a limit being reached around 12.5 to 13 years, although such a limit is not suggested by other data, notably for France (Ducros and Pasquet, 1978; de La Rochebrochard, 1999).

This secular decline has often been attributed to improved nutrition among infants and adolescents. The hypothesis, which cannot be verified directly, appeared to be supported by the idea that the onset of ovarian function is triggered when the proportion of body fat rises beyond a certain threshold in the course of a young girl's growth (Frisch and Revelle, 1970). Other elements, such as the cessation of menstruation in women whose body mass had fallen for some reason or another or in populations subjected to extremes of undernutrition or stress further strengthened this hypothesis. It has been largely discredited, however, by subsequent research. For example, growth of *skeletal mass* has been shown to be a much better predictor of puberty (Ellison, 1982), so much so that the body fat increase is now thought to be more a consequence of ovarian function onset than one of its causes (Wood, 1994). The possibility remains, however, that nutrition contributes to ovarian functioning in general and to onset of menarche in particular.

Differences between individuals remain, of course. The French data analyzed by de La Rochebrochard (1999) show an age distribution at first menses ranging from 9 to 18 years, although the ages between 11 and 14 years account for around 90% of cases.

The distribution of age *at menopause* is considerably more dispersed. The four most frequent years together account for only 40% to 50% of cases. On either side of the modal value, which is usually situated around 50 years, the distribution extends up to approximately 55 years in one direction and down to 35 to 40 years in the other (although cases of earlier menopause do occur) (see Chapter 33 and Treolar, 1974). The reasons for these interindividual variations are poorly understood. The only clearly documented factor is smoking, which is thought to bring forward age at menopause by between 1 and 2 years. For any individual woman, no correlation has been found between age at menarche and age at menopause (Treolar, 1974). Genetic factors may be implicated in the differences between individuals and in the mean values for different populations.

The predominant influence affecting age at menopause, however, is the development of treatments for perimenopausal conditions, which can substantially delay the onset of menopause, so much so that it becomes almost unobservable. However, because these are hormonal treatments that *also* have a contraceptive role, it does not follow that they extend the woman's fertile period, the risks of conception being extremely limited or even nonexistent.

II. FECUNDABILITY

We saw in the previous chapter that fecundability varies between couples and that a distribution of fecundability can be constructed from the distribution of the waiting times to conception. Figure 30–4 in the previous chapter showed that for a mean fecundability of 0.25, the individual values could range between 0 and approximately 0.6. A fecundability of 0.5 gives a mean time to conception of 2 months, whereas for a fecundability of 0.1 it reaches 10 months. Couples are thus very unequal in regard to the risk of conception—what are the main factors responsible for these differences?

For a particular couple, fecundability represents a combination of dimensions that are *biologic* (the respective physiologic capabilities of the man and the woman)[1] and *behavioral* (the frequency of sexual intercourse and its distribution within the menstrual cycle). Most of these results were obtained from populations in which contraceptive practice can be assumed to be negligible, especially at the start of marriage. For newly married couples, we can also assume a sexual activity of at least average level, in the absence of any direct information about the sexual behavior of these populations. In the next section we give some results from contemporary surveys on the sexual activity of couples.

Mathematic models have been constructed that describe the evolution of fecundability with the *woman's age*, by neutralizing the effect of coital frequency. In practice the assumption is made that intercourse takes place within each cycle at least once during the fertile period (i.e., at the point of ovulation). Under these hypotheses and for premenopausal women, Wood and Weinstein (1988) estimate total fecundability (the probability of conceiving regardless

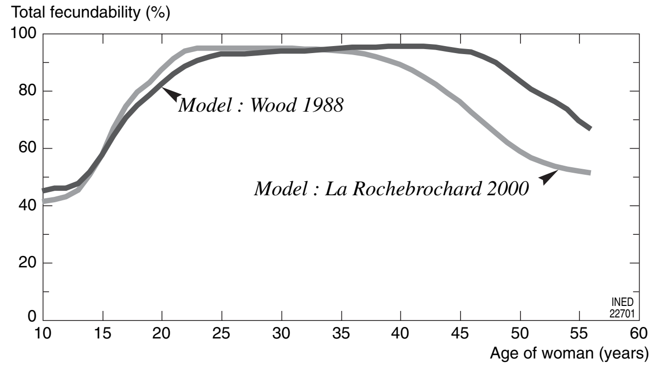


FIGURE 31–1 Age-specific total fecundability (women), per cycle of exposure to risk of conception (woman and man not yet sterile, cycle with sexual intercourse at point of ovulation) Source: La Rochebrochard Élise de, 2000. *Modélisation de la baisse de la fertilité avec l'âge de la femme et l'âge de l'homme* Paris, Université Pierre et Marie Curie, 438 p. (biomathematics thesis); Figure 9.4.

of the outcome of the conception) at close to 0.95 between 25 and 45 years; de La Rochebrochard (2000), who takes into account the reduced risk of ovulation in the years leading up to menopause, finds that fecundability actually starts to decrease between 35 and 40 years (Figure 31–1).

Turning from *total* fecundability to *recognizable* fecundability, we note that the latter is calculated by including intrauterine mortality occurring before a pregnancy is detectable by delayed menstrual period or a pregnancy test (i.e., in the first 2 weeks of gestation). Mortality in these first 2 weeks is particularly hard to observe and often it is necessary to use indirect methods. Working again on the assumption that sexual intercourse occurred at least once during the fertile period, the estimates proposed for recognizable fecundability were in the order of 0.3 to 0.4 (Leridon, 1977). More recently, de La Rochebrochard (2000) has suggested a value not exceeding 0.24 (between 25 and 35 years), which in terms of live births would give an *effective* fecundability not exceeding 0.21 at the most propitious ages (Figure 31–2): This is a low value compared with those obtained by direct demographic observations, which are approximately 0.25 (Leridon, 1977).

If age is a determinant of fecundability, this is because certain biologic parameters vary with age—the frequency of anovular cycles, the mean length and variance of menstrual cycles, the fertilizing quality of sperm, the risk of genetic defects in the embryo, and so forth. These same factors could also explain the differences between individuals. Any exceptionally long waiting time to conception should alert to the possibility of ovulatory difficulties, spermatic abnormalities, or problems in coitus. These causes are analyzed

[1] Forms of incompatibility may also exist between the aptitudes of the two partners that result in the sterility of a couple in which the man and the woman are nonetheless fertile individually.

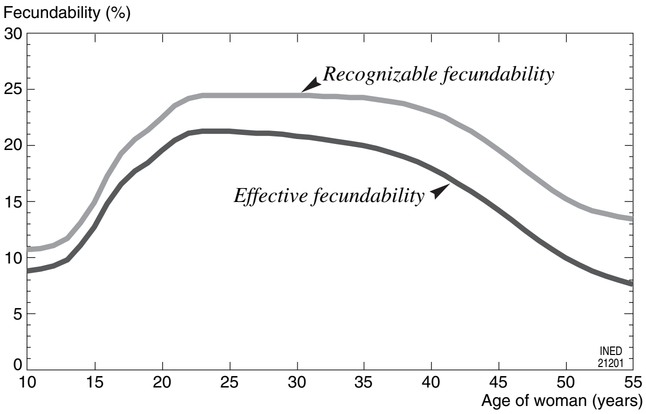


FIGURE 31–2 Recognizable fecundability and effective fecundability, by age of woman, per cycle of exposure to risk of conception (woman and man not yet sterile, cycle with sexual intercourse at point of ovulation). Source: La Rochebrochard Élise de, 2000. *Modélisation de la baisse de la fertilité avec l'âge de la femme et l'âge de l'homme* Paris, Université Pierre et Marie Curie, 438 p. (biomathematics thesis); Figure 9.14.

in the Chapter 33, since in practice it is not possible to know beforehand whether a couple seeking help for infertility is totally sterile or merely subfecund.

III. FREQUENCY OF SEXUAL INTERCOURSE

A comparison of results from surveys carried out in nine European countries in the early 1990s showed little between-country variation in the reported mean frequency of intercourse: In all countries, the weekly mean (for men and women in couples, aged between 18 and 50–75 years depending on the survey) is between 1.7 and 2.1, equivalent to 6.8–8.4 acts of coitus per 4-week period[2] (Sandfort *et al.*, 1998). The contrasts are greater in the developing countries. According to World Health Organization (WHO) surveys conducted at the same time, the mean frequency reported by women in couples (15–49 years) for the previous month ranged between 1.5 and 5.7 in six African countries and between 2.5 and 7.6 in five countries of Asia and Latin America (Caraël, 1995). Not only are the between-country disparities large but the mean values are significantly lower than those for the developed countries—an important reason is that a large proportion of the women interviewed (between 34% and 58% of the Africans) reported having had no intercourse with their husband in the previous month. This behavior can be attributed to numerous factors, including polygamy and widespread recourse to prostitutes, the existence of periods when intercourse is discouraged or proscribed (pregnancy, postpartum abstinence related to breast-feeding, terminal abstinence at the end of reproductive life), and economically motivated periods of separation.

Frequency of intercourse usually declines with age (for people living in couples), but the important variable in fact seems to be length of union, at least in the union's early years (Udry, 1979; Leridon, 1993). In the French Analyse des Comportements Sexuels en France (ACSF) survey (Leridon, 1973), people aged 25 to 34 years and living as a couple reported a mean frequency of intercourse over the past 4 weeks of 14.6 when the couple had existed for less than 6 months, 11.7 when this duration was between 6 months and 1 year, and 9.6 for durations between 10 and 14 years (Figure 31–3). The most significant decline thus occurs during the first 12 months. The age-specific data indicate that the age effect is apparent primarily after 50 years (for women): Thus, it has only a weak influence on the risk of conception during the reproductive period.

Naturally, allowance must also be made for reporting problems, which can distort comparisons between populations. Generally speaking, people questioned on this subject tend to give answers that are biased in the direction of normative considerations, corresponding to behavior in periods of full activity and overlooking the existence of periods when sexual activity has been reduced for various reasons (health problems, pressure of work, etc.).

IV. BREAST-FEEDING

As was seen in the previous chapter, breast-feeding is important in lengthening birth intervals. Without any effective substitute for the infant's nourishment, breast milk is essential for its survival—the death of the mother or a new conception too soon would endanger the life of the infant. Multiple births are a definite additional handicap. A study in Senegal (Cantrelle and Leridon, 1971) reported mortality levels over the first 3 months four times higher for twins than for other children and more than twice as high in the first year (for a more complete analysis, see Pison, 1989).

Breast milk production is dependent on the peptide hormone prolactin, which also has a role in inhibiting ovulation after delivery. Prolactin levels and milk yield depend in turn on the suckling stimulus—it has been established that more frequent suckling delays the resumption of ovulation and is therefore an important factor in variations in the duration of postpartum

[2] In practice, the question was usually asked for the 4 previous weeks.

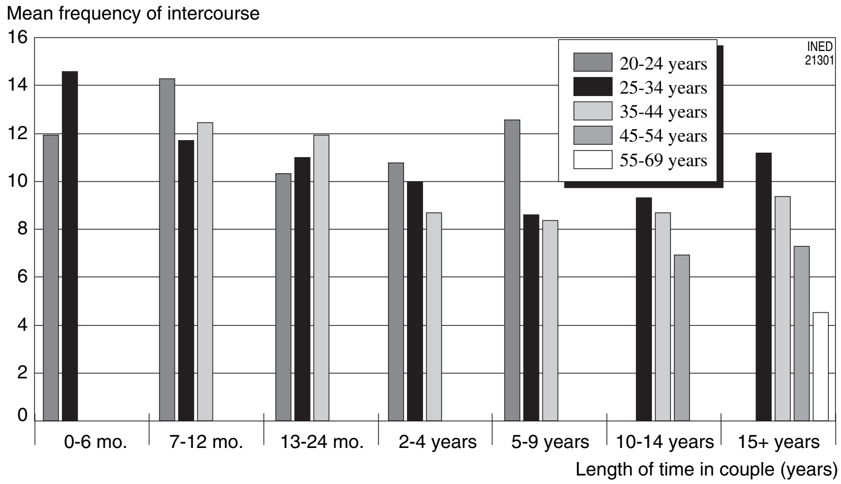


FIGURE 31–3 Mean frequency of intercourse over previous 4 weeks by age and length of time in couple: France, ACSF survey (men and women 1-year sexually active). Source: Leridon Henri, 1993. La fréquence des rapports sexuels. Données et analyses de cohérence, *in:* Michel Bozon and Henri Leridon (eds.), Sexualité et sciences sociales. Les apports d'une enquête, *Population*, vol. 48(5), p. 1381–1408.

TABLE 31–1 Breast-feeding Intensity and Duration of Postpartum Amenorrhea (Java)

Breast-feeding intensity	Minutes per suckling	Number of sucklings per 24 hours	Total duration per 24 hours	Median duration of amenorrhea, months	*Subjects*
1. Low	5.12	8.23	41.81	11.47	*41*
2. Medium low	9.15	7.31	66.33	14.01	*81*
3. Medium high	8.53	8.85	73.64	17.65	*199*
4. High	8.11	11.32	89.19	20.35	*61*

Source: Jones, 1988.

amenorrhea between women and between populations (Konner and Worthman, 1980). Table 31–1 presents an example of the observed relationship between suckling frequency and duration of postpartum amenorrhea (note that although the return of menses is not an absolute marker for the resumption of ovulation, the *mean* durations of amenorrhea and of anovulation are similar).

In this sample of rural women in Java, the total duration of suckling varies in a ratio of one to two between the four groups (formed for the purposes of the analysis), and this is associated with a difference of almost one to two in the median duration of amenorrhea. For this reason alone the birth interval increases by 9 months between the first and fourth groups. Thus it is clear that the correspondence between *mean duration of breast-feeding* and *mean duration of amenorrhea*, as described in the previous chapter, can only be approximate.

One reason for a low suckling frequency is that the infant has begun to receive supplementary food. Studies that take account of whether breast-feeding is partial or full have in fact shown the risks of resumption of ovulation and of conception to be significant even when breast-feeding is only partial, before the child has been fully weaned (see for example John *et al.*, 1987).

All these ingredients contribute to the existence of large differences in breast-feeding durations between populations, as can be seen from Table 31–2. Table 31–2 shows results for developing countries from two large-scale international programs—the World Fertility Survey (WFS), conducted for the most part between 1974 and 1981, and the Demographic and Health Survey (DHS), for which the data presented here concern the years from 1986 to 1997. Large disparities are observed, because the mean duration of breast-feeding varies from 5 months in Costa Rica to 29 months in Bangladesh. When data from both sources exist for a country, they show that breast-feeding practice has not decreased across the periods (the trend appears rather to be one of increase).

The subsequent columns of Table 31–2 confirm (as was seen in the previous chapter) that the mean durations of amenorrhea are always shorter than those of breast-feeding; they range from 5 months (Brazil) to

TABLE 31–2 Mean Durations of Breast-feeding, Amenorrhea, Postpartum Abstinence and Postpartum Sterility (World Fertility Survey and Demographic and Health Survey)

Country	Year	Mean duration				
		Breastfeeding (WFS)	Breastfeeding (DHS)	Amenorrhea (DHS)	Abstinence (DHS)	Nonsusceptible period (DHS)
Africa						
Burkina Faso	1993	—	25.1	16.5	19.0	22.1
Central African Republic	1995	—	21.2	15.2	12.7	17.5
Côte d'Ivoire	1994	—	20.4	13.6	14.8	17.9
Egypt	1992	—	18.4	8.2	3.0	8.7
Kenya	1977	15.7	—	—	—	—
Lesotho	1977	19.5	—	—	—	—
Madagascar	1992	—	20.2	13.2	7.0	14.5
Namibia	1992	—	16.9	10.9	10.7	15.3
Niger	1992	—	20.4	15.2	3.9	15.7
Rwanda	1992	—	26.4	16.8	2.7	17.1
Senegal	1978 and 1993	18.5	20.1	15.1	7.5	16.9
Sudan (north)	1978	15.9	—	—	—	—
Uganda	1995	—	19.5	13.4	5.5	14.5
Zambia	1992	—	18.3	12.7	8.3	14.7
Latin America						
Brazil (north-east)	—	—	9.7	4.9	3.7	6.1
Colombia	1991	9.2	13.9	6.5	5.7	9.0
Costa Rica	1976 and 1995	5.0	—	—	—	—
Dominican Republic	1976	8.6	—	—	—	—
Guatemala	1975	—	19.8	11.2	5.5	13.0
Guyana	1977	7.2	—			
Haiti	1977 and 1995	15.5	17.3	12.1	7.0	14.2
Jamaica	1975	8.1	—	—	—	—
Mexico	1976	9.0	—	—	—	—
Panama	1975	7.4	—	—	—	—
Paraguay	1979	11.4	—	—	—	—
Peru	1977	13.1	—	—	—	—
Trinidad and Tobago	1977	8.0	—	—	—	—
Venezuela	1977	7.4	—	—	—	
Asia						
Bangladesh	1975 and 1994	28.9	283	11.5	3.2	12.1
South Korea	1974	16.3	—	—	—	—
Fiji	1974	9.9	—	—	—	—
India	1993	—	261	9.5	5.4	11.1
Indonesia	1975 and 1994	23.6	239	10.4	4.7	11.3
Jordan	1976	11.1	—	—	—	—
Kazakhstan	1995	—	13.7	7.4	4.1	8.3
Malaysia	1974	5.8	—	—	—	—
Nepal	1976	25.2	—	—	—	—
Pakistan	1975	19.0	—	—	—	—
Philippines	1978	13.0	—	—	—	—
Sri Lanka	1975	21.0	—	—	—	—
Syria	1978	11.6	—	—	—	—
Thailand	1975	18.9	—	—	—	—
Turkey	1993	—	13.3	5.6	2.7	6.1
Yemen	1992	—	16.8	9.3	—	

WFS, World Fertility Survey; DHS, Demographic and Health Survey.

Sources: WFS: Benoît Ferry and David P. Smith, 1983. *Breastfeeding differentials* London, WFS, 89 p. (Comparative Studies, no. 23); DHS Country Reports (1986–1997). Calverton: Macro International Inc.

about 17 months (Rwanda). The effective duration of the *postpartum nonsusceptible period*, which refers to the period of zero fecundability following a birth, depends on the length of breast-feeding, but also, in some situations, on the duration of sexual abstinence after the birth of a child. When this duration is clearly shorter than that of the amenorrhea, it has little impact on the mean duration of postpartum sterility; in some cases, however, as in Burkina Faso and Namibia, it helps to extend it significantly.

V. CONTRACEPTION

The effective risk of conception at any point in time also depends, once we abandon the natural fertility framework, on couples' behavior in respect of contraception. The *effectiveness of contraception* is defined as the complement to 1 of the ratio between the *residual fecundability* p_r (with contraception) and the *natural fecundability* p_n:

$$E = 1 - \frac{p_r}{p_n} \qquad \text{(Eq. 1)}$$

The value of E is often expressed as a percentage. Thus, a method that is 99% effective reduces the risk p_n to 1% of its initial value.

1. Measurement of Contraceptive Failure Rate (Homogeneous Situation)

As defined above, contraceptive effectiveness measures the *reduction in risk* (of conception) induced by use of a contraceptive method; this measurement is thus independent of the amount of exposure to the risk without any contraception. The failure rate, by contrast, uses the number of *pregnancies that actually occur* in a given period—so it is in fact a fertility rate. It can also be calculated just for the periods of exposure to the risk of conception (the case of the Pearl index), thus modifying its nature considerably.

From the definition given above, it follows that:

1. The actual number of conceptions will depend both on E and on p_n. For example, if $p_n = 0.3$, a method with a 90% effectiveness will leave a risk of conception of 0.03 (or 3%) per month;
2. The notion of contraceptive effectiveness applies to a well-defined unit of time, that of fecundability—the month (or the menstrual cycle), and excludes any other;
3. Over a longer period, such as the year, the monthly risks are combined, so that an apparently high effectiveness (like 90%) will have left a nonnegligible number of conceptions by the end of 12 or 24 months (respectively, 31% and 48% in the selected example).

TABLE 31–3 Monthly Distribution of Conceptions (per 1000 Women). Homogeneous Case: Monthly Probability of 30% or Monthly Probability of 3% (Contraceptive Method 90% Effective)

Month	At 30% risk		At 3% risk	
	Not yet conceived	Conceiving in the month	Not yet conceived	Conceiving in the month
1	1000	300	1000	30
2	700	210	970	29
3	490	147	941	28
4	343	103	913	27
5	240	72	886	27
6	168	50	859	26
7	118	36	833	25
8	82	25	808	24
9	57	17	784	24
10	40	12	760	23
11	28	8	737	22
12	20	6	715	21
13	14		694	
Total		986		306

Thus, in our example, a method that is 90% effective gives 3% conceptions per month and 31% at the end of 1 year (Table 31–3). The complement to 100 of E (10%) has no particular significance in terms of failure rate and hence it is incorrect to assert, as unfortunately often happens, that when a method with a 90% efficacy is used by 100 women, 10 will become pregnant within a year.

Now let us examine in more detail what is involved here. For a group of N couples, observed during a certain number of months after adopting a particular contraceptive method, let

C_i be the number of conceptions during month i,
N_I be the number of women still in observation during month i,
S_i be the number of women at risk of conception during month i,

We note that N_i decreases progressively due to the women lost to follow up, and that S_i also decreases for the same reason but primarily as and when the women conceive and thus enter the nonsusceptible period.

The percentage of women conceiving during the n *first months* (if no woman has had the time to conceive more than once) will be

$$T_n = \frac{\text{Total number of conceptions}}{\text{Mean number of women under observation}} \times 100 = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} N_i} \times 100\, n \qquad \text{(Eq. 2)}$$

If all the women are still under observation after n months, Equation 2 reduces to

$$T_n = \frac{\sum C_i}{N} \times 100 \qquad \text{(Eq. 3)}$$

If none of the conceptions is wanted, T_n *represents a contraceptive failure rate*, calculated for period n. In the case where exits from observation during period n are significant, using Equation 3 instead of Equation 2 will result in under-estimating the true failure rate.

Can we define a relationship between T and E? In the homogeneous case and again taking account of only one conception per woman, the procedure is straightforward. The monthly number of conceptions (C_i) is obtained by multiplying the number of women at risk (S_i) by the risk of conception (residual fecundability p_r):

$$C_i = S_i p_r = S_i p_n (100 - E)/100$$

if E is expressed in percentage terms. If the number of women under observation (N) stays constant during the observation period, the number of women still at risk (i.e. who have not yet conceived) at the start of month i is simple to calculate:

$$S_i = N(1 - p_r)^{i-1}$$

The formula for summing the terms of a geometric progression then gives:

$$\sum_{i=1}^{n} C_i = P_r N \sum (1 - p_r)^{i-1} = N\left[1 - (1 - p_r)^n\right]$$

and Equation 3 becomes:

$$T_n = 100\left[1 - (1 - p_r)^n\right] \qquad \text{(Eq. 4)}$$

which establishes a relationship between the rate of conception T and the residual fecundability (and hence also the contraceptive effectiveness E).

Let us again use the elements in Table 31–3:

- When there is no contraception, Equation 3 gives as rate of conception for the first year:

$$T = \frac{986}{1000} \times 100 = 98.6\%$$

- With a contraceptive that is 90% effective the rate will be:

$$T = \frac{306}{1000} \times 100 = 30.6\%$$

- We can check that an identical rate is obtained with Equation 4:

$$T = 100\,[1 - (1 - 0.3)^{12}] = 98.6\%$$

and

$$T = 100\,[1 - (1 - 0.03)^{12}] = 30.6\%$$

But the same is not true for the heterogeneous case, i.e. when there is couple-to-couple variation in natural fecundability or (to an even greater degree) in contraceptive effectiveness. For in this case *the couples with the highest risk conceive soonest* so that the mean risk of those who have not yet conceived after month i decreases. We will give an example of this with the Pearl index.

2. The Pearl Index

In the 1930s, Pearl (1932) proposed a method for calculating *a rate of conception that takes into account only the periods of exposure to the risk.* Once conception has occurred, the woman is in effect protected against any new conception, for a period at least as long as the pregnancy and sometimes appreciably longer. Paradoxically, therefore, a conception actually helps to reduce the rate of conception observed over period n as a whole. In the Pearl index, the denominator is formed by the sum of the S_i, rather than the N_i:

$$R_n = \frac{\text{Total number of conceptions}}{\text{Number of months of exposure to conception}} \times 1200 = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} S_i} \times 1200 \qquad \text{(Eq. 5)}$$

This is a rate of conception, expressed for an annual period (per 100 women-years). The first point to note is that the *maximum theoretical value of this index is not 100 but 1200.* Assume that each woman conceives at each ovulation ($p = 1$): In these conditions the number of conceptions C_i would always be equal to S_i and R would be equal to 1200. In practice, p is not equal to 1. As we have seen, when no contraception is being used, it is between 0.2 and 0.3. In the homogeneous case, we can check that:

$$R = 1200\ p$$

i.e. that the rate of conception is given directly by p (apart from the coefficient 1200), as in fact results from the definition of R.

With $p = 0.3$, R is 360, and with $p = 0.15$ (a 50% level of effectiveness of the contraception), R still has a value of 180. In a situation of heterogeneity, the R values corresponding to identical values for mean fecundability would be even higher.

The relation $R = 1200\ p$ is only valid in conditions of homogeneity, except for an observation limited to a single cycle; in this case, under conditions of heterogeneity, we still have:

$$R_1 = 1200\ \bar{p}$$

Beyond a single cycle, the relation no longer holds, because of the progressive selection that operates, since it is the most fecund women who conceive soonest. Table 31–4 gives an example of the effects of this selection, taken from an American survey. Calculated over the first 12 months, the index R was 36%; over the first 24 months, less than 30%; and after 4 years, 26%.

This dependence of R on the length of observation is a drawback if we wish to compare two separate surveys that use different lengths of observation. For this reason *it is imperative* always to mention the exact length of the observation and to compare only values that are calculated over the same period (preferably 12 months).

Assuming we have a distribution of p_r we can try to establish a relation between R and E using an *empirical* approach. In accord with the conclusion of the previous paragraph, the relation must be calculated for an observation period of a specific length. This was the approach used by Henry. Table 31–5 shows the relations he established between values of R_{12} (12 months) and E.

TABLE 31–4 Effect of Length of Observation on the Pearl Index (R)[a]

Length of observation	R (per 100 women-years)
1 month	42.4
6 months	37.1
12 months	35.7
18 months	30.8
24 months	29.4
36 months	27.3
48 months	26.2

[a] Results calculated on the same group of women.

Source: Charles F. Westoff, Robert G. Potter, Philip C. Sagi, and Elliot G. Mishler, 1961. *Family growth in metropolitan America*. Princeton, Princeton University Press, 433 p.

The relation becomes looser as E decreases and the number of conceptions rises, but it is clear enough for the usual range of effectiveness E (90%–100%).

As long as the probability of conceiving in 12 months remains low, which in practice means levels of effectiveness of at least 95%, the proportion of women who conceive in 12 months (T_{12}) is very close to R_{12}. The number of women exposed to the risk (S_i) decreases only very slowly and Equations 2 and 3 differ little in the case of $n = 12$. It is therefore correct, in this case *and in this case only*, to interpret the Pearl index as the proportion of women who conceive in 12 months. We saw earlier that this would not be true for higher values of risk of conception.

Even in this case, however, it must be noted that the effectiveness of the method used is not equal to the complement at 100 of the failure rate. If, for example, we obtain a Pearl index value of 4 (per 100 women-years), it would be incorrect to conclude that the method has an effectiveness of 96%—according to Table 31–5, its effectiveness is in fact 99%.

3. Fertility Rates, Failure Rates and Fecundability

Mention can be made of a final specificity of the indices T and R—they refer to a cohort of women, characterized by the fact of *all being exposed to the risk of conceiving* (even if this risk is low) *at the start of observation*. This is what distinguishes them from conventional fertility rates, in which the number of births (or conceptions) is divided by *all* the women of a given age, whether they are exposed to the risk or not.

The distribution of conceptions, as presented in Table 31–3 for example, is equally uninformative as to the value of the conventional fertility rate, and, vice versa, knowing this fertility rate is of no help in evaluating fecundability. The fertility rate (number of

TABLE 31–5 Relation between the Pearl Index calculated for 12 months (R_{12}) and effectiveness of contraception (E)

R_{12} (for 100 women-years)	E %
1.5 to 2	99.5
3 to 4	99
6 to 8	98
12 to 21	95
about 30	90

Louis Henry, 1968. Essai de calcul de l'efficacité de la contraception, *Population*, vol. 23(2), p. 265–278.

births or conceptions per 1000 women under observation during one year) is in fact determined by several parameters:

- Fecundability, which determines the time necessary to conceive;
- The duration of gestation and of the postpartum nonsusceptible period during which the woman is unable to conceive. This nonsusceptible period is short when the woman does not breast-feed (approximately 6 weeks) but can last up to a year when she breast-feeds for 2 years (as observed in many African and Asian societies);
- Fetal mortality, which is responsible for further lengthening the interval between live births.

If we set aside sterility (i.e., take into account only women who are still fertile) the fertility rate is equal to the reciprocal of the interbirth interval. This result is consistent with what we would expect intuitively—a birth every 2 years is the same as a fertility rate of 0.5 (or 500 per 1000).

Let us examine a simple case in which we will disregard fetal mortality and assume that all the women have the same fecundability p and the same nonsusceptible period m. The interval between two births will be equal to the sum of the waiting time to conception and the nonsusceptible period (including in the latter the length of gestation), that is:

$$I = \frac{1}{p} + m$$

The fertility rate (annual) of women who are still fertile will be, when I is expressed in months:

$$F = \frac{12}{I}$$

Table 31–6 shows that even in this simplified case the values of the indices T (conception rate for first 12 months), R (Pearl index), F (fertility rate) and $1 - E$ are all different.

4. Other Problems

We have already identified a difficulty in the measurement of the contraceptive failure rate—namely that this measurement (even when we express it in annual terms) depends on the length of observation. A distinction also has to be made between the *theoretical* effectiveness and the *clinical* effectiveness (or *use* -effectiveness) of any one method. The former is evaluated in ideal laboratory conditions of use and assuming full compliance with instructions. The latter is measured under more realistic conditions for a real population whose members are not of above average motivation or education. The difference between these measurements is likely to be greater the more demanding the method is to use. This is why the effectiveness of the same method when used by two separate populations can depend in large part on the cultural distance between these two samples. The temperature method, for example, when practiced correctly by a population of highly motivated couples can be as effective—if not more so—as the pill used carelessly or half-heartedly. This is why it is important in a data collection to be completely rigorous and not to exclude a failure by blaming it on the carelessness of the subject. It is clear that if all failures are excluded, as tended to be done in the early evaluation studies of the pill and the intrauterine device (IUD), an effectiveness of 100% will always be obtained.

More recently, a third notion has been introduced: that of *extended use-effectiveness*, to take into account the contraceptive failures occurring after the discontinuation of a method that has been not well tolerated. This amounts to incorporating another dimension of contraception, namely its *acceptability.* What, after all, is

TABLE 31–6 Calculation of the Different Indices in Two Simple Cases[a]

Index	Monthly risk of conception:	
	$p = 0.25$ (without contraception)	$p = 0.0125$ (contraception with 95% effectiveness)
Mean time to conception ($1/p$)	4 months	80 months
Birth interval (I)	16 months	92 months
Annual fertility rate ($12/I$)	75 p. 100	13 p. 100
Number of women conceiving during first 12 months of exposure to the risk (T)	97 p. 100	14 p. 100
Pearl index for first 12 months (R)	300 p. 100	15 p. 100
$1 - E$	100 p. 100	5 p. 100

[a] Homogeneous groups, total nonsusceptible period of 12 months.

TABLE 31–7 Effectiveness, Failure Rates and Continuation Rates for Various Contraceptive Methods

Method	Failure rate[a]		Use-effectiveness, %	Continuation rates, %	
	Theoretical	Clinical		1 year	2 year
Tubal sterilization	0.04–0.08	0.04–0.08	~100	—	—
Vasectomy	0.15	0.15–1.0	99.9	—	—
Injection	?	0.30–2.50	99.5–99.9	40–85	25–80
Oral contraceptives					
Combined	0.03–0.10	1–7	98–99.5	40–80	30–70
Sequential	0.20–0.56				
Minipill	1.60–3.00				
Temperature	0.07–1.30	1–25	90–99.5	—	—
Intrauterine device	1.00–3.00	1–10	98–99.5	50–85	40–75
Diaphragm	1.50–3.00	3–35	90–99	—	—
Condom	1.50–3.00	3–35	90–99	—	—
Vaginal (chemical)	(1.8)	2–40	50–99	—	—
Calendar rhythm (Ogino)	(3.8)	15–50	50–95	—	—
Withdrawal	(3–10)	(10–100)	—	—	—
No contraception	—	100–300	0	—	—

[a] Number of conceptions per 100 years of exposure to the risk. Whenever possible, the rate was calculated for the first 12 months (R_{12}).

Source: Henri Leridon, 1977. *Human fertility: the basic components*. Chicago, The University of Chicago Press, 202 p.

Table based on Christopher Tietze, 1971. Effectiveness of contraceptive methods, *in:* Egon Diczfalusy and Ulf Borell, *Control of human fertility*, p. 303–320. New York, Wiley, 354 p. (Proceedings of the 15th Nobel Symposium, Lidingo). and various issues of *Population Reports*, 1973–1975.

the value of a method that is completely effective if it cannot be tolerated for longer than two or three months? And what else could explain that women requesting abortion are often found to be former pill users, if it is not the low tolerance for this type of contraception by some women?

Table 31–7 illustrates these different problems by giving fairly large ranges of failure rates (theoretical and clinical), use-effectiveness, and the 1- and 2-year continuation rates, for the various contraceptive methods.

Life table methods (a classic tool of demography) allow a maximum use of all the available information, notably when discontinuations and interruptions are frequent (see, for example, Potter, 1960).

5. One Abortion Does Not Mean One Averted Birth

When a woman decides to have an abortion, she averts a live birth. Yet it is inaccurate to think that if, for example, 100,000 abortions are carried out each year in a population, 100,000 births will have been averted in this population. This is because the interval between two births is equal to the sum of the waiting time to conception ($1/p$), a gestation time (g_m), and a *postpartum* nonsusceptible period (t_m). Under natural fertility conditions, plausible values for these three component parts would be 5, 9, and 10 months, or 24 months in total. In the case of abortion, the last two components are reduced to approximately 3 and 1 months, respectively, making a total interval of 9 months instead of the 24 months for a live birth: Therefore 24/9 = 2.67 abortions are needed to avert one live birth. The ratio is, however, close to 1 where there is no breast-feeding (t_m is now only 2 months) and where a contraceptive method is used (which increases the ratio $1/p$). In general, the number of births averted by an abortion is equal to:

$$\frac{[1/p(1-E)] + g_m + t_m}{[1/p(1-E)] + g_a + t_a}$$

where g_a and t_a are, respectively, the durations of gestation and of the *postpartum* nonsusceptible period in the case of abortion.

Acknowledgment

Translated from French by Godfrey I. Rogers.

References

Cantrelle Pierre and Leridon Henri, 1971 Breast-feeding, child mortality and fertility in a rural zone of Senegal, *Population Studies*, vol. 25(3), p. 505–533.

Caraël Michel, 1995 Sexual behaviour, *in:* John Cleland and Benoît Ferry, *Sexual behaviour and AIDS in the developing world*, p.75–123. London, Taylor and Francis, 244 p.

Ducros A. and Pasquet P., 1978 Évolution de l'âge d'apparition des premières règles (ménarche) en France, *Biométrie humaine*, vol. 13(1–2), p. 35–43.

Ellison Peter T., 1982. Skeletal growth, fatness and menarcheal age: a comparison of two hypotheses, *Human Biology*, no. 54, p. 269–281.

Ferry Benoît and Smith David P., 1983 *Breastfeeding differentials* London, WFS, 89 p. (Comparative Studies, no. 23).

Frisch Rose E. and Revelle Roger, 1970 Height and weight at menarche and a hypothesis of critical body weights and adolescent events, *Science*, no. 169, p. 397–399.

Henry Louis, 1968. Essai de calcul de l'efficacitéde la contraception, *Population*, vol. 23(2), p. 265–278.

John Meredith, Menken Jane, and Chowdhury Alauddin, 1987 The effects of breastfeeding and nutrition on fecundability in rural Bangladesh: a hazards-model analysis, *Population Studies*, vol. 41(3), p. 433–446.

Jones Robert E, 1988. A biobehavioral model for breastfeeding effects on return to menses post-partum in Javanese women, *Social Biology*, vol. 35(3–4), p. 307–323.

Konner Melvin and Worthman Carol, 1980 Nursing frequency, gonadal function, and birth spacing among Kung hunter-gatherers, *Science*, no. 207, p. 788–791.

La Rochebrochard Élise de, 2000. *Modélisation de la baisse de la fertilité avec l'âge de la femme et l'âge de l'homme* Paris, Université Pierre et Marie Curie, 438 p. (biomathematics thesis).

La Rochebrochard Élise de, 1999 Les âges à la puberté des filles et des garçons en France. Mesures à partir d'une enquête sur la sexualité des adolescents, *Population*, vol. 54(6), p. 933–962.

Leridon Henri, 1977. *Human fertility: the basic components.* Chicago, The University of Chicago Press, 202 p.

Leridon Henri, 1993. La fréquence des rapports sexuels. Données et analyses de cohérence, *in:* Michel Bozon and Henri Leridon (eds.), Sexualité et sciences sociales. Les apports d'une enquête, *Population*, vol. 48(5), p. 1381–1408.

Mosley Henry, 1979. The effects of nutrition on natural fertility, *in:* Henri Leridon and Jane Menken (eds.), *Natural fertility/fécondité-naturelle*, p. 83–105. Liège, Ordina Éditions, 556 p.

Pison Gilles, 1989. Les jumeaux en Afrique au sud du Sahara: fréquence, statut social et mortalité, *in:* Gilles Pison, Étienne van de Walle and Mpembele Sala Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 245–270. Paris, INED, PUF, 446 p.

Potter Robert G., 1960 Length of the observation period as a factor affecting the contraceptive failure rate, *Milbank Memorial Fund Quarterly*, vol. 38(2), p. 140–152.

Querleu Denis, Spira Alfred, and Leridon Henri, 1990 Epidémiologie de la fertilité, *in: Encyclopédie Médico-Chirurgicale (Paris)*, Gynécologie, 738, C^{10}, p. 1–6.

Sandfort Theo, Bos Henny, Haavio-Mannila Elina, and Sundet Jon Martin, 1998 Sexual practices and their social profiles, *in:* Michel Hubert, Nathalie Bajos and Theo Sandfort (eds.), *Sexual behaviour and HIV/AIDS in Europe*, p. 106–164 London, UCL Press, 442 p.

Tanner James M., 1962 *Growth at adolescence*, 2nd ed. Oxford, Blackwell Scientific Publications, 325 p.

Tietze Christopher, 1971. Effectiveness of contraceptive methods, *in:* Egon Diczfalusy and Ulf Borell, *Control of human fertility*, p. 303–320. New York, Wiley, 354 p. (Proceedings of the 15th Nobel Symposium, Lidingo).

Treolar Alan E., 1974 Menarche, menopause and intervening fecundability, *Human Biology*, no. 46, p. 89–107.

Udry Richard J., 1979 Changes in the frequency of marital intercourse from panel data, *Achives of Sexual Behaviour*, no. 9, p. 319–325.

Westoff Charles F., Potter Robert G., Sagi Philip C., and Mishler Elliot G., 1961. *Family growth in metropolitan America.* Princeton, Princeton University Press, 433 p.

Wood James W, 1994 *Dynamics of human reproduction: biology, biometry, demography.* New York, Aldine de Gruyter, 653 p.

Wood James W. and Weinstein Maxine, 1988. A model of age-specific fecundability, *Population Studies*, vol. 42(1), p. 85–113.

CHAPTER

32

Fetal Mortality

CATHERINE GOURBIN

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Fetal mortality has until now usually been studied as an attribute of other demographic phenomena, whether for its effect on fecundability in the analysis of fertility (Henry, 1964; Wood, 1989) or as a risk factor in the context of infant mortality. It acquired some importance through the concept of perinatal mortality (stillbirths plus infant deaths at 0–6 days) that Peller (1948) introduced to avoid the errors caused by unreliable data on infants' vital status at birth. However, studies on this subject have generally been based on civil registration data and the fetal deaths taken into account are those subject to compulsory declaration.

On that last point, note that registration of fetal deaths was slow to be introduced compared with that of live births. The explanation for this lack of interest is to be sought in the historic origins of civil registration. It is generally accepted that this developed from the parish registers in which baptisms, marriages and burials were recorded (see Chapter 121). The act of baptism allowed the infant to be given the status of a person (or more exactly, a Christian) by freeing it of original sin. Similar thinking informed legislation regulating the declaration of events for civil registration, whereby the newborn child, through birth certification, acquired a legal identity. The latter establishes rights with regard to filiation, inheritance, and settlement but is conditioned by the fact of the child being born alive and viable. Not until the 19th century[1] did most countries adopt the criterion of *gestation duration* and the limit of *28 weeks* (counted from the date of the last menstrual period), which is considered the length of time after which a fetus becomes capable of autonomous existence (Empire français, 1805).

Contemporary developments such as new patterns of fertility behavior (birth control, smaller families), the increasingly technical character of childbearing in the areas of reproduction (*in vitro* fertilization) and medical care for the newborn, the ethical debates over embryo status are combining to give the levels and determinants of fetal mortality greater priority as research topics.

The importance of standardized statistics in this field was illustrated in 1950, when the World Health Organization (WHO) Committee of experts on health statistics recommended that late fetal mortality statistics include only fetal deaths occurring after week 28 of gestation (WHO, 1950a), thereby reaffirming the measures adopted in 1925 by the League of Nations Health Office. Concurrently with this decision, the WHO acknowledged that fetal mortality was a major problem whose full extent was not reflected in the existing official figures. The representatives of the American National Committee on Population Statistics who appeared before this committee quoted the frequency of fetal deaths for 1948 in New York city, where 67% of recorded fetal deaths occurred before week 20 of gestation, 12% between weeks 20 and 28, and 18% after week 28 (WHO, 1950b). The limit of 20 weeks' gestation was in fact the legal limit for declaration of stillbirths in New York. Influenced by this example, the WHO proposed classifying fetal deaths according to the gestation duration at

[1] Except in Sweden and Norway, where the first legal provisions for declaration of stillbirths date from 1686 and 1797, respectively.

which they occur, with the following divisions: *early fetal mortality* (less than 20 weeks), *intermediate fetal mortality* (20–27 weeks), and *late fetal mortality* (28 weeks or more).

I. FETAL MORTALITY LEVELS

Evaluation of fetal mortality levels is necessarily limited by data availability. Conventional demographic statistics derived from the stillbirths recorded by civil registration systems pertain to late fetal mortality only. Many countries use the criterion and limit of 28 weeks' gestation, but the substitution between criteria proposed by the WHO (1000 g in weight, 35 cm in length, or 28 weeks of gestation, for the production of international statistics) creates problems for comparability of the levels, because the limits are not equivalent (Gourbin, 1995). Further definitional disparities arise from the adoption of different limits for the same criterion. Mention was made above of the case (noteworthy) of many states in the United States registering fetal deaths after 20 weeks of gestation from the end of the 1940s, but more recently other countries such as Portugal and Finland have adopted the WHO recommendations (WHO, 1977) for the compilation of national statistics, which provide for reporting all births with a weight of at least 500 g (or 22 weeks of gestation or 25 cm in length).

1. Late Fetal Mortality

Because civil registration declaration of stillborn children is compulsory in developed countries, it can generally be assumed that the number of fetal deaths at 28 weeks or more of gestation is known, at least in those countries with a long tradition of keeping statistical records. Table 32–1 shows the levels of late fetal mortality (or stillbirths, often referred to in French as *mortinatalité*) calculated from these data for the various world regions.

To facilitate international comparisons, the WHO recommends standardization of perinatal mortality rates for weight or gestational duration, including in the calculation only infants with a birth weight of at least 1000 g or a minimum gestation of 28 weeks. In practice, it is usual for the denominator to include the fetal deaths recorded by civil registration

TABLE 32–1 Late Fetal Mortality in the World (per 1000 Births)

Country	Late fetal mortality	Country	Late fetal mortality	Country	Late fetal mortality
Africa		**Australasia**		**Europe (cont.)**	
Mauritius 1996–1997	12.8	*Australia 1996*	4.1	**France 2000**	**4.6**
Egypt 1995	3.8	*New Zealand 1996–1997*	3.1	*Germany 1997*	*4.3*
Tunisia 1997	11.0			*Greece 1997*	*6.1*
		Asia		Hungary 2001	*5.6*
North/Central America		*Azerbaijan 1995–1996*	4.5	*Iceland 1994–1997*	*3.3*
Canada 1995	6.1	*China–Macao 1993*	22.0	**Ireland 1999**	**4.2**
Cuba 1996	10.3	*Japan 1997*	2.8	**Italy 1998**	**3.7**
El Salvador 1996–1997	*5.4*	*Kazakhstan 1997*	9.1	*Latvia 1995–1997*	*9.1*
Guatemala 1993	17.1	*Malaysia 1997*	4.4	*Lithuania 1995–1997*	*5.3*
Mexico 1997	6.4	*Philippines 1993*	5.2	**Luxembourg 2000**	**4.6**
Panama 95–96	6.3	*Tajikistan 1994*	8.4	**The Netherlands 1999**	**4.6**
Puerto Rico 1996–1997	11.2	*Uzbekistan 1997*	6.9	*Norway 1996–1997*	*4.2*
United States 1998 ≥28 wk	3.4			*Poland 1997*	4.9
United States 1998 ≥20 wk	6.7	**Europe**		**Portugal 1999**	4.3
		Austria 2001	2.6	*Romania 1997*	6.3
South America		*Belgium 1995*	4.8	*Russian Fed. 1995*	7.4
Argentina 1997	11.4	*Bulgaria 1997*	7.5	*Slovakia 1994–1995*	4.0
Brazil 1998	9.7	*Croatia 1996–1997*	4.4	*Spain 1996*	3.9
Chile 1997	4.8	*Czech Rep. 1996–1997*	3.8	**Sweden 2000**	**3.9**
Ecuador 1997	13.5	**Denmark 2000**	3.8	*Ukraine 1997*	6.7
Venezuela 1996	21.4	*Estonia 1995–1997*	7.8	*United Kingdom 1997*	5.4
		Finland 2000	**2.8**	*Yugoslavia 1995*	6.2

Sources: *Items in italics:* Demographic Year Book, 1998, United Nations; *in* roman: National Population Statistical Yearbooks; **bold**: Buitendijk *et al*., 2003.

plus all live births regardless of weight or gestation duration.

In Western and Central Europe and the other industrialized countries,[2] late fetal mortality at the end of the 1980s was usually around 4 per 1,000 to 6 per 1,000, levels that are extremely low and, it may be noted, consistently lower than those for infant mortality (Masuy-Stroobant, 1994). This situation is the result largely of more accessible and effective care during pregnancy and at birth, although also of a favorable socioeconomic context. Coverage appears to be complete, and if any bias does occur it tends toward overestimation of late fetal mortality when the legislation specifies the signs of life by which to recognize a live birth[3] (presence of heart beat or respiration in Austria and Switzerland) or when infants are included who were live born but died before their birth was registered, or in the 24 hours after birth (Gourbin and Masuy-Stroobant, 1995).

Data on late fetal mortality are conspicuously lacking in Africa and some South American and Asian countries due to civil registration systems that are still defective or to the lack of statutory provision for the declaration of stillbirths. Local surveys have attempted to estimate levels of perinatal mortality; conducted in the 1970s, these studies situated late fetal mortality at between 18 per 1,000 (Kinshasa, Zaire) and 95 per 1,000 (Mulamfashi, Zambia). However, these surveys were carried out on women giving birth in hospitals thus it would be unsafe to generalize these levels. Furthermore, differences in collection procedures and high levels of drop out from the initial samples undermine the validity of the results and make comparison impossible (Nzita Kikhela, 1986).

Using the surveys conducted in the *World Fertility Survey* (WFS) program, Casterline (1989) has attempted to evaluate levels of fetal mortality, whether the event is classified as a miscarriage (before 28 weeks of gestation, the pregnancy having been recognized by the mother) or a stillbirth (28 weeks and over). For an observation period limited to the 5 years preceding the survey, comparison of 22 African, South American, and Asian countries shows the percentage of pregnancies ending with fetal death (miscarriage plus stillbirth) to vary between 5.6% (5.2 + 0.4) in Ghana and 20.4% (17.8 + 2.6) in Guyana. In countries where induced abortion is authorized, an induced abortion is less likely to be deliberately misreported as a spontaneous abortion and so the figures will be more reliable. Overall these estimations appear to give 50% to 80% coverage of fetal deaths compared with the results obtained from intrauterine mortality life tables (see below). Despite results that it might have been fruitful to develop in subsequent research, those in charge of the *Demographic and Health Survey* (DHS), which was partly a continuation of the WFS, decided to discontinue collection of this kind of data, judging that the costs outweighed the potential returns (Casterline, 1989).

[2] Excluding the newly industrialized countries of Asia.

[3] The live-born child who dies a few moments later without having presented these signs will be recorded as stillborn on condition of satisfying the criteria for stillbirth declaration.

2. The Risks of Fetal Death in Pregnancy

Although levels of late fetal mortality can be evaluated in countries with a long tradition of statistical record keeping, the situation is different when analysis encompasses the entire duration of pregnancy. Estimation of the risk will depend closely on the study protocols for early pregnancy detection. Before reviewing these, a short description of the hormonal responses initiated by conception is useful for understanding the difficulties associated with measurement of very early fetal loss.

Pregnancy Detection

Ovulation in the human female occurs on average about 15 days after the start of the last menses.[4] Between 36 and 40 hours before ovulation there is a surge in production of LH (luteinizing hormone) that unblocks the meiotic process of the ovum. A fertilized ovum moves through the fallopian tube to the uterine cavity, reaching it 3 to 4 days later. Finally, 1 week after ovulation, the embryo attaches to the lining of the uterine wall. This implantation is accompanied by an increase in production of the hormone hCG (human chorionic gonadotrophin).

Early detection of pregnancy (and its development[5]) is based on identification of the surges in hormone production. Research reviewed by Wilcox *et al*. (1993) involved establishing the serum presence of a protein known as *early pregnancy factor* that is detectable within 24 hours of fertilization. Measurement of this factor is problematic, however, and doubts exist over its specificity as a marker of pregnancy. What happens to ova in the first week after fertilization remains uncertain, and the study by Hertig (1959) to determine the risks of subclinical loss remains unique in this field, both for

[4] To be exact, between the 8th and 24th days.

[5] *In vitro* pregnancies are not considered here.

the information obtained and for the methods used. This study was conducted on women undergoing hysterectomy for reasons other than pathologic states that could prevent fertilization. The patients were asked to have sexual intercourse around the presumed date of ovulation and the operation was carried out in the days that followed. Any product of conception that existed was removed. Histologic examination was performed to determine the egg's developmental stage and the normal or abnormal character of its evolution. The results obtained, however, have been subject to differing interpretations (Leridon, 1973). In summary, of all the conceptions observed, 37% would result in a live birth, whereas half of all fertilizations, that is potential pregnancies, would end in fetal death before the end of week 4 of gestation. Over and above the ethical problem raised by the methods used to obtain these results, a number of criticisms can be made:

Measurement of the rates is based on too few observations;

The study conditions: Hysterectomy is rarely performed on women younger than 30 years, yet risks of fetal death are known to increase with age;

The hypotheses underlying the results: (1) All morphologically abnormal conceptions end with expulsion of the fertilized ovum and (2) these are the only conceptions to end in this way. Although the first hypothesis appears correct in most cases, it is unsafe to assume normal embryo development solely on the basis of an oocyte's normalcy at a very early stage of development (Kline *et al*., 1989).

Notwithstanding these criticisms, Hertig's study drew attention to the extremely high levels of intrauterine mortality during the first weeks of pregnancy. It also showed (or confirmed) that the series of risks associated with the different stages of the gestation period declines exponentially at the start of the pregnancy. Although other studies have confirmed this hypothesis, estimations of total risk levels and of variation in fetal mortality rates by gestation duration remain unclear, the results obtained being heavily influenced by the aims and protocols of individual studies (Modvig *et al*., 1990).

II. INFLUENCE OF DATA COLLECTION METHODS

The most frequently made criticism of retrospective studies concerns the bias caused by misreporting of events. In the field of fetal mortality, the three most common biases are the following:

A fetal death can be forgotten, this recall failure depending on socioeconomic factors, the gestational age at which loss occurs, and the length of time since the event[6];

Irregular menstrual cycles can be confused with spontaneous abortion;

An induced abortion can be presented as a spontaneous abortion.

In addition, retrospective studies that cover women's complete reproductive histories (Tietze, 1948; Risch *et al*., 1988) and calculate the risk of fetal death in relation to all recorded pregnancies are liable to overestimate fetal mortality by overrepresenting the least fertile women if these are trying to achieve a desired number of children (the problem is identical in panel surveys).

The recruitment of women for *prospective surveys* depends on whether the study aims to identify very early fetal losses or those occurring after a first missed menstrual period. In the former case, the participants are all volunteers, want to have a baby, and agree to submit to the strict biologic observation[7] necessary for very early detection of pregnancy (Edmonds *et al*., 1982; Wilcox *et al*., 1988, 1993). Pregnancy is detected by measuring urinary concentration of the hCG secreted by the placenta after implantation.[8] Detection of this hormone is the basis of early pregnancy diagnosis and the accuracy of the result depends on the specificity of the assay used.[9] In the second case, the study uses women who enter observation at the point when their pregnancy is clinically recognized (medical visit or ultrasound scan), or who agree to keep an accurate menstrual record and to report any pregnancy that is followed by a recognized fetal death

[6] On the latter points, the comparison by Wilcox and Horney (1984) of data from a study of spontaneous abortions recorded in hospital when they occurred with those obtained by retrospective questioning of the same women, show that 25% of subjects did not recall having suffered a spontaneous abortion. This difference was directly related to the gestation duration at which death took place and to the length of time since the event: the coverage achieved was 82% if the fetal death took place in the 10 years before the survey, 73% if it occurred more than 20 years earlier.

[7] Urine specimens had to be taken every morning and stored in the refrigerator for collection once a week by the laboratory doing the analyses.

[8] Then implantation occurs, a large and rapid rise is observed in hCG circulating rates. This hormone, with others, stimulates hormonal production and maintains the pregnancy.

[9] It is noteworthy that the results of the studies conducted before the mid 1980s are better than those obtained subsequently, these differences apparently being due to the non-specific nature of the assay (Wilcox *et al*., 1993).

TABLE 32–2 Estimation of Risks of Fetal Death by Recognized Pregnancy in Various Studies (per 100 Pregnancies)

Sources	Date and place of study	Study type	No. of pregnancies	Risk of fetal death, %
Tietze *et al.*, 1948	1937–1946 United States Baltimore	Prospective, prenatal visit	10,397	12.2
		Retrospective	30,133	17.0
Leridon, 1976	1971 France Créteil	Prospective, prenatal visit	1,304	11.5
		Retrospective	2,248	15.3[a]
Wilcox *et al.*, 1981	1961–1970 United States Minnesota	Prospective, menstrual cycle monitoring	1,493	13.1[b]
Edmonds *et al.*, 1982	1980–1981 England Southampton	hCG measurement	118	62.0
Wilson *et al.*, 1986	1981–1983 Canada Vancouver	Prospective, first ultrasound scan	747	2.9
Risch *et al.*, 1988	1974–1978 Canada and United States	Retrospective	6,282	12.8
Santow and Bracher, 1988	1986–1987 Australia	Retrospective	6,213	15.1
Wilcox *et al.*, 1993	1983–1985 United States Columbia	hCG measurement	197	32.0
Gourbin, 2000	1984–1988 Hungary	Statistical registration records	708,950	12.1

hCG, human chorionic gonadotropin.
[a] Induced abortions are excluded from the denominator.
[b] Until week 20 of gestation.

(even if the facts have not been clinically diagnosed) (Wilcox *et al.*, 1981).[10] Some of these prospective studies have been used for constructing intrauterine mortality life tables.

Table 32–2 shows the differences in estimation of risk according to the method used. The mortality rates were calculated by dividing the number of fetal deaths by the number of recognized pregnancies. Except for the study by Harvey Risch *et al.* (1988), retrospective surveys find levels of risk of fetal death higher than those estimated in prospective studies. Heterogeneity among individuals is partly responsible for these differences—women who have had the desired number of children avoid becoming pregnant again, whereas women who have suffered repetitive spontaneous abortions go on trying in the hope of a favorable outcome. In other words, the number of pregnancies will be influenced by the outcomes of previous pregnancies. Furthermore, some of the fetal deaths reported over the woman's reproductive life in retrospective studies have occurred before any medical visit or too early in gestation for hospitalization to be necessary. In most prospective studies, however, entrance to observation is at the initial prenatal visit. Since fetal death occurs mainly in the first weeks of pregnancy, the levels of intrauterine mortality will depend closely on how early this visit takes place. Lastly, it can be noted that a study based on a systematic recording of fetal deaths in Hungary shows a level of risk of fetal mortality similar to those observed in prospective studies (Gourbin, 2002).

[10] The level of fetal mortality in the former case will be closely related to the gestation duration recommended for the start of prenatal visits (thus conditioning an earlier or later recognition of the pregnancy) or for the first ultrasound scan. In the second case (self-observation), the results can be biased by the greater or lesser vigilance (or ability) of the women in spotting the signs of pregnancy, or by irregular menstrual cycles.

III. INTRAUTERINE MORTALITY RATES BY GESTATION DURATION

Although prospective surveys are more reliable for estimating fetal death frequencies, we also need to know the distribution of these deaths at the different stages of pregnancy in order to analyze the variation in the risk of fetal death by gestation duration and to evaluate the total level of intrauterine mortality. This approach has been used in some studies and has led to the construction of intrauterine mortality life tables, the first of which, published by French and Bierman in 1962, remains a reference in this field (Kline *et al.*, 1989; Leridon, 1977). In theory, the life tables should be constructed using probabilities that allow for the different pregnancy outcomes that can be competing with death—each of these outcomes is exclusive, so if one occurs the other cannot. This is the case between fetal mortality and induced abortion in early pregnancy, therapeutic abortion during pregnancy, and live birth at full term. In the six series of fetal death probabilities presented here (Table 32–3), French and Bierman exclude live births from the denominator, Ehrardt (1963) estimates the probabilities (see below), Shapiro *et al.* (1962) do not give a clear explanation of the method used, Taylor (1960) and Goldhaber and Fireman (1991) compute daily proba-

TABLE 32–3 Risk of Fetal Death by Gestation Duration in Different Studies (for 1000 Recognized Pregnancies)

Gestation duration, weeks	Survey					
	French and Bierman (1962) Kauai Island, Hawaii, 1953–1956	Schapiro *et al.* (1962) New York, United States, 1958–1959	Erhardt (1963) New York, United States, 1960	Taylor (1970) California, United States, 1959–1966	Goldhaber and Fireman (1991) California, United States, 1981–1982	Gourbin (2002) Hungary, 1984–1988
4–7	108.1	13.7	82.4	61.2	23.1	20.1
8–11	69.9	58.8	67.1	48.7	56.3	40.8
12–15	44.8	40.0	28.2	25.3	33.3	20.2
16–19	13.3	14.0	10.7	11.0	11.5	9.3
20–23	8.5	6.5	8.9	7.7	4.4	7.1
24–27	3.2	3.6	2.1	3.3	2.4	4.0
28–31	3.0	2.2	4.3	3.6	1.5	1.8
32–35	3.0	2.6	2.0	3.3	1.5	2.0
36–39	3.4	4.0	6.9	4.2	1.0	2.9
40+	6.8	9.3	10.8	14.3	2.5	3.0
Total risk	239.6	145.8	182.4	170.0	130.8	121.0

bilities,[11] and Gourbin (2002) estimates single-decrement probabilities of fetal loss to allow for induced abortions.

Whichever method is used, determining the risks encounters two major problems:

The number of fertilizations cannot be known because of the difficulty, if not impossibility, of detecting the very earliest stages of pregnancy, which is when large numbers of zygotes[12] are lost;

The risk of death by gestational age is usually calculated for the gestation duration at which the fetus (or embryo) is expelled. However, death can occur well before this expulsion and the resulting difference between fetal age (length of time between fertilization and fetal death) and gestational age can be very large (Boué and Leridon, 1971); fetal age is extremely hard to ascertain, and the risks of fetal death are usually calculated for the gestation duration at the time of expulsion.

[11] Estimation of the rates included extrauterine pregnancies for the tables of French and Bierman and Shapiro; conversely, women whose pregnancy ended with fetal death in the 2 days following the medical visit were excluded for the tables of French and Bierman, Taylor, and Goldhaber and Fireman.

[12] Development of the fertilized egg in the uterus is reflected in a changing terminology: From fertilization to day 19, we speak of an egg or a zygote, from day 20 to month 3 is the period of embryogenesis, and from month 4 to term is the period of fetal development. For the sake of simplicity the term *fetal death* is mainly used in this chapter.

From week 4 of gestation, the total risk of fetal mortality varies between 24% and 12%, a disparity that is due mainly to the level of mortality between 4 and 8 weeks. The lower fetal mortality rates between 4 and 7 weeks reported in studies by Goldhaber and Fireman and Gourbin, are evidence of an underrecording of fetal deaths at these low gestation durations. After 8 weeks of gestation, the six series of mortality probabilities follow the same pattern of intrauterine mortality, characterized by a probability of fetal death that is very high at the start of the pregnancy and declines rapidly up to week 20. It remains fairly constant between gestational weeks 24 and 36 before rising slightly at the higher gestation durations. The series of mortality probabilities published by French and Bierman, which is considered the most reliable, especially for the early stages of pregnancy, is a good example of this mortality pattern—the risks of fetal morality decline in a completely linear fashion up to week 20, then stabilize, before increasing after week 40. It should be noted, however, that the more recent studies by Goldhaber and Fireman and Gourbin find only a small increase in risk at the end of normal pregnancy duration, probably as a consequence of medical interventions for pregnancies that have gone beyond the forecast term. Ehrardt assessed intrauterine mortality during the early weeks by extrapolating the level of intrauterine mortality in weeks 8 to 11 (judged to be reliable) to calculate the rates for 0 to 4 weeks and 4 to 7 weeks. For the period 0 to 4 weeks the value of the rate was 11%, corresponding to a total risk for the pregnancy of 29%, a level that is still below the 32% estimated by Allen Wilcox *et al*. (1988) from hCG concentrations and for the period from 3 weeks' gestation (or 1 week after fertilization) to the end of the pregnancy. Although this estimation is considered one of the best currently available, the frequency with which zygotes are lost during the very first week of their development remains an unknown.

IV. THE DETERMINANTS OF FETAL MORTALITY

Research in this field was for long limited to work by the United Nations (1954) and Morris and Heady (1955) on the determinants of late fetal mortality and that by Shapiro *et al*. (1962) and Taylor (1970) who, when constructing their intrauterine life tables, highlighted the role of biologic variables (mother's age, parity) in the etiology of fetal mortality. The 1970s and 1980s saw a growth in studies on perinatal mortality, most of which failed to distinguish between late fetal mortality and early neonatal mortality, the two phenomena being thought to have identical determinants. Gradually, however, more research has been done into the biologic and social determinants of fetal mortality, stimulated by new models of family formation, improved obstetrical and neonatal techniques, and spectacular advances in fetal medicine. Due to the scarcity of data, however, this research usually concerns only late fetal mortality.

1. Biologic Factors

Genetic Abnormalities

The literature of demography (relatively poor in this respect) and epidemiology identify chromosomal and genetic anomalies, which are usually incompatible with fetal development, as the main causes of first-trimester intrauterine mortality (Simpson and Carson, 1992; Kline *et al*., 1989; Boué and Leridon, 1971). According to these authors, 50% to 60% of first-trimester fetal deaths are due to chromosomal anomalies, of which the most common are autosomal trisomies (responsible for approximately 25% of deaths) and monosomies X (about 20% of deaths). Although several studies have reported broadly similar frequencies, underreporting is potentially large given that successful karyotyping is influenced by the type of tissue used for culturing and by the length of time until expulsion, which may be long if the embryos with anomalies have died very early in gestation. In addition, these figures do not include deaths in the first week after fertilization, although elimination of the zygote in this period probably results from severe chromosomal abnormalities[13] (Kline *et al*., 1989). Once through the first trimester of pregnancy, the malformations that can affect embryos are usually compatible with their development (although these anomalies have a higher frequency among fetal deaths than among liveborn infants [EUROCAT, 1995]) and other factors, biologic and social, acquire a determinant role in explaining fetal mortality.

2. Other Biologic Factors

The other biologic factors relate to the mother or to the infant itself and are chiefly age, reproductive history, and health status of the mother; the sex of the fetus; multiple pregnancies; and retarded intrauterine growth. The studies used in the construction of intrauterine mortality life tables also analyzed the relationship between maternal characteristics and fetal mortality, thus making it possible to evaluate their impact on mortality before 28 weeks of gestation.

Maternal Characteristics

Fetal mortality varies considerably over the mother's reproductive career and this variation in risk can be examined using the conventional characteristics of *age* and *parity*. Their influence on late fetal mortality was evaluated in 1938 by Yerushalmy (quoted by Morris and Heady, 1955); Shapiro *et al*. (1962) emphasized their importance throughout fetal life by describing the increasing risk of death with age and parity. Other aspects of women's reproductive life are also implicated.

Maternal age[14]. The risk of fetal mortality is lowest (around 11 per 1,000) when the mother is aged 20 to 24 years. This risk doubles between ages 20 and 40. Harvey Risch *et al*. (1988) have even reported a risk of fetal death before 28 weeks multiplied by 3 for women aged 45 compared with women aged 20. Similarly, a study describing fetal mortality profiles by gestation duration, reports risks of fetal mortality after 28 weeks of gestation multiplied by 4 for women over 40 compared with women aged 20 to 24 years, and a risk multiplied by 6 at the start of pregnancy (8–11 weeks) (Gourbin, 2002).

This increase is due partly to the specific effect of aging on the woman's reproductive capital and the accumulation of pathologic risks or morbid conditions,

[13] Experimental studies carried out using mice have shown an incomplete chromosome set to be responsible for the loss of zygotes in this period.

[14] The impact of paternal age on fetal mortality has rarely been studied. It seems however that it plays a role too (Rychtarikova *et al*., 2004).

but also to the effect of heterogeneity described earlier. The risk is also higher among very young women (under age 18) though the relationship is hard to evaluate because the small numbers that can be observed mean that the probabilities of dying are subject to large random variations.

Parity. The risk of fetal mortality is slightly higher in primiparous women, falls back at birth orders two and three, and then rises steadily in the higher parities (Pinelli, 1984). Some studies have claimed a doubling of risk from the sixth birth (Heady *et al*., 1955; United Nations, 1954). The differences in risk between first births and second and third births are not always obvious (Leridon, 1976; Bross and Shapiro, 1982), and regardless of the birth order considered, the results have to be qualified to allow for the difficulty of distinguishing the effects of age from those of parity (Wilcox and Gladen, 1982; Risch *et al*., 1988). A study conducted in Finland on early and intermediate fetal deaths found a higher risk only when the mother was aged 25 or over with at least of parity three (Hemminki *et al*., 1983). Similarly, an increased risk is observed for primiparous women at age 40 and older. These types of combinations (low age and high parity, or vice versa) appear to be the most problematic for fetal survival, suggesting that the association between parity and fetal mortality reflects other risk factors.

Interval between live birth and fetal death. Family planning programs have often drawn attention to the U-shaped curve of risks associated with the length of interval since previous pregnancy. Risks are particularly high for intervals of less than 9 months (Casterline, 1989) and even 6 months (10.9 per 1,000 versus 5.5 per 1,000 for the 13- to 24-month interval) (Kallan, 1992) and at the other extreme for intervals of over 60 months. In the former case, excess fetal mortality is explained by the fact that repeated and closely spaced pregnancies prevent the mother from recovering a satisfactory physiologic state (Gribble, 1993; Dupin and Rimbault, 1978). In the latter case, excess risk arises from subfecundity or an impaired endocrine function that prevents the mother from carrying a pregnancy to live birth.

Previous fetal deaths. The above argument would explain the close association observed between the number of previous fetal deaths and the risk of recurrent fetal death. Leridon (1987) notes that in a third pregnancy the risk of fetal death is around 20% to 25% if one of the earlier pregnancies has ended with intrauterine death; a similar risk (20%) was observed by Risch *et al*. (1988). Gourbin (2002) found risk of fetal death between 16 and 28 weeks of gestation more than doubled for women who had suffered one previous fetal death and more than quadrupled for those who had suffered two or more previous fetal deaths, compared with women who had never had a fetal death in their previous pregnancies. This repetition of the fetal death outcome in successive pregnancies can be seen as evidence of the heterogeneity already described (i.e., of interindividual variability in the capacity to conceive and produce a live-born baby). In this model, fetal death is a risk marker for a higher frequency of fetal death in future pregnancies. However, the occurrence of a fetal death is also a potential risk factor, in that fetal death can be a source of infections responsible for spontaneous abortions in future pregnancies.

Previous induced abortions. Most studies show that in the countries where abortion is legal, carrying out an induced abortion does not increase the risk of fetal death in subsequent pregnancies (Frank *et al*., 1991; Chung *et al*., 1982), except however in the countries where traumatic techniques are in use (WHO, 1979; Kline *et al*., 1978). The risks described in the latter case were thus due to the methods used to perform this intervention, the problem being especially acute in the countries which had legalized induced abortion very early (at the end of the 1950s) (WHO, 1979; Kline *et al*., 1978). Results from the study using Hungarian data (Gourbin, 2002) show risks of fetal death before 28 weeks of gestation at least 60% higher for women who have already had an induced abortion compared with women who have not, and the risk is more than double for women who have had two or more previous induced abortions. Today, the move away from techniques involving excessive and violent dilations of the cervix has reduced the risks of cervical incompetence, which is a risk factor for spontaneous abortion and prematurity. In the countries where abortion is illegal, recourse to clandestine abortion carries a risk of infections which are potentially serious, with consequences for the woman's subsequent reproductive career.

Maternal health status. In addition to risk behaviors such as smoking and drinking (discussed later), a number of pathologic conditions can affect fetal survival. These may be anatomic (serious uterine malformation), metabolic (diabetes), autoimmune (maternal-fetal blood incompatibility,[15] lupus erythematosus), chronic (renal or cardiac insufficiency), plus infections. Not all contribute equally to the etiology of fetal mortality, but the importance of some of them has been quantified. Uterine malformation, which occurs in 0.1% of women, causes spontaneous abortions in as

[15] Fetal-maternal blood incompatibility usually involves Rhesus sensitization: if the mother is Rhesus negative (Rh–; recessive characteristic) and the father Rhesus positive (Rh+; dominant characteristic), the mother who has been immunized against Rh+ in a previous pregnancy possesses antibodies that will cross the placenta and cause hemolysis of the fetal red blood cells of group RH+.

many as 25% of cases (Simpson and Carson, 1992). Even when well-controlled, diabetes still accounts for nearly 7% of late fetal deaths (Thomas, 1989). Pregnancy toxemia (arterial hypertension and edemas caused by pregnancy) and eclampsia (convulsions due to raised blood pressure) remain major causes of fetal and perinatal mortality. Of infectious diseases present in developed countries, toxoplasmosis and cytomegalovirus infections are responsible for intrauterine deaths or embryopathies if the nonimmune mother becomes infected during pregnancy. The figures are not comparable, however, with those for infectious diseases in developing countries, where malaria and sexually transmitted diseases are the main pathologies implicated in fetal death.

Malaria. Malaria can lead to fetal death through several clinical mechanisms.

1. Bouts of fever in the pregnant woman can cause a fetal tachycardia that results in death.
2. The severe anemias malaria produces can lead to increased fetal morbidity due to iron deficiencies.
3. In the case of malaria, the parasite *Plasmodium falciparum* is able to cross the placenta and cause it to dysfunction, thus preventing adequate nutrition of the fetus. This species of plasmodia is especially serious in a pregnant woman because it is estimated that nearly half of malarial episodes with *P. falciparum* that occur during pregnancy result in fetal death (Gentilini, 1993).

Sexually transmitted diseases. Gonorrhea and syphilis have direct consequences for women's reproductive careers, the former causing sterility by tubal obstruction, the latter increasing significantly the risk of fetal death. The lack of standardized data makes estimations difficult. Nevertheless, according to the WHO, more than a third of pregnancies in women suffering from untreated syphilis end with a late fetal death (Rowe, 1994).

Fetal Characteristics

Sex. According to a survey conducted by the WHO in several countries on the determinants of perinatal mortality (WHO, 1978), the infant's sex has an uncertain role in late fetal mortality, because the results vary from country to country. However, an analysis of the data for England and Wales by Waldron (1983) shows that the excess male mortality existing before 1960 subsequently declined and the differential is now almost nonexistent. The hypothesis put forward to explain this decline is of a change in the causes of excess male mortality. Because boys usually have a higher birth weight than girls (for the same gestational age), their delivery can take longer, thus increasing the risks of fetal distress and death during birth. This would imply that boys have gained most from the improved obstetric techniques that are responsible for a reduction in mortality from hypoxia and trauma. Early and intermediate fetal mortality by sex is harder to study because of the technical difficulty of determining fetal sex in the case of spontaneous abortions, difficulties caused by the excessive interval between fetal death and expulsion, or by an X monosomy[16] that leads to identifying a fetus as male. On this last point, improved identification techniques now make it possible to detect real X monosomies or those due to prior loss of another X or Y chromosome.

Multiple-fetus pregnancies. Multiple-fetus pregnancies are associated with increased risks of prematurity and growth retardation and late fetal mortality. Kiely (1990) used data for New York City between 1978 and 1984 to analyze perinatal mortality in multiple births and found late fetal mortality rates three times higher than those in single births. Similarly, a study of late fetal mortality in multiple births using data for 1994 for England and Wales showed a risk four times higher than for simple births (Office for National Statistics, 1994). Placental structure may be a contributory factor in this excess fetal mortality, with a monochormial placenta for a multiple-fetus pregnancy increasing the risk of *in utero* death or malformations.

Fetal development. Intrauterine growth retardation is the most common complication of elevated maternal arterial blood pressure during pregnancy, with a frequency roughly double that of normotensive women (Boutroy, 1995). The prognosis is more unfavorable the earlier in gestation the retardation begins and usually leads to the birth being induced.

3. Social Factors

Fetal mortality is no exception to the social inequalities in mortality patterns observed at all ages. Here we consider only those variables that are generally considered to be the most relevant (and the easiest to observe) for identifying the factors associated with the greatest risks to fetal survival. It is now accepted that the characteristics should be primarily those of the mother, since she alone has access to prenatal care facilities. Lack of data means that the available results usually apply to late fetal mortality only.

Legitimacy. That children born out of wedlock and without paternal recognition (illegitimate) suffer high excess mortality has long been known. The increased

[16] It can be recalled that the pair of sex chromosomes is XX for females and XY for males.

risk is due partly to the mother's social isolation, combined with less favorable economic conditions. This risk persists despite changes in patterns of behavior that have led to a growing number of extra-marital births and greater social acceptance of the phenomenon. Statistics derived from civil registration data (1985–1990) show that with the exception of Denmark, Iceland, and Sweden, all the European countries have excess late fetal mortality for extra-marital births, the relative risks ranging between 1.3 and 1.9 (except for Spain where it is 2.7). At present the inequality is even greater for late fetal mortality than for infant mortality (Burban, 1996). At shorter gestation durations (8–20 weeks), unmarried women have risks of fetal mortality double those for married women (Carlson *et al*., 1999, Gourbin, 2002).

Immigrant or nonimmigrant status, culture and social group. Studies have reported higher than average late fetal mortality for women of North African origin in France, for women of Pakistani origin in England, and for black women in the United States (Saurel-Cubizolles *et al*., 1986; Macfarlane and Mugford, 1984; Buck *et al*., 1995). With the role of biologic factors rejected, the first explanation put forward is that of membership of socially disadvantaged strata. However, the fact that these differences are found to hold after socioeconomic level is controlled for has opened the way for other explanatory hypotheses. These are based on the individual's cultural environment and more specifically on the different health-related values and norms that are transmitted to children in the home. The family equips its members with a model of reactions that reflects their group of origin. Geographic origin and *race* (as used in the United States to distinguish individuals by their skin color or geographic origin) are thus indicators of varying precision and relevance for the existence of behavioral differences.

Maternal education. The mother's educational level is an indicator of her ability to adopt health-promoting behaviors during pregnancy and make effective use of prenatal care services (Rutter and Quine, 1990; Masuy-Stroobant, 1988). Most studies report an excess late fetal mortality of between 30% and 80% for infants of the least educated women (Cnattingius and Haglund, 1992; Notkola and Valkonen, 1989). On the other hand, the study by Taylor that analyzed fetal mortality from month 4 of pregnancy detected no significant association with the mother's educational level at any gestation duration. This result could be explained by the small number of fetal deaths in each of the groups examined. Moreover, the two studies conducted using Czech and Hungarian data, each including several thousand fetal deaths, found higher risks of second-trimester fetal mortality for women of low educational attainment (Carlson *et al*., 1999; Gourbin, 2002).

Social class. Social inequality in mortality has often been equated with membership of a particular social class (or social group or socio-occupational group). Commonly defined in terms of occupational groups (England, Finland), the social classes influence mortality through other characteristics they are assumed to reflect; the educational levels and incomes associated with different occupations thus serve as indirect indicators of the material environment and health behaviors of women and households (on this subject see the Chapter 54). In England, the first country to define social class categories along these lines,[17] analysis of differential mortality is based on the father's characteristics, thus excluding from analysis all children born out of wedlock. A difference in late fetal mortality between extreme social classes is observed regardless of period, country and gestation duration. For England the relative risks when comparing late fetal mortality in social class V (highest mortality) with social class I (lowest mortality) are approximately 1.7 for 1934 and 1.5 for 1989. In Finland, a study by Hemminki *et al*. (1980) between 1973 and 1975 reports a risk of fetal mortality before 28 weeks of gestation of 5.9 per 1,000 for the highest class, compared with 8.5 per 1,000 for the class composed of unskilled workers (representing a relative risk of 1.5). However, although membership of a disadvantaged social group appears to be a risk for the fetus, the concept's vagueness and multidimensional nature make it hard to elucidate the causal mechanisms responsible for this increase in risks: Does it reflect barriers (financial and educational) to getting access to care, inappropriate health practices, the mother's employment situation, or a combination of several of these different factors?

Mother's employment situation and occupation. The influence of mother's employment on fetal mortality cannot be judged without reference to social constructs of female employment and without knowing the characteristics of the work environment. Until the middle of the 20th century, surveys conducted on the mother's *employment situation* and her *status in the profession* usually reported adverse effects. Since then, as a consequence of far-reaching changes in the structure of women's occupations and in the characteristics of economically active women, the influence of female employment on pregnancy has changed. Thanks to

[17] Elaborated in 1911 by Stevenson and based on the father's occupation and the social position it conferred, the classes were formed by grouping the occupations according to their common characteristics and the differences in mortality rates of each occupation (Quoted by Macfarlane and Mugford, 1984).

legislative measures to protect mothers during pregnancy, women employees benefit from certain rights and are better informed about prenatal care facilities (Saurel-Cubizolles and Kaminski, 1986). Regarding the actual *occupation* of the mother, findings vary depending on the indicators used to collect the information: occupation narrowly defined, socio-occupational category, or industry sector. The imprecision of information about *occupation* (inaccurate reporting, work characteristics often unspecified) impedes reliable evaluation of differences in risks. Results from several studies, however, indicate risks of spontaneous abortion (defined in these studies as fetal death before week 28 of gestation) that are significantly higher when the difficulty of the work (posture, heavy lifting, stress) is taken into account (Bryant and Love, 1991; McDonald *et al*., 1988). The *socio-occupational categories* were constructed to examine the relationship between different occupations and social behavior, but in most studies occupations are combined at very broad levels of aggregation. The resulting categories are too large and too heterogeneous to give an accurate representation of the associated risks. The *industry sector* refers more specifically to the risks related to the occupational environment, but inadequate information about the products with which individuals come in contact means that it is seldom possible to infer the toxicity of particular products.

Toxic substances and occupational environment. Research into substances with adverse effects on reproductive health has progressed since the discovery (in 1961) of thalidomide's role in causing congenital malformations. It has been extended to cover substances present in the occupational environment, and studies conducted mainly in developed countries have increased understanding of the risks associated with an array of toxins. These risks include impaired fertility in the mother (or father), and teratogenic or mutagenic effects leading to congenital malformations and spontaneous abortions. Unfortunately it is extremely difficult to isolate the contribution of the toxic products and distinguish it from other risk factors (behavioral), and the difficulty of evaluating the dose-response relationship is a further obstacle to identifying a link between polluting agent and health status. Are the effects cumulative (when exposure precedes pregnancy, but causes an accumulation in body tissues or damage to the reproductive system with potentially detrimental consequences for a future pregnancy), or are they adverse only if exposure occurs actually during pregnancy? In the latter case, are the effects constant or do they depend on the stage of pregnancy at which exposure occurs (Selevan, 1991)? In studies of the relationship between occurrence of spontaneous abortion and exposure to toxic products in the occupational environment, the most commonly observed group are women working in hospitals, where two substances have been identified as potentially harmful to fetal survival—antineoplastic agents, responsible for spontaneous first-trimester abortions, and ethylene oxide (better known as ether), with a risk of fetal death increased by 130% for the first toxin and by 60% for the second[18] (Hatch and Marcus, 1991). Several studies have also found harmful effects for organic solvents, such as toluene, which become more marked with daily exposure. They are also thought to have a renal toxicity associated with a maternal hypertension that is detrimental to the fetus. High exposures, like those experienced by house painters, can have a toxic effect on the male reproductive system and mutagenetic effects during spermatogenesis (Baker, 1994). Results from a recent study have shown that women exposed to solvents during pregnancy have a higher risk of giving birth to a child with multiple congenital malformations or malformations of the digestive system. Other research has tried to determine the toxicity of specific products (cadmium, lead, pesticides) on fetal development, but the results are often contradictory, due to imprecise information on exposure, the difficulty of forming control groups, or variables that are too numerous to control. Finally, a toxic substance may go unidentified if its effects cause the death of the zygote before the pregnancy has been recognized.

Other risk situations, both common and rare, are also possible. The risks of genetic mutations in the case of fetal exposure to ionizing radiation are known: they are greatest at the start of pregnancy, during the period of embryogenesis when the organs are being formed. If a therapeutic radiograph has been performed before the pregnancy was recognized, it is advisable to evaluate the dose of radiation that has been administered. It should be noted, however, that these risks are very low in the case of routine radiography. In exceptional situations, involving massive irradiation of the fetus (Hiroshima atomic bomb, Chernobyl nuclear accident), increases in perinatal mortality (Kato, 1971; Korblein and Kuchenhoff, 1997) and of morbidity in the early years of life (cancers of thyroid, acute leukemias) have been documented (Leenhardt and Aurengo, 2000; Noshchenko *et al*., 2001). Although these effects appear undeniable in the case of exposure to high doses of radiation, determining a threshold dose beyond which radiation can be teratogenic or cause intrauterine death proves difficult.

[18] These products are thought not to act, however, on the germinal cells or to have cumulative effects.

4. Behavioral Factors

Discussion will be limited here to alcohol consumption and smoking, both of which are hard to evaluate. The factors that determine these behaviors are social and cultural (education, income, social class, religion), and psychological (acceptance of the pregnancy, in particular). Drinking and smoking during pregnancy are usually reported to have a probable effect on fetal development, manifested for the former by retarded psychomotor development termed the *fetal alcohol syndrome* and for the latter by low birth weight. The effects appear less clear regarding a relationship between smoking or drinking and fetal death, and the results are the subject of much debate, the main problem being over the choice of variables to be controlled (since these behaviors are often associated with other risk factors) (Simpson and Carson, 1992; Kline *et al*., 1989).

Smoking. Cigarette smoking is a known risk factor for retroplacental hematomas and *placenta praevia*[19] (English and Eskenazi, 1992) that can lead to fetal death. But the direct relationship that has sometimes been observed between smoking and fetal mortality either disappears when alcohol consumption is controlled for (Harlap and Shioni, 1980) or is found to be completely nonexistent (Axelsson *et al*., 1984).

Alcohol. A number of studies have reported a correlation between increased fetal loss and alcohol consumption during pregnancy (Harlap and Shiono, 1980), possibly due to formation of a retroplacental hematoma (Kaminski *et al*., 1976; Little *et al*., 1986). However, a threshold of consumption beyond which a harmful effect can be observed proves hard to define (Kline *et al*., 1989).

This short literature review has made it clear that most first-trimester fetal deaths result from chromosomal and genetic abnormalities that are incompatible with fetal development. The *age of the mother* at fertilization, associated with an aging of her reproductive capital, is partly responsible for these anomalies in women aged 35 and over, although malformations can also result from exposure to certain toxins. Except for these effects on embryo structure, the weight to be assigned to the various factors of fetal mortality after the first trimester of pregnancy is hard to evaluate. Positive or nonexistent associations are reported for the same factor by different studies, reflecting inadequate sample size or the number of variables examined. Furthermore, the distinction between broad categories of variables is purely theoretical, and some variables relate to dimensions that are as much social as biologic. Thus maternal age and parity are not merely indicators of a woman's physiologic maturity and reproductive capacity; they also reflect the fertility behavior of her social milieu. Further, behaviors relating to health and illness may have been formed during childhood in the family home. At what level does each of these variables operate? To what extent are drinking and smoking among the mechanisms whereby variables of social identification contribute to the risk of fetal death? Generally, women who have experienced a fetal death have multiple handicaps that are both socioeconomic (low educational and income levels, unfavorable material conditions) and demographic (teenage pregnancies, fertility patterns characterized by high parity and short birth intervals). Viewed thus, fetal death can be seen as the culmination of a complex process involving the interaction and combination of different factors whose individual effects are hard to isolate.

[19] Nicotine and carbon monoxide can cause a risk of hypoxia through a reduced placental blood flow. This risk may be compensated for by placental hypertrophy resulting in a larger insertion.

V. HEALTH-CARE PROVISION AND FETAL MORTALITY

The decline in late fetal mortality (the only one for which long series of statistics are available) has been steady in the industrialized world since the end of World War II. It results from a combination of demographic factors (completed fertility at second child, hence fewer high-parity births and at high ages) and above all from major advances in the fields of prevention, fetal surveillance, and obstetric and neonatal techniques, which have led to an extreme medicalization of childbirth and to increasingly common intervention practices, not always with due attention to the medium- and long-term consequences for parents and children.

1. Legislative Measures and Recommendations for Pregnancy Protection[20]

The introduction of social protection systems and preventive health care policies has been responsible for an improvement in working conditions during pregnancy and for more effective prenatal monitoring. Working conditions for pregnant women are regulated in most industrialized countries.[21] The most common

[20] This paragraph focuses specifically on the industrialized countries, due to the lack of information about conditions elsewhere.

[21] It can be noted however that at the end of the 1980s Denmark, Sweden, and Norway, countries where levels of late fetal mortality have been low for many years, had no mandatory protection of pregnant women at work. There were instead merely recommendations (Demont *et al*., 1990).

measures are exemption from night working, change of work place if it represents a physical (i.e., toxic or mechanical) hazard for the mother or unborn child, permission to make prenatal medical visits in work time. In the case of sickness (with a medical certificate), the woman is put on sick leave and is entitled to receive all or part of her salary for however long its duration (WHO, 1988).[22]

Medical supervision of pregnancy can be compulsory or based on government recommendations (Denmark) or be the responsibility of occupational organizations (The Netherlands) (Demont *et al*., 1990). The minimum statutory number of compulsory *prenatal visits* is usually smaller than that recommended by the medical profession, three or four instead of nine, the latter figure being close to the average number of visits actually carried out (usually 12). It is generally recommended that the first visit take place in the first trimester of pregnancy, and some countries offer a financial incentive to this end. In France, where such a payment exists, a survey conducted in 1987 found that 4.3% of mothers in the most disadvantaged social strata nonetheless did not receive this allowance because they had failed to respect the conditions (Blondel and Saurel-Cubizolles, 1991). The proportion was 21% in the United States in 1993 (National Center for Health Statistics, 1995), reaching 34% in the black population. Medical check-ups can be carried out by midwives, general practitioners or specialists. The trend is to a greater role for doctors, in particular obstetricians. A system of home visits to pregnant women by midwives or specialist nurses is also operated in some countries (WHO, 1988). Prenatal visits are also an opportunity for women to be informed and educated about pregnancy and birth, and to receive advice about their personal life styles and the preventive behaviors they should adopt.

2. Medical Visits

Prenatal visits vary in content, and from the many studies carried out to validate the usefulness of one or other test it is hard to define a minimum standard of care that would give the best chances for fetal survival and development. Among the tests offered, however, one constant element is ABO and Rh blood group determination. Administration of anti-D gamma globulins to Rh– women has made it possible to prevent fetomaternal sensitization and to reduce the number of perinatal deaths from this incompatibility (0.06 per 1,000 of perinatal deaths in the United Kingdom in 1983). Screening for infections (syphilis, toxoplasmosis, rubella) that carry a high risk for the health of the fetus is almost systematic in Western Europe and helps to avoid harmful consequences such as congenital defects, either through prevention (vaccination against rubella for women without protective immunity) or by therapeutic abortion. Prenatal visits during pregnancy are the occasion to monitor for the appearance of hypertension and of toxoplasmosis in nonimmune mothers. Depending on the country, serologic tests can also be performed for hepatitis B, listeria, cytomegalovirus or HIV[23] (human immune deficiency virus) on women at risk. Because the 1970s, ultrasound scans have become a routine procedure for monitoring fetal survival and development, as well as for accurately estimating gestational age,[24] observing certain fetal malformations and as an aid to amniocentesis. The recommended number of scans during pregnancy is two, and the maximum three, although the actual number is often higher. The practice of measuring α-fetoprotein between gestational weeks 16 and 20,[25] lowered levels of which are associated with an increased risk of trisomy 21, is gradually being extended beyond its indications (family history of trisomy 21, high maternal age). The risk scoring sometimes used to identify problem pregnancies seems of questionable effectiveness given its arbitrary construction (confusion of determinants and intermediate variables, arbitrary weights assigned to these variables) and poor predictive power[26] (Alexander and Keirse, 1989).

Hospitalization for birth,[27] by ensuring the presence of medical personnel and facilitating any necessary intervention, has contributed to reducing the risks of fetal death during birth. Fetal cardiac monitoring during labour is now routinely used in the great majority of births (more than 75% in the United States [Ruzek, 1991]) to detect fetal distress. Although the practice of fetal monitoring is associated with a significant fall in fetal mortality during labor[28] (Vintzileos

[22] These measures seldom extend to self-employed groups and female farm workers.

[23] Vertical transmission of HIV can be greatly reduced by appropriate therapy during pregnancy (European Collaborative Study, 2001).

[24] An ultrasound scan performed before 12 weeks allows the term to be defined to within 1 week and should therefore avoid inductions being carried out too early.

[25] An error in gestational age modifies the results, and 10% of abnormal results could be normal if an error of 1 week is made (Alexander *et al*., 1989).

[26] This technique is based largely on the woman's reproductive history and is almost completely inoperative for nulliparous subjects.

[27] In the industrialized countries over 98% of births now occur in a maternity clinic or hospital, the exception being The Netherlands, where more than 30% of births take place at home.

[28] The saving appears to be in the order of 1 per 1,000: Studies conducted in Denmark and at Boston produced the following results: the number of *intrapartum* deaths went from 1.7 per 1,000 in the period when births with monitoring were only 9% of cases, to 0.3 per 1,000 in the later period when they were 93% (Thoulon *et al*., 1995).

et al., 1995), it is also associated with a significant increase in instrumental intervention in birth (forceps or caesarian delivery)[29] (Thoulon *et al*., 1995; Parer, 1990). Intrauterine growth retardation is often accompanied by fetal distress before the forecast term for birth: A study conducted in France in 1993 of obstetricians and neonatologists found that in such cases, 14% of obstetricians and 36% of neonatologists would decide to perform a caesarian delivery before the end of gestational week 27 (Bréart, 1996), this decision is justified by the improved care techniques for very premature babies. The progress in techniques is even more striking in treatments for subfecundity. Use of *in vitro* fertilization and, in particular, ovulation induction is becoming increasingly widespread, often resulting in multiple pregnancies (triplets or more) with a high risk of fetal and infant mortality. Embryo reduction and selective abortions performed in the first trimester of multiple pregnancies can improve the survival chances of the remaining embryos (Boulot *et al*., 1995); but the reduction in risk of fetal death thus achieved is merely fictitious, since the embryos that have been suppressed are never included in the calculation of these risks.

3. Prenatal Diagnosis and Therapeutic Abortion

Like embryo reduction though on a larger scale, prenatal diagnosis and the therapeutic abortion it sometimes leads to has generated extensive discussion, less for its statistical repercussions on measurement of fetal mortality than for its ethical and psychological implications.

Increased knowledge in the field of genetics, identification of the association between maternal age and trisomy 21, systematic use of ultrasound to detect morphologic abnormalities, and improvements in prenatal diagnostic techniques, have all contributed to make screening a common practice. In the case of trisomy 21 (or Down syndrome), screening should be offered to mothers aged 32 or over in the United States, 35 in most European countries and in Canada, and 38 in France and Italy. Although it is now accepted that the risk of trisomy 21 increases exponentially after age 35, no policy of systematic screening has been introduced (Aymé, 1996; Reid, 1990). As well as ultrasound, screening uses the biologic detection techniques of fetal karyotyping, fetal blood testing, and serum markers in the mother. Those involving fetal karyotyping are the most commonly used, because not all malformations can be identified by fetal blood testing, while serum markers[30] are at present used only for a more effective targeting of the populations at highest risk for trisomy 21 and who are thus possible candidates for amniocentesis (Haddow *et al*., 1994).

Fetal karyotype testing. Amniocentesis (sampling of amniotic fluid) is the oldest technique for detecting the presence of congenital malformations. It is performed in the second trimester of pregnancy, usually around week 15 or 16. The results are only obtained after 3 to 4 weeks, the time to culture the cells removed, which can mean a fairly late abortion if this course of action is decided upon. The risk of fetal death associated with amniocentesis varies between 0.5% and 1%. A more recent technique is *chorionic villus sampling* (sampling of placental tissue guided by ultrasound), which became more widely used in Europe in the early 1980s. It has the advantage that it can be performed earlier in gestation (between 10 and 14 weeks) and only a few days are needed to obtain the diagnosis. The risk of fetal death appears similar to that for amniocentesis (Aymé, 1996).

Once diagnosis has been made, the decision to terminate the pregnancy will depend on the severity of the disorder. In the case of anencephalia, for example, therapeutic abortion is performed unless the parents refuse. The problem is more delicate (both for medical teams and for parents) in the case of malformations where the severity of the handicap cannot be evaluated. This type of problem arises with trisomy 21, which is frequently associated with other severe malformations (primarily cardiac) that seriously compromise the vital prognosis. However, what should the decision be when there is only this chromosomal abnormality or when prenatal diagnosis shows a limb is missing? This is an area in which the answers are far from codified and clinical practice varies when parents request termination of the pregnancy (Papiernik, 1990).[31]

Data collected by the fourteen regional EUROCAT (European Registration of Congenital Anomalies) registers in the countries where abortion is legal show that when malformations are diagnosed, the percentages of terminated pregnancies vary from over 85% in cases of anencephalia and Down syndrome to around 58% in cases of omphalocele (Table 32–4). It should be stressed that practices of prenatal diagnosis and therapeutic

[29] This increase can be avoided if the monitoring is combined with a microanalysis of the fetal blood.

[30] Three serum markers are used: chorionic gonadotrophin hormone, α-fetoprotein, and estriol. Combining the results obtained for these facilitates identification of indications for amniocentesis.

[31] Parents and medical teams have a role in this decision-making, but levels of financial support for, and social attitudes to, the handicapped are also factors for consideration.

TABLE 32–4 Number of Cases of Anomalies Observed, Frequency of Prenatal Diagnosis (PND) and Postdiagnosis Therapeutic Pregnancy Interruptions (TPIs) Recorded in 14 EUROCAT Registers

Anomaly	No. of cases		% with PND		% TPI		% TPI/PND	
	1980–1984	*1990–1994*	*1980–1984*	*1990–1994*	*1980–1984*	*1990–1994*	*1980–1984*	*1990–1994*
Anencephaly	564	449	80	97	53	84	67	87
Spina Bifida	700	641	33	64	19	45	58	69
Bilateral renal agenesis	164	250	29	79	15	52	51	66
Omphalocele	201	334	29	73	23	43	77	58
Down syndrome	809	2109	14	47	13	41	93	88

EUROCAT, European Registration of Congenital Anomalies.
Source: EUROCAT, 1997.

abortion do not differ between the regions in which the registers are operated and the rest of the country. The prevalence of malformations appears greater, however, because data collection is more complete.

Given that maternal age is a risk factor for trisomy 21 and probably also for other chromosomal malformations, can the trend to later entry into motherhood be expected to produce an increase in these malformations? A simulation performed by Cornel *et al*. (1993) based on the situation prevailing in 1992 in The Netherlands found that if coverage for trisomy 21 screening was unchanged (around 50% for women aged 36 and over), the prevalence of this pathology among live births would rise from 1.4 per 1,000 to 1.8 per 1,000 in 2001.

A very general definition of therapeutic abortion is the voluntary interruption of a pregnancy on serious medical grounds not later than the date at which civil registration declaration of fetal death becomes compulsory (usually 28 weeks of gestation). Two major exceptions to this rule are France and Finland. In France, the 1975 abortion law allows for therapeutic interruption of pregnancy at any gestation length, on condition that the pathology involved is sufficiently serious and untreatable. This decision (taken by at least two doctors) concerns abnormalities that are thought to be incompatible with the infant's survival, but nearly 20% of the therapeutic abortions recorded in the French EUROCAT[32] registers are performed after the legal limit for declaration of 28 weeks of gestation. This is a large proportion,[33] and it may affect the statistics for late fetal mortality, because these deaths are not systematically reported for civil registration. If none of these deaths were reported, adding them to those declared to civil registration for the three regions covered by the French registers in the period 1980 to 1992, causes late fetal mortality to rise by 0.7%. This increase is not negligible given the low levels of late fetal mortality in developed countries (and corresponds to a relative increase of 8%). In Finland, all fetal deaths must be reported after week 22 of gestation although therapeutic abortion is allowed up to week 24. Disregarding the limit for civil registration declaration, estimation was made in Finland of the effect on perinatal mortality of therapeutic abortions performed after week 16 of gestation. Working from the hypothesis that of the therapeutic abortions, only one third of diagnoses would actually have resulted in perinatal death, Gissler *et al*. (1994) estimate the increase in the level of perinatal mortality at 1%, had these therapeutic abortions not been performed.

Large disparities are observed between countries and between social groups in the same country, as reflected in mortality indices that are consistently higher for the most disadvantaged categories. Several elements can be advanced as partial explanations for these inequalities:

Advice on personal health practices during pregnancy, distributed in public health campaigns and in prenatal visits, is more easily followed by mothers whose educational level enables them to understand the benefit of these preventive behaviors. The policies of positive discrimination sometimes advocated (Ruzek, 1991) are of unproven effectiveness (Blondel and Bréart, 1992);

Although medical costs are reimbursed as long as the woman follows the prenatal care program of her country's social protection system, any additional examinations (to reassure the mother-to-be or

[32] All fetal deaths after 20 weeks of gestation and all therapeutic abortions, regardless of the gestational duration at which they are carried out, are recorded.

[33] In the United States, 0.01% of abortions are performed after week 24 of gestation (20 weeks' minimum gestation being the legal limit in most States for declaration of a fetal death). Nearly half involve teenage pregnancies; the rest are for fetuses suffering from serious congenital malformation.

anticipate possible morbid conditions for the fetus) are accessible only to women in favorable economic conditions. In France, for example, very sharp sociocultural inequalities are observed in use of prenatal diagnosis for trisomy 21 among women under the age at which it is recommended and with no other risk factors and who therefore have to pay for the test. Similarly, despite free provision and systematic promotion of health care facilities after age 38, social inequalities in access to prenatal diagnosis remain (Moatti *et al*., 1990). The problem is thus one of information (as when the woman accepts without questioning a doctor's decision that a pregnancy interruption is unjustified) and understanding (in her perception of the handicap, for example).

The increasing complexity of medical techniques accentuates inequalities in use of health care facilities. The recent trends observed in treatment for subfecundity, however, could distort perception of these inequalities if it was confirmed that better educated women make greatest use of these reproductive techniques which usually carry high risks of fetal mortality (Bréart, 1996).

VI. INDUCED ABORTION AND FETAL DEATH

Induced abortion (see Chapter 35) can be considered either as a potential risk factor for fetal death[34] or, on the contrary, as an event that tends to reduce intrauterine mortality by removing a certain number of high-risk pregnancies that otherwise would have ended with fetal death. In the mid 1990s, an estimated 46 million induced abortions were performed in the world annually, of which more than 40% are thought to have been clandestine (Alan Guttmacher Institute, 1999). Induced abortion is certainly one of the oldest forms of birth control, one practiced in all societies albeit to varying degrees before modern methods of contraception made possible an effective regulation of fertility.

1. Induced Abortion as a Risk Factor for Fetal Mortality

A prime consideration when assessing the risks is whether abortion is legal. In countries where abortion is legal, the risk is associated chiefly with the methods used. Abortion is currently legal in more than 30 countries (available on request or for a wide range of medical and social reasons). Most of the developed countries are in this category, as are China, India, South Africa, Tunisia, and Turkey. Induced abortion is usually carried out in the first trimester of pregnancy, but whereas in most of the developed countries it is viewed as a backup to failed contraception, in Central and Eastern Europe it has been used since legalization as an instrument of family planning (Blayo, 1993). In countries where abortion is still subject to numerous legal restrictions, women use unsafe practices that can be a source of infections leading to maternal death, repeated fetal loss in subsequent pregnancies, or secondary sterility. In 42 countries situated mainly in sub-Saharan Africa and the Middle East, abortion is prohibited unless there is a threat to the woman's life. In the remaining countries, abortion is allowed on a variety of grounds, both medical (physical or mental health of the mother) and legal (in rape or incest cases) (United Nations, 1992, 1995).

The countries where abortion is prohibited are also those where information on fetal deaths, even at advanced gestational ages, is nonexistent. Consequently it is not possible to estimate the effect of clandestine abortion practices on intrauterine mortality. The most commonly proposed evaluations are for maternal mortality: WHO estimates the proportion of maternal deaths that result from this kind of abortion at 13% (WHO, 1989). Proportions as high as 31% have been reported for Matlab in Bangladesh and over 50% in Ethiopia (United Nations, 1992), figures that may imply equally high levels of uterine damage with adverse implications for subsequent reproductive careers.

The laws on pregnancy termination allow for it to be carried out in the best medical conditions. The Central and Eastern European countries liberalized abortion at the end of the 1950s,[35] and were followed nearly 15 years later by most of the industrialized countries.[36] Usually performed in hospitals or clinics, abortion has become safer and physically less traumatic, thanks to the development of suction curettage in particular. For early terminations (before week 7 of amenorrhea) a choice is now available (e.g., Belgium, England and Wales, France, Sweden) between a med-

[34] This term, it can be recalled, includes all fetal deaths (i.e., spontaneous abortions [fetal deaths occurring before the limit for compulsory civil registration declaration] + stillbirths [fetal deaths declared for civil registration]).

[35] Abortion was legalized in the USSR in 1921. Prohibited in 1936, it was made legal again in 1955 (Blayo, 1993).

[36] The usual reasons for which it is authorized are as follows: When the mother's life or her mental or physical health are threatened, on eugenic grounds (genetic diseases transmitted by the parents, or malformation), legal (rape, incest), and because of economic and social circumstances.

ication-based method (RU 486 + prostaglandins) or surgery, which are equally effective. The possible consequences for subsequent reproductive careers of induced abortions by vacuum aspiration or curettage[37] have been studied and found to be the same as those after normal delivery. A slightly higher proportion of extrauterine pregnancies (not significant) was observed, however, among women with a history of pregnancy termination and is perhaps indicative of postoperative infection (Buekens and Kaminski, 1996). But higher risks of prematurity, low birth weight and, to a lesser extent, fetal death have been previously reported in the Central and Eastern European countries, as a result of the excessive cervical dilation caused by the abortion methods in use until recently in these countries.[38]

2. Induced Abortion and Reduced Fetal Mortality

Conversely, induced abortion can serve to avoid pregnancies with high risk for fetal death when it is performed on women whose biologic and social characteristics are unfavorable for fetal survival. These variables can only be identified in the countries where induced abortion is legalized and where its registration is accompanied by collection of various demographic and socioeconomic information (some European countries and the United States).

The characteristics of women who have abortions differ sharply between the countries of Central Europe and the others. In the former, abortion was used to limit family size once certain parity has been reached, whereas in the latter most women who have abortions are young, nulliparous and unmarried (Blayo, 1991).

Woman's age. In the nineties, women younger than 25 years accounted for the highest frequency of abortions in Sweden and Finland (44%) but especially in Norway, England and Wales, and the United States (respectively 52%, 54%, and 60%). This contrasts with the situation in the Czech Republic, Hungary, but also in France, where abortion is a more generalized phenomenon that concerns all age groups (around 20% by age group), with a proportion of around 20% in 2000 for women after 35 years (HCSO, 2002). As regards the frequency of induced abortions after age 35, in Hungary, and until 1988, it was more or less identical to that observed before age 25, and thus marks both beginning and end of the reproductive calendar.

Parity. Differences in parity are more sharply drawn, since 63% of the abortions carried out in Hungary in 2000 were on women with at least two previous pregnancies. In Norway, Finland, and Sweden, on the other hand, induced abortion is more frequent among nulliparous women (roughly 50%). This pattern is even more marked in the United States and in England and Wales, where nearly 60% of women having abortions are nulliparous.[39] France occupies an intermediate position, closer to that of Hungary, since the proportion of women who have had two previous pregnancies account for over 44% in 1995 (INED, 1999). It should be remembered, however, that the differences observed between countries may also reflect disparity in the classificatory criteria used for this characteristic.[40]

Marital status. In the 1990s, the same contrasts are observed when we consider the mother's marital status at the time of the abortion—England and Wales and the United States are at one extreme, the Czech Republic is at the other. In the former countries, induced abortion involves overwhelmingly unmarried women (74% for England and Wales and 67% for the United States), whereas three fourths of women who abort in the Czech Republic are married, regardless of the year considered. Hungary shared the same distribution pattern until the end of the 1980s, when the proportion of married women fell sharply, to stand at 45% in 2000. The same change is also observed for the two previous characteristics and is probably the result of improved access to chemical and mechanical means of contraception; its effect is to bring Hungary gradually into line with the countries of Western and Northern Europe. In Finland and Norway, the frequency of induced abortion among married women has remained stable at around 30%. In France, the trend to a greater frequency of induced abortion among married women has slowly been reversed in the years since abortion was legalized and unmarried women now account for the larger share.[41]

[37] For pregnancy interruption using RU486 plus prostaglandins, it is too early for any possible consequences to be assessed.

[38] The total induced abortion rate remains extremely high in these countries. It still stood at over one abortion per woman in the countries of Central Europe in 1990 and reached record levels of four per woman in Russia and six in Romania (Blayo, 1993).

[39] Of women who abort, 21% in the United States and 26% in England and Wales are of at least parity two. But among women who do not want any more children sterilization is much more common in England and Wales and the United States, than in the other countries considered.

[40] Number of previous pregnancies in France, number of previous births (or previous deliveries) in England and Wales, Finland, Norway, Czech Republic, and Sweden, number of previous live births in the United States, and number of infants still alive in Hungary.

[41] The quality of registration, particularly in the early days of legalized abortion, can be questioned, however, because abortions performed in private clinics may have gone unrecorded.

The legal status of abortion, ranging from total prohibition to full legalization, has a series of consequences for the woman's future reproductive health that extend from secondary sterility to no observable effect. In countries where abortion has been legalized, women using pregnancy termination usually present characteristics associated with an unfavorable pregnancy outcome—women aged over 35 in the countries of Central Europe, young unmarried women in the other countries. Consequently, when it is performed in hygienic settings, using up-to-date techniques that minimize the risk of infection and traumatic lesions, pregnancy termination may indeed reduce the number of spontaneous fetal deaths since it is more frequent in the groups where exposure to this risk is highest.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The relatively recent growth of interest in fetal mortality, linked to new models of family formation (in which fertility regulation achieves a majority of wanted births) and development of increasingly effective medical techniques in the areas of reproduction, fetal care, and birth, has exposed the gaps in our knowledge as regards both levels and determinants of fetal mortality. Data on fetal mortality before 28 weeks are seriously deficient in the industrialized countries and is nonexistent elsewhere. Identification of pregnancies and fetal deaths in the subclinical phase (before any delay in menstrual period) requires surveillance protocols that are too restrictive to be implemented on large populations. What could be envisaged, however, is the creation of a system for collecting information as of the first prenatal visit (a pregnancy *log book* containing the basic biologic and social characteristics of the mother, fetal death record). In the countries where an initial obstetric visit is compulsory in the first trimester of pregnancy, a system on these lines would contribute to a more accurate evaluation of levels of intrauterine mortality after gestational week 12 and identification of the risk factors for fetal death during pregnancy, although remember that women in the most disadvantaged social strata are usually the last to make use of health care facilities. Improving information on fetal mortality is more problematic in developing countries. As the quality of civil registration systems gets better, levels of late fetal mortality will be known; for the moment, however, both for this and for early and intermediate fetal mortality, surveys are perhaps the only means to improve our knowledge of the phenomenon. However, another problem immediately arises—these are typically the countries where pregnancy interruption is prohibited and where, in the absence of means of contraception, women have recourse to clandestine abortions with a high risk for their subsequent reproductive health. The consequences of such practices for reproductive health, unfortunately unmeasured, should be made the object of in-depth studies. A modest first step in this direction would be to conduct surveys on women who are hospitalized for secondary complications from a clandestine abortion.

Acknowledgments

Translated from French by Godfrey I. Rogers. The research for the preparation of this chapter was funded by the *Fonds de développement scientifique de l'Université catholique de Louvain* and by *Project INTAS* 93-1617.

References

Alan Guttmacher Institute, 1999. *Sharing responsibility: women, society and abortion worldwide*, New York, Alan Guttmacher Institute, 57 p.

Alexander Sophie and Keirse Marc J. N. C., 1989. Formal risk scoring during pregnancy, *in*: Iain Chalmers, Murray Enkin and Marc Keirse (eds.), *Effective care in pregnancy and childbirth, vol. 1: pregnancy*, p. 347–365. Oxford, Oxford University Press, 791 p.

Alexander Sophie, Stanwell-Smith Rosalind, Buekens Pierre and Keirse Marc J. N.C., 1989. Biochemical assesment of fetal well-being, *in*: Iain Chalmers, Murray Enkin and Marc Keirse (eds.), *Effective care in pregnancy and childbirth, vol. 1: pregnancy*, p. 456–476. Oxford, Oxford University Press, 791 p.

Axelsson Gösta, 1984. Selection bias in studies of spontaneous abortion among occupational groups, *Journal of Occupational Medicine*, vol. 26, p. 525–528.

Aymé Ségolène, 1996. Dépistage prénatal et inégalités sociales, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui*, p. 521–538. Louvain-la-Neuve, Academia-Btuylant, L'harmattan, 560 p. (Chaire Quetelet 1994, Louvain-la-Neuve, 12–14 September 1994).

Baker Edward L., 1994. A review of recent research on health effects of human occupational exposure to organic solvents, *Journal of Occupational Medicine*, vol. 36(10), p. 1079–1092.

Blayo Chantal, 1991. Les modes de prévention des naissances en Europe de l'Est. *Population*, vol. 46(3), p. 527–546.

Blayo Chantal, 1993. L'accès à l'avortement: contexte légal, pratique et fréquence de recours. *Entre nous. La revue européenne de planification familiale*, no. 22–23, p. 6–9.

Blondel Béatrice and Bréart Gérard, 1992. Home visits for pregnancy complications and management of antenatal care: an overview of three randomized controlled trials. *British Journal of Obstetrics and Gynaecology*, vol. 99, p. 283–286.

Blondel Béatrice and Saurel-Cubizolles Marie-Joséphe, 1991. An indicator of adverse pregnancy outcome in France: not receiving maternity benefits, *Journal of Epidemiology and Community Health*, vol. 45, p. 211–215.

Boué Joëlle and Léridon Henri, 1971. La mortalité intra-utérine d'origine chromosomique, *Population*, vol. 26(1), p. 113–138.

Boulot P., Dechaud H., Giacalone P. L., Pereira M., Hedon B., Montoya F. and Molenat F., 1995. Réduction embryonnaire et grossesses triples, *in*: Alain Treisser and Bernard Blanc (eds.), *25es Journées nationales de la Société française de médecine périnatale, Paris, 1995*, p. 207–221. Paris, Arnette Blackwell, 325 p.

Boutroy M. J., 1995. L'enfant de mère hypertendue, *in*: Alain Treisser and Bernard Blanc (eds.), *25es Journées nationales de la Société française de médecine périnatale, Paris, 1995*, p. 49–58. Paris, Arnette Blackwell, 32 p.

Bréart Gérard, 1996. Prématurité spontanée, prématurité induite, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui*, p. 539–550. Louvain-la-Neuve, Academia-Btuylant, L'harmattan, 560 p. (Chaire Quetelet 1994, Louvain-la-Neuve, 12–14 September 1994).

Bross Dean S. and Shapiro Sam, 1982. Direct and indirect association of five factors with infant mortality. *American Journal of Epidemiology*, vol. 115(1), p. 78–91.

Bryant Heather E. and Love Edgar J., 1991. Effect of employment and its correlates on spontaneous abortion risk, *Social Science and Medicine*, vol. 33(7), p. 795–800.

Buck Germaine M., Shelton James A., Mahoney Martin C., Michalek Arthur M. and Powell Etta J., 1995. Racial variation in spontaneous fetal deaths at 20 weeks or older in upstate New York, 1980–86, *Public Health Reports*, vol. 110, p. 587–592.

Buekens Pierre and Kaminski Monique, 1996. Méthodes contraceptives et avortement, *in*: Marie-Josèphe Saurel-Cubizolles and Béatrice Blondel (eds.), *Santé des femmes*, p. 81–100. Paris, Flammarion, XVIII + 386 p.

Buitendijk Simone, Zeitlin Jennifer, Cuttini Marina, Langhoff-Roos and Bottu Jean, 2003. Indicators of fetal and infant health outcomes. *European Journal of Obstetrics, Gynecology and Reproductive Biology*, vol. 111, p. S66–S77.

Burban Anne, 1996. Les naissances hors mariage en Europe, une inégalité sociale persistante en matière de mortalité infantile malgré l'évolution des législations et des mentalités, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui*, p. 215–242. Louvain-la-Neuve, Academia-Btuylant, L'Harmattan, 560 p. (Chaire Quetelet 1994, Louvain-la-Neuve, 12–14 September 1994).

Carlson Elwood, Hoem Jan. M. and Rychtarikova Jitka, 1999. Trajectories of fetal loss in the Czech Republic, *Demography*, vol. 36(3), p. 327–337.

Casterline John B., 1989. Collecting data on pregnancy loss: a review of evidence from the World Fertility Survey, *Studies in Family Planning*, vol. 20(2), p. 81–95.

Chung Chin S., Smith Roy G., Steinhoff Patricia G. and Mi Ming-Pi, 1982. Induced abortion and spontaneous fetal loss in subsequent pregnancies, *American Journal of Public Health*, vol. 72(6), p. 548–554.

Cnattingius Sven and Haglund Bengdt, 1992. Socio-economic factors and feto-infant mortality, *Scandinavian Journal of Social Medicine*, vol. 20(1), p. 11–13.

Cornel Martina C., Breed Anton S. P. M., Beekhuis Johan R., Mte Meerman Gerard J. and ten Kate Leo P., 1993. Down syndrome: effects of demographic factors and prenatal diagnosis on the future livebirth prevalence, *Human Genetics*, vol. 92, p. 163–168.

Daly C., Heady J. A. and Morris J. N., 1955. Social and biological factors in infant mortality. III. The effect of mother's age and parity on social-class differences in infant mortality. *The Lancet*, 26 February, p. 445–449.

Demont F., Soutoul J. H., Pierre F., Decaux A., Grandbastien B. and Gicquel M., 1990. Exploration de la législation périnatale en Europe, *in*: Francis Puech and Bernard Blanc (eds.), *20es Journées nationales de la Société française de médecine périnatale, Lyon, 1990*, p. 85–96. Paris, Arnette, 132 p.

Dupin Henry and Raimbault Anne-Marie, 1978. *Les troubles nutritionnels chez la mère et l'enfant. Epidémiologie et prévention*. Bar-le-Duc, éditions Saint Paul, 128 p.

Edmonds Keith, Lindsay Kevin S., Miller John F., Williamson Elsbeth and Wood Peter J., 1982. Early embryonic mortality in women. *Fertility and Sterility*, vol. 38(4), p. 447–453.

Empire Français, 1805. De la paternité à la filiation, *in: Code civil*, titre 7, p. 315–318.

Erhardt Carl L., 1963. Pregnancy losses in New York City, 1960, *American Journal of Public Health*, vol. 53(8), p. 1337–1352.

EUROCAT, 1997. *Report 7–15 Years of surveillance of congenital anomalies in Europe 1980–1994*. Brussels, European Registration of Congenital Anomalies Working Group/Institute of Hygiene and Epidemiology, 2 volumes 182 p. + 357 p.

European Collaborative Study, 2001. HIV-infected pregnant women and vertical transmission in Europe since 1986. *AIDS*, vol. 15, p. 761–770.

Frank Peter I., McNamee Rosanne, Hannaford Philip C., Kay Clifford R. and Hirsch Sybil, 1991. The effect of induced abortion on subsequent pregnancy outcome, *British Journal of Obstetrics and Gynaecology*, vol. 98, p. 1015–1024.

French Fern E. and Bierman Jessie M., 1962. Probabilities of fetal mortality, *Public Health Reports*, vol. 77(10), p. 835–847.

Gentilini Marc, 1993. *Médecine tropicale*. Paris, Médecine-Sciences, Flammarion, 928 p.

Gissler Mika, Ollila Eeva, Teperi Juha and Hemminki Elina, 1994. Impact of induced abortions and statistical definitions on perinatal mortality figures. *Paediatric and Perinatal Epidemiology*, vol. 8(4), p. 391–400.

Goldhaber Marilyn K. and Fireman Bruce H., 1991. The fetal life table revisited: spontaneous abortion rates in three Kaiser permanent cohorts, *Epidemiology*, vol. 2(1), p. 33–39.

Gourbin Catherine and Masuy-Stroobant Godelieve, 1995. Registration of data: are livebirths and stillbirths comparable all over Europe, *Bulletin of the World Health Organization*, vol. 73(4), p. 449–460.

Gourbin Catherine, 1995. Critères d'enregistrement des naissances vivantes et des mort-nés en Europe, *in*: Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démo-sociales en Europe*, p. 219–242. Louvain-la-Neuve, Academia, Paris, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Gourbin Catherine, 2002. *Les déterminants biologiques et sociaux de la mortalité fœtale. Le cas de la Hongrie*. Louvain-la-Neuve, Academia, Paris, L'Harmattan, 262 p.

Gray Ronald, 1979. Biological factors other than nutrition and lactation which may influence natural fertility: a review, *in*: Henri Léridon and Jane Menken (eds.), *Patterns and determinants of natural fertility. Proceedings of a seminar on natural fertility*, p. 217–251. Liège, IUSSP, Ordina éditions, 566 p.

Gribble James N., 1993. Birth intervals, gestational age, and low birth weight: are the relations confounded? *Population Studies*, vol. 47(2), p. 133–146.

Haddow James E., Palomaki Glenn E., Knight George J., Cunningham George C., Lustig Linda S. and Boyd Patricia A., 1994. Reducing the need for amniocentesis in women 35 years of age or older with serum markers for screening, *New England Journal of Medicine*, vol. 330(16), p. 1114–1118.

Harlap Susan and Shiono Patricia H., 1980. Alcohol, smoking and incidence of spontaneous abortions in the first and second trimester, *The Lancet*, July 26, p. 173–176.

Hatch Maureen and Marcus Michele, 1991. Occupational exposures and reproduction, *in*: Michele Kiely (ed.), *Reproductive and perinatal epidemiology*, p. 131–142. Boston, CRC Press, 534 p.

Heady J. A., Daly C. and Morris J. N., 1955. Social and biological factors in infant mortality, II: variation of mortality with mother's age and parity. *The Lancet*, 19 February, p. 395–397.

Hemminki Kari, Kyyrönen Pentti, Niemi Marja-Liisa, Koskinen Kari, Sallmén Markku and Vainio Harri, 1983. Spontaneous abortions in an industrialized community in Finland, *American Journal of Public Health*, vol. 73(1), p. 32–37.

Hemminki Kari, Niemi Marja-Liisa, Vainio Harri, Saloniemi I. and Hemminki Elena, 1980. Spontaneous abortions by occupation and social class in Finland, *International Journal of Epidemiology*, vol. 9(2), p. 149–153.

Henry Louis, 1964. Mortalité intra-utérine et fécondabilité, *Population*, vol. 19(3), p. 899–940.

Hertig Arthur T., Rock John, Adams E. C. and Menkin Miriam C., 1959. 34 fertilized human ova, good, bad and indifferent recovered from 210 women of known fertility, *Pediatrics*, vol. 23(1), p. 202–211.

Hungarian Central Statistical Office, 2002. *Demographic Yearbook 2001*. Budapest, HCSO, 495 p.

Institut National d'Études Démographiques, 1999. *Statistiques de l'avortement en France. Annuaire 1995*. Paris, INED, 141 p.

Kallan Jeffrey E., 1992. Effects of interpregnancy intervals on preterm birth, intrauterine growth retardation, and fetal loss, *Social Biology*, vol. 39(3–4), p. 231–245.

Kaminski Monique, Rumeau-Rouquette Claude and Schwartz Daniel, 1976. Consommation d'alcool chez les femmes enceintes et issue de la grossesse, *Revue d'Epidémiologie et de Santé Publique*, vol. 24, p. 27–40.

Kato H., 1971. Mortality in children exposed to the A-bombs while in utero, 1945–1959, *American Journal of Epidemiology*, vol. 93(6), p. 435–442.

Kiely John, 1990. The epidemiology of perinatal mortality in multiple births, *Bulletin of the New York Academy of Medicine*, vol. 66(6), p. 618–637.

Kline Jennie, Stein Zena and Susser Mervyn, 1989. *Conception to Birth. Epidemiology of Prenatal Development*. New York, Oxford University Press, 433 p. (Monographs in Epidemiology and Biostatistics, vol. 14).

Kline Jennie, Stein Zena, Susser Mervyn and Warburton Dorothy, 1978. Induced abortion and spontaneous abortion: no connection? *American Journal of Epidemiology*, vol. 107(4), p. 290–298.

Körblein A. and Kuchenhoff H., 1997. Perinatal mortality in Germany following the Chernobyl accident, *Radiation and Environmental Biophysics*, vol. 36(1), p. 3–7.

Leenhardt Laurence and Aurengo André, 2000. Post-Chernobyl thyroid carcinoma in children, *Bailliere's Best Practice and Research. Clinical Endocrinology and Metabolism*, vol. 14(4), p. 667–677.

Leridon Henri, 1977. *Human fertility*: *the basic components* Chicago, University of Chicago Press, 208 p.

Leridon Henri, 1976. Facts and artefacts in the study of intrauterine mortality: a reconsideration from pregnancies histories, *Population Studies*, vol. 30(2), p. 319–335.

Léridon Henri, 1987. La mortalité fœtale spontanée: rôle de l'âge de la mère, de la parité et des antécédents d'avortement, *Journal de biologie, obstétrique et biologie de la reproduction*, no. 16, p. 425–431.

Lindbohm Marja-Liisa, Hemminki Kari and Kyyrönen Pentti, 1984. Parental occupational exposure and spontaneous abortions in Finland. *American Journal of Epidemilogy*, vol. 120(3), p. 370–378.

Little Ruth E., Asker Rafida L., Sampson Paul D. and Renwick J. H., 1986. Fetal growth and moderate drinking in early pregnancy. *American Journal of Epidemiology*, vol. 123(2), p. 270–278.

Macfarlane Alison and Mugford Miranda, 1984. *Birth counts: statistics of pregnancy and childbirth.* London, HMSO, 2 vol., 345 p. + 310 p.

Masuy-Stroobant Godelieve, 1988. *Inégalités sociales et santé de l'enfant. Une enquête en Hainaut sur le comportement préventif des mères*. Louvain-la-Neuve, CIACO, 283 p.

Masuy-Stroobant Godelieve, 1996. Santé et mortalité infantile en Europe. Victoires d'hier et enjeux de demain, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui*, p. 337–366. Louvain-la-Neuve, Academia-Btuylant, L'harmattan, 560 p. (Chaire Quetelet 1994, Louvain-la-Neuve, 12–14 September 1994).

McDonald Alison D., McDonald J. C., Armstrong Ben, Cherry Nicola M., Côté Rosane, Lavoie J., Nolin A. D. and Robert D., 1988. Fetal death and work in pregnancy. *British Journal of Industrial Medicine*, vol. 45, p. 148–157.

Moatti Jean-Pierre, Le Gales C., Julian C., Lanoe J. L. and Aymé Ségolène, 1990. Analyse coût-bénéfice du diagnostic prénatal des anomalies chromosomiques par amniocentèse, *Revue d'épidémiologie et de santé publique*, vol. 38(4), p. 309–321.

Modvig Jens, Schmidt Lone and Damsgaard Mogens T., 1990. Measurement of total risk of spontaneous abortion: the virtue of conditional risk estimation, *American Journal of Epidemiology*, vol. 132(6), p. 1021–1038.

Morris J. N. and Heady J. A., 1955. Social and biological factors in infant mortality, I: objects and methods, *The Lancet*, 12 February, p. 343–349.

NCHS, 1995. *Monthly vital statistics reports (Centers for Disease Control and Prevention)*, vol. 44(3), Supplement.

Noshchenko Andrey, Moysich Kirsten B., Bondar Alexandra, Zamostyan Pavlo, Drosdova Vera and Michalek Arthur, 2001. Patterns of acute leukaemia occurrence among children in the Chernobyl region, *International Journal of Epidemiology*, vol. 30(1), p. 125–129.

Notkola Veijo and Valkonen Tapani, 1989. Socio-economic differences in stillbirths and infant mortality in Finland 1976–82, 14 p. *in: Yearbook of Population Research in Finland*, vol. 27, 1989. Helsinki, The Population Research Institute, 89 p.

Nzita Kikhela, 1986. *La mortalité périnatale à Kinshasa: niveaux, déterminants et familles à risque*. Louvain-la-Neuve, Institut de Démographie, Université Catholique de Louvain, 290 p. (doctoral thesis in demography).

ONS, 1995. *Mortality statistics. Childhood, infant and perinatal.* London, Office for National Statistics, 155 p. (Series DH3 no. 28).

Papiernik Émile, 1990. Diagnostic anténatal et problèmes éthiques, *Chirurgie*, vol. 116, p. 541–545.

Parer Julian T. and Livingston E. G., 1990. What is fetal distress, *American Journal of Obstetrics and Gynecology*, vol. 162, p. 1421–1427.

Peller Sigismund, 1948. Mortality, past and future. *Population Studies*, vol. 1(4), p. 405–456.

Pinnelli Antonella, 1984. Facteurs différentiels de la mortalité fœto-infantile en Italie, *in*: UN/WHO/CICRED (ed), *Socio-economic differential mortality in industrialized societies*, vol. 3, p. 227–243. Paris, CICRED, 361 p.

Reid Margaret, 1990. Pre-natal diagnosis and screening: a review, *in*: Jo Garcia, Robert Kilpatrick and Martin Richard (eds.), *The politics of maternity care. Services for childbearing women in twentieth-century Britain*, p. 300–323. Oxford, Clarendon Press, 346 p.

Risch Harvey A., Weiss Noel S., Clarke Aileen and Miller Anthony B., 1988. Risk factors for spontaneous abortion and its recurrence, *American Journal of Epidemiology*, vol. 128(2), p. 420–430.

Rowe Patrick J., 1994. Sexually transmitted diseases: a challenge to reproductive health, *in*: Jitendra Khanna Paul F. A. van Look and P. David Griffin (eds.), *Challenges in reproductive health research: biennial report, 1992–1993*, Geneva, World Health Organization, 202 p. (Special Programme of Research, Development and Research Training in Human Reproduction).

Rychtarikova Jitka, Gourbin Catherine and Wunsch Guillaume, 2004. Paternal age and child death: the stillbirth case, *European Journal of Population*, vol. 20(1), p. 23–33.

Rutter Derek R. and Quine Lyn, 1990. Inequalities in pregnancy outcome: a review of psychosocial and behavioural mediators, *Social Science and Medicine*, vol. 30(5), p. 553–568.

Ruzek Sheryl, 1991. Women's reproductive rights: the impact of technology, *in*: Judith Rodin and Aila Collins (eds.), *Women new reproductive technologies: medical, psychosocial, legal and ethical dilemmas*, p. 65–87. New Jersey, Lawrence Erlbaum Associates, 171 p.

Santow Gigi and Bracher Michael, 1988. *Do gravidity and age affect pregnancy outcome?.* Canberra, The Australian National University, Research School of Social Sciences, 30 p. (Australian Family Project, Working Paper no. 4).

Saurel-Cubizolles Marie-Josephe and Kaminski Monique, 1986. Work in pregnancy: its evolving relationship with perinatal outcome (a review), *Social Science and Medicine*, vol. 22(4), p. 431–442.

Saurel-Cubizolles Marie-Josephe, Stengel Bénédicte, Kaminski Monique and Blondel Béatrice, 1986. Surveillance prénatale et issue de la grossesse selon l'origine des femmes nées hors de France métropolitaine, *Journal de gynécologie, obstétrique et biologie de la reproduction*, vol. 15, p. 19–27.

Selevan Sherry G., 1991. Environmental exposures and reproduction, *in*: Kiely Michele (ed.), *Reproductive and perinatal epidemiology*, p. 115–130. Boston, CRC Press, 543 p.

Shapiro Sam, Jones Ellen W. and Densen Paul M., 1962. A life table of pregnancy terminations and correlates of fetal loss, *The Milbank Memorial Fund Quarterly*, vol. 40(1), p. 7–45.

Simpson Joe L. and Carson Sandra, 1993. Biological causes of fetal loss, *in*: Ronald Gray, Henri Léridon and Alfred Spira (eds.), *Biomedical and demographic determinants of reproduction*, p. 287–315. Oxford, Clarendon Press, 482 p.

Taylor William F., 1970. The probability of fetal death, *in*: F. Clarke Fraser and Victor A. F. C. McKusik (eds), *Congenital malformations: proceedings of the 3rd International Conference on congenital malformations, The Hague, Sept. 1969*, p. 307–320. New York, Excepta Medica, XVI + 450 p. (International Congress Series, no 204).

Thomas Karl, 1989. *Gynécologie Obstétrique. Conception Contraception*. Brussels, Cercle Médical Saint Luc, 129 p.

Thoulon J. M. and Laville M., 1995. Le monitorage électronique fœtal per-partum, *in*: Alain Treisser and Bernard Blanc (eds.), *25[es] Journées nationales de la Société française de médecine périnatale, Paris, 1995*, p. 81–96. Paris, Arnette Blackwell, 325 p.

Tietze Christopher, 1948. An investigation into the incidence of abortion in Baltimore. *American Journal of Obstetrics and Gynecology*, vol. 56(6), p. 1160–1162.

United Nations, 1954. Mortalité fetale, mortalité des enfants du premier âge et mortalité des jeunes enfants, *Études Démographiques*, vol. 1(1–3), 146 p.

United Nations, 1992–1995. *Abortion policies. a global review*, vol. 1–3), New York, United Nations, 158 p. + 244 p. + 236 p.

United Nations, 1994. *Demographic yearbook 1992*. New York, United Nations, 823 p.

Vintzileos Anthony M., Nochimson David J., Guzman Edwin R., Knuppel Robert A., Lake Marian and Schifrin Barry S., 1995. Intrapartum electronic fetal heart rate monitoring versus intermittent auscultation: a meta-analysis. *Obstetrics and Gynecology*, vol. 85(1), p. 149–155.

Waldron Ingrid, 1983. The role of genetic and biological factors in sex differences in mortality, *in*: Alan Lopez and Lado Ruzicka (eds.), *Sex differentials in mortality: trends, determinants and consequences*, p. 141–164. Canberra, Australian National University, Department of Demography, 498 p. (Miscellaneous Series, no. 4).

WHO, 1950a. *Rapport sur les travaux de la première session, Sous-Comité de la Définition de la mortinatalité et de l'avortement*. Geneva, WHO, Comité d'experts des statistiques sanitaires, 21 p. (Doc.WHO/HS/STDEF/9).

WHO, 1950b. *Définitions proposées pour les termes de 'Naissance vivante' et 'Décès fœtaux'—Rapport présenté par les États-Unis d'Amérique, Sous Comité de la Définition de la mortinatalité et de l'avortement*. Geneva, WHO, Comité d'experts des statistiques sanitaires, 6 p. (Doc.WHO/HS/STDEF/3).

WHO, 1970. *The prevention of perinatal mortality and morbidity*. Geneva, WHO, 60 p. (Technical Report Series, no. 457).

WHO, 1977. *Manuel de la classification statistique internationale des maladies, traumatismes et causes de décès, Neuvième révision, 1975*, vol. 1. Geneva, WHO, 782 p.

WHO, 1978. Sex differentials, *in: Report on social and biological effects of perinatal mortality*, vol. 1, p. 77–85. Geneva, WHO, 202 p.

WHO, 1979. Gestation, birthweight, and spontaneous abortion in pregnancy after induced abortion, *The Lancet*, 20 January, p. 142–145.

WHO, 1988. *La maternité en Europe*. Geneva, WHO, 167 p. (La Santé Publique en Europe, no. 26).

WHO, 1989. *Abortion: a tabulation of available data on frequency and mortality of unsafe abortion*, 2nd ed, Geneva, WHO, 36 p.

Wilcox Allen J. and Gladen Beth C., 1982. Spontaneous abortion: the role of heterogeneous risk and selective fertility. *Early human development*, vol. 7, p. 165–178.

Wilcox Allen J. and Horney Louise F., 1984. Accuracy of spontaneous abortion recall, *American Journal of Epidemiology*, vol. 120(5), p. 727–733.

Wilcox Allen J., Treloar Alan E. and Sandler Dale P., 1981. Spontaneous abortion over time: comparing occurrence in two cohorts of women a generation apart, *American Journal of Epidemiology*, vol. 114(4), p. 548–553.

Wilcox Allen J., Weinberg Clarice R., Baird Donna D. and Canfield Robert E., 1993. Endocrine detection of conception and early fetal loss, *in*: Ronald Gray, Henri Léridon and Alfred Spira (eds.), *Biomedical and demographic determinants of reproduction*, p. 316–328. Oxford, Clarendon Press, 482 p.

Wilcox Allen J., Weinberg Clarice R., O'Connor John F., Baird Donna D., Schlatterer John P., Canfield Robert E., Armstrong Glenn and Nisula Bruce C., 1988. Incidence of early loss of pregnancy. *New England Journal of Medicine*, vol. 319(4), p. 189–194.

Wilson Douglas R., Kendrick V., Wittmann B. K., and McGillivray B., 1986. Spontaneous abortion and pregnancy outcome after normal first-trimester ultrasound examination, *Obstetrics and Gynecology*, vol. 67(3), p. 352–355.

Wood James W., 1989. Fecundity and natural fertility in humans, *in*: Stuart R. Milligan (ed.), *Oxford review of reproductive biology*, vol. 11, p. 61–109. Oxford, Oxford University Press, 458 p.

CHAPTER

33

Sterility

Causes and Treatment

HENRI LERIDON

Institut national d'études démographiques (INED), Paris, France

In most traditional societies, procreation is regarded as part of every individual's natural vocation, and sterility is experienced as a deep personal misfortune. A sterile couple is unable to fulfill one of its key social functions and the sterile partner (when he or she can be identified, or failing this the partner to whom the sterility is attributed) bears the responsibility for inflicting this reproductive failure on the other partner. In the low-fertility posttransition industrialized societies, social pressures to reproduce are of course weaker, yet sterility continues to be experienced as a tragedy by many of the couples it afflicts.

We begin by reviewing the problems associated with the observation and measurement of sterility, and then consider the biologic and social factors responsible for different types of sterility and examine the effectiveness of the treatments currently available.

I. FROM INFERTILITY TO STERILITY

Although there is little consensus over the definition of sterility, the most frequently quoted *estimations* of the scale of the problem converge to a remarkable degree. One often reads statements like "10% to 15% of couples are sterile." But their authors not only fail to specify exactly what is meant by *sterile*, they also often they use the impersonal form ("it is estimated that . . .") without giving details of the source or when a source is given it usually turns out to be completely unscientific or merely refers to another, equally unscientific source. The reason for this is quite simply the lack of reliable epidemiologic surveys on sterility, apart from some work carried out in a number of countries as part of demographic surveys. The results from these surveys in no way corroborate the estimations quoted above.

Even without specific surveys on the subject, the demographic sources supply one extremely simple indicator of sterility, namely *the proportion of marriages still infertile* (no live-born child) after 10, 15, or 20 years or when the woman reaches age 45 or 50. Two objections immediately come to mind. First, some couples will have remained *voluntarily* infertile; second, some types of sterility will have been overcome by specialist treatment. The first argument implies that there are fewer physiologically sterile couples than infertile couples (i.e., who remain childless); the second argument, conversely, implies that in the history of the cohort under examination, more couples have suffered from sterility than have remained permanently infertile. A simple way of overcoming these difficulties is to start by examining populations that dispose of only crude and inefficient fertility control techniques and nonexistent or extremely inadequate treatments for difficulties in conceiving.

An example is provided by the populations of preindustrial Europe. The evidence for the French provinces in the 18th century, for instance, is that *in general fewer than 5% of marriages* remained childless. Another example comes from developing countries that have not yet begun the fertility transition. Several

TABLE 33–1 Proportion of Couples Still Childless at 45 Years (Age of Woman) in Selected Countries

Population	Birth cohorts (C) or years of marriage (M)	Age at marriage, years	Proportion still childless at 45 years, %
France preindustrial	M: 1670–1769	20–24	4
France preindustrial	M: 1670–1769	25–29	6
France	C: 1935–1939	20–24	≤6
France	C: 1935–1939	25–29	<10
United States (white population)	C: 1930–1935	<45	7
Finland	C: 1931–1935	<45	6
Bangladesh	C: 1925–1930	<45	3
Mexico	C: 1925–1930	<45	5

Source: Henri Leridon, 1982. Stérilité, hypofertilité et infécondité en France, *Population*, vol. 37(4–5), p. 807–836.

dozen of these countries participated in the World Fertility Survey in the 1970s. In three fourths of them the proportion of married women who remained childless at 40 to 44 years *was less than 4%* and none exceeded 7% (Vaessen, 1983).[1] Table 33–1 shows a selection of the proportions obtained in a number of countries.

It is hard to see how the situation could be significantly different in the developed countries, and indeed all the evidence is that they do not have a higher frequency of unsurmounted sterility. Our first source for this is census data. In 1975, in the United States and France, the proportion of women remaining infertile among married women aged 35 to 39 years was 7% to 8%, and this proportion undoubtedly includes cases of *voluntary* infertility. Our other source is survey data. In France, the INED-INSEE surveys of 1978 and 1988 provided confirmation for the view that fewer than 6% of couples experienced permanent involuntary childlessness, although many more reported having had difficulty conceiving at one time or another (Leridon, 1982, 1991).

That last point is the key to much of the confusion. For if we count as sterile every couple who starts to express impatience because a conception (not necessarily the first) is proving slow to be achieved, the proportion of *sterile* couples will far exceed 5%. Indeed, it is not fanciful to think that their number will grow as the scope for obtaining medical advice expands. In France, as early as the 1970s, *roughly one in ten couples had been to a doctor for this reason*. We return to this point later.

The proportions of couples who are involuntarily childless at a particular age can thus be treated as estimations of *primary sterility* at this age. When age 50 (for women) has been reached, we can speak of *total sterility*, because the couple has never produced a child during its reproductive life span. In the case of a couple who has already had one or several children but is unable to have any more, the sterility is said to be *secondary*; it thus includes the possible consequences of delivery. Such cases seem to be fairly widespread in Africa, where, as mentioned above, levels of primary sterility are generally low.

[1] The sample analyzed included a few African countries. A more recent study, based on surveys carried out in Africa in the 1980s, has confirmed these results. See Larsen, 1994.

II. STERILITY AND AGE

Although some couples have the misfortune to discover that they are sterile very early on, all couples become sterile eventually. As we saw in Chapter 30, menopause forms a natural limit to the reproductive span. However, it is a *lagged* indicator because although mean age at menopause is around 50 years, mean age at last birth is closer to 40 years in populations subject to a *natural fertility* regime (see for example the table compiled by James Wood, 1994, p. 442). This time lag can have two sources, however. First, it may result from the onset of *permanent sterility* several years before the completion of menopause (menopause is not considered to be complete until after a minimum length of time without menstruation, usually set at 1 year), which could be attributable to an anovulatory phase preceding the final cessation of menses. Second, it can be assumed that the time needed to conceive increases with age, as too does the risk of fetal mortality. This means that a woman who was still *fecund* at a particular age might not have enough time to conceive a child who would be live-born in the years remaining until menopause (or permanent sterility). In terms of live births, which are usually the only observations available to the demographer, the two explanations are confounded and any estimation of the proportion of women already sterile at a particular age incorporates these two dimensions.

Several authors have proposed estimations for this *permanent sterility* based on the proportion of women still infertile beyond a certain age (Vincent, 1950; Henry, 1961; Leridon, 1977; Trussell and Wilson, 1985; Larsen and Menken, 1989). Figure 33–1 presents two types of estimation, the first for the proportion of sterile couples among newly married couples, by age at marriage (we are thus dealing with primary sterility), the second for the whole population of married couples, by the age of the woman (the sterility can therefore be secondary). On Figure 33–1A I have plotted an estimation by Trussell and Wilson (1985) and my own (Leridon, 1977), plus an exponential function fitted to the latter, for *newly married couples* at age x (age of the woman); the proportion of couples already sterile is given by:

$$y = 0.370 \exp (0.109x)$$

Figure 33–1B concerns *all the women* in couples at an age x; the proportion of couples already sterile is given by:

$$y = 0.427 \exp (0.113x)$$

From these results, it appears that approximately 10% of newly married couples are already sterile when

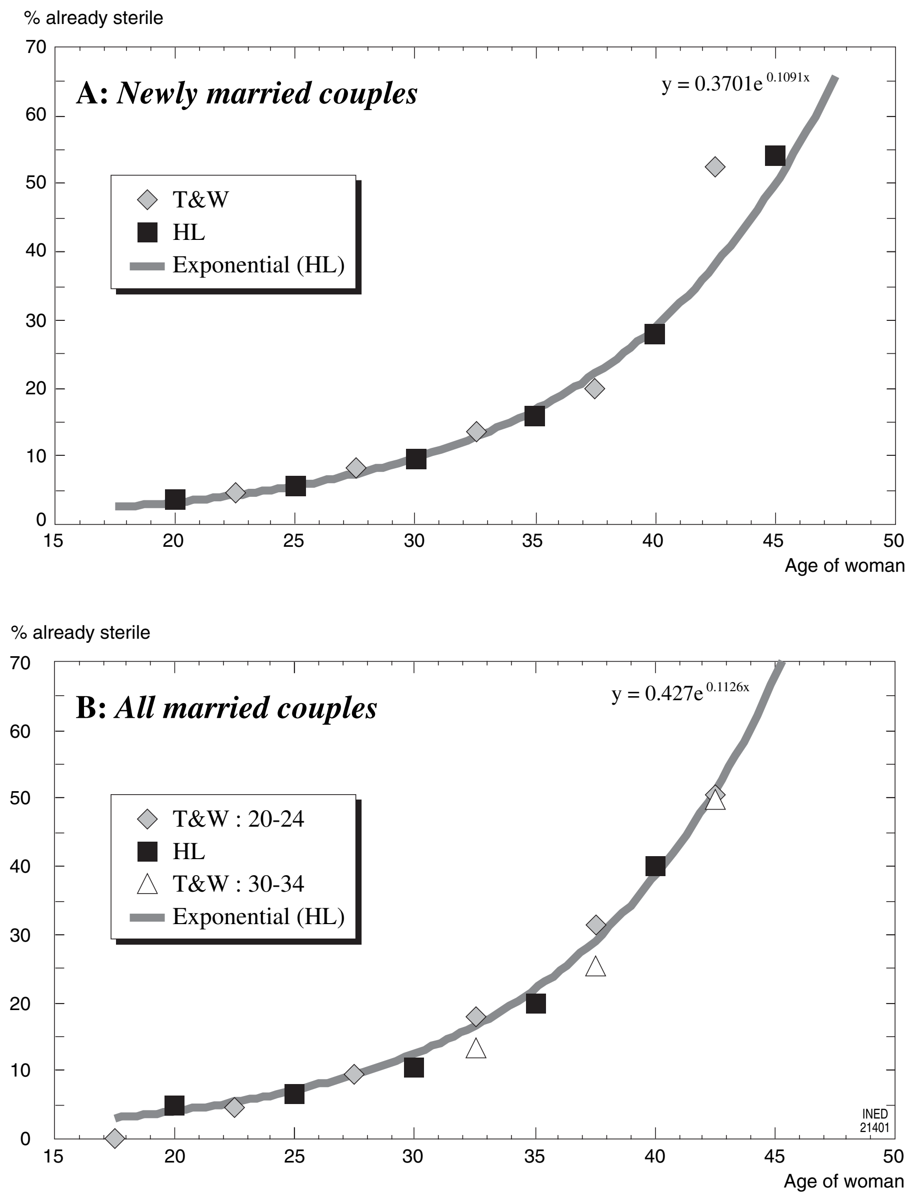


FIGURE 33–1 Proportion of couples already sterile. Source TRUSSELL James, WILSON Christopher, 1985. Sterility in a population with natural fertility, *Population Studies*, vol. 39(2), p. 269–286; LERIDON Henri, 1977. Sur l'estimation de la stérilité, *Population*, vol. 32, Special issue, p. 231–248.

the woman is 30 years old, 17% at 35 years, and 29% at 40 years. If cases of secondary sterility are included, the respective proportions rise to 12%, 22%, and 39%. In contexts where no treatments are available, this sterility would be permanent; in contemporary populations, the existence of effective treatments will have the effect of reducing these values.

III. THE CAUSES OF STERILITY

After having been assumed to be negligible, male sterility is currently thought to be the problem in three to four 10 cases of sterility. Some authors favor use of a third category to accommodate cases where responsibility is shared: "In roughly 30% of barren marriages, the husband has been found to be the significant factor. In another 20%, he plays an important contributing role" (Kempers, 1979, p. 273; see also Sherris and Fox, 1983). The idea is an *addition of two subfecundities* (of the man and of the woman) leading to an apparent sterility. Indeed, pursuing this line of reasoning a little further, it could be argued that in most cases the doctor is treating not a *single case of sterility* but *one* or *two* cases of subfecundity, thereby making illusory the distinction between *female sterility* and *male sterility*. Ideally, we would be able to say for a particular couple, that the difficulties are 30% due to the woman and 70% due to the man, but this is virtually impossible to establish in practice and contributes little to resolving the problem.

For want of a better solution let us accept this distribution of roles. We must now turn our attention to identifying the possible causes of sterility in one or other of the partners.

In men, cases of secretory azoospermia seem to be relatively rare, and involve mainly congenital abnormalities (chiefly cryptorchism). More frequently diagnosed conditions include excretory azoospermias, resulting from tubercular or gonococcal infections, while still more common are complex oligoasthenoteratospermic conditions, which also result from viral or other infections, hormonal imbalances, histories of cryptorchism, or which are attributed (rightly or wrongly) to presence of a varicocele.

In women, researchers are agreed that the main cause is *tubal disease*. This usually results from infections of various origins: puerperal infections (postpartum or postabortal), which thus cause secondary sterility, and pelvic inflammatory disease resulting from sexually transmitted diseases (gonococcal and, increasingly, chlamydial infections). If we include in this category most *endometriosis*, we can *explain* 20% to 30% of all sterility (including male). A further 20% to 30% is accounted for by *ovulatory problems* and *hormonal imbalances* (Sherris and Fox, 1983). This covers anovulation with genetic or autoimmune (congenital) origins, failure of follicle to rupture, and, especially, polycystic ovarian disease. Hormonal disorders are also responsible for failures in fertilization. In most studies there is a residual proportion, in the order of 20% to 40%, of unexplained sterility (termed *idiopathic*).

TABLE 33–2 Distribution of Causes of Involuntary Infertility

Cause	Couples, %
Ovulatory disorder	20–29
Tubal abnormality	15
Endometriosis	6
Cervical factors	3
Spermatic abnormalities	21–24
Other male causes	2
Problems of coitus	6
No cause identified	18–28

Source: Denis Querleu, Alfred Spira, Henri Leridon, 1990. Epidémiologie de la fertilité, in *Encyclopédie Médico-chirurgicale (Paris)*, Gynécologie, 738, C^{10}, 3–1990, p. 1–6.

Table 33–2 summarizes the distribution of the causes of *infertility*, which also includes the cases of total sterility. This breakdown is necessarily approximate because it is obtained by combining nonexhaustive studies. More recent results proposed by Snick *et al*. (1997) broadly confirm the proportions given in the table.

Note that most factors mentioned above can also explain a rise in fetal mortality, which in addition has a large genetic component (involving mainly chromosomal injuries occurring at random during fertilization and early cellular division, more than hereditary anomalies).

At another level, these dysfunctions can originate in exposure to polluting agents in the environment or in the workplace. The adverse effect of smoking has also been proved as has that of female sexual mutilation (see Sherris and Fox, 1983).

Finally, childbirth can also be responsible for secondary sterility (as was noted earlier), in particular through the risk of puerperal infection. The conditions of childbirth are thus determinant. The risk is very low when childbirth takes place under full medical supervision. At the other extreme, for Europe between the 17th and 19th centuries, we have estimated at 2% to 5% the risk of becoming sterile as a direct result of childbirth (Leridon, 1977). For similar reasons, abortion performed in unsatisfactory conditions can have the same type of consequences.

IV. MEASURING TREATMENT EFFECTIVENESS

The first point to note is that the *chances of success for the treatments* that are available vary greatly depending on the origin of the sterility. For the main cause (ovulatory disorders) probabilities of success in the order of 30% to 40% are usually advanced. However, not every pregnancy achieved after a hormone treatment should be automatically attributed to the treatment. To interpret the reported success rates correctly we need to know (as well as the length of treatment) the selection conditions for the couples receiving treatment. Selection operates in three ways—through the initiatives made by couples (changing doctors, abandoned attempts, delays, etc.), through the choices made by the clinician (diagnostic techniques, type of treatment, etc.), and through the time elapsed since first trying to become pregnant. The third criterion is especially important, because as this period lengthens so the group of couples who have still not conceived includes a growing proportion of authentically sterile (nil fecundability) couples and of subfecund couples with very low fecundability.

Let us assume that we observe a population of nonsterile couples, with a mean fecundability of 0.25, from the point in time when they first try to conceive (Table 33–3A, based on Schwartz, 1981).

After 2 years, 3% of couples will still not have conceived, and the mean fecundability of these couples will be 0.08 (a third of the level in the initial population): This group would thus appear to be eligible to be treated for sterility; yet, *57% of them will go on to conceive successfully in the following 12 months* without any treatment. Consequently it is with 57%, rather than with 0%, that we should compare the success rate obtained in the group of couples who received treatment (Leridon and Spira, 1984).

The reasoning set out above is only valid for couples who are not totally sterile. When a cause of total sterility (complete azoospermia, complete tubal blockage) can be diagnosed with certainty, any pregnancy has to be credited to the treatment. This may be the case in tubal surgery, artificial insemination with donor, or *in vitro* fertilization (IVF), on condition that only couples who are totally sterile have been selected. Yet, with IVF, for example, this is often far from the case; the technique is also used on women with an *unexplained*

TABLE 33–3 Evolution of a Population Including Fecund Couples and Sterile Couples (Fecundability Initially Distributed According to a Beta Function with Parameters $a = 2.75$ and $b = 8.25$)

Time elapsed (in months)	Number of couples who have not yet conceived	Mean fecundability of these couples	Proportion of sterile couples, %	Proportion conceiving before 1 year, %
A: No sterile couples				
0	1000	0.250	0	90.1
1	750	0.229	0	88.4
2	578	0.212	0	86.7
3	456	0.196	0	85.0
6	248	0.162	0	80.0
12	99	0.120	0	71.0
24	29	0.079	0	57.3
36	12	0.059	0	47.7
48	6	0.047	0	40.8
60	4	0.039	0	35.6
B: Plus 5% sterile couples				
0	1053	0.237	5.0	85.6
1	803	0.214	6.6	82.6
2	631	0.194	8.4	79.4
3	509	0.176	10.4	76.1
6	301	0.133	17.6	65.9
12	152	0.078	34.9	46.2
24	82	0.028	64.6	20.3
36	65	0.011	81.5	8.8
48	59	0.005	89.8	4.2
60	57	0.003	93.0	2.5

sterility, and there is no certainty that the pregnancies obtained can all be attributed to IVF.

Table 33–3B is constructed on the same base as Table 33–3A but to the initial population have been added 5% of totally sterile couples. After 2 years, 8% of the couples still infertile and within this group, 65% of couples are totally sterile. Consequently there is even more reason for these couples to be offered treatment promptly, although it can be noted that a further 20% will conceive unaided in the following 12 months.

V. HOW EFFECTIVE, IN PRACTICE, ARE THE VARIOUS INFERTILITY TREATMENTS?

In the 1960s and 1970s, treatment was mainly *hormonal* or, in the case of tubal obstruction, *surgical*. The success rates reported in the literature were often between 30% and 40% in both cases, without precise indications being given about the duration under consideration or the selection conditions for the women treated. Working on a large sample of couples consulting after at least 1 year of infertility, Collins *et al*. showed in 1983 that the pregnancy rates for the treated women were in fact little higher than those for the women not treated, being, respectively, 41% and 35%. It is true that the first group selected women known to have indications for treatment, hence signs of impaired fecundity, while the second group could include more women presenting no obvious sign of low fecundity. That said, the study provided a clear illustration of the process described above and showed the need for caution when interpreting success rates for these treatments; these results were moreover confirmed by later studies (see for example Collins *et al*., 1995; Snick *et al*., 1997).

The 1980s and 1990s saw a rapid development of the techniques of *artificial insemination by donor* (AID) (in the case of male sterility) and *in vitro fertilization with embryo transfer* (IVF). The maximum success rate per AID cycle is 15%, corresponding to 7% in live births (CECOS, 1998). However, repeated attempts are possible, and the cumulative rate can then reach 40% or 50% for conceptions and 20% to 25% for births. In IVF, the conception rate achieved with egg harvest (oocyte retrieval after hormonal stimulation of the cycle) is around 20% (FIVNAT, 1998a) and approximately 15% as live births. This technique can also be repeated, although because it is much more complicated and invasive it is unusual for more than three attempts to be made. The success rate (in births) then reaches 34% (FIVNAT, 1998b).

AID and a proportion of IVFs are being replaced by the new technique of ICSI (intracytoplasmic sperm injection), which involves injecting a single sperm directly into the inner structure of the egg. The success rate seems to be slightly higher than with IVF.

Two further points must also be made. First, all these rates decline as the woman's age increases, especially after 40 years—medically assisted procreation is only a partial remedy for the effects of aging on fecundity. Second, both hormonal treatments and IVF frequently result in multiple births, the increased number of which shows up even in national-level figures.

Lastly, mention can be made of the nonmedical alternative to infertility treatment, adoption, which, though its legal framework varies between countries, is present everywhere.

VI. ESTIMATING THE DEMAND FOR TREATMENT

According to the most recent French survey, 15% of couples aged 25 to 45 years (who had already tried to conceive) had sought medical advice at least once for a *sterility problem* (Leridon, 1991). Can this proportion be taken to represent the *demand for treatment*? Definitely not, insofar as the initiative of some of these couples could be judged premature (less than 1 or 2 years of trying for a pregnancy, for example) or at any rate not a justification for immediately starting on a major treatment—a third of couples who reported having sought medical advice had done so before 1 year of infertility and were thus being overimpatient. Data from an earlier (1978) survey allow us to approach the problem from a different angle (Leridon, 1982). At that time only 10% of married women reported a fertility problem, but for the reasons given above this estimate may be more realistic; it also seems more consistent with results from American surveys, according to which approximately 5% of married women (15–44 years) had sought medical help during the previous 5 years (Mosher and Pratt, 1990).

In the 1978 French survey, "Did you ever take longer than you would have liked to get pregnant, or not manage to at all?" Of these, 3.7% were still infertile at the time of the survey; 14.7% had eventually conceived (one fifth of these after treatment or surgery). The second group was thus made up mainly of subfecund subjects and, related to this, those with correctable sterility.

Usually (approximately six in 10 cases), the difficulties were apparent from the first attempt to become pregnant. This enables us to estimate the mean fecundability of the subfecund couples. We do this by com-

TABLE 33–4 Distribution of Time to Conception in a Group of Subfecund Couples

	Time to conception (months)						Mean time	Median
	0–11	12–23	24–35	36–47	48+	Total		
Observed[a]	36	23	16	10	15	100	26.5	17
Model[b]	37	21	13	8	21	100	33.6	8

[a] Distribution of the interval between marriage and first birth, reduced by 16 months (mean duration of pregnancy, 9 months; mean voluntary waiting period at start of marriage, 7 months) for couples who have had difficulties conceiving (INED-INSEE survey, 1978; Author's computation).

[b] Distribution of the time to conception in a population whose fecundability is distributed according to a beta function of mean 0.04 and standard deviation 0.02.

paring the mean interval between marriage and first birth for the women reporting difficulty in conceiving their first child, with the mean for those did not report any difficulty: the figures are 42.5 months for the former and 21.8 for the latter, giving a difference of *20 months*. This interval can be broken up into three components:

- A *voluntary waiting time*, from use of birth-spacing contraception
- The *necessary time to conception*, after contraception has been discontinued
- The *duration of pregnancy*, corresponding on average to 9 months

We can reasonably assume that the voluntary waiting time is the same among the subfecund as among the others, since the difficulties in conceiving do not actually become apparent until after this period. Mean fecundability—the monthly risk of conception when no contraception is used—is well known for *normal* populations, that is, those not subject to selection: It is usually around 0.25, giving a mean time to conception of 5 to 6 months (see Chapter 30); the mean interval between marriage and first birth is 21.8 months for couples not reporting a problem, so we can estimate at: 22 − 6 − 9 = *7 months, approximately, the voluntary waiting time* at the start of marriage.

For the subfecund group, mean time to conception would thus be: 42.5 − 7 − 9 = 26.5 months. The earlier studies showed that the fecundability of the couples could be fitted to a beta function (Pearson I) with two parameters a and b. An advantage of this function is that the fecundability of couples who have not yet conceived at duration n is also distributed according to a beta function, with parameters a and $b + n$. Thus it is not unreasonable to assume that the fecundability of the group of subfecund couples is also distributed according to a beta function and to try to estimate the mean fecundability. What is needed, therefore, is a beta function that gives a distribution of the times to conception as close as possible to that observed for the subfecund group. Strictly speaking, the latter distribution is not known, since the voluntary waiting period at the start of marriage is not known for each couple *individually*: The mean has merely been estimated (indirectly) at 7 months. Accordingly we simply make a uniform reduction in the interval between marriage and first birth of 7 months for the voluntary wait and of 9 months for the duration of pregnancy.

The result appears in Table 33–4. A beta distribution of mean 0.04 (parameters, 3.80 and 91.20) is found to give the best fit.

The main inconsistency is over the long periods (in excess of 4 years), which makes a large difference to the mean value, but it must be noted that the observed distributions are necessarily truncated here, because the time elapsed since marriage is not always adequate. With that qualification made, the fit is good, and we can reasonably conclude that *the mean fecundability of couples self-reporting as subfecund is in the order of 0.04*.[2]

This result is directly applicable in our attempt to estimate the demand for treatment to combat sterility. It would scarcely be reasonable for this estimation to include couples who identify themselves as subfecund and who may have sought medical help, but who have been trying to conceive for a few months only. What should the minimum wait be before proceeding with a medical intervention? There is probably no single answer to this question, and the speed and precision of the diagnosis will vary with each individual situation. *A minimum wait of 2 years is often specified*. Yet as Table 33–4 shows, 58% of subfecund couples conceive before 2 years (59% in our population had in fact conceived within 2 years). So *only the others* (42%) *should* be considered as clinically *sterile* and thus be accepted as possible candidates for treatment. If the waiting time for treatment is increased to 4 years (to allow for the investigations necessary before deciding on a

[2] Comparable to the result after 4 years in Table 33–3.

surgical intervention, for example), the proportion of subfecund couples who have still failed to conceive unaided and who therefore ought to be offered treatment stands at only 21%.

The difference between these two figures (42 and 21%) is worth considering. It results from those couples who satisfy the usual definition of sterile (2 years without conception) but who then manage to become pregnant—usually without medical assistance—in the following 2 years. Would complicated treatment or potentially costly operations have been justified in such cases? How is the choice to be made between an additional waiting period (which in any case cannot be predicted at the individual level) and an *immediate cost*? The question obviously only arises for subfecund couples, whose chances of conception are not completely nil but who as we have seen make up the great majority of potential candidates for treatment.

VII. THE DEMAND FOR TREATMENT: STOCKS AND FLOWS

Let us recapitulate the example analyzed above:

- 3.7% of couples suffer from sterilities that are not overcome, with 2.6% of couples remaining childless;
- 14.7% of couples report difficulty in conceiving, 8.2% of them from the first pregnancy.

In the latter group, those who have experienced difficulty from the first pregnancy form the subfecund group that we have just analyzed. Between 21% and 42% of them may be candidates for treatment, depending on whether the requirement is set at 2 or 4 years of previous infertility, representing 1.7% to 3.4% of all couples. Couples who have experienced difficulties achieving a pregnancy other than the first probably have a higher mean fertility, because their mean interval between marriage and first birth is much closer to that of *no-problem* couples, than to the previous couples. Let us accept that 10% to 20% of them would be eligible for a treatment[3]: 0.7% to 1.3% of all couples would thus be involved on the grounds of a secondary sterility.

In all, *6% to 8% of couples are potential candidates for a treatment* against sterility, and this breaks down as follows:

- 2% to 3% of couples who will remain totally infertile unless treated successfully,
- 1% of couples who have a secondary sterility,
- 2% to 3% of subfecund couples, who experience serious difficulty achieving a first pregnancy,
- Approximately 1% who are secondary subfecund, whose difficulties become apparent only after one or more births.

If it is accepted that all cases of tubal sterility (20%–30% of infecund couples) are indications for IVF, and if we add to this approximately 10% of other indications, we have 2% to 3% of all couples who are potential candidates for this kind of intervention.

In France as a whole, 300,000 to 400,000 couples are formed each year: *7000 to 12,000 could therefore be expected to request IVF each year*. In fact, in 1997 more than 20,000 first egg harvests (on women who had never previously had one) were carried out, and 3200 first AID: the indications have thus been significantly extended, well beyond the hypotheses adopted here. Increasingly, sterile couples are systematically offered IVF, especially using the ICSI procedure, regardless of the origin of the sterility.

[3] If the mean Fecundability of this group was 0.10—which is already a fairly low level—15% of them would still have not conceived after 2 years, and fewer than 5% after 4 years.

Acknowledgment

Translated from French by Godfrey I. Rogers.

References

Bari Kolata G., 1978. Infertility: promising new treatments, *Science*, vol. 202(4364), p. 200–202.

Collins John A., Burrows Elizabeth A., Willan Andrew R., 1995. The prognosis for live birth among untreated infertile couples, *Fertility and Sterility*, vol. 64(1), p. 22–28.

Collins John A., Wrixon William, Janes Lynn B., Wilson Elaine H., 1983. Treatment-independant pregnancy among infertile couples, *New England Journal of Medicine*, vol. 309(20), p. 201–206.

FIVNAT, 1998a. Bilan général FIVNAT 1997, *Contraception, Fertilité, Sexualité*, vol. 26(7–8), p. 463–465.

FIVNAT, 1998b. Influence du rang de la tentative en FIV, *Contraception, Fertilité, Sexualité*, vol. 26(7–8), p. 466–472.

Henry Louis, 1961. Some data on natural fertility, *Eugenics Quarterly*, vol. 8(2), p. 81–91.

Kempers Roger D., 1979. Female contraception and sterilization, *in*: Edward E. Wallach, Roger D. Kempers (eds.), *Modern trends in infertility and conception control*, p. 331–332. Baltimore, The Williams and Wilkins Co., 439 p.

Larsen Ulla, Menken Jane, 1989 Measuring sterility from incomplete birth histories, *Demography*, vol. 26(2), p. 185–202.

Larsen Ulla, 1994. Sterility in Sub-saharan Africa, *Population Studies*, vol. 48(3), p. 459–474.

Leridon Henri, Spira Alfred, 1984. Problems in measuring the effectiveness of infertility therapy, *Fertility and Sterility*, vol. 41(4), p. 580–586.

Leridon Henri, 1977. Sur l'estimation de la stérilité, *Population*, vol. 32, Special issue, p. 231–248.

Leridon Henri, 1982. Stérilité, hypofertilité et infécondité en France, *Population*, vol. 37(4–5), p. 807–836.

LERIDON Henri, 1991. Stérilité et hypofertilité: du silence à l'impatience?, *Population*, vol. 46(2), p. 227–248.

MOSHER William D., PRATT William F., 1990. *Fecundity and infertility in the United States*. Hyattsville, NCHS, 9 p. (Advance Data, no. 192).

QUERLEU Denis, SPIRA Alfred, LERIDON Henri, 1990. Epidémiologie de la fertilité, in *Encyclopédie Médico-chirurgicale (Paris)*, Gynécologie, 738, C^{10}, 3-1990, p. 1–6.

SCHWARTZ Daniel, 1981. Importance de la durée d'inféconditá dans l'appréciation de la fertilité d'un couple, *Population*, vol. 36(2), p. 237–250.

SHERRIS Jacqueline D., FOX Gordon, 1983. Infertility and sexually transmitted diseases. a public health challenge, *Population Reports*, *Serie L*, no. 4, p. 127.

SNICK H. K., SNICK T. S., EVERS J. L., COLLINS John A., 1997. The spontaneous pregnancy prognosis in untreated subfertile couples: the Walcheren primary case study, *Human Reproduction*, vol. 12(7), p. 1582–1588.

TRUSSELL James, WILSON Christopher, 1985. Sterility in a population with natural fertility, *Population Studies*, vol. 39(2), p. 269–286.

VAESSEN Martin, 1983. *Childlessness and infecundity*. Voorburg (Netherlands), IIS, 44 p. (WFS Comparative Studies Note, 2 Series, no.31: Cross National Summaries).

VINCENT Paul, 1950. La stérilité physiologique des populations, *Population*, vol. 5, no.1, p. 45–64.

WOOD James W, 1994. *Dynamics of human reproduction: biology, biometry, demography*. New York, Aldine de Gruyter, 653 p.

CHAPTER

34

Natural Fertility and Controlled Fertility

Levels and Models

HENRI LERIDON

Institut national d'études démographiques (INED), Paris, France

I. THE "INTERMEDIATE VARIABLES"

With their article "Social structure and fertility: an analytical framework," published in 1956, Kingsley Davis and Judith Blake opened a new field of research. After analyzing and classifying a large corpus of anthropological studies, the authors identified several levels of variables that influence fertility. These were the intermediate variables, through which social and cultural factors operate to influence fertility; the social norms, closely linked to the social organization an individual belongs to, which govern fertility and marriage behavior and the other intermediate variables; and the structural characteristics that define the position of the individual within society. This classification proved analytically fruitful, and the long list of intermediate variables proposed by Kingsley Davis and Judith Blake drew attention to a number of factors that had previously been unjustly neglected, such as age at the beginning and end of the reproductive period, pregnancy wastage, periods of sexual abstinence, plus of course the variables that describe the regime of conjugal unions (age at marriage, divorce, remarriage, etc.).

A serious omission from their list, however, and that for almost 20 years continued to receive little or even no attention, was the key role of breast-feeding, despite research that had established its importance in influencing the duration of postpartum amenorrhea. This variable had to be reintroduced into Blake and Davis' analytical framework, and more importantly into the questionnaires employed in demographic surveys from the mid 1970s.

Table 34–1 presents a classification of all the variables that are currently thought to be potential determinants of fertility.

1. Recombining the Intermediate Variables

The list in Table 34–1 is of course intended primarily as an analytic framework. It is not possible to assemble at any one time all the information pertaining to a given population, and the impact of the different variables depends greatly on the general fertility level, according to whether or not couples are already controlling fertility. In many countries, the process of reproduction is largely dominated by behavioral variables (desire to have children, number of children wanted, use of contraception, etc.), which makes it reasonable to draw on theories from sociology and economics for the keys to explain the observed behavior. But it was not always so, and is still not so in some regions of the world.

The response of demographers has been to concentrate first on describing the situations of natural fertility, in which couples individually do not adopt behavior deliberately intended to space or limit the

TABLE 34–1 Determinants of Fertility: a Classification

A. THE INTERMEDIATE VARIABLES
These are the variables, to use the terminology of Davis and Blake (1956), through which the cultural factors operate to influence fertility. It will be noted that the authors' framework did not include the duration of breastfeeding.
1. Biologic capability
a. Ages at puberty, menopause, and permanent sterility
b. Monthly probability of conception (fecundability)
c. Intrauterine mortality (spontaneous)
d. Duration of postpartum nonsusceptible period
2. Conjugal regime
a. Age at marriage (or entry into sexual union)
b. Frequency of permanent celibacy
c. Frequency of, and ages at, death, widowhood, or divorce
d. Frequency and ages of remarriage
e. Periods of separation of partners (seasonal work, illness, etc.)
3. Other sociocultural variables
a. Frequency of intercourse (affecting 1b)
b. Duration of breastfeeding (affecting 1d)
c. Sexual taboos and restrictions (affecting 1d)
4. Fertility regulation variables
a. Individual action on length of union (affecting 2)
b. Action on exposure to conception: abstinence, use of contraception (affecting 3 and 1b)
c. Action on pregnancy outcome: induced abortion (affecting 1c)
d. Voluntary sterilization (of one partner) (affecting 1a)

B. SOCIOECONOMIC STATUS VARIABLES
These variables describe an individual's socioeconomic context (his or her *milieu*) and may explain (through the variables in groups A-2, -3, -4) differences in fertility levels. In this sense, therefore, they are external to the previous framework.
a. Ethnic origin, cultural group
b. Religion
c. Educational level
d. Socio-occupational status
e. Economic status, income
f. Family structure
g. Place of residence (urban, rural)
NB: Ronald Freedman's *Community-level Variables* (1974) could be included here. These variables seek to describe the individual's immediate environment: proximity of school, clinic, shops, administrative services, access to electricity, public transport, etc. The necessary information can be collected directly by the survey taker for all the inhabitants of the same neighborhood or same village.

C. THE *ENVIRONMENT* VARIABLES
These variables describe some of the constraints faced by individuals. They act mainly on the variables of group A-1, or describe general structural changes.
a. Health
b. Nutrition
c. Age structure
d. Political system

D. THE PSYCHOSOCIOLOGIC VARIABLES
These variables describe in a direct way individual attitudes towards fertility problems or towards the intermediate variables, relative to the norms prevailing in their social group (see B).
a. Desires with respect to fertility (number of children, sex, spacing, and so forth)
b. Attitudes toward marriage and the family
c. Attitudes toward sexuality
d. Attitudes toward birth control

Source: Henri Leridon, Laurent Toulemon, 1997. D*émographie. Approche statistique et dynamique des populations.* Paris, Economica, 440 p. (Économie et statistiques avancées); adapted from Davis, Kingsley, and Blake, Judith, 1956. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol. 4(4), p. 211–235.

number of births. In such a context, the biologic variables of reproduction assume their full importance, though they are never the only ones involved—fertility is always part of a specific social context, and customs in respect of marriage, sexuality, and breast-feeding also play a crucial role. The concept of natural fertility (*fécondité naturelle*) was introduced by Louis Henry in 1953, who referred to "the fertility that a human population would have if it made no conscious attempt to limit the number of births," and in 1961 he added that such fertility "is the one most influenced by physiological factors." In practice, it must also be pointed out that Louis Henry was referring basically to legitimate fertility, thus nuptiality was left out of his definition. With a view to removing some of the ambiguity I have given the following definition (Leridon, 1989):

> A natural fertility regime characterizes a situation in which the level of fertility results, above all, from a combination between these physiological factors of reproduction and behavior that is simply consistent with the norms of the social group. In particular, it is excluded that any preferences that individuals and couples may have as regards the size of their family can influence their age at marriage, the spacing of their births, and their completed family size (p. 33).

It was seen in Chapter 30 that by breaking up the interval between two successive conceptions it is possible to identify some intermediate variables. We can also attempt to reconstruct the process, on the basis of these elements, to obtain the observed (or observable) fertility. This procedure has given rise to various attempts to model the reproductive process.

2. The Early Approaches

Using the previously identified variables, efforts focused on constructing models that would describe real processes. The early models did not include the risk of sterility and merely made the assumption that fecundability tended to zero beyond a certain age. In addition, they only applied to women who were already married (i.e., assumed to be exposed to the risk of conception).

Henry (1957) was the first to develop a purely algebraic approach. This involved setting the following integral equation, if continuous notation is being used (or the equivalent recurrent equation, if using discrete notation):

$$C(x) = p(x)\left[1 - \int_0^G C(x-g)k(x-g,g)dg\right]$$

where:

—x age (or the duration since marriage, which is taken to have occurred at an exact age);
—C(x) number of conceptions at age x (per woman);
—p(x) fecundability at age x (identical for all the couples of age x);
—k(x − g, g) probability of still being in the nonsusceptible period (see chapter 31) at duration g, for a conception at age x − g;
—G maximum duration of the nonsusceptible period.

The nonsusceptible period is the sum of the length of the gestation period and the longer of the two postpartum periods of amenorrhea and abstinence. The formula given above makes no distinction between pregnancies that result in live births and the others; this can be done in a subsequent stage by splitting the function k (into k_v and k_a) and the function C (into C_v and C_a).

We denote as Q(x) the number of conceptions before x. If marriage occurs at age x_0, the mean rate of conception from age x_0 to age x will be:

$$\frac{Q(x)}{x - x_0} = \int_{x_0}^{x} C(t)dt$$

Since all women are assumed to be at risk of conceiving at the start of marriage, the rate is at its maximum during the first month. This value is never again attained, since at any given time a certain proportion of the women will have a nonsusceptible status, i.e. will have nil probability of conceiving.

The mean interval between marriage and the first conception is simply equal to 1/p(x) for women married at age x (p varying little during the early years). If $\bar{p}$ denotes mean fecundability for the group of ages under consideration, v the proportion of conceptions that lead to live births, and $\bar{g}$ the mean length of the nonsusceptible period for all conceptions, the mean interval between live births is:

$$\bar{i} = \frac{1}{v}\left(\frac{1}{\bar{p}} + \bar{g}\right)$$

These formulas are valid only in the simplified case where all women are assumed to have the same functions p and k. For a heterogeneous population, the results are more complex. The interval between marriage and first conception, for example, will be equal to the harmonic mean of $f(p)$, if $f(p)$ is the distribution function of fecundability between all the women of a particular age:

$$m = \int \frac{1}{p} f(p)dp \neq \frac{1}{\bar{p}}$$

Several authors have preferred to use the theory of Markov chains (see especially Potter and Sakoda, 1966; Sheps and Menken, 1973). This involves defining a set of possible states (e.g., fecund, pregnant, postpartum nonsusceptible) and the probabilities of transition between states, and then calculating the length of time spent in each state and the number of entries and exits for each state. The time spent in state *I* can depend on state *I* and on $I+1$, but does not usually depend on the external time (the total time elapsed since the start of the process, which in this case is age). The main advantage of the method is that it lets us analyse the convergence of the process towards the stationary state and to describe the latter, over the period when fecundity varies little (20–35 years). Inclusion of fetal mortality and sterility is not problematic.

3. Simulations Using the Monte Carlo Method

Simulations using the Monte Carlo procedure are useful in several cases where the models mentioned above are inadequate. Greater complexity can be added to make the model more realistic, for example, by introducing the probabilities of marriage, and of divorce or widowhood. The various functions can be made to depend on age or duration of marriage or both. The heterogeneity of real cohorts in relation to these variables can be introduced. Particularly valuable is the applicability of the analysis to populations practicing birth control. This method enjoyed great popularity in the 1960s and 1970s (see Ridley and Sheps, 1966; Jacquard, 1967; Barrett, 1969; Horvitz *et al.*, 1971; Dyke and MacCluer, 1973; Jacquard and Leridon, 1973), but was somewhat discredited by the construction of excessively complex models. It remains useful, however, for carefully targeted uses, such as evaluating the effect of a particular variable.

The life of an individual is simulated by applying to him or her, during each unit of time (month or trimester), the probabilities appropriate to his or her situation during the period in question. Figure 34–1 gives the schematic representation of such a model, which also incorporates variables for fertility control. A circle corresponds to a calculation (increment change in a duration variable, for example), a triangle to a "yes/no" test, and a rectangle to a random drawing. Consider a woman aged 15, still single. Step one is to

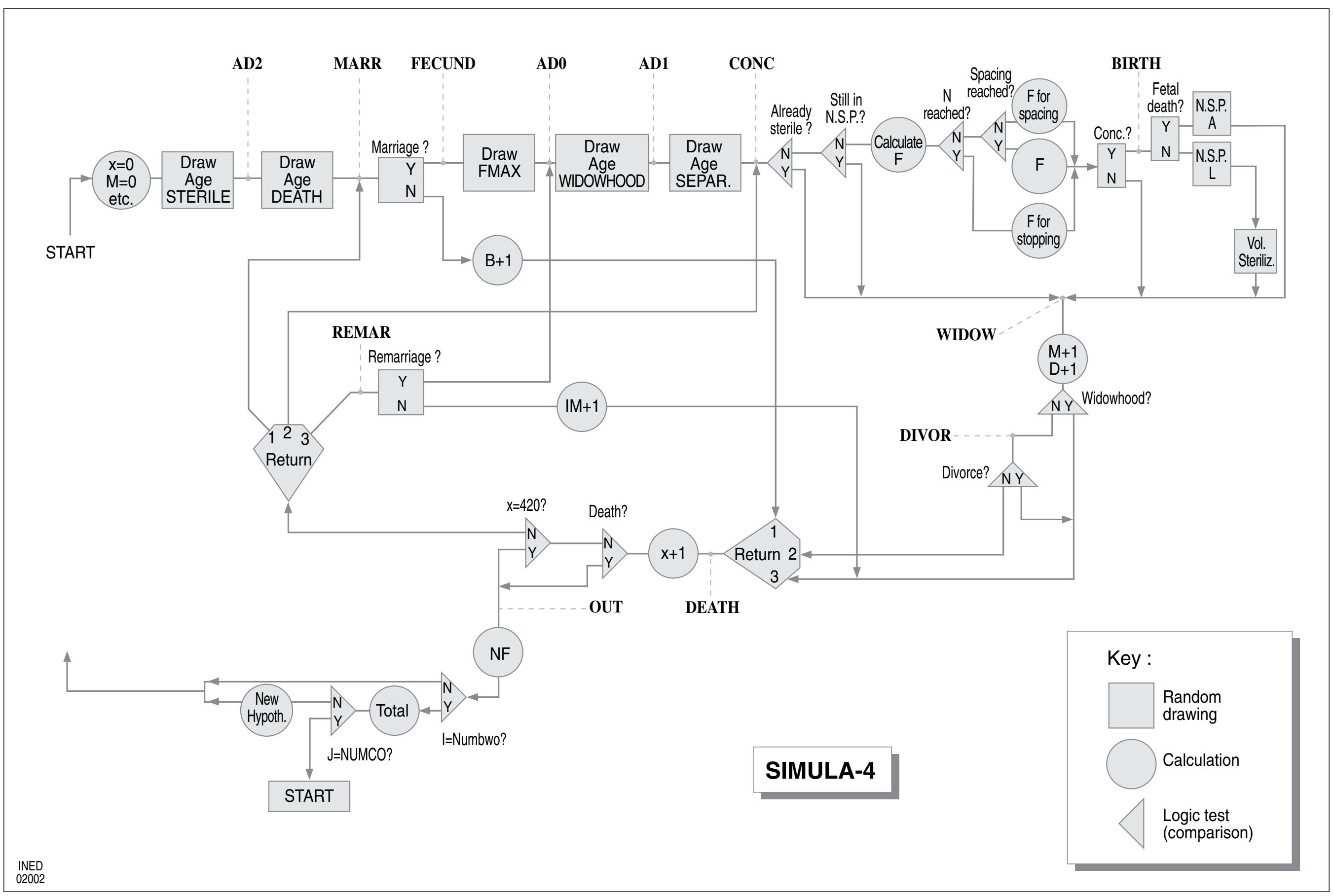


FIGURE 34–1 Schematic flow chart of the Simula program. Source: Leridon (1977).

determine by random drawing in distribution functions adapted to the population under examination, an age at death and an age of onset of permanent sterility, the assumption being made that these two events are unrelated to the reproductive history. The same procedure is followed for age at widowhood or at divorce (after marriage). The woman is then exposed to the risk of marriage during the first period (first month of her 15th year), then in each successive period. As long as the woman is unmarried, we stay on a short program loop in which we simply check that age at death and birthday 50 are not reached. Once marriage occurs, the woman is presumed to be exposed to the risk of conception. We check that she is not in a nonsusceptible period (which cannot be the case before the first pregnancy) and is not already sterile (permanently), and calculate her fecundability (which varies from one woman to another and with age); we draw to determine whether or not there is a conception, and draw again to determine the outcome of the pregnancy. If the couple is planning its fertility, a special module allows contraception to be included by reducing the value of fecundability, by a greater amount if the desired number of children has already been reached (stopping contraception) than if it has not yet been reached (spacing contraception). The length of the nonsusceptible period is then determined from an appropriate distribution (depending on the outcome of the pregnancy and the duration assumed for breast-feeding). The main loop is ended by a decision on voluntary sterilization, and the checking of ages at widowhood, divorce, and death. In the case of widowhood or divorce, the process starts again at the stage of marriage.

This process is repeated several thousand times, building up a sample of birth histories comparable to what would be obtained with a retrospective survey in a real population and which can be analyzed in the same way. As in any sample, the results have a margin of uncertainty (variance), which represents a disadvantage but also an advantage: a disadvantage because to reduce this margin of error it is necessary to work with large numbers; an advantage because it is also possible to calculate the variance estimates by repeating the process with the same parameter values.

The following sections contain examples of results obtained with a model of this kind.

4. The Princeton Indexes

As part of a vast project of comparative research on the demographic transition in Europe, Coale and his colleagues at Princeton proposed using a set of three indexes to separate the respective roles of marriage and marital fertility in the evolution of general fertility (Coale and Watkins, 1986). These were in fact the only parameters that could be analyzed given the available data. The approach is based on a double standardization, with calculation of a proportion of married women standardized for age, using reference fertility rates as weighting coefficients, and an index of marital fertility standardized for age, using the age-specific marriage rates as weighting coefficients. The general fertility index is then equal to the product of these two indexes and can be used to compare the observed fertility with that of a reference population. For each age a, let:

$w(a)$ number of women in the population;
$w_m(a)$ number of married women in the population;
$h(a)$ fertility rate in a reference population (fertility of Hutterite women);
$f(a)$ fertility rate in the population;
$f_m(a)$ marital fertility rate in the population

All births are assumed to occur within marriage. The indexes calculated are:

Index of marriage:

$$I_m = \frac{\sum w_m(a)h(a)}{\sum w(a)h(a)}$$

Index of marital fertility

$$I_g = \frac{\sum w_m(a)f_m(a)}{\sum w_m(a)h(a)}$$

Index of general fertility

$$I_f = \frac{\sum w(a)f(a)}{\sum w(a)h(a)} = I_m I_g$$

(we can verify that I_f is equal to the product of $I_m I_g$ by developing the two formulas and, bearing in mind the assumption of no extramarital fertility, $w_m f_m = wf$).

The fertility of the Hutterites (an anabaptist sect living in the northern United States) was chosen as the standard since this population has one of the highest fertility levels on record, along with that of female French Canadians in the 17th and 18th centuries (see Table 34–4).

5. The Coale and Trussell Schedules

As the first stage of research into age-specific model schedules of fertility, Ansley Coale began by examining age patterns of marriage. It is obvious that when the overwhelming majority of births occur within marriage, the first part of the age-specific (general) fertility curve depends essentially on that of marriage. A comparison of the age distributions of first marriage frequencies in various populations led Coale (1971) to conclude that they could all be deduced from each other by means of the following transformation:

$$G(a) = CG_S\left(\frac{a - a_0}{k}\right)$$

$G(a)$ proportion ever-married at age a;
G_s proportion observed in a standard population;
C proportion ultimately (at age 50) ever-married;
a_0 age at earliest marriage;
k defines the *pace* of nuptiality (the number of years of nuptiality equivalent to 1 year in the standard population S).

The function G_s selected represents the situation in the Swedish population in the mid 19th century, after dividing by a factor C_s (proportion ultimately married), so that its value reaches 1.0 40 years after the earliest marriages.

If we are only interested in the tempo of nuptiality, $G(a)$ depends on just two parameters: a_0, which is easily located, and k.

Likewise, the frequency function $g(a)$ can be deduced from the function $g_s(a)$. Ansley Coale also defined an algebraic expression of $g(a)$, which is a double exponential:

$$g(a) = \frac{k_1}{k}\exp\left[-\frac{k_2}{k}(a - a_0 - k_3 k) - \exp\left\{-\frac{k_4}{k}(a - a_0 - k_5 k)\right\}\right]$$

(with: $k_1 = 0.19465$, $k_2 = 0.174$, $k_3 = 6.06$, $k_4 = 0.2881$, $k_5 = 6.06$),

hence:

$$g(a) = \alpha \exp[\beta - \gamma a - \exp(\beta' - \gamma' a)]$$

Subsequently, Coale (1977) linked this formulation for g(a) to the notion of *marriage circles* (*cercles de nuptialité*) that Henry had developed in 1968.

Turning next to age-specific fertility, he merely defined the *model schedules* without taking into account the actual level of fertility. Given the function $G(a)$ defined earlier (proportion ever-married at age a), we can write:

$$f(a) = G(a)\, r(a)$$

$f(a)$ fertility rate at age a;
$r(a)$ marital fertility rate at age a.

This formula assumes there are no:

—extramarital births
—union dissolutions, through widowhood or divorce.

These two restrictions can be removed by slightly modifying functions $G(a)$ and $r(a)$.

As was done for the proportion ever-married, the marital fertility rates $r(a)$ are defined on the basis of a model schedule, representing a situation of natural fertility (no contraception). The relation is as follows (Coale and Trussell, 1974):

$$\frac{r(a)}{n(a)} = M\exp[m\nu(a)]$$

$n(a)$ natural fertility rate at age a;
M coefficient of total reduction in fertility;
m parameter representing the size of the fertility reduction resulting from use of contraception;
$v(a)$ modulating factor of m, by age a.

Since we are concerned only with the age pattern of fertility, not with its level, M can be disregarded. Its inclusion in Equation 2 expresses the fact that even when no attempt is being made to control births (m = 0), levels of fertility can vary widely. For this case, however, Coale admits a unique profile by age and hence a single-year schedule $n(a)$. The schedule $v(a)$ is also unique. This profile is intended to allow for the fact that when fertility is voluntarily controlled, the fertility rates that experience the largest reduction are those for older women who are attempting to limit the size of their family.

Two standard schedules are thus selected, for $n(a)$ and $v(a)$. Their M_T and V_T values are given in Table 34–2.

Ultimately, therefore, $r(a)$ depends only on parameter m (and the scale factor M).

Finally, by setting:

$$\sum_a f(a) = 1$$

we obtain:

$$f(a) = G_S\left(\frac{a - a_0}{k}\right) n_T(a)\exp[m\nu_T(a)]$$

TABLE 34–2 Values of by Age

	20–24 years	25–29 years	30–34 years	35–39 years	40–44 years	45–49 years
$n_T(a)$	0.460	0.431	0.396	0.321	0.167	0.024
$v_T(a)$	0	–0.316	–0.814	–1.048	–1.424	–1.667

The final set of schedules thus has three specified parameters: a_0, k, and m. The published tables (Coale and Trussell, 1974) are ordered by the values of these parameters, but other entries can also be used (various parameters that are easily calculated from observed distributions):

—Mean age of mother at birth (MEAN);
—Standard deviation of age at birth (STDEV);
—Ratio of fertility at ages 15–19 to fertility at ages 20–24 years (R1);
—Median age (MED);
—Coefficient of asymmetry (SKEW);
—Ratio of the average parity of women at 15–19 to that of women at 20–24 (PAR1);
—Same ratio for women at 20–24 and 25–29 (PAR2).

Using these entry parameters, we can look for the pattern that is closest to an actual distribution and thus determine, for example, the value of parameter m. That this parameter is a reasonably good indicator of fertility control in the population can be verified by comparing it with the value of the fertility rate at ages 40 to 44 or ages 35 to 39: the value of m rises steadily when these rates fall. Figure 34–2 compares the values of m with those of the average age at last birth calculated for a set of German villages studied by Knodel (1986), for marriages between 1740 and 1900. The experience of each village over time can thus be followed. We see that at the onset of the transition the mean age at last birth often rises slightly while m falls; next, the mean age falls steadily while m increases. Thus we go from a mean age close to 40 years with $m = 0$, to an age of around 38 years when m is 0.5.

6. The Bongaarts Model

Bongaarts (1976, 1983) has suggested a simplified model, based on a multiplicative decomposition of the fertility rate at a given age (or of the sum of these rates):

$$\text{TFR} = C_m C_c C_a C_i \text{ TBF}$$

where:

C_m index of nuptiality (1 if all women of reproductive age are married, 0 if there is no marriage);
C_c index of contraception (1 if there is no contraception, and 0 if all fertile women use a method that is 100% effective);
C_a index of abortion (1 if there is no abortion, 0 if all pregnancies are interrupted);

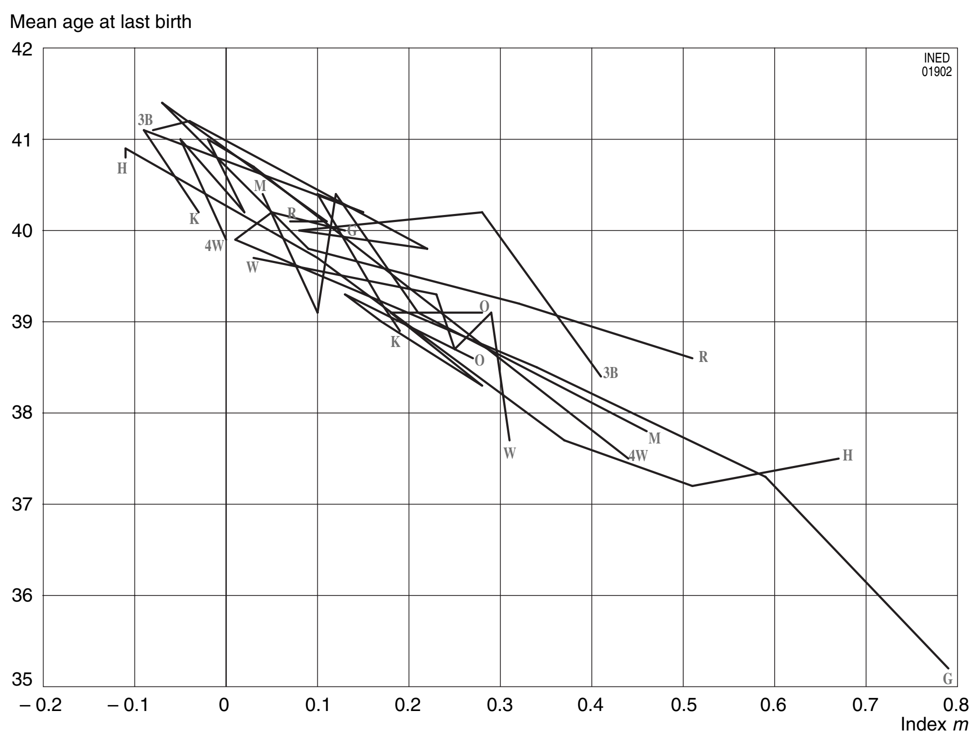


FIGURE 34–2 Correlation between mean age at last birth and the m index (German villages, marriages between 1740–1749 and 1875–1899). Source: Leridon (1989), based on Knodel (1986).

C_i index of postpartum nonsusceptibility (1 if there is no breast-feeding and no abstinence, 0 if the duration of sterility is infinite);
TBF total fecundity at biologic maximum (in this case, the sum of the age-specific rates is equal to 15.3 children).

Each of these indexes can also be expressed as the ratio of two fertility rates, with and without the effect of the factor under consideration:

C_m = TFR/TM = Total fertility rate/Marital fertility rate
C_c C_a = TM/TNM = Marital fertility rate/Natural fertility rate
C_i = TNM/TBF = Natural fertility rate/Total fecundity at biologic maximum

For example, for the sum of the rates (equal to the total fertility rate or the completed fertility) we obtain the following values:

	TFR	=	C_m	x	C_c	x	C_a	x	C_i	x	*TBF*
Crulai (Normandy) (marriages 1674–1742)	5.83		0.566		1.000		1.000		0.673		15.3
Hutterites (USA) (marriages 1921–1930)	9.15		0.733		1.000		1.000		0.816		15.3
Bangladesh (1976)	6.73		0.853		0.928		1.000		0.556		15.3
South Korea (1970)	3.90		0.586		0.780		0.848		0.658		15.3
France (1972)	2.40		0.519		0.334		0.973		0.930		15.3

Table 34–3 presents the results of calculating the three indexes C_m, C_c, and C_i, in various countries (World Fertility Surveys [WFSs] of the 1970s). Since total fertility rate (TFR) is also available in these surveys, the value of the *total fecundity at biologic maximum* (TBF) can now be calculated, so that the equation is no longer overdetermined. The values are seen to range from 10.0 to approximately 16.5, instead of the expected 15.3. The equation is thus imperfect, due to marked variation between populations in the factors omitted from the three C coefficients, errors in estimating each coefficient, disruptions caused by interactions between factors, or a combination of these explanations. Bongaarts' formula should thus only be used to obtain rough estimates of the different variables.

II. RESULTS: LEVELS OF FERTILITY

1. Observed Levels of Natural Fertility

Many studies in historical demography, notably those using the family reconstitution method, have made it possible to measure with precision the level of fertility in populations where fertility regulation was either absent or ineffective. Table 34–4 presents some results from this research.

These data show that even within a geographically limited area (France) and during the same period, natural fertility levels can vary widely between human population groups. The results of the simulations presented in the next section confirm that this variability can be explained by variables other than those of fertility regulation. No direct conclusion can therefore be reached (from the differences in the levels of overall fertility) concerning the wishes of couples in these populations regarding the desire to limit or not the number of their children.

2. Study of Possible Levels for Natural Fertility

The simulation models presented earlier can be used to explore the possible levels for natural fertility, by combining the values of the various parameters as observed or estimated for historical populations. Table 34–5, constructed using the SIMNAT model, which is a simplified version of SIMULA (developed by the present author from Jacquard's model) (see Figure 34–1), displays the number of children per woman under the stated assumptions and with no mortality or other causes of premature union dissolution.

According to this model, a woman who remained fecund and married between ages 15 and 45, would have, on average, 17.5 children, if she did not breast-feed. If the effect of sterility is introduced, the maximum is reduced to 15.2, a value that is close to Bongaarts' *biologic maximum fecundity*. In practice, there is no known population in which all women marry at age 15 and do not breast-feed their children. Thus the number of 15 children, although possible at an individual level, is not possible at the level of an entire population or subgroup. The values observed for real populations are between five and 10 children per woman and are thus obtained by combining the

TABLE 34–3 Parameter Values of the Bongaarts Model in Various Populations

Country	Marriage index (Cm)	Contraception index (Cc)	Infecundability index (Ci)	Total fertility rate (TFR) observed	TBF (estimated)
AFRICA					
Ghana	0.84	0.94	0.64	6.52	12.90
Kenya	0.81	0.95	0.69	7.71	14.53
Lesotho	0.80	0.96	0.64	5.71	11.62
Senegal	0.92	0.98	0.65	7.27	12.41
Sudan (North)	0.79	0.98	0.69	6.32	11.84
Regional mean Africa	*0.83*	*0.96*	*0.66*	*6.66*	*12.66*
AMERICAS					
Colombia	0.68	0.78	0.82	6.25	14.37
Costa Rica	0.65	0.45	0.89	4.12	15.84
Dominican Republic	0.77	0.80	0.81	7.05	14.13
Guyana	0.74	0.74	0.87	5.07	10.65
Haiti	0.69	0.89	0.69	5.85	13.81
Jamaica	0.74	0.68	0.85	5.06	11.83
Mexico	0.76	0.88	0.79	7.40	14.01
Panama	0.71	0.60	0.80	5.24	15.37
Paraguay	0.68	0.78	0.78	5.71	13.81
Peru	0.71	0.92	0.70	6.90	15.09
Trinidad and Tobago	0.69	0.58	0.87	3.48	10.00
Venezuela	0.76	0.75	0.79	7.45	16.54
Regional mean Americas	*0.72*	*0.74*	*0.81*	*5.95*	*13.79*
ASIA					
Bangladesh	0.90	0.94	0.52	6.04	13.72
Fiji	0.71	0.69	0.81	4.50	11.33
Indonesia	0.80	0.77	0.55	4.64	13.70
Jordan	0.84	0.93	0.77	9.33	15.51
Korea (South)	0.63	0.78	0.66	4.96	15.29
Malaysia	0.68	0.77	0.88	4.97	10.79
Nepal	0.85	0.98	0.56	6.10	13.08
Pakistan	0.83	0.98	0.64	6.30	12.11
Philippines	0.66	0.78	0.74	5.87	15.40
Sri Lanka	0.58	0.78	0.60	3.87	14.24
Syria	0.79	0.96	0.78	9.09	15.37
Thailand	0.69	0.71	0.6	4.93	15.72
Regional mean Asia	*0.75*	*0.84*	*0.68*	*5.93*	*13.85*
Overall mean	0.75	0.82	0.73	6.11	13.62

World Fertility Survey, rural areas only.

Source: Susheela Singh, John B. Casterline, and John Cleland, 1985. The proximate determinants of fertility: Sub-national variations, *Population Studies*, vol. 39(1), p.113–136.

hypotheses of early or late marriage and short or long breast-feeding. These two variables alone can thus account for a doubling of fertility levels, all of which, however, are considered to represent *natural* fertility regimes.

3. From Natural Fertility to Controlled Fertility

By introducing into these models some fertility control parameters (desired number of children, effectiveness of stopping or spacing contraception), it is also possible to evaluate the fertility-reducing impact of the contraceptive methods used. The advantages include that of being able to measure the impact both at couple level (how many couples would have more children than they want?) and at the aggregate level (by how much is fertility reduced?). Thus use of a contraceptive that is 95% effective, following the birth of the second child, will greatly reduce the overall fertility of the population (from eight to three children, for example) yet will leave a large majority of couples dissatisfied, because they will have had at least one more child than expected!

TABLE 34–4 Data on Natural Fertility

Place	Period	Mean number of children	
		Completed families of women married age 20[a]	All women surviving to age 50[b]
French Villages			
Ile-de-France	M 1740–1779	10.1	6.1
Tourouvre (Perche)	M 1665–1714	8.6	6.0
Crulai (Normandy)	M 1674–1742	8.3	5.6
Thézels-Saint-Sernin (Quercy)	M 1700–1791	6.6	3.7
Populations of European origin			
Hutterites (United States)	M 1921–1930	10.9	9.5
Canada	M 1700–1730	10.8	8.0
Amish (United States)	G 1900–1920	7.0	6.3
Non-European populations			
Martinique (French Caribbean)	G 1914–1928	7.9	5.4
Sine Saloum (Senegal)	N 1963–1965	6.1	6.7
Villages near Bombay (India)	N 1954–1955	5.4	5.3

M, year of marriage; G, years of birth of women; N, years of birth of children
[a] Sum of marital fertility rates ages 20–49 years.
[b] Sum of general fertility rates ages 15–49 years.
Source: Adapted from Leridon, 1977.

TABLE 34–5 Completed Family Size with Different Levels of Natural Fertility (No Mortality and No Marriage Dissolutions)

Permanent sterility[a]	Postpartum nonsusceptible period[b]	Marriage at age 15	Distribution of marriage: early[c]	Distribution of marriage: late[d]
Normal	No breast-feeding	15.2	12.8	8.1
	Short breast-feeding	12.3	10.3	6.5
	Long breast-feeding	10.0	8.4	5.3

Shared assumptions:
Mean fecundability = 0.25 at ages 20–24 years (with a fecundability of half this value, the results have to be multiplied by a coefficient of around 0.8). Fecundability varies with age and between women; mean intrauterine mortality: 15% (varying with age).
Additional assumptions:
[a] Normal sterility: proportion of permanently sterile couples by age, as estimated by Louis Henry (mean age at onset of sterility: 41 years).
[b] Length of postpartum nonsusceptible period: with no breast-feeding: mean = 2 months; with short breast-feeding (15 months average): mean = 7 months; with long breast-feeding (24 months average): mean = 13 months
[c] Early marriage: Asiatic and African type: mean age: 19 years, no permanent celibacy.
[d] Late marriage: European type: 25 years, proportion never married: 8%
Source: Leridon, Henri, 1977. *Human fertility. The basic components*. Chicago, The University of Chicago Press, 202 p.

Table 34–6 shows the stages of the passage from a situation of noncontrolled fertility (in this example the situation in France in the middle of the eighteenth century) to a contemporary regime, with fertility in the region of two children per woman. This relies on the assumption that after the birth of their second child, couples use contraception that is 99% effective, this measure of effectiveness including the possible recourse to induced abortion. A 90% effectiveness, though it might appear already high, would lead to an average close to four children, well in excess of the two children desired.

TABLE 34–6 Comparison of Various Fertility Regimes (France, Historical and Contemporary): Results of the SIMULA Model

	Historic regime (no voluntary fertility limitation)	Contemporary regime (desired number of children: 2). Effectiveness of contraception:		
		0	90%	99%
Life expectancy (female)	25 years	70.5 years	75 years	75 years
Probability of survival at 15 years	0.46	0.95	0.98	0.98
Mean age at marriage	25 years	21.5 years	21.5 years	21.5 years
Probability of conception/month	0.25	0.25	0.025	0.0025
Duration of breast-feeding	2 years	1 year	0	0
Mean number of live births, per woman at 15 years	4.1	8.6	3.8	2.0
Mean number of children surviving to 15 years, per woman at 15 years	1.9	8.4	3.8	2.0
Net reproduction rate	0.93	4.1	1.8	0.95
Annual rate of natural increase	–0.2%	+4.5%	+2.0%	–0.2%
Proportion of married women having more than desired number of children	—	98%	86%	18%

Source: Henri Leridon, 1977. *Human fertility. The basic components*. Chicago, The University of Chicago Press, 202 p.

Acknowledgment

Translated from French by Godfrey I. Rogers.

References

Barrett John C., 1969. A Monte Carlo simulation of human reproduction, *Genus*, vol. 25(1–4), p. 1–22.

Bongaarts John, 1976. Intermediate fertility variables and marital fertility rates, *Population Studies*, vol. 30(2), p. 227–242.

Bongaarts John, and Potter Robert G., 1983. *Fertility, biology and behavior: an analysis of the proximate determinants*. New York, Academic Press, 230 p.

Coale Ansley J., 1971. Age pattern of marriage, *Population Studies*, vol. 25(2), p. 193–214.

Coale Ansley J., 1977. The development of new models of nuptiality and fertility, *Population*, vol. 32, special issue L. Henry, p. 131–154.

Coale Ansley J., and Trussell, James, 1974. Model fertility schedule, *Population Index*, vol. 40(2), p. 185–258.

Coale Ansley J., and Watkins Susan Cotts, 1986. *The decline of fertility in Europe*. Princeton (New Jersey), Princeton University Press, 484 p.

Davis Kingsley, and Blake Judith, 1956. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol. 4(4), p. 211–235.

Dyke Bennett, and MacCluer, Jean W. (eds.), 1973. *Computer simulation in human population studies*. New York, Academic Press, 518 p.

Freedman Ronald, 1974. *Community-level data in fertility surveys*. London, WFS, (WFS Occasional Papers).

Henry Louis, 1953. Fondements théoriques des mesures de la fécondité naturelle, *Revue de l'Institut International de Statistique*, vol. 21(3), p. 135–151.

Henry Louis, 1957. Fécondité et famille. Modèles mathématiques, *Population*, vol. 12(3), p. 413–444.

Henry Louis, 1961. Fécondité et famille. Modèles mathématiques II, *Population*, vol. 16(1), p. 27–48.

Henry Louis, 1961. La fécondité naturelle. Observations, théorie, résultats, *Population*, vol. 16(4), p. 625–636.

Henry Louis, 1968. Problèmes de la nuptialité: Considérations de méthode, *Population*, vol. 23(5), p. 835–844.

Horvitz D. G., Giesbrecht F. G., Shah B. V., and Lachenbruch P. A., 1971. POPSIM, a demographic microsimulation model, *in*: IUSSP (ed.), Actes du Congrès international sur la population, London (vol. 1), p. 95–106. Liège, IUSSP, 844 p.

Jacquard Albert, 1967. La reproduction humaine en régime malthusien: un modèle de simulation par la méthode de Monte Carlo, *Population*, vol. 22(5), p. 897–920.

Jacquard Albert, and Leridon Henri, 1973. Simulating human reproduction: How complicated should a model be? *in*: Bennett Dyke, Jean W. MacCluer (eds.), *Computer simulation in human population studies*, p. 241–250. New York, Academic Press, 518 p.

Knodel John, 1986. Demographic transition in German villages, *in*: Ansley J. Coale, Susan Watkins (eds.), *The decline of fertility in Europe*, p. 337–389. Princeton, Princeton University Press, 484 p.

Leridon Henri, 1977. *Human fertility. The basic components*. Chicago, The University of Chicago Press, 202 p.

Leridon Henri, 1989. Fécondité naturelle et espacement des naissances, *Annales de démographie historique* 1988, p. 21–33.

Leridon Henri, Toulemon Laurent, 1997. *Démographie. Approche statistique et dynamique des populations*. Paris, Economica, 440 p. (Économie et statistiques avancées).

Potter Robert G., and Sakoda James, 1966. A computer model of family building based on expected values, *Demography*, vol. 3(2), p. 450–461.

Ridley Jeanne C., and Sheps Mindel, 1966. An analytic simulation model of human reproduction with demographic and biologic components, *Population Studies*, vol. 19(3), p. 297–310.

Sheps Mindel, and Menken Jane, 1973. *Mathematical models of conception and birth*. Chicago, University of Chicago Press, 428 p.

Singh Susheela, Casterline John B., and Cleland John, 1985. The proximate determinants of fertility: Sub-national variations, *Population Studies*, vol. 39(1), p. 113–136.

CHAPTER

35

Induced Abortion

GIANPIERO DALLA ZUANNA

Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy

ABORTION DURING THE DEMOGRAPHIC TRANSITION

Voluntary abortion (referred to here for reasons of convenience as abortion) is the end point of a complex succession of acts and decisions which can be mapped as indicated in the schema below:

From Conception to Abortion: Stages in Decision Making

Conception	Abortion	
	Yes	No
Wanted	A	B
Unwanted	C	D

Studying abortion, involves studying the four situations (A, B, C, and D) represented in the table within the population, in an attempt to clarify the underlying mechanisms of the different choices (or of nonchoice) by individuals and couples. Situation A, which is undoubtedly uncommon, occurs when, after a child has been willingly conceived, the couple decides to prevent the birth. For example, the couple may decide that the result of conception does not correspond to their expectations. This may be a reason for selective abortion (depending on sex) or a so-called *therapeutic* abortion (following an antenatal examination showing congenital malformations).

The incidence and relative weight of the three other situations are determined by the population's position on the path of demographic transition. During the *pretransition stage*, most couples consider conception within marriage as a normal thing, and the arrival of a child is rarely perceived as a threat to the well-being of the parents or to children already born. Moreover, health and hygiene are generally deplorable, and for a woman an induced abortion is more of a risk than childbirth. During this pretransition stage, undesired conception (which at certain periods or in certain circumstances may be quite frequent, especially outside wedlock) leads to unwanted births, which, in certain social circumstances may result in the abandonment (exposure) of the child or even to infanticide.

During the *demographic transition*, a growing number of couples wish to limit the number of children they have. In countries that experienced an *early demographic transition* (EDT) and in which the drop in fertility took place essentially before or at the time of World War I, lack of access to effective contraceptives led to an increase in the number of undesired births and at the same time to an increase in the number of abortions which, even if illegal, had become less dangerous for mothers, thanks to progress in medicine and hygiene. The combination of these factors led to an increase in the frequency of abortions. In these same countries, during the subsequent phase (*posttransition*), recourse to abortion declined as more effective methods of contraception were made available.

Table 35–1 shows a number of empirical examples. These data represent those countries in which abortion is available on demand or for socioeconomic reasons

TABLE 35–1 Fertility and Abortion when Abortion becomes Legal on Economic and Social Grounds or on Demand, and Abortion Rates in 1987

Country	Year abortion liberalized	Year TAR[a] calculated	Year TFR[a] calculated	TFR[b]	TAR[a]	TAR[a] towards 1987
Bulgaria	1956	1960	1956	2.14	1.02	2.26
Canada	1969	1973	1973	2.33	0.31	0.36
China	1974	1979	1973	4.74	1.32	1.36
Cuba	1965	1970	1968	4.30	1.41	2.05
Czechoslovakia	1962	1965	1962	2.35	0.92	1.70
East Germany	1972	1975	1972	1.77	0.88	0.82
Denmark	1973	1977	1973	1.93	0.85	0.64
Finland	1970	1973	1970	1.83	0.78	0.41
France	1975	1978	1975	1.96	0.48	0.46
Iceland	1975	1980	1975	2.61	0.37	0.42
Italy	1978	1982	1978	1.92	0.69	0.54
Hungary	1956	1960	1956	2.40	2.68	1.32
Japan	1948	1952	1950	3.59	1.42	0.65
Mongolia	1989	1991	1992	4.60	1.54	—
Netherlands	1981	1985	1981	1.56	0.18	0.18
Norway	1975	1979	1975	1.99	0.62	0.59
Poland	1956	1965	1956	3.30	0.85	0.52
Singapore	1974	1980	1974	2.38	1.01	1.05
Switzerland	1974	1980	1974	1.89	0.72	0.69
USA	1973	1976	1973	1.89	0.85	0.94

[a] TAR, total abortion rate (average no. of abortions per woman), see note 1.
[b] TFR, total fertility rate (average no. of children per woman).
Source: For abortion: United Nations, 1992, 1993, 1995. Abortion policies: a global review. New York, United Nations, vol. I: Afghanistan to France, 158 p; vol. II: Gabon to Norway, 243 p.; vol. III: Oman to Zimbabwe, 236 p.; for fertility: United Nations, Demographic Yearbook, different years.

and data on fertility and abortion are available for the years immediately following the liberalisation of abortion laws. Excluding Hungary from the calculation, the correlation coefficient between fertility and total abortion is 0.753, and the two indicators are linked by the following linear relation:

$$\text{TAR} = 0.125 + 0.285 \times \text{TFR}^{1}$$

Thus presented, the trend appears excessively simple. In reality, the pattern has varied considerably from one country to another. Although there has been insufficient research in this area, it has been determined that during the demographic transition, the balance between abortion and traditional contraception has varied considerably (Frejka, 1985; Mundigo, 1993). Generally speaking in countries where abortion was legalized at a time when fertility was still high, abortion seems to have played a predominant role to the extent of having been the principle method of birth control. In the other countries, abortion was for most couples a resort used only when contraception had failed; limitation of births essentially involved periodic abstinence or coitus interruptus or (in only few countries) use of condoms.

It may appear as though the correlation between the two measures is high simply because levels of abortion and fertility are themselves high where the rates of pregnancy are high. In reality, abortion rates were high where legislation was introduced while fertility was also high and it is in these same countries that several decades after legislation, abortion levels remain the highest (the final column in Table 35–1) although fertility has dropped substantially. It appears that in countries where abortion was legalised while fertility

[1] The usual measurements of *abortion* are very simple: the two most common indicators are the *abortion rate* (induced abortions/women aged 15–49 or 15–44) and the *abortion ratio* (induced abortions/pregnancies), in which the denominator is the total number of induced abortions and of births in the same period. Obviously, both these measurements give different and complementary values for the incidence of induced *abortion* and are used in conjunction to describe the phenomenon. The *abortion rate* fails to take into account the age structure of the female population of childbearing age. To offset this drawback, it is possible to use the *total abortion rate* (TAR), or average number of abortions per woman, which is analogous to the total fertility rate (TFR), calculated as the sum of the *abortion* rates per age (or age groups). If the age structure of women of childbearing age is roughly rectangular, it is possible to estimate TAR by multiplying the *abortion rate* by the number of years involved (for example, by multiplying, as in Table 35–1, the abortion rate of women aged 15–49 by 35). If there is little variance in the age groups structure of the countries being compared, they may be thus transformed into TAR to render the measurement more significant.

remained high, the spread of contraception was less rapid than in countries where contraception was introduced while abortion remained illegal.

However, this chart is rather superficial and cannot fully represent the significance of the differences observed in levels of abortion from one population to another. For example, in the Soviet Union, and more generally in the Slav countries, or in Japan, the success of abortion as a means of birth control may be linked to factors other than the mere legislative calendar: possibly, among these populations, the pretransition *cultural environment* was more favorably disposed toward abortion than in the other EDTs (Frejka, 1985; Blayo, 1993; Avdeev *et al*., 1994; Otha and Sawayama, 1994; Dalla Zuanna, 1996; Stloukal, 1996).

Another challenge to the linear representation of the trend comes from *the still high levels of abortion in some western EDTs, despite the widespread availability of modern contraceptive techniques* (Dalla Zuanna, 1996). In the early 1990s, in France, Denmark, Sweden, the United Kingdom, and Norway, to cite but a few, although 70% of couples used modern contraceptive techniques (sterilization, the pill, intrauterine devices [IUDs]), the average number of abortions per woman was still 0.3 to 0.4. As young nulliparous women largely sought these, the explanation may be found in the young age at which they begin having sexual relations, when contraceptive protection is often inadequate. Among couples who have already had all the children they want, abortion is much less common in these countries than in those EDTs in which modern contraceptive techniques are less widespread. However, the existence of a direct relationship between the availability of modern contraceptives and low levels of abortion is debatable.

To conclude, the positioning of abortion during the demographic transition in EDTs is controversial. Causal structures of the simple type,

increase in undesired conceptions $\rightarrow$ *more abortions*

or

wider availability of modern contraceptive techniques $\rightarrow$ *less abortions*

must be viewed with caution and evaluated within diverse cultural and social contexts. As we will discuss in subsequent sections, in accounting for differences in abortion, it is not sufficient to consider merely those variables directly linked to reproductive behavior.

In countries that experienced a late demographic transition (*LDT*), the role played by abortion is different because the widespread desire to limit the number of births in marriage appeared at a time when modern contraceptive techniques were already (at least from the technical angle) available. The diversity of conditions is due to four factors: *traditional views of abortion; current legislation; the availability of means of contraception; and the capacity/willingness of couples to practice effective contraception*.

The observation made of the Slav countries remains valid: When traditional moral values do not disapprove of abortion and when legislation is (either formally or de facto) liberal before fertility begins to decline, the number of abortions may be high, although modern contraception is fairly widespread, as for example in Cuba, China, or Viet Nam (Feng, 1994; Farina, 1996). However, in most LDTs, abortion legislation is very restrictive (Table 35–2), even if the predominant religions are morally tolerant toward it (whether during the entire pregnancy, as in the case of Buddhism and Hinduism, or during only the first 4 months of pregnancy, as in the case of certain Islamic schools of thought). In these countries, abortion is rarely as frequent as in Japan or in the Slav countries. However, in many LDTs with restrictive legislation and in the throes of demographic transition, clandestine abortions are far from negligible, despite the wide availability of modern family planning techniques. This is attributable to the fourth of the factors: the capacity/willingness of couples to practice effective contraception. Indeed, some couples who want to limit births may find it difficult to adopt the rational choice (even when there are facilities to assist with family planning): the use of effective contraception. In numerous LDTs (especially in Latin America), the early age of marriage and low use of contraception to space births mean that couples achieve the desired number of children while the mother is still young and will remain fertile for another 20 years; in addition, gender relations are such that the couple's equilibrium is not very compatible with women taking responsibility for contraception (Pinnelli and Milan, 1996; Paiewonsky, 1999). Lastly, the manner in which birth control is

TABLE 35–2 Proportion of Developing Countries with Authorized Abortion (Early 1990s)

Recognized grounds for abortion	Proportion, %
To save the mother's life	91
To preserve the mother's physical health	52
To preserve the mother's mental health	35
Sexual violence or incest	26
Fetal malformation	23
Request	6

Source: WHO, 1994. *Abortion: a tabulation of available data on the frequency and mortality of unsafe abortion*, 2nd ed. Geneva, World Health Organization.

introduced to a community rarely respects the culture of the target population and may give rise to suspicion or even refusal among potential users.

As we have already observed, in many LDTs, the combination of the factors described seems to lead to situations in which abortion is widely practiced and the proportion of illegal abortions, often carried out in deplorable sanitary conditions, is high. Nevertheless, caution is required because it is often difficult to estimate the scale of illegal abortion. Estimates by different sources and using different methods yield very different results, whether for abortion itself or for its consequences, particularly maternal mortality. Consequently, the data available must be taken with caution, even when they come from the most reputable organizations or from solemn declarations made at major international conferences (Mauldin, 1994); this is all the more so since they are not necessarily exempt from the constraints of diplomacy and political compromise. This same caution is necessary when examining the data gathered by Singh and Henshaw (Table 35–3), most of which concern LDTs and adopt the average hypothesis proposed for the early 1990s by the World Health Organization (WHO, 1994).

Thus, there is no doubt that in many LDTs, abortion has become a dramatic problem since the decline in fertility. As a result, research on abortion has begun to seek data, focusing on LDTs at the expense of EDTs (Table 35–4). Studies have multiplied in a number of directions, although it is possible to identify three major ones: (1) estimation of the number of abortions in the absence of satisfactory registration; (2) analysis of the pathways to abortion, using a statistical–quantitative approach; and (3) favoring qualitative techniques and adopting an anthropologic approach to examine an individual's decision as the expression of a choice that has been historically determined by particular sociodemographic contexts. We will describe the *how things stand* from these three angles, emphasizing methodologic aspects rather than results. To do so, we will sort the possible themes, setting aside aspects that have been much debated recently, such as

TABLE 35–3 Estimate (Circa 1990) of No. of Abortions by Degree of Safety for Mothers, Abortion Rates, and Ratios

Regions	No. of abortions by sanitary conditions			Rates[a]	Ratios[b]
	Total	Good conditions	Poor conditions		
World total	***49,666,000***	***29,666,000***	***20,000,000***	**1.20**	**26**
Developed countries	17,769,000	15,409,000	2,360,000	2.10	55
Developing countries	31,897,000	14,257,000	17,640,000	0.97	20
AFRICA	3,820,000	30,000	3,790,000	0.80	11
East Africa	1,350,000	—(1)	1,350,000	0.93	12
Central Africa	190,000	—	190,000	0.37	5
North Africa	550,000	30,000	520,000	0.49	10
Southern Africa	240,000	—	240,000	0.72	14
West Africa	1,490,000	—	1,490,000	1.10	14
ASIA	24,362,000	15,132,000	9,230,000	0.98	22
Eastern Asia	11,820,000	11,820,000	—	1.07	31
Southeast Asia	4,262,000	1,412,000	2,850,000	1.18	25
South-central Asia	7,300,000	1,300,000	6,000,000	0.80	15
Western Asia	980,000	600,000	380,000	0.86	16
EUROPE	3,300,000	3,040,000	260,000	0.87	35
Eastern Europe	2,036,000	1,926,000	110,000	2.53	60
Rest of Europe	1,264,000	1,114,000	150,000	0.42	21
Former USSR	11,580,000	9,500,000	2,080,000	6.47	76
LATIN AMERICA	4,775,000	155,000	4,620,000	1.33	28
Caribbean	325,000	155,000	170,000	1.20	29
Central America	890,000	—	890,000	0.96	20
South America	3,560,000	—	3,560,000	1.48	32
United States and Canada	1,719,000	1,719,000	—	0.79	28
OCEANIA	110,000	90,000	20,000	0.52	17

[a] Estimated number of abortions per woman (abortions/W_{15-49}) × 35.

[b] Ratio of the number of abortions to the number of live births plus abortions: 100 × abortions/(births + abortions)

(1) When the numbers are low or unreliable, they have not been included or estimated.

Source: Susheela Singh and Stanley Henshaw, 1996. The incidence of abortion: a worldwide overview focusing on methodology and on Latin America, *in:* IUSSP (ed.). *Seminar on Socio-cultural and political aspects of abortion in a changing world, Kovalam, Trivandrum, India, 25–28 March 1996,* p. 1–33. Liège, IUSSP, discontinuous paging.

TABLE 35–4 Research Work on Abortion by Field of Interest, Selected Years Since 1972

Research	1972	1975	1980	1985	1990	1995	1996
Methodologic research (EDTs and LDTs)	5	6	5	1	2	2	2
Research on EDTs	60	41	39	25	25	21	21
Research on LDTs	6	6	13	7	12	16	16

EDT, countries experiencing an early demographic transition; LDT, countries experiencing a late demographic transition.
Source: Author's calculation based on figures provided by the *Population Index's* references.

induced early abortion by the suction method carried out at the first sign of a delayed period regardless of whether the cause is the fertilization of an ovule or sex-selective abortion depending on the sex of the children already born or that of the fetus.

1. Estimation of the Number of Abortions in the Absence of Regular Registration

The techniques used to estimate the number of abortions vary, particularly depending on the data available. We will adopt the classification of Singh and Henshaw (1996) to present those techniques of greatest interest from a methodologic angle.

a. Retrospective Surveys (Random or Selected Populations)

Estimating the number of abortions by retrospective survey, which might appear to be the simplest approach, is actually flawed by women's systematic tendency to report a lower number of induced abortions than those they have actually had: Whenever the results of such surveys are compared with those from other sources, a systematic underevaluation, which may exceed 50%, appears even in countries in which abortion has long been legal (Casterline, 1989; Henshaw, 1989; Jones and Forrest, 1992; Huntington *et al*., 1996). This is clearly linked to the psychological and ethical implications of abortion, which may cause interviewed women to conceal something they consider to be a failure, a fault, or at the very least a private matter, especially if they rejected their pregnancy for nontherapeutic reasons (Atkin, 1989). In contrast, the data provided by retrospective surveys are more reliable in countries such as Romania, Turkey, or China, where abortion is more often used as a method of birth control.

However, significant improvements may be achieved by complying with a number of recommendations on how to handle sensitive, strictly personal questions or matters that verge on illegality (or that are illegal). The question of the number (and date) of induced abortions may, for example, be included in a questionnaire on sexual behavior that interviewees fill in themselves and return to the interviewer in a sealed envelope. Similarly, honesty during the interview is improved if questions on induced abortion are included as part of a questionnaire on health, rather than on reproduction. Finally, better results are obtained when questions regarding abortion are administered during a gynecologic interview of women consulting gynecologic and obstetric services; statements are obviously more truthful and complete when they are made in connection with concrete health problems (Bonarini, 1984).

The problem of underrecording would not be so important if the tendency to conceal were independent of the women's characteristics. Unfortunately, this is not the case. Some surveys, based on hospital abortion registers, make it possible to check the truthfulness of retrospective health histories. For example, in Estonia, it has been found that failure to report an abortion is relatively more frequent among Estonian women than among Russian women, women older than 40 years, those who have had an abortion after the third month of pregnancy, unmarried women, and women who have more than three children (Anderson *et al*., 1994). In the United States, omission is more frequent among nonwhite women, the less educated, and those whose previous abortion was a long time ago (Udry *et al*. 1996).

b. Estimates Based on Hospital Registers of Abortion Complications

In countries where abortion is illegal and it is possible to carry out surveys (even nonexhaustive ones) on reasons for admission to hospital, it is possible to estimate the average number of abortions per woman (n_{av}) by the following formula:

$$n_{av} = \frac{n_e \times H_{av}}{H_{ac}} \times K \qquad \text{(Eq. 1)}$$

in which n_e is the average number of children per woman; H_{av} the number of hospital admissions for abortion, whether induced or not; H_{ac} the number of childbirths in the same hospitals and K a coefficient that takes into account spontaneous abortions, the proportion of abortions not involving admission to hospital, the proportion of deliveries in hospital, and prevalence of contraception (WHO, 1994; Singh and Wulf, 1991, 1994). Estimation of coefficients H_{av} and K in this formula is quite complicated because, for each country, adjustments must be made to take account of their specific situation (for further details, see the articles cited). The estimate is a very crude one, although its advantage is that it uses data generally available in countries where, for example, surveys similar to Demographic and Health Surveys (DHSs) of women's fertility have recently been carried out. A considerable proportion of the estimates in Table 35–3 have been produced using this formula, which, when compared with other estimation techniques, yields results that are generally deemed acceptable.

c. Estimates Based on Bongaarts' Model for Analyzing Intermediate Variables of Fertility

We know that in Bongaarts' model (see Chapter 34), the average number of children per woman n is derived from the maximum number of children n_{max} a woman may have (generally estimated to be 15) using the following expression (Bongaarts and Potter, 1983):

$$n = n_{\max} \times C_m \times C_c \times C_a \times C_i \qquad \text{(Eq. 2)}$$

in which the four coefficients C measure the reducer effects of marital fertility C_m, contraception C_c, induced abortion C_a, and postpartum amenorrhea C_i, the last of which in practice combines the effects of breast-feeding and postpartum abstinence. This method may be used only if there are few births outside wedlock and thus makes it possible to estimate abortion only among married women. Without going into the technical details of the estimation of the coefficients, note that it is fairly easy to measure the average number of children per woman, C_m, C_c, and C_i on the basis of different types of surveys of reproductive behaviour (*World Fertility Survey* [*WFS*], *DHS, or Fertility and Family Survey* [*FFS*][2] to cite but the most well-known and accessible). This leaves C_a as the only unknown value in the Equation 2. This method of measuring abortion (with a number of variations) has been and still is widely used (Frejka, 1985; Foreit and Norman, 1992; Johnston and Hill, 1996) because it has the advantage of requiring no data on abortion. In addition C_a (between 0 and 1) is of important demographic significance because it measures the proportion of potential fertility lost through induced abortion; comparison of C_a and C_c (the coefficient of contraception) also makes it possible to speculate as to the relative importance of these two factors in birth control.

For example, for 1972 in Poland, using data from the World Fertility Survey and other official data on n, the Equation 2 may be quantified as follows (Frejka, 1985):

$$n = n_{\max} \times C_m \times C_c \times C_a \times C_i$$

$$2.3 = 15.3 \times 0.52 \times 0.41 \times C_a \times 0.95 \rightarrow C_a = 0.74$$

This means that in this particular case, induced abortion reduces theoretical fertility by 25%. The estimate for C_a using official sources (administrative statistics) would be 0.88. Thus, at the beginning of the 1970s in Poland there was an unrecorded level of abortion, as the article referred to also points out.

The method's main technical shortcoming is that any inaccuracy in the estimation of coefficients other than C_a directly affects the value of C_a and hence the number of abortions. The problem is perhaps C_c (the contraception coefficient), especially in countries where methods such as periodic abstinence and coitus interruptus, whose efficacy (which largely depends on the level of motivation of those who practice them) is difficult to determine, are widespread. Moreover, C_a may also be affected by the weight of factors the model fails to take into consideration because they are assumed to be invariable whatever the population, (such as the level of sterility and the frequency of sexual relations). The model may be applied only in countries where having children outside marriage is proscribed; consequently, the coefficient C_a represents only abortions among married women, unless very precise data are available for unmarried women.

If we do not have values for all the coefficients in Bongaarts' model, it is possible to adopt the same *residual* logic to estimate abortion, by reducing or grouping the intermediate variables required for the estimate. For example, one can estimate the average number of abortions per woman in the years immediately after liberalization using data solely on fertility, by the regression line for the data in Table 35–1 (paragraph 1). However, this has the drawback of not taking contraception into account. Whatever the case, such methods produce results that are heavily dependent on simplistic and often unacceptable hypotheses (Bonarini, 1992).

d. Estimates Based on the Trend in the Proportion of Repeated Abortions

If *abortion* remains roughly constant over a ten year period, the proportion of *repeat abortions* is heavily dependent on the level of *abortion* (Tietze and Jain, 1978).

[2] Fertility and Family Surveys, collected in Europe during the 1990s.

Using data for 12 countries with poor statistics, Stanley Henshaw (1989) made an empirical estimate for the following quadratic equation that, for the 12 countries, fits the data quite well ($R^2 = 90\%$).

$$a_{15-44} = 1000 \times \frac{A}{F_{15-44}} = 135.9 \times P^2 + 4.8 \quad \text{(Eq. 3)}$$

in which a_{15-44} is the abortion rate between the ages of 15 and 44, A the number of abortions, F_{15-44} the number of women aged from 15 to 44, and p the proportion of repeated abortions.

If we have reliable data on the proportion of abortions ranking above 1 (from hospital surveys, for example), the equation allows us to estimate the abortion rate without requiring any other data.

However, as the author himself emphasizes, the model has been highly criticized from a theoretical angle (Henshaw, 1989; Singh and Henshaw, 1986). First, it is applicable only to countries in which there is no decisive variation in abortions over time, otherwise the estimates obtained reflect the average level of abortion during the last 10 years. Second, the database used (a few EDTs in the 1970s and 1980s) is very limited in time and space; if we apply the model to other countries, we implicitly assume that the ratio of overall abortions to repeated abortions is the same as in the countries (and the periods) used to estimate the Equation 3. The problem with this hypothesis is that because the ratio of overall abortions to repeated abortions may vary from one population to another due to differences in the internal heterogeneity of populations (Dalla Zuanna and Garofolin, 1991; Bonarini, 1992; Singh and Henshaw, 1996). It is possible to clarify this with an extreme example of maximum heterogeneity: Let us assume that within a given population 1% of women use abortion only as a means of birth control and the remaining 99% have never aborted; the abortion rate will be relatively low (10/1000), but virtually all these abortions will be repeated abortions. Moreover, in countries where there is no regular registration of abortion (i.e., precisely those for which an estimative method has to be used), to determine the proportion of repeated abortions in many cases the only source is hospital data on women hospitalized for abortion-related complications, who do not constitute a representative population. Finally, as we have already pointed out, reconstituting abortion histories is a sensitive matter because women may not mention their previous experiences of abortion. Consequently, estimation of the proportion of repeat abortions on the basis of declarations made by women themselves may underestimate the number.

In spite of its somewhat fragile foundations, this method may be used to provide an idea of the level of abortion and, in countries with regular statistics, to speculate as to the possible level of underrecording. For example, in 1987, in two regions of Italy, official data produced very similar abortion rates for women aged 15 to 44: 8/1000 in the Veneto and 10/1000 in Calabria, at the lower end of the ladder of Italy's 20 regions. However, the proportion of repeated abortions was 27% in Calabria and only 15% in Venezia. If we apply the Equation 3, it gives an abortion rate of 8/1000 for Venezia (equal to that of the official statistics), but for Calabria the figure is 15/1000 (i.e., 50% higher than the official figure). This estimate concurs with other indices, suggesting that not legal induced abortions are common in southern Italy.

To close this section, let us revert to the problem posed in the first paragraph. It is difficult to estimate the number of clandestine abortions and the results depend heavily on the method adopted. If we partly follow the recommendations on estimation of maternal mortality made by Parker Mauldin (1994), to obtain scientifically valid indicators of clandestine abortion we must (1) not adopt a single method but rather compare the results of different approaches; (2) for each method used, not accept a one-off estimate but provide an idea of the scale of possible variations by altering the basic hypotheses; (3) strictly control the data sources and examine the possible effects of errors of observation.

2. Study of the Causes of Abortion: A Statistical and Quantitative Approach

The complexity of the causal chain leading to induced abortion has already been emphasized: To fully to understand the phenomenon, it would be necessary to take into account the entire process leading to the decision. Reverting to the initial scheme (Table 35–1), we would need to distinguish between the causes of a conception (whether wanted or not) and the causes of a decision to terminate a pregnancy by induced abortion rather than trying (or accepting) to allow the pregnancy to run its course to birth (whether wanted or not). Unfortunately, in practice it is far from simple to do so, if only because, as we have already seen, abortion does not easily lend itself to observation through a conventional quantitative survey. In addition, as we will explain more fully in the next paragraph, causal analysis is marred by the lack of a theoretical explanation for abortive behavior. In most cases, studies are merely descriptive and are intended at the very most to highlight categories among whom the risk of abortion is high or low. Even the most sophisticated analyses, capable of correctly isolating the effect of some factors on abortion, ultimately serve only to clarify the description rather than to improve our ability to explain, simply because we have no behavioural hypothesis to verify. Rather than *causal analyses of abortion*, it would be more accurate to talk of *differential analyses of abortion*.

Thus, we may distinguish two groups of analytic techniques. The purpose of the first group is to measure abortion by isolating different groups of women and analysing the differences observed.

As we explained in note 2 above, the two most commonly used measurements are the *abortion rate* (the number of abortions per woman) and the *abortion ratio* (the number of abortions per pregnancy or per birth). The great majority of demographic studies on abortion are based on these measurements in terms of age, marital status, the number of previous births, race, religion etc.; the sole constraint is the availability of data to permit classification, using these same criteria, both of women who abort (the numerator) and of all women or women who give birth (the denominator). These measurements are also the basis of the compilations regularly published by the Alan Guttmacher Institute until the late 1990s (*Abortion, a world review*) and by the United Nations (in particular, *Abortion Policy, a global review*, published in three volumes in 1992, 1993 and 1995) and of the series generally available in those countries with regular registration systems. As we have already seen, we may go beyond the mere calculation of abortion rates or ratios and assess the net impact on abortion of a number of potential explanatory variables. Studies based on *logistic regression*, in which the *abortion/pregnancy* ratio is modelled so as to calculate the relative risk of abortion for each category (or cross-classification) of explanatory variables are increasingly common (see, for example, Feng, 1996). In order to apply this technique a set of data on pregnancies classified by outcome (abortion or birth) and by explanatory variable is required; this is generally based on survey data, although it may also be based on comparison between registers of abortions and births, provided the explanatory variables may be classified in the same way for both events.

The second group of techniques used to study differential abortion (of which demographers, unfortunately, make less use) adopts the case-control analysis. Women who have had abortions are the *cases* and those who have not had abortions (properly selected at random) the control group; a variable is linked to abortion if it is not similarly distributed among the cases and the control group. In contrast to the methods described previously, it is possible to use this technique even if the overall population distribution (or the distribution of women having had a child) in terms of the variables concerned by the differential study is unknown: it is sufficient to know the distribution among a sufficiently representative control group. In addition, where case-control analyses are concerned, we may call on proven techniques to highlight the effect on abortion of each variable, regardless of composition in terms of the other possible explanatory variables. Examples of this type of differential analysis of abortion (in most cases carried out by epidemiologists) show the importance of careful choice of method for selecting the control groups (La Vecchia, 1986; Skjeldestad, 1994; Houghton, 1994). In the same scheme as case-control studies, although simpler to handle, we have research based on comparison of the distribution (by the available variables) of women who have had an abortion and the total population (see, for example, Henshaw and Kost, 1996).

3. The Causes of Abortion: An Anthropologic Approach

Anthropologists have long been interested in abortion (and in everything connected with human reproduction). Nonetheless, two factors have recently oriented these topics in a new direction: realization of the growing importance of abortions in the LDTs and the specific and longstanding interest of anthropologists in matters concerning women, and more recently, gender relations. In her introduction to the collection of articles *Abortion from a Cross-cultural Perspective* (1996, p. 481) Barbara Rylko-Bauer makes the following conclusion:

The value of an anthropologic and cross-cultural perspective on abortion may not lie so much in measuring precise, quantifiable features (*although this should be an integral part of such work when possible*)*, but rather in:* (1) *documenting the socio-cultural, economic, ideological and political complexity of abortion across cultures, as these case studies have done; and* (2) *in identifying the processes and contexts that define abortion at the various levels* (*individual, community, national*) *of analysis and how they articulate with each other. The end result will provide a more realistic picture of abortion world-wide and will focus attention on the women who have remained invisible for far too long, and whose lives, after all, are ultimately affected by the debates, policies, and studies of abortion.*

From a methodologic perspective, anthropologic studies are quite distinct from those described so far. As is also apparent from the above quotation, anthropologists emphasize complexity, interaction between the different levels of analysis and more generally the impossibility of understanding abortion outside the framework of relations between body–spirit, man–woman, family–community and community–state. They set out to interpret each experience of abortion as the response by an individual will to a set of social, historic and cultural determinants. In contrast, the quantitative studies described above take another (and thus perfectly complementary) tack: once the variables that may be linked to abortion have been identified, we adopt the implicit hypothesis that there is no other source of heterogeneity, in other words that the individuals characterized by a given vector of explanatory variables are equal in terms of the risk of abortion. Another fundamental difference of this method is that anthropologists choose a deductive approach: the sole purpose of field study is to validate hypotheses, adopted *a priori*, for the behavior of individuals and communities. In contrast, for the most inductive of the statistical-descriptive studies presented above, the objective, which is often implicit, is to *let the data speak for themselves*.

The methods of field research too are different. As emphasized in the passage cited above, the prime concern of anthropologic research is not to measure abortion as a phenomenon. First and foremost it is concerned by the collection of detailed data on the lives of women (and of men) who decide to abort (or not to abort), information on their family, professional or social environment, on the moral values of their friends and relatives and of the community to which they belong. Consequently, an anthropologic study requires lengthy and detailed observation of the community in question, involving the reconstitution of lengthy life histories, bringing together key witnesses, the dynamics of focus group discussions, in other words groups of women with common characteristics (such as those who have recently had abortions or given birth) with whom the investigator leads a discussion in order to bring out the chain of events and factors responsible for their having aborted or having given birth.

Examples of the application of this information gathering technique to the study of abortion may be found in the compilation of articles referred to above (Rylko-Bauer, 1996) and in the proceedings of an IUSSP seminar (1996). Moreover, more and more studies of abortion during the pretransition stage adopt historico-anthropologic approaches, as may be seen from a review of the papers presented to the seminar organized in October 1994 by IUSSP on *Abortion, infanticide and neglect in the Asian past: Japan in Asian comparative perspective*, and in the results of research into abortion in the Soviet Union during the communist era (Avdeev *et al*., 1994). Clearly, these are very recent studies whose results have not yet been confirmed, and *the behavioral dimension of induced abortion remains largely a question mark* (Mundigo, 1996, p.15). The two most noteworthy arguments in favour of a behavioral theory of induced abortion are perhaps the emphasis placed on the dialectic between religious, individual and collective norms and on negotiation between men and women over decisions in respect of contraception (and the methods) and the outcome of a pregnancy; in this regard, the study made by Georges (1996) on Greece is highly instructive.

In comparison with these studies, conventional differential analyses seem to have very little to offer, particularly in terms of a genuine explanation. Nevertheless, anthropologic studies themselves have limits. By necessity, anthropologists deal with very small groups of individuals who, moreover, are not chosen at random. For example, in an article entitled, *Abortion Policy and Practice in Greece*, the author restricts himself to two samples of married women on the island of Rhodes (i.e., 26 and 62 women in each group) chosen on a voluntary basis after most of the women selected by random sampling had refused to participate in the study (Georges, 1996). In actual fact, the well-known complexity and specificity of the various situations that may account for abortion prevent us from easily accepting any abusive extrapolations from observations that can only be highly specific.

As we have already observed, the deductive nature of anthropologic studies means that hypotheses are predetermined before being tested against observations in the field; consequently, the results of empirical verification may to some extent themselves be predetermined. Thus the sentence of Barbara Rylko-Bauer, referred to above, places a strong emphasis on the key role of women in the abortion process. This insistence may lead us to forget that

> High levels of abortion may signal a crisis of cultural representation of the unborn children, which assumes a "relational" nature, since for anyone who has an abortion the subjective right to be born is subordinated to other rights of the parents (or of the mother), or alternatively springs from society's willingness to accept it, specifically expressed by the parents (or woman)[3] (De Sandre, 1995, p. 19).

In this particular instance, the rich tradition of anthropologic studies of the cultural representation of reproduction might provide the tools for interpreting and more deeply understanding the problems referred to by De Sandre. Possibly, the different *cultural representations* of birth and nonbirth may hold one of the keys to interpreting the different combinations

[3] *Elevati livelli di abortività possono essere un segnale della crisi della rappresentazione culturale del figlio-nascituro, che assume una natura "relazionale"* [*virgolette dell'autore*], *nel senso che, per chi abortisce, il diritto soggettivo alla nascita è subordinato ad altri diritti dei genitori* (*o della madre*) *o, alternativamente, sorge sulla base dell'accoglienza sociale specificatamente manifestata dai genitori* (*o dalla donna*).

between abortion/contraception noted from one country to another during the different stages of the demographic transition.

To conclude, the anthropologic approach and the differential statistical–quantitative approach may mutually reinforce each other; as has already been observed, a number of recent scientific initiatives, especially under the aegis of IUSSP, seem to have taken this direction. Interaction between these, and other disciplines may perhaps allow us to build, stone by stone, more satisfactory explanations for the phenomenon of abortion.

Acknowledgment

The editors express their gratitude for this complimentary translation by the WHO.

References

ANDERSON Barbara, KATUS Kalev, PUUR Alan and SILVER Brian, 1994. The validity of survey response on abortion: evidence from Estonia, *Demography*, vol. 31(1), p. 115–132.

ATKIN Lucille, 1989. Psycho-social aspect of data collection, *in*: Francine Coeytaux *et al*. (eds.), 1989. *Methodological issues in abortion research*, p. 81–84. New York, Population Council, 124 p. (Proceedings of a seminar organized by the Population Council in collaboration with the International Projects Assistance Services and the World Health Organization, New York, December 12–13, 1989).

AVDEEV Alexandre, BLUM Alain and TROITSKAJA Irina 1994. Histoire de la statistique de l'avortement en Russie et en URSS jusqu'en 1991, *Population*, vol. 49(4–5), p. 903–934.

BLAYO Chantal, 1993. Le rôle de l'avortement dans les pays d'Europe centrale et orientale, *in*: UIESP (ed.), *International Population Conference, Montreal 1993*, vol. 1, p. 235–251. Liège, IUSSP, 662 p.

BONARINI Franco, 1984. *Problemi di analisi e risultati sulla epidemiologia dell'aborto in base a dati di reparti ostetrico-ginecologici*. Padova, Università di Padova, 43 p. (Rapporti Tecnici del Dipartimento di Scienze Statistiche dell'Università di Padova, no. 1).

BONARINI Franco, 1992. *Mutamento dei fenomeni demografici in Italia dopo la legalizzazione dell'aborto*. Milan, Libreria Universitaria CUESP, 88 p. (Quaderni del Gruppo Nazionale di Ricerca su Epidemiologia dell'aborto su dati di cliniche).

BONGAARTS John and POTTER Robert G., 1983. *Fertility, biology and behaviour. An analysis of the proximate determinants*. New York, Academic Press, 230 p.

CASTERLINE John B., 1989. Collecting data on pregnancy loss: a review of evidence from the World Fertility Survey, *Studies in Family Planning*, vol. 20(2), p. 81–95.

COEYTAUX Francine, LEONARD Ann and ROYSTON Erica (eds.), 1989. *Methodological issues in abortion research*. New York, Population Council, 124 p. (Proceedings of a seminar organised by the Population Council in collaboration with the International Projects Assistance Services and the World Health Organization, New York, December 12–13, 1989).

DALLA Zuanna Gianpiero (ed.), 1996. *Contraccezione e aborto alle soglie del 2000. Paesi ricchi e paesi poveri a confronto*. Rome, Dipartimento di Scienze Demografiche, Università di Roma La Sapienza, 225 p.

DALLA Zuanna Gianpiero and GAROFOLIN Patrizia, 1991. L'abortività ripetuta nel Veneto negli anni 1978–86, *in*: Gianpiero Dalla Zuanna and Paola Giorio (eds.), *Studi sull'aborto volontario nel Veneto*, p. 151–168. Venice, Giunta Generale, 241 p. (Collana di formazione della Giunta Regionale, no. 13).

DALLA Zuanna Gianpiero, 1996. Paesi ad antica transizione, *in*: Gianpiero Dalla Zuanna (ed.), *Contraccezione e aborto alle soglie del 2000. Paesi ricchi e paesi poveri a confronto*, p. 155–188. Rome, Dipartimento di Scienze Demografiche, Università di Roma La Sapienza 225 p.

DE SANDRE Paolo, 1995. Continuità e discontinuità dei recenti processi demografici: introduzione, *in*: SIS (Società Italiana di Statistica), *Proceedings of the meeting Continuità e discontinuità dei processi demografici, 20–21 aprile, Arcavacata di Rende (CZ), Italy*, p. 11–24. Cosenza, Rubettino, 672 p.

FARINA Patrizia 1996. Asia orientale e meridionale, *in*: Gianpiero Dalla Zuanna (ed.), *Contraccezione e aborto alle soglie del 2000. Paesi ricchi e paesi poveri a confronto*, p. 73–121. Rome, Dipartimento di Scienze Demografiche, Università di Roma La Sapienza, 225 p.

FENG Wang, 1994. *The rise of abortion in modern China*. Liège, IUSSP, 18 p. (Paper presented at the IUSSP seminar on ⟨Abortion, infanticide and neglect in population history: Japan in Asian comparative perspective⟩, Kyoto, October 20–21, 1994).

FOREIT Karen and NORMAN Dorothy, 1992. A method for calculating rates of induced abortion, *Demography*, no. 29, p. 127–137.

FREJKA Tomas, 1985. Induced Abortion and Fertility, *Family Planning Perspective*, vol. 17(5), p. 230–234.

GEORGES Eugenia, 1996. Abortion policy and practice in Greece, *Social Science & Medicine*, vol. 42(4), p. 509–519.

HENSHAW Stanley, 1989. Estimating incidence of abortion from repeat abortion histories *in*: Francine Coeytaux, *et al*. (eds.), *Methodological issues in abortion research*, p. 33–38. New York, Population Council, 124 p. (Proceedings of a seminar organized by the Population Council in collaboration with the International Projects Assistance Services and the World Health Organization, New York, December 12–13, 1989).

HENSHAW Stansley and KOST Kathryn, 1996. Abortion patients in 1994–1995: characteristics and contraceptive use, *Family Planning Perspectives*, vol. 28(4), p. 140–147.

HOUGHTON Anita 1994. Women who have abortions: are they different? *Journal of Public Health Medicine*, vol. 16(3), p. 296–304.

HUNTINGTON Dale, MENSCH Barbara and MILLER Vincent, 1996. Survey questions for the measurement of induced abortion, *Studies in Family Planning*, vol. 27(3), p. 155–161.

IUSSP, 1994. *Abortion, infanticide and neglect in the Asian past: Japan in Asian comparative perspective*. Liège, IUSSP, multiple paging. (Papers presented at the seminar in Kyoto, October 20–21 1994).

IUSSP, 1996. *Socio-cultural and political aspects of abortion in a changing world*. Liège, IUSSP, multiple paging. (Papers of the seminar of Kovalam, Trivandrum, Inde, March 25–28, 1996).

JOHNSTON Heidi and HILL Kenneth, 1996. Induced abortion in the developing world: indirect estimates, *International Family Planning Perspectives*, vol. 22(3), p. 108–114.

JONES Elise and FORREST Jacqueline, 1992. Underreporting of abortion in survey of US women: 1976–1988, *Demography*, vol. 29(1), p. 112–126.

LA VECCHIA Carlo, 1986. Characteristics of women undergoing induced abortion: results of a case-control study from Northern Italy, *Contraception*, vol. 32(6), p. 637–649.

MAULDIN W. Parker, 1994. Maternal mortality in developing countries: a comparison of rates from two international compendia, *Population and Development Review*, vol. 20(2), p. 413–421.

MUNDIGO Axel, 1993. Health and social aspects of induced abortion: an overview of research needs, *in*: IUSSP (ed.), *International*

Population Conference, Montreal, 1993, vol. 1, p. 203–208. Liège, IUSSP, 662 p.

Mundigo Axel, 1996. The challenge of induced abortion research: trans-disciplinary perspectives. A background paper, *in*: IUSSP (ed.), *Seminar on Socio-culutral and political aspects of abortion in a changing world, Kovalam, Trivandrum, India, March 25–28, 1996*, p. 1–27. Liège, IUSSP, multiple pages.

Otha Motoko and Sawayama Mikako, 1994. *An analysis of the motivation for mabiki and abortion as related to child rearing customs in early modern Japan: research based on non-quantitative evidence of the history of the people mentality.* Liège, IUSSP, 32 p. (Paper presented at the IUSSP seminar on 〈Abortion, infanticide and neglect in population history: Japan in Asian comparative perspective〉, Kyoto, October 20–21, 1994).

Paiewonsky Denise, 1999. Social determinants of induced abortion in the Dominican Republic, *in*: Axel Mundigo and Cynthia Indriso (eds.), *Abortion in the developing world*, p. 131–150. Geneva, World Health Organization, New Delhi, Vistaar Publications, 498 p.

Pinnelli Antonnella and Milan Giulia, 1996. America Latina, *in*: Gianpiero Dalla Zuanna (ed.), *Contraccezione e aborto alle soglie del 2000. Paesi ricchi e paesi poveri a confronto*, p. 123–154. Rome, Dipartimento di Scienze Demografiche, Università di Roma La Sapienza, 225 p.

Rylko-Bauer Barbara, 1996. Abortion from a crosscultural perspective: an introduction, *Social Science & Medicine*, vol. 42(4), p. 479–482.

Singh Susheela and Henshaw Stanley, 1996. The incidence of abortion: a worldwide overview focusing on methodology and on Latin America, *in*: IUSSP (ed.). *Seminar on 〈Socio-cultural and political aspects of abortion in a changing world,〉 Kovalam, Trivandrum, India, 25–28 March 1996*, p. 1–33. Liège, IUSSP, discontinuous paging.

Singh Susheela and Wulf Deirdre, 1991. Estimating abortion lLevels in Brazil, Colombia and Peru, using hospital admission and fertility survey data, *International Family Planning Perspectives*, no. 17, p. 8–13.

Singh Susheela and Wulf Deirdre, 1994. Estimating levels of induced abortion in six Latin American countries, *International Family Planning Perspectives*, no. 20, p. 4–13.

Skjeldestad Finn, 1994. When pregnant—why induced abortion? *Scandinavian Journal of Social Medicine*, vol. 22(1), p. 68–73.

Stloukal Libor 1996. Eastern Europe's abortion culture: puzzles of interpretation, *in*: IUSSP (ed.), *Seminar on Socio-cultural and political aspects of abortion in a changing world, Kovalam, Trivandrum, India, 25–28 March 1996*, p. 1–11. Liège, IUSSP, multiple paging.

Tietze Christopher and Jain Anrudh, 1978. The mathematics of repeat abortion: explaining the increase, *Studies in Family Planning*, no. 9, p. 295–299.

Udry Richard, Gaughan Monica, Schwingl Pamela and van den Berg Bea, 1996. A medical record linkage analysis of abortion underreporting, *Family Planning Perspectives*, vol. 28(5), p. 228–231.

United Nations, 1992, 1993, 1995. Abortion policies: a global review. New York, United Nations, vol. I: Afghanistan to France, 158 p; vol. II: Gabon to Norway, 243 p.; vol. III: Oman to Zimbabwe, 236 p.

WHO, 1994. *Abortion: a tabulation of available data on the frequency and mortality of unsafe abortion*, 2nd ed. Geneva, World Health Organization, 234 p.

C H A P T E R

36

Demography and Sexuality

MICHEL BOZON
Institut national d'études démographiques (INED), Paris, France

In common with the sociology of the family and the anthropology of kinship, demography only very belatedly turned its attention to sexuality, which is ironic given the discipline's comprehensive treatment of couple formation and nuptiality, divorce and partnership breakdown, fertility, and contraception. Not examining sexuality explicitly amounts tacitly to taking it as a natural and relatively fixed element of human behavior. Only with the advent of modern contraception and, more important, the emergence of the acquired immunodeficiency syndrome (AIDS) epidemic, did demographers develop instruments for detailed observation of sexual behavior, thereby acknowledging that it was a phenomenon subject to meaningful variation. The historic context for demographers to direct their attention to sexual behavior was the transformation and decline of nuptiality in developed countries; it was not directly linked to earlier efforts to explain and follow fertility reduction in developing countries.

The reasons for this difficulty in considering sexuality are discussed and illustrated in the first part of this chapter. The next section argues that demography is in fact technically well equipped for a rigorous and critical approach to observation of sexuality. Three fields are then identified where analytical observation of sexual behavior from a demographic standpoint (sometimes in conjunction with other social sciences) can increase understanding of major societal processes. Analysis of sexual histories elucidates contemporary shifts in sexual standards and changes in family and partnership behaviors. Analysis of differences between men and women in sexual attitudes, behavior, and experiences is of fundamental importance to the study of gender relations. The fertility decline and contraceptive revolution have had a profound impact on women's sexual lifestyles; freed from the constraints of reproduction, women have acquired an unprecedented room for maneuver in their inevitable confrontation with men. Finally, cross-cultural analysis of sexual behavior reveals unexpected differences between cultures, in the sphere of representations but also in the practices and context of sexual activity.

I. DEMOGRAPHY'S DIFFICULT INCLUSION OF SEXUALITY

In its early days, demography followed the course traced by institutional statistics—at the end of the 19th century the discipline dealt only with the phenomena that were objectified and recorded by the official statistical system. In addition, because actual sexual behavior was unrecorded, it was ignored by demography. Initially, therefore, sexuality was approached only through its consequences and institutional expressions—civil registration data were used to study fertility and nuptiality and to measure extrainstitutional behavior such as illegitimate fertility. Before long, however, demographers developed an analytic approach for separating the various demographic factors and variables, most of which were in interaction with each other. But this new approach did not immediately cause demographers to direct their attention to sexual behavior. Typical in this respect was the

Italian statistician and demographer Corrado Gini. In the 1920s Gini introduced the concept of fecundability, defining it as

> The probability for a married woman to conceive during a month, in the absence of any Malthusian or neo-Malthusian practice intended to limit procreation (Gini, 1928).

This probability obviously depends on the frequency of intercourse, yet Gini does not make this point, thereby implicitly assuming that all married couples have sexual intercourse, that the frequency of their relations is unconnected to such variables as age and duration of union, and that the probability of conceiving is independent of the frequency and forms of sexual activity. None of these assumptions is correct—a couple may have no sexual relations, sexual activity does vary with age and duration of union (as we shall see later), and it can take a variety of forms, regardless of whether contraception is being used. Corrado Gini, like the early sociologists and anthropologists, implicitly adopts a naturalistic model of sexuality, which holds sexual activity to be a natural given for human kind, fixed in its forms of expression and solely biological in its determinants. This essentialist ideology of the early demographers was certainly a more serious impediment than inadequate sources. Demography began to examine sexuality and make it an object of study only when demographers had adopted the social constructionist stance characteristic of modern social science, which emphasizes instead the social and cultural construction of human behavior, and hence the social frameworks for sexual interaction, the varied repertoires of human sexual practices and their relative independence from biology.

The sexual life of married couples, brought to light by historic demography, might also have drawn the attention of demographers. Historic demographers established early on that fertility had begun to decline in many countries well before the advent of modern means of contraception (the pill, the intrauterine device [IUD], surgical sterilization, the condom etc.). In France for instance, couples were actually *malthusian* or *neo-Malthusian* before Malthus, in that from the end of the 18th century they were deliberately limiting their progeny, using methods other than the *moral restraint* advocated by the English clergyman. In Latin countries like France, this conscious control of fertility has been attributed to the practice of withdrawal or *coitus interruptus*. But how was such a radical change and constraint introduced into the exercise of the most *legitimate* sexuality, that of married couples? With Philippe Ariès and a few others, Jean-Louis Flandrin is one of the small number of French historians to have examined the question. From his analysis of confessors' manuals, theological writings, juridical records and popular culture, he concludes that the range of sexual expression was much more diverse in the 17th and 18th centuries than is generally recognized. A particularity of Northern and Western Europe was the invention, at the end of the Middles Ages, of a pattern of late marriage, in which women seldom married before 25 (and men even later). The question to be addressed therefore concerns the response of young people to this constraint, given that illegitimate births—subject to severe social censure—remained rare. Two hypotheses can be advanced. One is prolonged sexual abstinence, with young people waiting until marriage or thereabouts to commence sexual relations; the other involves recourse to sexual practices that carried no risk of pregnancy. Some historians favor the first hypothesis, invoking the possibility of a strict policing of *courtship* in premodern societies. Others, including Jean-Louis Flandrin, and with whom we tend to agree, argue for the second hypothesis:

> Unmarried men of all ages must have had to seek sexual satisfaction with prostitutes, or in adultery, or in "unnatural" practices such as homosexuality, bestiality, masturbation, heavy petting, etc. [. . .]. There was also, increasingly frequently, between the sixteenth and mid-twentieth centuries, withdrawal. [. . .] Before the large-scale introduction of contraception to marriage, initiation to these practices was done "in sin" [i.e. in the context of illegitimate relations] (Flandrin, 1981, p. 348 and p. 125).

In addition, in several regions of France communities appear to have tolerated a degree of premarital promiscuity, including fondling and nongenital practices. Because individuals were already familiar with these practices, so in the privacy of their marriage bed, couples were able to start reducing their fertility, despite the hostility of the Church:

> Jean-Louis Flandrin and John T. Noonan have clearly established the great importance attached by the Church from the 1850s onwards to the new scourge referred to as "conjugal onanism." By this euphemism were designated all the techniques whereby the married couple could obtain pleasure without a risk of pregnancy. […] The authors have provided an exhaustive catalogue of this "cheating": withdrawal, mutual masturbation (termed "ignoble service"), orogenital caresses, anal intercourse (A. Corbin, 1984, p. 238).

The term *coitus interruptus* used by historic demographers would thus appear to be a blanket term for a wide range of practices that allowed sexual activity to occur without risk of pregnancy well before the introduction of modern contraceptive methods. These practices, however, long remained invisible to demographers.

A third occasion for demography to direct its attention to sexuality was the pioneering article by Kingsley Davis and Judith Blake *Social Structure and Fertility: an Analytic Framework* (1956), which heralded the arrival of *intermediate variables* in fertility studies. Inspired primarily by anthropologic research, the authors proposed analysis using a set of variables that not only had a direct effect on fertility patterns but were themselves determined by the classic socioeconomic characteristics. In this perspective, the effect of educational level, for example, would be considered not for its direct impact on fertility but for its influence on determinants such as age at marriage or the incidence of breast-feeding. Blake and Davis group the variables thus defined into three broad categories: factors affecting exposure to intercourse; factors affecting exposure to conception; factors affecting gestation and successful parturition.

The first category contains, to begin with, the variables characterizing types of union (age at marriage, etc.), then those governing exposure to the risk of pregnancy for couples: voluntary and involuntary abstinence (the latter due to illness or temporary separation, for example), and frequency of intercourse. The second category includes sterility and subfecundity, and the forms of contraception, including those that are neither mechanical nor chemical, which the authors describe thus:

> Means of contraception other than mechanical and chemical include the "rhythm method" (which can also be classed as voluntary abstinence), withdrawal, simulated intercourse without penetration, various perversions, etc. (Davis and Blake, 1956, p. 212, note 3).

Unfortunately the authors barely expand on these points in their article, and instead concentrate mainly on postpartum abstinence. This analytic framework later strongly influenced the design of questionnaires used in national fertility surveys in the 1970s, but nearly always with the same lack of attention to sexual relations. Other reproductive models developed in the 1960s and 1970s did include an exploration of the link between coital frequency and fecundability, although lack of empirical data prevented any firm results being reached (see, for example, Bongaarts and Potter, 1983). This stream of demographic research was linked to the expansion of family planning in the developing countries and it focused mainly on the variables that were susceptible to interventions, which was believed not to be the case for sexuality.

Lack of empirical data and the implicit essentialist ideology of the demographers were durable obstacles to any demographic investigation of sexuality. Two things helped to change this situation. The first was the gradual emergence of the sample survey, and notably the longitudinal survey, able to capture more intimate and informal behavior, as one of demographers' main sources, while at the same time use of civil registration data declined. The second was a series of societal changes, including the transformation of partnership and family behavior, the widespread adoption of modern contraception, and the emergence of the AIDS epidemic, which motivated demographers to include questions on sexuality in their surveys and then to collaborate with scholars from other disciplines in conducting specific investigations into sexual behavior (Caraël, 1995; Hubert *et al*., 1998). Demographic interest in sexual behavior, which became apparent around the time of the *second demographic transition* (van de Kaa, 1987), focuses less on reproductive sexuality than on sexual behavior as a component of relationships, whether conjugal or not, and as an element of individual well-being and health in the broadest sense.

II. DEMOGRAPHIC OBSERVATION OF SEXUALITY: CONCEPTS, INDICATORS, LIMITATIONS

Demography's principal contribution to social scientific study of sexuality is to introduce a rigorous analytical viewpoint for the description of sexual phenomena. It does this without closing its eyes to the difficulties, limitations and biases that affect observation in this field; indeed, these are integral to the analysis. Observation in this context means the *construction of observation protocols* (i.e., of questionnaires) because the self-reports of the actors themselves are the only basis for apprehending sexual behavior (Bozon, 1995 and 1998a).

Any observation protocol rests on a handful of key concepts, which, reflecting demography's strong quantitative orientation, are operationalized as indicators. The key concepts in demographic analysis of sexual behavior are sexual history, sexual relations, and type of sexual activity. Standard procedures of demographic analysis, such as precise dating of events, recording of durations, and assigning of observed behavior to reference periods, are also important in the demography of sexual behavior.

The start of an individual's sexual history corresponds to sexual initiation, which in a sociodemographic perspective is understood as the onset of sexual activity with a partner. For the analysis of the social construction of sexuality, this threshold is considered more significant than the period of infantile sexuality, to which psychoanalysts, in contrast, attach great importance (Bozon and Leridon, 1996). Sexual initiation coincides with neither the biological transition of menarche (first menstrual period), nor the socially recognized threshold of couple formation. Particular attention is given to age at first intercourse and the characteristics of the first sexual partner, and also to the length of time between first intercourse and first union formation, known as the duration of sexual exposure before first regular partnership—in the case of virginity at couple formation, this duration is nil. In practice, it is always shorter for women (Macro International, 1994; Caraël, 1995; Bozon and Kontula, 1998; Bozon, 2003).

In addition to the characteristics of sexual initiation, a retrospective sexual history is characterized using two general indicators—number of sexual partners and length of time in sexual relationship. These two indicators are quite distinct. Most of a life spent in sexual relationship is usually though not necessarily associated with a small number of partners or indeed with a single partner. Conversely, a large number of partners can correspond to a relatively short total length of time in sexual relationship. It should be noted that number of sexual partners and time spent in sexual relationship differ clearly from classic demographic indicators such as number of conjugal partners or time spent in union. Classic demographic observation of the course of individuals' unions may involve description of periods of life in a couple, with successive partners. In demographic analysis of sexual behavior, by contrast, an individual's relations with partners can be thought as simultaneous (simultaneous *multipartnership*, i.e., having simultaneous multiple partners). More generally, a sexual partner is not necessarily a spouse, whereas in some cases a spouse may not become a sexual partner from the start of the union or may cease to be one after a certain time even though the couple still exists.

The sexual history of each individual thus comprises sexual relationships alternating through time with periods of sexual inactivity. Sexual relationships have a beginning, a development and an end; they can be categorized according to the type of bond between the partners—stable couple, one-night-stand, noncohabiting steady partner, holiday relationship, casual partner, and so on. Each relationship must be assessed in terms of its degree of exclusivity (or inversely its simultaneity with other relationships) and can be characterized by the types of contraception or protection used. The periods of sexual inactivity, meanwhile, correspond to two quite different situations—periods with no sexual partner or periods when the individual is in a union but has no sexual relations with his or her partner (spacing or cessation of sexual relations). For any given individual it is in fact hard, except in the case of major physiological problems, to conclude for a definitive cessation of sexual activity—indirect statistical evaluations are derived from the proportion of individuals who have not had any sexual activity in the previous year or in the previous 5 years.

Not all types of sexual activity are equivalent. An indicator can be constructed to express variability in the intensity of sexual activity, usually evaluated from frequency of intercourse (during the previous month or during an average month). Another issue of interest is the range of sexual activity, which can be explored from many different angles. When considering sexuality in its procreative dimension, it is important to show the diversity of the forms of contraception employed and how this varies over the life course. Surveys on young people, for example, must draw attention to forms of nonpenetrative sexual activity. Sexuality can also be analyzed in a more hedonistic perspective, from the angle of diversity of sexual repertoire. An increase in the range of sexual practices (used at least once in the lifetime), such as oral sex, anal penetration and masturbation has been observed between different generations of adults in developed countries. Questions about sexual practices can also take a recent sexual relation as reference (last intercourse): Questioning can examine the attainment of orgasm by one and the other partner. Achieving orgasm can become the object of an expectation or even a norm only in a historicocultural context where sexuality is losing its procreative function and where power is shared more equally between the partners. Last, sexual practices can be approached from the health angle. In the historic context of the AIDS epidemic, various protective or nonpenetrative sexual practices are promoted via prevention campaigns—hence men who have sexual relations with other men are encouraged to use safer sex practices. What needs to be measured here is the gap between the messages of the public health campaigns, the reported intentions of the actors and their actual practices.

Sexual behavior research in demography explores the meanings, representations and norms relating to sexuality, insofar as they contribute to understanding of sexual behavior. For instance, knowing the feelings for a partner (in love/not in love) when sexual relations begin is valuable for situating this relationship in an individual's sexual history. Respondents are often questioned about their attitudes to fidelity and extramarital relationships. In this domain there is a fairly high consistency between norms and practices, though it is not systematic—young adults and in particular young couples are, for example, the most severe in their judgment of extra-marital relations and the least likely to have them. However, this does not mean that norms govern behavior: rather they provide an interpretation of behavior. In every setting and in every period, the domain of sexuality has been subject to norms, but they are losing their power as absolute and immutable laws, based on religious or social rules. Increasingly, normative orientations are flexible, transitory and adaptable; rather than setting the boundaries of behavior, they express the meaning of behavior and reflect the growing individualization of life histories. This is a general historic evolution, although change is proceeding more slowly in those cultures where sexual norms remain based on sacred and immutable foundations. In many cultures, moreover, the normative systems are still sharply differentiated along sex lines.

Data quality is of crucial importance for a demographic description of sexual behavior. In quantitative surveys on this topic, interviewer effects (particularly the interaction between the sex of interviewer and of respondent) need careful analysis. The questions dealing with sexual activity proper are not usually asked face-to-face. In the sphere of private life and sexuality, the honesty of answers cannot be verified or checked against external sources. The only way to assess the reliability of the data is by an analysis of the internal consistency of the questionnaires. Data on coital frequency have been shown to be acceptably reliable (Leridon, 1996), as also have those for partner numbers (Leridon, 1998). The discrepancy observed in all surveys between the number of partners reported by men and women is evidence less of bias than of a gender difference in the way that sexuality is lived—men report more partners because they attach importance to the number of sexual encounters, whereas women ignore the experiences which have lacked emotional involvement and which as far as they are concerned did not count.

If demography can be used to produce a rigorous description of sexual behavior, it is nonetheless important not to abstract sexuality from its broader nonsexual environment. Demographic study of sexual behavior is not to be confused with sexology or sex therapy. Detailed analysis of sexual behavior does not create a new discipline—it contributes to demographic and sociological investigation of conjugal relationships, family behavior and gender relations, notably for the purpose of international comparison.

III. SEXUALITY AND THE TRANSFORMATION OF CONJUGAL TRAJECTORIES

Conjugal trajectories have undergone multiple transformations in recent decades. The contemporary decline of formal marriage in developed countries, the rise of informal cohabitation, and the growing proportion of women living in lone-parent families following union breakdown, bring into focus the elementary forms of relationships between two individuals prior to union formation, and the crystallizing but also disruptive potentials of sexual partnership. The conditions in which a sexual bond is established, its course over time, its degree of exclusivity, and the living arrangements that accompany it must be examined to understand the growing importance of *infraconjugal* or nonconjugal relationships before, during and after more conventional conjugal relationships.

Young adults in developed countries can have durable sexual relationships with stable partners, without even considering cohabitation. Today, the relationship is created by the start of the sexual bond, thus marking the first step in the couple's formation. The initial sexuality no longer depends for its legitimation on the power of a social institution (marriage), which in the process also legitimated any children that were produced. This reversal in the order of operations, such that sexuality precedes its social acknowledgment, confers a different meaning on sexual activity. The intensity and continuity of sexual activity gradually establishes the couple, signaling the desire for it to continue and perhaps for children to be produced. Another fast-increasing situation is that of women with children, who after a union breakdown, experience an essentially permanent phase of lone-parent family living. A large proportion of these single mothers are also in a stable but noncohabiting relationship with a new sexual partner. Sexual life and family life are here sharply compartmentalized. The increase in conjugal dissolution thus favors the emergence of new attitudes, in which sexuality and couple no longer coincide.

For a given individual, being in a sexual partnership does not necessarily preclude other relationships—a situation that is expressed by the notion of simultaneous *multipartnership*. This distinguishes sexual partnership from marriage-type relationships, which in nonpolygamous societies are exclusive of other such relationships and therefore can only be experienced sequentially. Description of a society's conjugal and family behaviors cannot be limited to the official and visible bonds. Treating sexual partnerships, even of the most transient kind, as a form of relationship implies acknowledging the existence of a hidden polygamy in nonpolygamous societies. The AIDS epidemic has been a powerful impetus to systematic examination of individuals' secondary and parallel relationships. A society also reveals much about itself in the forms of simultaneous multiple partnerships that it tolerates, even when these concern only a minority. Stable sexual multipartnership, common in the Caribbean and found also in some European countries, involves the coexistence of several sexual relationships over a long period—and is always characterized by a hierarchical ordering of the partners. Also observed is having multiple partners at separation, which is tolerated when it corresponds to the transition from one partner to the other. Finally, a form frequently encountered in developed societies is casual sexual multipartnership, which is accepted as a rare and covert exception in an otherwise stable conjugal context.

A sociodemographic approach offers a new perspective on homosexuality, which avoids explaining homosexual behavior either as the expression of a nature that defines and delimits a discrete population category, or as the simple consequence of socialization within a minority culture. Sociodemographic descriptive analysis has established that the life histories of homo–bisexuals are not significantly different from those of heterosexuals. In particular, they experience periods of life in couples, with same- or other-sex partners, as well as periods of separation and isolation (Messiah and Mouret-Fourme, 1996; Schiltz, 1997).

Studying the complexity of partnership and sexual histories and how they are linked to family and conjugal behaviors is in fact to assess the place sexuality occupies in the lives of individuals (Bozon, 1998b). Individuals' profiles are formed very early, at the time of sexual initiation. The priority that some individuals give to conjugal relationships, leads them not to put sexual activity at the forefront of their lives; sexual pleasure is not sought for its own sake and sexuality is important chiefly for its symbolic role in the life of a couple that is itself reasonably orderly. Individuals with this conjugal orientation tend to be those who have had a late sexual initiation. For them, sexuality is merely one component of the couple's life and has a secondary place among their preoccupations. For other individuals, by contrast, whose sexual careers have usually started young, sexual activity is in some sense independent and a source of essentially personal gratification—in this perspective, stimulating desire within the couple and acquiring new sexual partners are valued for their own sakes, and partnership histories are likely to be more *eventful*, containing more separations and parallel adventures. This coexistence in the population of contradictory attitudes (conjugal orientation/individual orientation) is long-standing, but in the past it remained below the surface. The conjugal and family changes of the 1970s and 1980s occurred in part because these contrasting attitudes toward sexuality began to find open expression in much more diverse conjugal behaviors and itineraries.

The sociohistoric contexts specific to groups of birth cohorts thus help to construct and shape the sexual experiences of those cohorts. A sexual life that began in a post-AIDS world cannot be analyzed in the same terms as a sexuality that began in the 1950s at the time of the baby boom. The availability of treatments for AIDS (from 1996 on) has changed the lives of persons living with AIDS and the general conditions of sexual experience, but more so in the developed countries than in the developing world. Whatever the period, however, a major component in the study of sexuality is the state of gender relations in the period under consideration.

IV. GENDER RELATIONS AND SEXUALITY

Surveys of sexual behavior invariably record systematic differences in the responses of men and women. This cannot be attributed simply to reporting effects. These divergences are the expression of differences both in norms and attitudes, and in behavior. Some even take the form of inconsistencies, when the same event is the subject of contradictory accounts. Such inconsistencies are nonetheless highly revealing of what has changed and what has persisted in gender relations.

Women's sexuality continues to be expressed in commitments and frameworks very different from those of men. The transformations of female sexuality observed in different countries since the 1960s and 1970s have not been reflected in greater tolerance of women in regard to men, but rather in a more demanding attitude toward their spouses or partners and an increased willingness to break off unsatisfactory relationships and to embark on new ones. In France, for example, the proportion of women who consider extraconjugal relationships unacceptable was higher in the 1990s than it was around 1970. If the sexual and conjugal itineraries of women have become more varied, this is not because of any change in their aspirations—their fundamental ambition remains that of living their sexuality within a stable couple. Women attach far more importance than men to the commitment or bond, be it conjugal or simply affectional, as the basis for the sexual relationship. Accordingly, sharp contrasts are observed in how sexual histories are experienced and interpreted.

When women are questioned about their first sexual intercourse, they often report that this was with a stable partner, with whom they were in love and who was older than them. Men usually begin their sexual lives younger, especially in the southern

European countries, and have more occasional partners with whom they are not emotionally involved. As a general rule, men have many more sexual partners than women before forming their first stable relationship.

Men living in a couple or in a stable relationship are consistently more likely than women to report having had a recent sexual partner outside their relationship. In line with this behavior, they are more ready to approve the statement that "having sex affairs during one's marriage is not wrong ('acceptable')" Unlike men, women are very reluctant to report autoerotic activity, and they are less likely to report having practiced oral sex or experimented with anal penetration. Other systematic differences are observed—men more often report use of condoms, including in last intercourse, while more women report having had a sexually transmitted infection (STI).

The sexual experiences of men and women tend to diverge sharply after age 50. The menopause is the end of a woman's fertile life and in many societies this was traditionally held to be the end of her sexual life, whereas a man faces no such symbolic limit. The transformations in the social position of women over the last 30 years, more than treatments for the menopause, have reduced the importance of this threshold and the existence of a female sexuality after age 50 has become more acceptable. At this age, women in couples in the 1990s have a much more active sexual life than their mothers had at the same age. But the inequalities between men and women persist—because of men's preference for younger women and women's longer life expectancy, women without spouses (due to divorce and, above all, widowhood) still have difficulty finding a new partner (Delbès and Gaymu, 1997).

These differences between men and women can be interpreted from two opposing perspectives. In a psychobiologic perspective, it could be argued that gender differentiation in sexual behavior derives primarily from the different roles of males and females in the reproductive process. But this explanation is inadequate, particularly for historic contexts where the development of medical contraception has freed the exercise of sexuality from the constraints of procreation. In a gender-relations perspective, the norms, representations, and behavior of men and women in the sexual domain are the results of an explicit socialization; they express one of the fundamental forms of male power.

Not all countries have completed the historic and demographic process that allows women's sexuality to be viewed as other than a strictly controlled social obligation to procreate (Tabet, 1985). The process is least advanced in Africa, even though fertility decline is now apparent in all countries of the continent (Locoh and Makdessi, 1995; Locoh and Vallin, 1997; Tabutin and Schoumaker, 2004). This is where girls' age at marriage is lowest, close to puberty in some rural zones (Niger, Mali), and where the proportion of a woman's lifetime spent in reproduction (pregnancy and breastfeeding) remains largest; it is here too that sterility carries the greatest stigma and women's sexuality after menopause is least accepted. Moreover, given the widespread practice of female genital cutting in Africa, questions must be asked about the place of pleasure in women's sexuality.

Most non-African countries have completed their demographic transition and achieved a twin stabilization of fertility and mortality at relatively low levels. The historic decline in fertility occurred earlier in the developed countries, where it was due primarily to nonmedical forms of contraception. These methods relied heavily on self-mastery by the man (Fisher, 2000), although fear of unwanted pregnancy and its consequences were still borne primarily by the woman. In France, for example, fertility decline began just before the Revolution of 1789 and extended over almost 150 years. In Britain, it was not until the second half of the 19th century (Szreter, 1996). Not until the 1960s did much more reliable modern forms of contraception become widely available, in a sociohistoric context characterized by greater independence and agency for women (Leridon, 1987). These changes prepared for the major transformations that revolutionized the way women actually lived their sexuality. In the developing countries, the shift from high fertility to fertility close to replacement level, is occurring much faster, in the space of three or four decades, causing serious adjustment problems, notably in the area of gender relations (see for instance, Martine *et al.*, 1998).

A new life course is emerging in which the reproductive thresholds play a less important role. The time spent either pregnant or breast-feeding has come to represent only a small proportion of a woman's total reproductive lifespan. Periods of active sexuality with no procreative objective are appearing—a conjugal sexual life that continues longer though produces few offspring; a juvenile sexual life that precedes partnership formation; a sexuality following partnership dissolution; an active postmenopausal sexuality. In this way sexuality is associated with more hedonist values of pleasure, well-being, and mutual support between partners. For women, sexuality can become, in ideal conditions, an intimate ritual of the egalitarian couple: it stops being considered a duty. Whether this view of the couple is fully endorsed by men is questionable. Even today the experience of men and women is largely contradictory.

The modernization of sexual attitudes, the development of contraception and the contemporary changes in the status of women, have produced no thoroughgoing upheaval in the representations and unequal roles of men and women in this field. Men tend to be viewed, and to view themselves, as free desire-driven subjects, whereas women are often still seen (and see themselves) as objects to be possessed. The persistence of an archaic sexual violence against women in all countries confirms that male power can take extreme forms when circumstances permit. Women's attitude remains that of trying to stabilize and regulate male desire by containing it within the framework of a relationship. Furthermore, due to their more sustained assumption of parental roles and the representations associated with them, women still tend to see sexual activity and even the relationship with their partner as being less important than longer-term bonds, like those of parent to child.

V. CROSS-CULTURAL COMPARISON OF SEXUAL BEHAVIOR: THE CONTRIBUTION OF DEMOGRAPHY

If demography is valuable for interpreting the historicocultural changes that have occurred in the field of sexuality, notably in the second half of the 20th century, it can also be used to characterize the differences of sexual climate or sexual culture observed between countries, both near and far, in the contemporary era. These comparisons are often done in an unscientific and unverifiable way, and using questionable national stereotypes—a group of people are broadly categorized as conservative and puritan, another as sexually open and sophisticated, yet another as having an unrestrained sexuality, and so on. One of the tasks for demographic analysis of sexual behavior is to introduce verifiable cultural characterizations, based on sets of comparable indicators. This approach is distinct from, although complementary to, that of anthropology, which is more concerned with capturing a culture in its entirety. The demographic approach, however, shifts emphasis away from a cultural explanation of differences, in that it attempts instead to decompose phenomena, by distinguishing how sexual behaviour in different cultures is influenced by variables or explanatory factors such as gender, rural–urban contrasts, marriage systems, and the social organization of age groups.

We have conducted a comparative analysis of the characteristics of sexual initiation of men and women in about 50 mainly European, African, and South American countries from the 1970s to the 1990s (for detailed figures, see Bozon, 2003). Comparison of the respective male and female ages at sexual initiation in older cohorts (born in the early 1950s) reveals three major traditional models of entry into sexual activity.

First is the large group of societies where parental and family strategies are designed to avoid any delay in female entry into conjugal, reproductive and sexual life through pressure to form a union as near puberty as possible with significantly older men, compounding gender domination by age domination. Although in such societies male may be later than female sexual initiation, males generally experience a significant period of premarital sexual activity. This model is traditionally found in sub-Saharan Africa (e.g., Mali, Senegal, and Ethiopia), and very likely also in the Indian subcontinent (represented in our sample only by Nepal).

This contrasts with a second group, which includes the Latin and Latin American cultures, where social control is exerted to delay women's union formation and entry into sexual activity by strenuous efforts to preserve their virginity, premarital loss of which breaks a cardinal rule and dishonors the family group and the spouse. By contrast, young males are urged to prove their manhood early on, either with prostitutes or older women, and their sexual initiation takes place well before that of females. This group of countries includes southern Europe (from Portugal to Greece and Romania), Latin America (Brazil, Chile, Dominican Republic), Thailand, and other Asian societies for which quantitative data are not available. In Brazil, for example, in the 45- to 49-year-old age group, the median age gap between first intercourse and first union is 7 years for males, but just 1 year for females. In the two models described, women are under great pressure to conform to social norms, but there are equally strong injunctions for males in due course to behave as *real men*.

There is a third group of societies, less homogeneous than the other two, characterized among the older cohorts by closely matched male and female timings of entry into sexual life. This may reflect later marriage and strict supervision of the conduct of young people, including males, as in Singapore or Sri Lanka (and doubtless also China and Vietnam). Late male and female sexual initiation is found in non-Latin Catholic societies like Poland and Lithuania. Finally, many European societies, like those of Northern Europe, but also Switzerland, Germany, and the Czech Republic, have for several generations been *gender-equal* in rather early sexual initiation timings.

There have been universal, but limited, changes in recent decades, as reveals an examination of the behaviors of the younger cohorts (born in the second half of the 1970s). In cultures where female is significantly earlier than male sexual initiation, the trend has reversed in two ways over the past 20 years, with a slight increase in female age at first coitus and a sharp drop in the male age. The female–male timing gap has narrowed in sub-Saharan Africa, as in Senegal, Togo, Cameroon, Mozambique, and Zimbabwe; likewise for Nepal. A new trend emerging in some countries (Kenya, Gabon, and Zambia) shows that male ages are now slightly earlier than females. In some societies, finally (Niger, Burkina Faso, Ethiopia), there is barely any discernible change. One common trait in all these trends is that young people are breaking free of a series of traditional controls: Males to become sexually active earlier, females to lighten the constraint of early marriage without necessarily relinquishing entry into sexual life. So, in eastern African countries, 40% to 50% of young women have at least 2 years of sexual activity before their first union. In some AIDS-ravaged countries, campaigns to delay first sexual intercourse may have had an inhibiting effect, particularly among females.

In the group of societies that prize early male and late female sexual activity, any changes have mainly concerned women—their comparatively late age at first intercourse has tended to fall, whereas the male age has remained unchanged. So, female–male timing gaps have narrowed in Brazil, Chile, and Spain, while in other countries—like Bolivia, Nicaragua, and the Dominican Republic—they are unchanged. In Portugal and Romania, sexual initiation timings are highly gender-specific, as in Italy, where female age at first intercourse has actually risen in the last decade. Even as virginity at marriage becomes a receding ideal and reality, a clear gender-based double standard still operates. Where male remains much earlier than female first intercourse, initiation by prostitution remains a significant factor, as in Central America and Thailand. But prostitution as a vehicle for first coitus declines where timing gaps narrow, as in Japan, where they have dropped from 26% in older cohorts to 11% in younger cohorts.

Where male and female ages at first intercourse were already close, they have fallen simultaneously: In many European countries and the United States, they have dropped by approximately a year in the last two decades, although age at first union—now very distinct from age at marriage—has trended upward over the same period. That this fall is not sharper may be due to a stabilizing and synchronizing effect of spreading secondary education on sexual initiation timings, which now occur within an increasingly shorter interval toward the end of secondary education. They mark the end of adolescence and entry into an age of personal free agency—youth—rather than transition to an increasingly later adulthood. The Nordic countries, especially Iceland, have become the *earliest* of developed countries, to the point where in some (Norway, Denmark), females enter their sexual life before males. These societies have long tolerated premarital relationships, and a long-standing and strongly egalitarian trend marks the behavior of men and women in the social, occupational and sexual spheres.

Another comparison, based on 15 surveys conducted in Africa and Asia and one in Rio de Janeiro, using a common protocol developed by the World Health Organization (WHO) also reveals variations of a sometimes-unexpected kind (Caraël, 1995). Clear cross-national differences are observed for the frequency of intercourse within married and stable cohabiting couples. Although fertility levels are high, African countries are characterized by infrequent and irregular sexual activity within stable couples, even when polygamous situations are excluded. This is a consequence of male migration and postpartum abstinence, which can last almost 2 years, in a context where fertility remains high; the irregularity also reflects the existence of noncohabiting couples. To a lesser degree, this low frequency is observed in Manila and Sri Lanka. Conversely, high levels of conjugal sexual activity characterize Rio de Janeiro and continental Asia (Thailand, Singapore). A comparative study was also performed on the proportions of men and women having sexual relations with casual partners or prostitutes, outside their stable partnership or conjugal relationship. Very high proportions are recorded for men in West and East Africa. The proportions are intermediate in Rio de Janeiro and in Thailand and very low in the rest of Asia. The role of commercial sex in casual relationships varies greatly between countries. A culture of using prostitutes exists among men in Tanzania (East Africa) and in Thailand, where prostitution accounts for almost all extraconjugal relations. In West Africa and Rio de Janeiro, on the other hand, commercial sex has much less relative importance in the casual relations of men in stable partnerships. Interpretation is problematic because of large cross-national variations in the social status of prostitution—it does not carry a stigma everywhere and does not always involve a monetary transaction.

Having simultaneous multiple partners is also observed in Europe though at much lower levels (Leridon, 1998; Hubert *et al*., 1998). France and Finland, one of the Nordic countries, report the highest frequencies, but with an interesting difference—in Finland this behavior is reported in almost equal proportions by men and women, whereas in France, as in most southern European countries, it is reported mainly by men. The same specificity of the two countries, and the same contrast between them, appears in the repertoire of sexual practices—in Finland this is quite varied for both men and women, whereas in France it is much less varied for women than for men. These variations should be interpreted not as sharp behavioral differences between French and Finnish women but as expressions of differing openness to women's sexuality.

Countries are also contrasted by their open and tolerant sexual climate. This climate is reflected in legislation on sexual orientation, the forms chosen for sex education, and the structures for the provision, and the distribution of contraception. It is noteworthy that a country like The Netherlands with a tradition of liberal legislation on sexual orientation and where discrimination against homosexuals is illegal, reports a higher proportion of homosexuals in the surveys than other European countries. Several countries have made provisions for the legal recognition of same-sex couples (Festy, 2001): Northern European countries being the first (Denmark in 1989), gradually followed by southern Europe (France in 1999). This translates and indeed strengthens a climate of tolerance.

Use of a sociodemographic perspective for the comparison of sexual behavior allows us to identify the points of cross-national difference more precisely and to suggest the main factors responsible. Of these, gender relations, in their specific national configurations, have a great impact, but primarily through their influence in each country on the social organization of unions and couples—the formal or informal character of unions, the proportion of noncohabitating couples, the independence of individual partners in couples, the degree of social control over women, the material possibility or impossibility for women to leave their partner and live alone, and acceptance of same-sex relationships. Norms and national sexual cultures operate more at a second stage to provide the form and justification for this social organization of relationships.

CONCLUSION

Demography does not hold a dominant position in social science research on sexuality; however, neither do any of the other disciplines (sociology, anthropology, history, social psychology, psychoanalysis, etc.) active in this field. It would be undesirable for a single discipline or even a subdiscipline to unify the investigation of sexuality by taking it for unique object and becoming the science of sexuality, in the way that oceanography is the science of the oceans, for example. Sexuality has a reality but it cannot be approached from a single perspective. The specific contributions of individual disciplines are as valuable as the interpretative incompatibilities sometimes observed between them. One of the strengths of demography is its attachment to the construction, application and generalization of rigorous indicators of sexual behavior, with which to achieve an objectification of sexuality, and substantiate the notion of behavioral diversity. A sharing of disciplinary perspectives is possible. The importance that demographers attach to the sexual and partnership history in the construction of behavior, creates affinities with more individual-oriented disciplines, such as psychoanalysis and psychology. Analysis of the links between gender relations and sexuality favors collaboration between demography and sociology. With anthropology, meanwhile, demography shares a deep interest in the social and cultural organization of marriage and unions. Sexuality is preeminently a relational activity and interdisciplinary object, whose full contours emerge only in the endless chain of interpretations produced by the different disciplines.

Acknowledgment

Translated from French by Godfrey I. Rogers.

References

Ariès Philippe, 1991. La contraception autrefois, *in: Amour et sexualité en Occident*, p. 115–131. Paris, Seuil, 335 p.

Berquó Elza, 1984. União dos sexos e estrategias reprodutivas no Brasil, *Revista Brasileira de Estudos de Populacão* (Campinas), vol. 1(1–2), p. 35–98.

Bongaarts John and Potter Robert G., 1983. *Fertility, biology and behaviour: an analysis of the proximate determinants*. New-York, Academic Press, 230 p.

Bozon Michel, 1995. Observer l'inobservable: la description et l'analyse de l'activité sexuelle, *in:* Nathalie Bajos, Michel Bozon, Alain Giami *et al*. (eds.), *Sexualité et sida. Recherches en Sciences Sociales*, p. 39–56. Paris, ANRS, 382 p.

Bozon Michel, 1996. Reaching adult sexuality. First intercourse and its implications: from calendar to attitudes, *in:* M. Bozon and H. Leridon (eds.), *Sexuality and the social sciences: a French survey on sexual behaviour*, p. 143–176, Dartmouth Publishing, Aldershot.

Bozon Michel, 1998a. Amour, désir et durée. Cycle de la sexualité conjugale et rapports entre hommes et femmes, *in:* Nathalie Bajos, Michel Bozon, Alexis Ferrand, Alain Giami and Alfred Spira (eds.), *La sexualité aux temps du sida*, p. 175–234. Paris, PUF, 494 p.

Bozon Michel, 1998b. La sexualité a t-elle changé? Regards sur l'activité sexuelle et sur ses significations à l'ère du sida, *in:* Nathalie Bajos, Michel Bozon, Alexis Ferrand, Alain Giami and Alfred Spira (eds.), *La sexualité aux temps du sida*, p. 11–34. Paris, PUF, 494 p.

Bozon Michel, 2001, Sexuality, gender and the couple: a sociohistorical perspective, *Annual Review of Sex Research*, vol. XII, p. 1–32.

Bozon Michel, 2003, At what age do women and men have their first sexual intercourse? World comparisons and recent trends, in *Populations et Sociétés*, no. 391, English version (can be downloaded on http://www.ined.fr)

Bozon Michel and Heilborn Maria Luiza, 1996. Les caresses et les mots. Initiations amoureuses à Rio de Janeiro et à Paris, *Terrain*, no. 27, p. 37–58.

Bozon Michel and Kontula Osmo, 1998. Sexual initiation and gender in Europe: a cross-cultural analysis of trends in the twentieth century, *in:* Hubert Michel, Bajos Nathalie, and Sandfort Theo, 1998 (eds.). *Sexual behaviour and HIV/AIDS in Europe. Comparisons of national surveys.* London, UCL Press, p. 37–67.

Bozon Michel and Leridon Henri, 1996. The social construction of sexuality, *in:* M. Bozon and H. Leridon (eds.), *Sexuality and the social sciences: a French survey on sexual behaviour*, p. 3–26, Dartmouth Publishing, Aldershot.

Caraël Michel, 1995. Sexual behaviour, *in:* John Cleland and Benoît Ferry, *Sexual behaviour and AIDS in the developing world*, p. 75–123. London, Taylor and Francis, 244p.

Corbin Alain, 1991. La petite Bible des jeunes époux, *in: Amour et sexualité en Occident*, p. 235–247. Paris, Seuil, 335 p.

Davis Kingsley and Blake Judith, 1955–56. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol. 4(4), p. 211–235.

Festy Patrick, 2001. The "Civil Solidarity Pact" (PACS) in France: an impossible evaluation, *Populations et Sociétés*, no. 369, English version (can be downloaded on http://www.ined.fr)

Delbès Christiane and Gaymu Joëlle, 1997. L'automne de l'amour: la vie sexuelle après 50 ans, *Population*, vol. 52(6), p. 1439–1484.

Fisher Kate, 2000. Uncertain aims and tacit negotiation: birth control practices in Britain, 1925–1950, *Population and Development Review*, vol. 26(2), p. 295–317.

Flandrin Jean-Louis, 1981. *Le sexe et l'Occident.* Paris, Seuil, 376 p.

Gini Corrado, 1924. Premières recherches sur la fécondabilité de la femme, *in: International Mathematical Congress, Toronto*, p. 889–892. Toronto, The University Press, 2 vol.

Herdt Gilbert, 1997 (ed.). *Sexual cultures and migration in the era of AIDS. anthropological and demographic perspectives*, Oxford, Clarendon Press, 256 p.

Hubert Michel, Bajos Nathalie, and Sandfort Theo, 1998 (eds.). *Sexual behaviour and HIV/AIDS in Europe: comparisons of national surveys.* London, UCL Press, 442 p.

Johnson Ann, Wadsworth Jane, Wellings Kaye, and Field Julia. *Sexual attitudes and lifestyles.* Oxford, Blackwell Scientific Publications, 1994. 528 p.

Laumann Edward, Gagnon John, Michael Robert, and Michaels Stuart, 1994. *The social organization of sexuality: sexual practices in the United States*, Chicago, University of Chicago Press, 718 p.

Leridon Henri (ed.), 1987. *La seconde révolution contraceptive: la régulation des naissances en France de 1950 à 1985*, Paris, PUF-INED, 380 p. (Travaux et Documents, Cahier no. 117).

Leridon Henri, 1996. Coital frequency: Data and consistency analysis, in M. Bozon and H. Leridon (eds.), *Sexuality and the social sciences: a French survey on sexual behaviour*, p. 203–226, Dartmouth Publishing, Aldershot.

Leridon Henri, 1998. Les nouveaux partenaires: estimation de leur nombre et de l'évolution au cours des années récentes dans une enquête transversale, *in:* Nathalie Bajos, Michel Bozon, Alexis Ferrand, Alain Giami and Alfred Spira (eds.), *La sexualité aux temps du sida*, p. 153–174. Paris, PUF, 494 p.

Locoh Thérèse and Vallin Jacques, 1997. Afrique noire: la baisse de la fécondité, *Population et sociétés*, no. 338, September, p. 1–4.

Loyola Maria Andrea, 1992. Sexualidade e Reproducão, *Physis. Revista de Saude Coletiva*, (UERJ, Rio de Janeiro), no. 1, p. 93–105.

Macro International, 1994. *Women's lives and experiences.* Washington, Macro International Inc., 48 p. (DHS Program).

Martine George, Das Gupta Monica, Chen Lincoln, 1998, (eds.). *Reproductive change in India and Brazil*, Oxford, Oxford University Press, 419 p.

Messiah Antoine and Mouret-Fourme Emmanuelle, 1996. Homosexuality, bisexuality: elements of sexual socio-biography, *in:* M. Bozon and H. Leridon (eds.), *Sexuality and the social sciences: a French survey on sexual behaviour*, p. 177–202, Dartmouth Publishing, Aldershot.

Parker Richard, 1991. *Corpos, Prazeres e Paixões. A cultura sexual no Brasil contemporaneo*. São Paulo, Best Seller, 295 p.

Schiltz Marie Ange, 1997. Parcours de jeunes homosexuels dans le contexte du VIH: la conquête de modes de vie, *Population*, vol. 52(6), p. 1485–1536.

Szreter Simon, 1996. *Fertility, class and gender in Britain, 1860–1940*, Cambridge, Cambridge University Press, 704 p.

Tabet Paola, 1985. Fertilité naturelle, reproduction forcée, *in:* Nicole-Claude Mathieu (ed.), *L'arraisonnement des femmes. Essais en anthropologie des sexes*, p. 61–146. Paris, Editions de l'EHESS (Cahiers de l'Homme).

Tabutin Dominique and Schoumaker Bruno, 2004. The Demography of Sub-Saharan Africa from the 1950s to the 2000s: A Survey of Changes and a Statistical Assessment. *Population E*, no. 3–4, p. 457–554.

Toulemon Laurent and Leridon Henri, 1992. Twenty years of contraception in France: 1968–1988, *Population: An English Selection*, vol. 4, p. 1–314.

van de Kaa Dirk J., 1987. Europe's second demographic transition. *Population Bulletin*, vol. 42(1), p. 1–59.

PART

III

SOCIOECONOMIC AND CULTURAL FACTORS OF FERTILITY

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Although fertility-control decisions are ultimately a matter of private sexual life, that does not make fertility strictly the affair of individuals and couples. Like all human behavior, reproductive behavior is influenced by the economic, social, political, and cultural context in which individuals and couples live. This vast topic has been addressed in a copious body of literature, which must be reflected here as exhaustively as possible.

In Chapter 37, Heather Joshi and Patricia David review the thinking of recent decades on the fertility impact of economic and social circumstances, and how the literature addresses seven key issues affecting developed and developing countries alike:

Why should fertility decline as development levels rise?
Can fertility decrease in poor countries?
Is the influence of women's education dependent on materialistic or cultural determinants?
Is reproduction a form of old-age security?
Why does fertility fluctuate in prosperous countries?
Do public policies promote faster fertility decline?
Can public policies prevent fertility declining below replacement level?

Connecting up with this initial survey, in Chapter 38 Ron Lesthaeghe and Karel Neels take two examples—France and Belgium—as the basis for a geographic interpretation of the fertility behavior changes observed over a 150-year transition, illustrating how new behaviors—demographic innovation—always arise at exactly the same places to spread nationwide following the same geographic dynamic. The same sociocultural patterns have driven the adoption of new behaviors in turn at the different stages of change.

In Chapter 39, which in a way sums up Section II, Bart de Bruijn reviews and comments on the many theories and explanatory models which have come out of post-18th century thinking on fertility trends and differentials, from Malthusian propositions to contemporary psychological approaches via the sociobiologic, economic, institutional, diffusionist, culturalist, and other theories, not, of course, forgetting demographic transition theory itself.

CHAPTER

37

The Social and Economic Context of Fertility

HEATHER JOSHI AND PATRICIA DAVID

Centre for Longitudinal Studies, Institute of Education, University of London, London, United Kingdom

Harvard School of Public Health, Boston, United States

Features of the world beyond the individual and the couple determine whether, why and how fertility is controlled. The economic and social systems provide the opportunities and constraints within which choices can be made. They also help determine whether people are aware that choice is possible, whether, as Coale (1973) put it, fertility is within the sphere of conscious control. The economic and social context may also help to set the standards of behavior between family members and the values to which they aspire. How far each sort of consideration applies to the explanation of fertility changes already observed, and how far social and economic variables can be used to predict, or manipulate, the future course of human reproduction remain controversial. This chapter is mainly concerned with the social and economic explanations of deliberate actions whose cumulated outcome is the rate of human reproduction. Leaving biologic factors aside, already discussed in Chapters 30 to 34, we treat fertility as a form of behavior rather than as a response to innate drive that has been *bred in* to survivors. We also leave aside, for the most part, the question of partnership formation and dissolution that are dealt with in Chapters 25 to 29.

Four topics will be considered. The first three sections elaborate some of the theoretical explanations of fertility offered in recent literature. First we describe a behavioral framework for the understanding of social reproduction (and make a passing reference to social biology). Next we review some approaches from just two of the social sciences, economics and sociology. Finally, we consider some contemporary issues, in both developing and industrial countries, on which empirical applications of these approaches have shed both heat and light. The questions selected are as follows:

Why does fertility decline with development?
Can fertility decline in poor countries?
Women's education: a materialist or ideational influence?
Is childbearing for old age security?
What makes fertility fluctuate in rich countries?
Are state policies effective in accelerating fertility decline?
Are state policies effective in preventing subreplacement fertility?

This is more an account of territory through which we have traveled than a complete *Michelin Guide*. We have not attempted a comprehensive coverage of topics or literature. We hope to spark enough interest to encourage the reader to discover the numerous authors we have neglected (particularly outside the Anglo-Saxon literature) and that what we do cover will show the way for further interdisciplinary endeavor.

I. SUPPLY AND DEMAND: A FRAMEWORK OF FACTORS INFLUENCING FERTILITY

The level of fertility in a society or a group is the cumulated outcome of deliberate actions (behavior) and biologic functioning, social pressures and chance operating for each pair of parents, actual or potential. The constraints that condition behavior are largely determined by the social environment they inhabit. Coleman and Schofield (1986) observed that demography still has no central paradigm linking processes at the microlevel with those operating at the societal level. Microlevel behavior must be involved if there is any change in fertility in the aggregate, at the macrolevel; but microlevel variation can also be observed, and linked to variations within an economic and social system that is in aggregate unchanging. The primary focus of most (nonbiologic) fertility theories is therefore on the individual.

Contextual factors external to the individual or couple that affect behavior and/or decision-making regarding family formation must work through mechanisms that operate at the individual level. One general framework that has proven useful is set out by Easterlin, Pollack, and Wachter (1980). They classify these mechanisms into those affecting the biologic production of births, *supply* or *natural fertility*, and *demand* factors, which may reduce wanted fertility below the biologically feasible maximum. Their approach, sometimes identified as that of the *Pennsylvania School*, is a general formulation not only because it allows for fertility sometimes being constrained by biologic limits, but also because it allows for the preferences, which define demand, themselves to be determined within the system.

The Pennsylvania approach is contrasted with the *Chicago-Columbia School* whose exponents include Becker (1981), Schultz (1981), and Cigno (1991). This group of writers adopt a *demand* approach to fertility, seeing decisions about the number, timing (and quality) of children as the outcome of rational decision making, maximizing the utility to be gained from fertility decisions subject to the resources available and the rates (*prices*) at which they may be transformed into competing sources of utility. It is assumed that tastes or preferences remain fixed, and are unified within couples. The reactions of the rational decision-maker to changes in *income* and *prices* are predicted.

One variant of the supply–demand framework is to investigate the constraints around which decisions are made. For Hobcraft and Kiernan (1995) the constraints on parenthood are biology, money, security, ideas, and especially time. For Folbre (1994), constraints are socially structured—by membership of social groups, such as social class, religious, or ethnic groups and gender. Membership of such groups overlaps, and thus individuals are subject to diverse, possibly conflicting pressures. In different circumstances different types of group membership may be dominant. Her book contains a very extensive review of the evolving structures over several continents and centuries. This recognition that the demographic actor's tastes and behavior are affected by group identity within broad societal influences helps to account for differential fertility within societies. It also suggests that group norms and interests may mediate the impact of social and economic considerations upon any given individual. The overlapping nature of group membership adds complexity to the notion of self-interest.

Another variant of the supply–demand framework, as adopted by Bulatao and Lee (1983) adds specific consideration of the reconciliation of supply and demand. The deliberate regulation of the *supply* of children is conditioned on both *demand* factors and the costs of, or obstacles to, fertility regulation. These include social influences on attitudes to, knowledge of, and access to contraception, abortion, or the practice of infanticide; the gender balance of power to make decisions regarding control of reproduction; and husband–wife communication regarding fertility regulation. Where births are *unwanted*, there is said to be *unmet need* for control of fertility.

The next two sections elaborate factors than can be thought of as influencing the supply–demand model of fertility decisions.

1. Supply

Supply factors affect the number of surviving children by regulating the length of the reproductive life span and the spacing of births within it. Beyond purely biologic factors, they include social constraints on the timing of entry into marriage or sexual union, and practices affecting the timing of parenthood. Constraints on behaviors directly affect the biology of childbearing, such as norms regarding breastfeeding and postpartum abstinence (Chapters 30 and 31), and use of contraception or abortion (Chapters 34 and 35). These regulate the pace of childbearing during the reproductive years. Supply factors also include features of the material environment (such as nutrition and hygiene) that may affect the chances of conception and spontaneous abortion.

Factors that affect survival of liveborn children, particularly where there are high rates of infant and child mortality, can be thought of as affecting the supply of surviving progeny. Infant mortality may indirectly affect subsequent fertility through shortened breastfeeding even if there is not any deliberate intention to replace a child who dies (Bongaarts, 1983). The end of childbearing is affected by social factors—also often group-specific—such as those affecting maternal survival and widowhood; norms governing separation, divorce, and remarriage; and those influencing deliberate means of terminating childbearing such as sterilization, terminal abstinence, or effective contraception.

Some *supply* determinants are more clearly the result of the behaviors of the individual, such as breastfeeding practices and the timing of marriage, variable across and often within societies around the world. Others, like child or maternal mortality levels, are not behavioral in nature, but are affected by the material conditions of life and available health technology. The social environment (a *macro* consideration) determines the resources available for the care of mother and child, and the knowledge (*technology*) of how this should be done, and sets norms for acceptable practices. The social environment also provides any knowledge about methods of birth control and affects their cost, social as well as monetary.

The technology of *producing* children of a given standard goes beyond biologic factors to the investment of resources made in children once born, building up their *human capital* of health and skills. Changes in any of this technology modify what reproductive outcomes are feasible and permit different outcomes to be chosen. The state of knowledge about producing everything else can also change, which will also affect the setting in which fertility decisions are made. Note, too, that in the production of successful surviving children, if nothing else, the venture is beset with risk and uncertainty. Outcomes could be quite different if it became possible to determine the sex of a child in advance or guarantee that a woman would still be able to conceive if she postponed motherhood until her 40s.

These factors, *proximate* determinants of childbearing, will be affected by the social and economic environment. Society, for example, can influence who enters sexual unions, at what age, and on what terms. In historic times in Europe, norms surrounding marriage acted as an important *safety valve* to control fertility (as well as containing sexuality), preventing the formation of new households when economic conditions were poor. Marriage is now less important with respect to childbearing in the United States and Europe, and other social norms regulate age at first birth. The first birth may precede formal marriage in many African societies, where proof of fecundity may be an important precondition to finalizing the marriage bond. Far more important to reducing the overall level of fertility is the deliberate regulation of fertility within marriage.

2. Demand

To economists at least, it seems natural to turn to the analogy of the buyer's side of a market once fertility enters the realm of control, not merely limited by biosocial *supply* constraints. When families can do little but accept however many (or few) children *God sends*, the question of whether they have adapted their behavior to achieve desired fertility does not arise. Once it becomes technically feasible, and thinkable, to reduce childbearing below the *natural* maximum, it becomes relevant to ask what fertility is desired, and to seek explanations for reproductive behavior in its motivation as well as its means.

People may be able to specify the fertility they would like to achieve by stating a family size, perhaps depending on sex composition. They may be able to distinguish between an *ideal* family and the number that they would prefer given existing circumstances. Even parents who embark on family formation with no clear goal, may come to weigh up the advantages and disadvantages of adding another birth: the *marginal child* (Leibenstein, 1974). The death of an existing child may motivate another birth as a *replacement* (Vallin and Lery, 1978), but the question still remains of why a particular target family size was desired, and whether it is worth the costs of achieving it. Those who decide to remain voluntarily childless (or *child-free* to put it more positively) can be thought of as taking the decision that even the first child is not worth having, but as being generally influenced by the same considerations as those who choose to limit a positive number of children (Keilman 1996).

The *demand* approach to fertility applies microeconomic theory of *rational decision making* to the whether and when to control fertility. The decision-maker is aiming to attain maximum feasible *utility* from the sources of satisfaction available, which include children. How much these are valued (i.e., how much satisfaction they provide) will be affected by the parents' *preferences*, or tastes. The parents' commitment to bringing a child into the world and bringing it up to adulthood is likened to investment in a consumer durable. The benefits expected must be weighed against the cost of achieving them, which is the forgone return from an alternative investment of their resources.

This reduction of complex, private, and moral decisions to the self-interested calculus of the counting house is not universally palatable. The theories do recognize that children are not exactly like physical assets, as they cannot be sold. They can also allow for parents acting in their children's, as well as their own, interests. Through the assumption of altruism, one decision-maker draws satisfaction from the well-being of another. Indeed Folbre (1994) caricatures the actors in neoclassical fertility theory as a family of *Rational Economic Men*, perfectly altruistic and harmonious at home, and single-mindedly self-seeking in the market place. Such agents need not be constantly adjusting their private lives to economic signals; it suffices that some economic changes induce some people to adjust their family formation behavior, for it to be worth thinking about what such adjustments might consist of. The assumption of a united household is being challenged by the application of game theory to bargaining models. This highly technical literature is not covered by the present review.

3. Social Biology: a Digression

Why is there a demand for children? Does biology help explain the demand as well as the supply side? The fundamental reasons people bear children are often assumed to lie in a biologic imperative to reproduce. Evolutionary biologists, working with animal models of reproductive behavior, have tried to extend these to the human organism. These theorists now propose that when it is in the individual's interest to overexploit existing resources, populations will grow too large for the available pool of resources and the population will experience excess deaths as a result (a *Malthusian check* on the growth of the population).

Lately, some of these theorists have examined reproductive strategies as adaptive mechanisms that individuals used to maximize the *fitness* of their genetic inheritance. These *life-history* theorists are concerned with understanding how alternative reproductive strategies evolve to affect reproductive success. Use of the term *strategy* suggests conscious choice, but its use by biologists does not make such an assumption.

One school of thought suggests that such *reproductive success* may be achieved by having fewer offspring, and investing more parental effort in their *quality*. This may be the best strategy for ensuring the long-term reproduction of their genetic traits (Chisholm, 1993). Attention is focused on how physiologic factors that directly affect reproduction have evolved, and why norms that affect birth spacing patterns, such as prohibitions on intercourse or prescribed breast-feeding durations, are found to vary across different cultural contexts (Wood, 1994). Why have these evolved in this particular way?

According to some life-history theorists, mortality risk is what determines these differences in the reproductive strategy that is adopted (Promislow and Harvey, 1990, 1991; cited in Chisholm, 1993). When mortality rates are high, the most efficient strategy is to maximize the number of offspring: a short-term strategy to ensure the survival of their genetic traits, because at least some of their offspring will survive to reproduce. When mortality is low, so the theory goes, the best strategy is to invest more parenting effort in fewer offspring: a long-term strategy to maximize the number of descendants in the distant future by maximizing the reproductive potential of their offspring.

Because theorists who focus on human fertility must contend with biologic mechanisms and psychological forces that drive human sexuality and control of conception or the number of surviving children, attempts are now being made to synthesize work from the fields of evolutionary biology and developmental psychology (Chisholm, 1993).

Do parents, in response to an innate drive to reproduce their own traits, adjust their family-building patterns in response to an assessment of risk or a calculation of the marginal returns on their investment, (Kaplan, 1994)? Is this how cultural differences in the *proximate determinants* of fertility evolve? Hobcraft and Kiernan (1995) wonder whether it is biologic drives that keeps reproduction going at all. Evolutionary biology can portray *regression to two-child norm* as a trade-off between carrying on the lineage or gene pool and the costs of childbearing. We turn to the social sciences for other perspectives.

II. ECONOMIC EXPLANATIONS OF FERTILITY BEHAVIOR

One feature of the environment is the economy, the availability of resources through production and exchange. The demand approach is to consider fertility decisions as being made by rational agents, setting advantages against disadvantages, benefits against cost. An objective, reflecting personal tastes and norms or values of the peer group or society at large, is maximized subject to the resources available, their relative prices and the technology. A rational choice would imply that taking any other decision would have left the parents worse off. Changes in the economic environment, if big enough, could lead to adjustment of family formation.

The *utility function* sets out the valuation the decision-maker puts on various sources of satisfaction. The objectives are essentially the fruits of reproduction, the fruits of production and the enjoyment of leisure. In *new home economics* models (as those originating in the Chicago-Columbia School are also known), the satisfaction parents derive from children depends on both their number and their quality. This key construct denotes health and skills, which will depend on how much parental time and other resources are invested in them. Parents have other sources of satisfaction in their own consumption of other goods. If they are altruistic toward their children they will also draw utility indirectly from their children's (and future descendants') standard of living. Parents therefore gain from an increase in either their own consumption, or *child services*, if there is an increase in their resources (*budget*) meaning they do not have to sacrifice any of the other. Whether they will choose to trade one for the other when the budget is fixed, depends on whether their relative valuation of the outcome exceeds the cost of achieving it. The valuation reflects individual (or socially conditioned) preferences. The question of relative costs is central to the analysis, as being the most likely front on which the economy may send signals to formers of families, who may react by engaging in *substitution*, a reaction to changes in relative prices.

The level of resources, the *budget*, is given by the couple's endowment of physical and human capital. The latter represents their earning potential that can be accumulated through past investments in their health and skills. The greater the resources at the couple's disposal, the larger the range of possible combinations of consumption and children they may choose. Increases in the budget lead to *income* effects in contrast to *substitution* or *price* effects. The range of opportunities may also be widened by improvements in technology, the rate at which resources may be transformed into either of the desired outputs, or indeed in the safety, reliability, and cheapness of fertility control. The range of outcomes chosen may be constrained by peer pressure or social prescriptions. Hence the contrast between neoclassical economists' interests in what choices people make with what Duesenbery (1960) described as sociologists' interests in why people have little choice to make. Because we are interested in apparently deliberate behavior in fertility control, we elaborate the economic framework a little further.

1. Benefits

Before examining what costs of childrearing might consist of, let us first list ways in which parents may benefit from having children:

Consumption-like benefits: the intrinsic value of the child to the parents, for its own sake;
Income earning: the value of the child's labor to the parents;
Security: especially in old age and in times of stress;
Perpetuation of the lineage;
Net cash benefits for children from the state

Blake (1972) adds to the customary economist's catalogue of utilities an item recognizing the social roles of parenthood. It confers status within the kinship system and provides channels for mutual emotional support between generations. Every sort of benefit varies according to the number of children already in the family, and across social settings. It may vary between males and females and be subject to numerous sources of uncertainty. Benefits may or may not take into account the welfare of the next generation. They need to be evaluated as a present value, giving a smaller weight to benefits in the distant future than to those nearer the present. This weight (*discount factor*) represents the couples' rate of time preference and any allowance they make to write down risky outcomes. A low discount rate and altruism toward future generations (which tend to go together) would increase the *present value* of any given stream of benefits expected, and tend to increase fertility. Conversely, a reduction in altruism (*intergenerational solidarity*), all else equal, would lead to lower fertility. This could be a special case of the sociological notion of *individualization* (Lesthaeghe and Surkyn 1988).

2. Costs

Among the costs of an additional child are as follows:

Physiologic wear and tear of pregnancy, childbirth, and lactation (maternal depletion), risks to the mother, and costs of obstetric care;
Value of goods and services spent on the child, and invested in *quality*;
Value of market time devoted to child-rearing: earnings forgone, usually by the mother;
Value of time diverted from nonmarket activities, such a leisure, sleep, care of existing children;
Any planned bequest (transfers of nonhuman capital; e.g., dowries)

Once again the importance of each item depends on the context. Uncertainties affect how much any one birth *project* might cost. A lower risk of infant and child mortality can raise the rate of return on investments in children considerably. Improved child survival is often thought to reduce fertility, by cutting the number of births needed to produce a given surviving family size. But the *demand approach* to fertility points out that the rising rate of return could induce increased *investment*. This could involve a higher number of births but the investment might also be in the quality of existing children (Schultz 1981). It is sometimes also assumed that *rational* parents will assess the child's own risk or future potential (a child's endowment) as they make decisions to invest in or curtail the care they give—their investment in child quality.

Even given the context, the cost of any particular child is not fixed, different levels and timing of investment will produce different levels of *quality*, and different children will need different *investments* to produce the same result. There is enormous variation in the world as to how much is invested in children, and also enormous variation in the amount of investment that is paid for by the family and the state, depending for example on how medical and educational services are financed. For the moment, we assume that the private costs of children (i.e., those not born by the state) are borne by the family. Distribution of costs between mothers and fathers is considered in subsequent sections. Note that other family members, particularly siblings, may bear some of the cost of an additional child.

The item within costs referring to the diversion of the mother's productive labor into reproduction has received massive attention. Childrearing is assumed to displace time from other work, and it is the woman whose time is assumed to be interchangeable between production and childrearing, rather than the man's, because that is how parents generally divide up the work. Almost universally, the woman is the primary care giver, taking responsibility for childrearing and domestic work as well as her biologic role of carrying and nurturing the newborn. The father, assuming he is around, takes the primary breadwinning role in the economy beyond the family. The traditional division of labor leads to gains from specialization, realized by both parties if resources are pooled. It is normal, natural, and, according to Becker (1991) and Cigno (1991), an efficient way of maximizing gains from specialization without running into coordination and enforcement problems that might beset a larger commune. However, the domestic division of labor between men and women is often reinforced by wage discrimination against women in the labor market, which will tend to discourage women's participation in the paid economy, and might not be economically efficient.

Whatever the merits or shortcomings of this interpretation of the sexual division of labor, wherever it exists, it implies a trade-off between women's paid work and reproduction, which had come to play a central role in economists' attempts to model fertility trends in developed countries. Because the family is assumed to be choosing both childbearing and the woman's labor force participation at the same time, some further factor determining them both must be identified if the *chicken and egg* problem of reciprocal causation is to be disentangled. Easterlin (1980) suggests that in the (then) contemporary United States the primary factor would be the generation's relative income and dismisses the *liberation* of the *new woman*, as a major explanatory force. This contrasts with the view of Davis (1984), but Easterlin's dismissal of the *importance of women's social and economic status as an engine of demographic change in the U.S.* is echoed by Joshi (1985,1990) also pointing to the very modest nature of women's advance into the labor market. For new home economics, the independent factor is the woman's wage if she's a potential labor force participant. It is the relevant price of her time, reflecting income foregone for every hour devoted to unpaid maternal duties. The higher the woman's wage, for example because of her education, the lower would fertility be predicted.

3. Preferences and Values

The set of *preferences*, which evaluates the objectives in the demand model can almost interchangeably be described as *tastes, desires, wants, needs, aspirations, goals, values, interests*, and *ideologies* (Pollak and Watkins, 1993, p. 481). One might add *norms*. The orthodox approach prefers to treat preferences as given outside the system (Stigler and Becker, 1977), as not being the source of change into which economists should inquire or hope to make predictions and that can either be safely ignored or left to other social sciences. Economists have therefore generally seen little point in subtle differentiation between influences on motivation.

The Pennsylvania School of economic demography is an exception. They declare a need for interdisciplinary dialogue to understand how preferences are formed and may themselves adapt to changes in the economic environment (endogenous preferences). They point out that the Chicago School's definition of the maximand as a set of unmeasurable commodities generated by household production would enable them to describe the results of a change in tastes as if it had been a change in unobservable household technology. Although the two interpretations are observationally equivalent, the conclusions to be drawn about whether a change is for the better are different. If preferences do not change and a technical change increases household production possibilities, the family must be at least no worse off. If household production changes because preferences change, the utility of the two situations is not comparable (Pollak and Watkins, 1993).

What can be studied about changing preferences are social mechanisms of diffusing new values or norms, and the role that economic change may play in setting the goals to which people aspire. The most well known hypothesis of an economic influence on agents' aspirations (endogenous preferences) is Easterlin's *relative income hypothesis*, which states that young people planning to start family life will attempt to achieve or improve upon the standard of living their own parents managed to set for them. This is also an example of a theory formulated at a macro rather than a micro level.

4. Gender Relations

So far the exposition has gone along with the assumption of *home sweet home*, the couple and the family acting in concert. This assumption deserves scrutiny. It would be reasonable if we knew that some mechanism was at work regulating the standards of private behavior, ensuring that family members treated one another fairly, if not lovingly, so that each party shared the *gains of marriage* for example. Religions have generally taken the lead at promulgating *family values*, and so, to a lesser extent have secular value systems. It was the early Catholic Church, for example, that insisted that each spouse should consent to a marriage. The state, in the form of social workers and family law, sets sanctions for those who break the rules, although it does not normally venture into private territory and it has changed the rule to allow more individual freedom. State regulation of the responsibilities of marriage partners and of parenthood is seen by Cigno (1991) as essential to ensure that contracts are enforced. If the sexual division of labor keeps a woman away from the labor market, she loses, forever, the opportunity to accumulate marketable skills, and is vulnerable should the marriage end without proper compensation.

As Folbre (1994) points out, parents do not necessarily share costs and benefits equally, particularly when the father parts company with the mother. If the father deserts without contributing to the child's maintenance, the mother is left with double demands on her time and fewer resources to invest in the child's *quality*. The absent father probably loses out on the joys of parenthood, but if he stays around he may get more than his fair share of the benefits. If he has patriarchal control over the family, he is likely to contribute less, at least of time and effort to domestic activities and be able to get the children, as well as their mother, to do his bidding.

Situations of marked difference between the economic and social status of men and women are likely to affect parents' perceptions of the value of a child's contribution—sons and daughters will be valued differently. Sons are also more valuable, provided parents have access to their earnings, wherever the earning opportunities for females are inferior to those for males. Daughters have an alternative value in principle as auxiliary household labor, but where marriage takes young women away to live with the bridegroom's family, it is sons' wives who make this contribution, and any dowry to be paid for daughters would be an example of a *bequest* adding to her cost.

Female participation in the labor force is often confused with the *status of women*, although they are not at all the same thing. Employment does not automatically entail autonomy, although it may help. Women's wages may be a better approximation, as they show upon what terms women had paid jobs and, unless someone has control over their income, the purchasing power that it permits them.

Blank (1993) points out that individuals are differently empowered to take decisions and make choices affecting their lives, for social, psychological, and economic reasons. Relatively low empowerment of women relative to men is a characteristic of gender relations in many societies, but there are differential levels of autonomy across different groups. As the distribution of empowerment changes, one should expect shifts in decisions about the formation of families.

5. Collective Interests

The distribution of costs and benefits depends not only on the division of labor between the sexes, but also on the family organization and availability of other *care givers* and the societal provision for bearing some of the costs of reproduction (Lloyd, 1994). The costs of fertility may be partly shared by the wider community, for example, in the form of free or subsidized health and education services or tax allowances and cash benefits for children in a modern welfare state. Similarly the collectivity may have an interest beyond the individual in the size of coming generations as contributors to the pension scheme or as home-grown members of the labor force or military recruits. As such, they bring benefits to other citizens besides their parents, particularly those with no children of their own, and particularly where immigrants would not be satisfactory substitutes. In countries where fertility is considered too high, this could be because the public sector is bearing more of the costs than the private decision-makers (e.g., in the Caribbean departments of France) and/or because people beyond the family have to bear adverse consequences of congestion, competition for scarce jobs or resources. Such spillover consequences for others than the decision-maker are technically known as *externalities* and constitute a reason why one might want public authorities to intervene in private decision making.

The effectiveness of public policy to change population trends is discussed in subsequent sections of this chapter.

III. A SOCIOLOGICAL VIEW

A sociologic perspective considers the context of social organization, cultural norms and the sanctions by which societies have attempted to regulate reproductive behavior. Culture, religion, and ideational change mediate the economic influences on fertility, and offer normative constraints on behavior.

Cultural sanctions on reproductive behavior are the meanings organized and shared by groups that emphasize criteria of right and wrong (Kreager, quoted in Kertzer, 1995). Ideology, custom and culture also reinforce the authority structure within the family—between parents and children and between men and women. The importance given in most societies to the regulation of sexuality and marriage and to family and kinship systems, which determine and are the result of reproductive behavior, is evidence that fertility processes are a central concern of cultural systems and social organization (Kertzer, 1995).

Fertility behavior is influenced by both sacred and secular values. The first statement of *demographic transition theory* by Notestein (1945) suggests that religious values and the moral codes predominant among pretransition societies are important in keeping fertility high. With the movement toward a secular society, brought about as by *modernization*, the taboos surrounding deliberate control of fertility break down, bringing on the transition to low fertility.

In this view, all aspects of life are integrated in traditional societies; family life, economics, and politics are all governed by religious norms. With the coming of modern industrial forms of economic organization, these aspects of social organization become differentiated, and individuals are freed from the all-encompassing constraints of the dominant value system. They gain more freedom to choose different values and beliefs, particularly those related to family life. The process of modernization and the development of capitalism in Europe were believed to be accompanied by a move toward secular individualism. Traditional religious ideas gave way to self-expression, in turn affecting the individual expression of sexuality and family values. Parenthood was no longer a dominant moral obligation dictated by religious or social values (Simons, 1986). Lesthaeghe (1983) has suggested that this emergence of individual autonomy in European cultures led to behavioral patterns oriented toward the welfare of nuclear households rather than toward larger kinship groups. As individuals perceived themselves to be more self-reliant, changes in family formation, dissolution, and reproduction took place. These changes, he suggests, involved an increased orientation toward child welfare, and this pattern contributed to reinforcing greater independence from the family itself, as economic production shifted from familial to nonfamilial modes.

The sociological view of fertility focuses attention on how particular forms of cultural and social organization lead to norms and values that influence when and how childbearing takes place. One well-known example of such social norms is those found in parts of West Africa, which regulate when sexual relations may begin following a birth. Such culturally prescribed practices may have arisen to protect the new infant from the competition of a sibling whose birth follows too closely and interrupts breast-feeding, which can lead to malnutrition and ill health or death. Although traditional values dictate abstinence until the child is weaned or walking, these norms may break down with the influx of new ideas or values, such as those accompanying the spread of Islam or the increasing influence of the West. Islam proscribed sexual relations for a 40-day period, much shorter than the traditional norms. With the advent of modern technologies, however, women are able to maintain the prescribed spacing of their births, while resuming sexual relations with their husbands, who may use the new norms to justify their own pressure to resume relations (Bledsoe *et al.*, 1993).

Religious values may be upheld only among small groups, who are able to maintain fertility norms quite distinctive from the society in which they function. One such group is the Hutterites, an Anabaptist religious sect living on the American Great Plains, whose religious principles prohibit contraception and lead to a completed family size of almost 10, a much higher level of fertility than of the surrounding society. Secular society will tolerate nonconformist fertility behavior as long as it remains the choice of a minority and does not threaten to destroy the legitimacy of the prevailing norms and values (Simons, 1982).

Chamie (1981) studied religious differentials in fertility among Lebanese women, and found that non-Catholic Christians had higher fertility than Catholics and both groups had lower fertility than Muslims. However, he also found substantial differences among the different Muslim groups, fertility highest among Shia women, slightly lower among Sunni Muslim women, and among the Druze somewhere between that of Catholics and non-Catholics. It seems it is not religious norms and values, *per se*, that lead to these differences, but other norms shared by different religious groups. Such would fit with Folbre's view that group membership helps to determine reproductive behavior. In some cases these pro-natal norms are strengthened by the minority status of the group, feeling the need to keep up their numbers in a somewhat hostile environment. Such an explanation has been offered for the relatively high fertility of Muslims in India and Walloons in Belgium.

There is also heterogeneity in the fertility of Israeli Jews, dependent not on ethnicity, but on the *religiosity* of certain groups (Friedlander and Feldmann, 1993). In Israel, a modern developed society by most standards, the fertility rate of around 2.8 births per woman is still well above most European countries. This national average disguises large differences in fertility between religious groups. The total fertility rate (TFR) is 4.5 among all religious Jews in Israel, while ultra-Orthodox Jews have very high fertility, a TFR of 7.5. This group of highly religious Jews shuns the values of the outside world, while sharing group norms that lead to early marriage and very high fertility. However, their high fertility does not entail extra cost to parents because family arrangements are such that younger children are looked after while women work and men spend their time in religious study.

Although rejecting the values of the outside world, this pattern of childbearing and rearing is supported by the wider society. The success of these pronatalist religious values is due in part to arrangements in the wider society, which provides economic supplements such as child allowances and housing assistance, and contributions from external religious groups, their own institutions and the government. In this case, it appears that the pressure on individuals to conform to religious sanctions and fertility norms is powerful, and as the size of this community increases, their family-building pattern could result in increased political power.

The historic evidence from the European Fertility Project (Coale and Watkins, 1986) exposed the shortcomings of Notestein's classic theory of demographic transition. This relied heavily on *modernization* and changes in the mode of production to explain fertility change. This project suggested that fertility was most strongly related to cultural boundaries, defined by language, ethnicity, or geography. Cleland and Wilson (1987) synthesized this historic evidence and contemporary findings of the World Fertility Survey program, conducted in less-developed countries around 1975. Their findings counter the economic view of fertility behavior as the product of rational decision-making. They present evidence that fertility decline was related less to socioeconomic factors than to *social variables* and that the diffusion of new ideas about the acceptability of modern birth control was its main determinant. The formation of the nation state in the process of modernization creates another mechanism for direct and indirect influence on reproduction, at macro- and microlevels (see Casterline, 1989 and subsequent sections).

Many attempts to explain demographic processes, including fertility behavior, have been post-hoc explanations for observed changes, mostly driven by the types of data conventionally collected in surveys (McNicoll, 1980). Lacking a theoretical perspective, these explanations have yielded a puzzling array of possible factors and configurations that may lead to reproductive change. McNicoll elaborates what has been called an *institutional theory* of fertility change. Unlike *ideational* theories, this allows some role for the vastly differing contexts, both historic and current, in which communication about fertility takes place. He argues that institutions rooted in the history of each society give rise to local patterns of social organization: the family and local community; law, and the local dimension of public administration; stratification and opportunities for mobility; the labour market; and state-individual relations. These institutional *endowments*, as well as the transactions individuals make in adapting their *hopes and expectations about the future* to the institutional constraints, together trace the path a particular society takes to fertility decline (McNicoll, 1994). Individuals are not merely passive receptors of social norms, but actively try to re-negotiate the institutional constraints in their own interest. Such a theory attempts, as Greenhalgh (1995) puts it, to *situate fertility* in its cultural, political, and economic context. Such an approach provides the possibility of testing hypotheses about how particular aspects of social organization translate into microlevel reproductive behavior, providing a role for individual action in shaping reproductive outcomes. Folbre (1994) would add collective action by interest groups to the script. Writers concerned with the influence of social values on fertility include Lesthaeghe and Surkyn (1988), who cite the demonstration by Simons (1982) that the baby boom and bust of 1945 to 1985 in England and Wales was foreshadowed by an indicator of religious adherence (Easter Day Communicants). Preston (1986) also believes that value changes play an important role in fertility trends, although he feels they can ultimately be traced back to an economic rationality.

Even in societies that have not yet undergone a fertility decline, people are not merely passive acceptors of normative constraints. Recent evidence from West

Africa suggests that individuals *tinker* with social norms regarding the spacing of births (Bledsoe *et al*., 1993). However, in West Africa cultural norms do not include the idea of *stopping* after a preferred family size has been reached. This suggests that although the means of fertility regulation are well known, it is changes in ideas regarding the acceptability or desirability of family limitation within marriage that leads to the *stopping behavior* characteristic of significant fertility declines.

The *social construction of the family* differs across cultures and may explain why fertility has not declined in some regions, even those where development is far advanced. Cleland and Wilson found that sub-Saharan African and Middle Eastern societies differ from other regions in fertility preferences. Socioeconomic differences in fertility preferences in these regions are more pronounced than divergences in fertility itself.

The social organization of the household—itself an outcome of *cultural* forces that lead to different configurations of household structure—influences marriage timing and the ability of the household to support dependents. Because these seem bound up with fertility differences, a key question, then, becomes, "What determines family systems?" In much of sub-Saharan Africa, the boundary of the family extends to a large kin or lineage group, which makes it more likely that the larger group to which individuals owe allegiance will retain much control over fertility behavior. Practices such as child fosterage, arrangements where the extended family provides child care, and state provision for education, "break[s] down household economic boundaries and spread[s] the impact of additional children on family resources across a wider kin network" (Lloyd, 1994, p. 11–12). Such institutional supports to high fertility may prevent the spread of the *small family norm* in these regions.

The *family mode of organization* framework hypothesizes that the extent to which the family organizes individual members' activities plays a role in fertility change (Axinn *et al*., 1992). This can occur in two ways. Fertility change may be caused by a changing demand for children as a result of desiring to maximize the outcome of the new, nonfamily organization of activities (structural economic change) or may be due to ideational change, a result of the diffusion of new ideas and fertility goals through alteration of the group (nonfamily) with whom family members have contact. In this way, this approach attempts to connect macrolevel social change to individual-level fertility decisions and behavior. This seems a promising way to approach a test of the rival economic and sociologic theories of fertility change, but there is still scant evidence on which to judge these hypotheses.

IV. QUESTIONS ABOUT FERTILITY AND SOCIETY

This section looks at some of the answers, and controversies, that social scientists have offered to certain questions. Answers are summarized at the end of each section.

1. Why Does Fertility Decline with Development?

One of the greatest paradoxes facing the *demand approach* to fertility is to reconcile the historic (at least) association of fertility decline with economic development. In the economist's view, the latter increases the resources available to potential parents. If their decision making resembles that of a consumer demanding a *normal good* (i.e., one for which demand increases as income increases), one might expect economic growth to raise the demand for progeny.

One interpretation is that positive income effects are outweighed by increasing costs of parenthood as economies develop. There are higher standards of consumption to be provided, fewer opportunities for child labor, and increasing investment in education and training required by an increasingly sophisticated economy.

Just as economic growth involves the accumulation of physical capital, it also requires increasing standards of skill and sophistication among the workers. New home economics appeals to its distinction between child quality and child quantity in this context. Economic development offers improved opportunities for investing in child quality (better survival, better schools, better employment prospects). This encourages parents to invest more resources per child than they would have done at a lower level of income and, given the right assumptions about relative income elasticities, to substitute quality for quantity. They cut family size to raise the quality of life for each child in the smaller family. This resembles propositions reached independently by evolutionary biologists. Although the detailed apparatus of this theory is difficult to prove or disprove, it is consistent with much historic experience. Fertility decline is associated with improvement of child health and the spread of schooling (investment in human resources) and with accounts of the prospective cost of education as being a motive for controlling fertility (Crook, 1997).

Other ways in which economic development may reduce the *demand for children* could be pressure of space in urbanization; competing employment opportunities for women, as the home and the workplace become separate; and the reduced risks in the economic environment against which a large family may be viewed as insurance. Women's employment and insurance are discussed in subsequent paragraphs.

The classic demographic transition theory was based on the idea that differences in fertility are rooted in economic structures. Traditional peasant economies are organized along familial lines, whereas capitalist economies are not. With increasing industrialization comes disintegration in the familial mode of production and the pronatalist structures it engenders. As industrialization spreads, economic *modernization* is accompanied by the diffusion of the small nuclear family norm. As we have seen, the findings of the Princeton European Fertility Project did not support the classic view of fertility transition. Change spread to all social groups too quickly to reflect real changes in structural economic conditions. A threshold was apparent: when fertility declined by 10%, it tended to proceed quickly to low levels (Watkins, 1986). The timing of fertility change seemed to follow cultural boundaries and to be associated with indicators of social development, such as education or literacy, rather than with economic development (Cleland and Wilson, 1987; Brass and Kabir, 1982), and its pace was closely associated with a shift from fundamentalist to secular world views (Lesthaeghe and Wilson, 1986).

Cleland and Wilson (1987) argue that the fundamental force behind fertility decline is not the changing balance of costs and benefits of childbearing, but a change in ideas and norms, acting independently of the level of economic development. They suggest that "the fundamental forces of change operate at the societal level" (p. 24), not at the microlevel of individual decision making. Differences between high- and low-fertility groups in the World Fertility Survey are not to be found in their reported preferences for children, but in couples' willingness or ability to regulate their fertility in accord with these preferences.

Fertility transitions, both historic and modern, are predominantly driven by *stopping* behavior. After a specified number of children, couples adopt birth control or otherwise limit the size of their families. One suggestion is that *modernization* (by this we mean the shift in economies driven by technological advances that result in increased output) brings about a spread of new ideas, a change in mentality, from a belief that family size is a result of chance to a clear notion of what family size *ought to be*, in van de Walle's phrase *numeracy about family size*. Means of regulating fertility are known in most societies, but "any contraceptive technique is probably most effective after the couple has reached the number of children they do not want to exceed" (van de Walle, 1992, p. 501).

How these new ideas arise and why new norms develop are still unanswered questions. Cleland and Wilson (1987) argue that the growth in human *mastery over nature* and widespread mortality decline has facilitated the spread of fertility control by *undermining the reproductive imperative*. Once the latent desire for lower fertility appears, it is the spread of knowledge regarding effective means of birth control that eliminates excess or *unwanted* births.

Development is likely to facilitate the spread of knowledge. Development is also likely to induce the latent desire for lower fertility, particularly if it is of a broad-based, rather than *trickle-down* nature. But economic development cannot be claimed as a necessary and sufficient condition for fertility decline as implied in the *Best Contraceptive* slogan.

Finally, the explanations of the link from development to fertility decline are multiple and not mutually exclusive: reduced benefits; increased costs; demand for quality; improved contraceptive technology or delivery; changed perceptions of what behavior is advantageous, appropriate, or even thinkable. These could act singly or jointly. What operates in any particular situation does not necessarily apply in others. Just as economic development cannot be simply replicated from one country to another, the course of demographic development can follow diverse paths, reflecting local features of both social and economic structures, and drawing on the technology and experience of those who have gone before.

2. Can Fertility Decline in Poor Countries?

A closely related debate is whether fertility decline is possible without economic development. The view that it is not, or at least not likely, goes back at least as far as the Bucharest Population Conference of 1974, when the slogan was coined, "Development is the Best Contraceptive." This echoes classic demographic transition theory that social and economic development is its prime cause. Development reduces dependence on children or raises the cost of rearing them. Thomas (1991) restated this slogan. He argues that programs offering the means of fertility control will find more willing takers if accompanied by measures to reduce the risks or provide alternative sources of insurance to poor parents.

In poor societies, the need for child labor, either for wages or in family enterprise, is one source of the demand for children. Insurance against risk in old age or in times of crisis, which children can provide, is another (Cain, 1982). One source of security, available to some, is land holding, but any impact this may have on the need for a large family through greater security may be offset by a demand for child labour. (For further discussion of old age security see subsequent sections.)

In the joint-family systems of much of Asia, old-age support is a family concern (Cain and McNicoll, 1987). Sons bear the primary responsibility for care of elderly parents, leading to a demand for sons, and a good probability that at least one will survive to support his parents in old age. Considerations of security, therefore, can outweigh the cost-benefit accounting, which might precede the decision not to have another child (Cassen, 1978).

Furthermore, the high mortality levels usually prevalent in the poorest societies may lead to more births in order to *replace* children who die to ensure that at least some children, especially sons, survive to adulthood. In poor societies, where education levels are low, the costs of raising children may be exceeded by the benefits those children will bring when they reach the age at which they are net contributors to the family income. For a number of reasons, therefore, when parental income is low, the emphasis is on child *quantity* rather than *quality*.

Nevertheless, traditions of filial support may be compatible with low fertility (Cleland, 1993; Weinstein *et al*., 1990). Examining fertility decline in Bangladesh, a range of factors thought to reduce demand for children were reviewed (Cleland *et al*., 1994). No significant changes in school enrollment, the value of children's labour, social security benefits from the state, female participation in the labor force, or improvements in the educational or social status of women were found. Conversely, some argue that economic reversals can induce fertility decline. For example, increasing landlessness could drive families from agriculture to the service and sales sector, where fewer children are needed to supply labor and costs of raising them increase. The Bangladesh study detected no significant deterioration in the economy to an extent that might decrease demand for children. Despite the evidence that factors affecting demand for children had not changed, fertility began to fall in the late 1970s, accelerating in the mid 1980s, reaching the level of about 4.5 children per woman by 1990. The authors suggest that the long-term decline in child mortality in Bangladesh led to a greater willingness to adopt innovative behavior. All that was necessary was access to the means of fertility control, provided by a strong government family planning program, which effectively removed barriers to use of contraceptives by delivering advice and supplies at the doorstep. Despite the evidence that demand for family planning was low before the initiation of the national program, its comprehensiveness appears to have maximized the spread of new behavior and sustained it.

These conclusions may be compared with those of Dow *et al*. (1994) for Kenya. Here too, the resurgence of a family planning program accompanies fertility decline, but a deterioration of economic conditions is judged to have triggered the motivation to seek its services. However, identifiable transfers (that is, support) to the elder generation did not appear to have changed.

In yet another context, Martine (1996) examined the fertility decline over a period of intense social change and deep recession in Brazil. Total fertility fell dramatically and very quickly from an average of six births per woman in the early 1960s to 2.5 births per woman in the mid 1990s. This was achieved primarily through high rates of abortion and sterilization. Martine traces increased motivation for fertility regulation to institutional changes in health and social security. The expansion of health services served to legitimize the use of birth control, even in the face of strong church opposition, and increased the level of knowledge about, and access to, effective means of regulating fertility. Government-sponsored attempts to *modernize* also accelerated rural–urban migration, and increased interest in consumer goods, which in turn increased the costs of childrearing. Mass media are also given credit for promoting consumption and eroding traditional ideas about male authority. Coverage of the social security system was expanded and a guaranteed income in old age established. This shifted responsibility for the elderly from the family to the state, and despite deliberate pronatalist elements of the social security system, such as child allowances, the impact of the system was in fact antinatalist. Changes in government policies played an unintended role in Brazil's fertility decline.

Martine (1996) places less weight than others on the argument that Brazil's economic difficulties lessened demand for children (Carvalho *et al*., 1981). Rather, his evidence suggests that a constellation of factors—structural and ideational—are at the heart of the sustained transition. The aggregate analysis is persuasive, but does not directly connect these changes in the larger society to mechanisms at the individual level. Can diffusion of new ideas be linked directly to changes in individual behavior?

To put it briefly, fertility has declined in poor countries and poor communities. This is not necessarily as a result of poverty. Ideational and structural factors short circuit any iron link between development and fertility decline.

3. Women's Education: Materialist or Ideational Influence?

One oft-cited benefit of development is the increased access to educational opportunities that may accompany changes in the demands of the labour market. How increased access to education affects the status of women in general, and their fertility, is a major concern of policy makers today.

It is well known that, for a number of reasons, educated women are more successful than their uneducated *sisters* in bearing and rearing healthy children. They are thus less likely to bear more children than their family size goals (see Cochrane, 1979 and Cleland, 1986, for evidence on maternal education and fertility; Cleland and van Ginneken, 1988 for a review of the evidence regarding child survival). The strong and nearly universal link between female education and fertility decline has provided further evidence on how and why reproductive change occurs. It suggests that the effects of education are both material and ideational in nature.

Education can indirectly influence fertility through several pathways: by affecting supply factors, the demand for children, and the *costs* of control. The mechanisms through which it works are unclear, partly because in some circumstances very small amounts of school attendance—just a few years—are sufficient to observe an effect. Given the unequal and very uncertain quality of education widely available in many societies, this effect can hardly be due to actual knowledge imparted and makes it unlikely that strikingly different attitudes toward the world are imparted and absorbed by girls in so short a time. Some have proposed that besides exerting a powerful independent effect on fertility, education (or *schooling*) may act to change reproductive behavior by increasing a woman's autonomy (her ability or freedom to make decisions about her own fertility) (see Basu, 1996, for a detailed exposition of these issues). Moreover, the relationship may partly reflect *selectivity* especially where the general education level is low. That is, educated women are more likely to come from families of higher status or wealth, who want to educate their daughters and also may impart an orientation to lower fertility. This confounds the relationship between fertility and education (Sathar, 1996). We now look at the evidence for all of these relationships.

Jeejebhoy (1995) provides us with an extensive review of evidence on the matter from demographic surveys. Education can influence the *supply* factors by breaking down traditional norms about breast-feeding and postpartum abstinence. Even a little education shortens the duration of breast-feeding. Breast-feeding may be incompatible with the employment educated women seek. They are also more likely to be aware of and able to afford alternatives to breast milk. As education increases, postpartum abstinence decreases, and this is thought to be due to a greater intimacy between husbands and educated wives. These education effects raise fertility, unless countered by effective birth control.

On the other hand, education raises the age at which a woman marries (Appleton, 1996), reducing the length of her reproductive life. The reasons are many: Educated women have a greater role in decisions about whom and when they marry; they are likely to have greater control over resources due to employment before marriage; and, in some settings, they are less marriageable than uneducated women (Jeejebhoy, 1995). However, postponement of marriage may simply be followed by couples *catching up* to those marrying earlier by increasing the pace of subsequent childbearing (Sathar, 1996).

Women's education, at least beyond a few years of schooling, can also reduce the number of desired children. Economic explanations emphasize the increased costs of an educated mother's time and increased *rationality* of decisions about family size and desired sex composition. When the education of women becomes widespread, the value of their labor increases. They become more active in the labor force, and more economically independent. This in turn can reduce the felt need for male children, either as labor or as security in old age (Cochrane, 1979).

Women's education can also reduce the obstacles to contraception, lessening the psychosocial *costs*. There is a direct relationship between education and the practice of contraception, due in part, perhaps, to more frequent interactions with the outside world (Basu, 1992). Thus, educated women are also more likely to know about birth control and to have better access to effective methods because they have more experience in dealing with the world beyond the household. In this sense, education may work by facilitating the spread of knowledge about reproduction.

Another mechanism proposed is that education increases a woman's status and autonomy, giving her more control over her own fertility and affecting her motivation to bear children. There is a good deal of evidence that a higher threshold level of education must be reached in highly stratified societies than in egalitarian ones before fertility starts to decline (Jejeebhoy, 1995). This may be especially true where gender stratification is strong, because it limits the domains over which educated women can make decisions and determines the amount of education necessary before they are taken independently. The social context also conditions the relationship between education and a woman's motives to bear children. Where her opportunities are greater, she may gain status in other ways than by bearing sons, and may need to depend on them less for security in old age. In gender-stratified societies, more education is necessary before women are confident of their own economic and social status to revise their fertility goals (Jeejebhoy, 1995). It may be that the education–fertility link is strongest where education actually can be translated into employment.

Education may also help to increase a woman's autonomy over her own body, including her right to refuse sex (Orobuyloye, 1995). It may also lead to increased husband–wife communication regarding fertility decisions and use of contraception. Evidence from Africa suggests that educated women are more likely to be in agreement with their spouse's fertility preferences than uneducated women (Bankole and Olalaye, 1995), but this does not always result in lower fertility. In Kenya and Ghana, for example, men and women share a common goal of high fertility.

Educated women are more likely to marry men who are also educated, and although women's education is a consistently stronger predictor of fertility than their husband's education, his role in shaping fertility behavior is clearly pertinent. Husbands' roles in family-building strategies and decisions are important, but the meager evidence available does not reveal any consistent associations between husband and wife preferences and other factors (Mason and Taj, 1987). Basu points out that both men's and women's fertility preferences are likely to be affected by the position of women in society, so that for example, if women are denied a place in the labor market, both husband and wife should want more sons. This may extend to the other actors in the household who influence family-building strategies, such as mothers-in-law (Basu, 1992). Others have suggested that a husband's non-family experiences (such as changes in residence, and work and leisure activities outside the family) have a positive effect on contraceptive use (Axinn, 1992). Opportunities for work and residence outside the family may depend to some extent on the husband's educational attainment. If research designs exclude men, important mechanisms involved in behavior change may be missed.

Jeejebhoy's extensive review of education-fertility linkages concludes that in societies with strong gender stratification, the most powerful effects of education are to reduce traditional inhibitions on the *supply* factors—to raise fertility. Only in more egalitarian settings do the fertility-inhibiting effects of education—on marriage age, family-size preferences and use of contraception—counterbalance or exceed these. This may explain why the relationship between education and fertility is less clearcut than the relationship between education and contraceptive practice. In some settings contraception may be initiated only after unwanted births have already occurred. Jeejebhoy (1995) argues that this may be the case where women achieve the power to make reproductive decisions rather late in their married lives, when desired family size has already been exceeded, rather than earlier in their life course.

In highly gender-stratified societies, religion and patriarchal kinship structures shape marriage practices and female economic roles and sometimes make them seem impervious to education, or at least resistant to its effects. Cultural systems and social organization that strongly influence reproductive behavior often have their roots in the predominant economic structures and modes of production, at least historically. For example, higher levels of literacy and employment among South Indian women do not explain observed differences in female autonomy—and fertility—between North and South India, which have deeper historic roots (Basu, 1992).

This may be one reason why community context—community-level effects—seems to operate when we look at the relationship between fertility change and education. Generally speaking, high societal levels of education are strongly correlated with low fertility at the aggregate level (Cleland and Jeejebhoy, 1996). However, the overall level of education in a society can exert a powerful effect on its own to influence the fertility behavior of individuals regardless of their educational attainment. Cleland and Jeejebhoy show that, in the Indian states of Kerala and Tamil Nadu, where literacy rates are high, fertility at the individual level is inversely related to years of schooling. Both states have much lower total fertility than the low-literacy states of Uttar Pradesh and Bihar. In the latter, however, there is no consistent relationship between a wife's education and fertility at the individual level. The contextual effect of education exerts a strong influence on fertility behavior that is independent of its individual effect.

A striking example of how contextual factors may change is illustrated in a study examining education and fertility decline in two Chinese provinces (Freedman *et al*., 1988). Early in the shift from moderate to low levels of fertility in both provinces the educational level of the production brigade had an effect on fertility that was independent of the effect of an individual woman's educational attainment. In rural areas, fertility of women at each level of education increased steadily with the proportion of the local population that was illiterate, and where the proportion illiterate was as high as 40%, the effect of individual educational level was inconsistent (as seen in Uttar Pradesh and Bihar).

Following large-scale campaigns and intensive family planning program effort, these local (contextual) differences were leveled, and the overall level of literacy in the area was no longer consistently related to the level of fertility. Moreover, the effect of education at the individual level also disappeared very shortly after inception of the government program. Where the government family planning program was strong, knowledge and supplies quickly became available to everyone in the area, and social differentials in fertility disappeared (Riley and Gardner, 1997). Government policies seem to have the power to diminish educational differences in contraceptive use and fertility.

LeVine *et al*. (1991), tried to investigate what processes at the individual level that lead to lower fertility might be affected by increased education. Their study in two Mexican communities found that mothers with more education used child health and family planning services more frequently than less-educated mothers. Both factors increased child survival and lowered fertility, but some intervening processes were found that provide clues to how increased schooling may translate into changes in fertility.

Using in-depth interviews and observation, they found that in contrast to less-educated women, more schooling led women to aspire to higher status for themselves. They were more likely to report making major household decisions jointly with their husbands. Educated women also spent more time talking to their infants and were more responsive to them than uneducated women. Such behavior has long-term effects on an infant's cognitive development. They expressed higher occupational aspirations for their children, while reportedly expecting less help from their children in old age. These findings confirm that schooling influences the attitudes of individuals, and also points to the effect education appears to have in increasing parental investment in a child's *human capital* or *quality*, directly raising the costs of children by demanding more of a mother's time. Whether new fertility preferences and practices preceded and promoted these educational differences in attitudes and behavior or resulted from them cannot be deduced, but this study is evidence of schooling's wide-ranging effects.

Finally, there are multiple mechanisms by which increased female education affects fertility behavior, and good evidence that its influence is both materialist and ideational in nature.

4. Is Childbearing for Old Age Security?

Caldwell's theory of the fertility transition is an elegant synthesis of material and social forces (Caldwell, 1982). In the process of development, not only does fertility decline but there is also an associated reversal of intergenerational wealth flows. In the high-fertility society, the elders are in control and draw economic support from their adult offspring. This creates a motive for high fertility, or at least an obstacle to the adoption of family limitation. The more surviving children, the more secure will be those parents who survive into old age. Those with few or no surviving children will be endangered themselves. In the course of development, children become a less reliable source of support; the gerontocracy weakens, especially as the younger generation gains skills and becomes more mobile. Even without enforcement problems, alternative sources of security become more attractive: physical assets, purchased insurance, the capital market and related institutions like credit unions or friendly societies, and, eventually, social insurance pensions providing old age security through the state. Caldwell notes that the transition does not just involve the cessation of wealth flows from younger to older generations, but a reversal. Parents invest more resources in the human capital of their offspring, and bequeath to the next generation whatever has not already been used up.

There is conflicting evidence on whether the wealth flows are always in the directions Caldwell posits (e.g., Dow *et al*., 1994). In rich counties there are gross flows in both directions, and mutual dependence, in some cases of the elderly and the middle-aged (Thane, 1989). In some poor settings, fieldwork among primitive groups in South America suggests children do not contribute more than they consume and points to other motivations or drives to reproduce (Kaplan, 1994).

People may still want to call on family members to provide physical care, companionship, or emotional support in old age, and in some cultures it is important to know that a descendant will be on hand to lead one's funeral rites. In India, it has traditionally been important for a son to survive to perform this function, but isolated, although significant, incidents have recently been reported of daughters taking this responsibility (P. Visaria, personal communication).

In more developed societies, with increased life expectancy, the issue of old age dependency gains salience over that of who takes care of the funeral. The establishment of collective institutions for the material support of the elderly in developed countries removes one reason for having children. Cigno (1991) points out that there must be other reasons keeping childbearing going at all. He also points to the irony that the pay-as-you-go pension systems, mostly established when the elderly were relatively less numerous than they are becoming, are facing severe difficulties because of the declining size of the contributing generations. If their existence is deterring childbearing, pension schemes can be said to contain the seeds of their own destruction. One may not need to have children to provide a private source of income in old age, but someone has to rear the next generation of contributors to the pension scheme. This externality has led to the suggestion that pension entitlements should be relinked, positively to the number of offspring produced (Demeny, 1987; Vallin, 1992).

At present, a large class of prolific pensioners are actually penalized in the cash support they can expect. Most state (and all private) pensions are positively related to the earnings. Women who have raised children often forgo at least some of their earnings during their labor force years, which tend to be reflected in lower pension (Joshi and Davies, 1994). Thus the opportunity costs of childbearing are carried on into old age, reinforcing low fertility and further undermining the long-term viability of the pension scheme. In the short term, elder dependents and child dependents appear to be in competition for the resources available. In the longer term, investment in the young is still essential to ensure resources for the old (Qvortrup, 1994).

Developed economies have many institutions other than parenthood to secure income and services for the aged. Less-developed societies have hitherto had relatively few aged people to support and these have not necessarily been dependent on their own offspring. The empirical importance of the old age security motive for childbearing has been inconclusively debated. Old age security is still a reason for reproduction in aging societies, but one that is more likely to be acted on collectively than individually.

5. What Makes Fertility Fluctuate in Rich Countries?

Has the fertility transition finished? Have the developed countries with their high-quality population, their highly developed state and market institutions, and their high levels of birth control technology reached a new low-fertility equilibrium? Will decline continue to be sustained below replacement and beyond or is it the nature of reproductive behavior in advanced societies to fluctuate?

Richard Easterlin's famous hypothesis is that in mature and stable posttransitional society, fertility would have a tendency to fluctuate, sometimes below, sometimes above replacement. As spelled out in his popular book, *Birth and Fortune*, the central mechanism is relative generation size and its influence on aspirations relative to earnings (Easterlin, 1980). Members of small cohorts would reap advantages in education and the labour market; young adults would find life relatively easy, so that their aspirations for living standards would be attained sooner, and permit the forming of larger families than generations who were more numerous relative to their predecessors. A key link in the chain is the assumption that standards are set by those the cohort enjoyed during their upbringing. This *relative income hypothesis* is an example of how preferences are determined within the system (*endogenous*). Easterlin argues that relative generation size determines a number of other series, from wives' labour force participation, to divorce, suicide, and crime. The evidence that relative generation size affects earnings is mixed (Wright, 1991). The evidence that a small generation became the parents of a baby boom is strongest for just one cycle in the United States and perhaps Australia, Canada (Wright, 1989), and France (Leridon, 1978; Chenais, 1983). Ermisch (1988) did find a modest contribution of relative generation size, among other variables, to his analysis of parity-specific birth rates in England and Wales between 1952 and 1985.

The prediction of self-perpetuating swings has not been borne out. By the 1990s, the children of the 1960 to 1970s *baby bust* should have been generating a new boom, but fertility in the United States has only moved from around 1.8 in the 1980s to around replacement (2.1) in 1990 to 1992 (Monnier and Guibert-Lantoine, 1995). Perhaps the conditions of economic stability on which the hypothesis was founded have not held out or perhaps the model never really worked.

A rival model, which also predicts fluctuating fertility in advanced economies, was proposed by Butz and Ward (1979) building on the microeconomic Chicago school model of Willis (1973). Their prediction was, contrary to Easterlin's, that fluctuations in fertility would be *counter-cyclical*: more births in hard times than times of prosperity. Their reasoning was based on the growing proportion of wives in the labour force, for whom increased wages would mean *substitution* away from childbearing. As the proportion of two-earner couples increased, presumed positive *income effects* of men's wages would be outweighed by the negative effect of rises in women's wages. The upswing of the 1950s baby boom would fit with a regime where income effects dominated. The end of the boom would be triggered by sufficient two-earner couples confronting a continued wage growth. The model was fitted to data for 1948 to 1956 and successfully predicted 1957 to 1975, but its fit to the earlier part of the century or to subsequent years is questionable (Olsen, 1994). Given continued rise in women's labour force participation, we should expect continued decline in fertility into the 1990s, but as we have seen, such a trend has reached a plateau (in most other advanced countries except Southern and Eastern Europe and Japan).

One reason for the failure of these models to find confirmation in statistical data may be the annoying habit of history not to stand still long enough for a cyclic process to display any regularity. Another view, perhaps another way of saying the same thing, is that there are other factors to be taken into account. These include the policies adopted by governments (described in subsequent sections), the growing compatibility of motherhood and employment, particularly via the purchase of child care, the growing instability of marriage, the dissociation of fertility from marriage, and a climate of uncertainty and global insecurity (Ermisch, 1990, Olsen, 1994). Most of these factors point to continued low fertility, if not continuing *flight from parenthood*. Hobcraft and Kiernan (1995) see "little prospect of a resurgence of parenthood" without the help of "biological and social pressure to reproduce."

A word is necessary about the role of contraceptive innovation in accounting for fluctuations in fertility in industrial countries. In particular it is very widely believed the introduction of the contraceptive pill brought about the 1960s baby bust, and that periodic scares about the health hazard of the pill bring about blips in the birth rate as users rush away from that type of birth control without substituting another. Murphy (1993) argues that changing attitudes to the contraceptive pill are a major element in short-term fluctuations in British fertility. In the work by De Cooman *et al*. (1987), evidence of *pill scares* was not abundant, but one year did show a slight excess of births, otherwise unexplained. An alert about the health effects of some types of pill in October 1995 was followed by a modest rise in both abortions and maternities in England and Wales (Wood *et al*., 1997).

The fact that there is a range of modern and traditional means of contraception should mean that changes in the price, availability or quality of one product would result in people choosing another method. Some of the boom births were described as *unwanted* in demographic surveys (Westoff and Ryder, 1977). Using demographic surveys for France, Leridon (1985) estimates that 21% of births in 1965 were not desired, and an equal number were not planned to arrive at that particular time.

It is not only the absence of effective methods of birth control that made this situation possible, but also a general context in which the child that was one too many was not altogether so . . . Contraceptive practice may have slackened during the 1960s due to the fact that couples' intentions had become less certain during this period.[1]

Over the years 1965 to 1975, as contraception improved, unwanted fertility was halved, accounting for a drop of 0.32 on the period fertility rate. However, most of the births in both 1965 and 1975 were planned. The fall in planned births accounted for 0.28 and in mistimed births by 0.21. Even the unwanted births could not have been too seriously unwanted as the knowledge, in principle, existed of how to prevent them. Even if the methods were somewhat old-fashioned, inconvenient, and/or unreliable they served to keep fertility low in the 1930s. One therefore needs an explanation of the motivation, as well as the means, to end the baby boom. Certainly, the demand approach would admit that improved means of birth control would tip the balance in favor of its use at the margin, and could theoretically have a large impact if the conventional methods were insuperably distasteful. There is also the argument that the users of birth control became more effective, or careful, as well as the methods themselves.

Reliable birth control, accessible to all women has a number of other consequences, which could in turn help bring about fertility decline. It has permitted widespread sex outside marriage. This increased opportunities to form, and move between, partnerships. It has increased the freedom of both men and women, which has arguably led to the revision of norms about the formation of sexual unions. The easier sexual mores added to other factors weakening the institution of marriage (and ironically led to extramarital births). With partnerships less solid, people become more cautious about committing themselves to parenthood (Ermisch, 1990).

Posttransitional fertility is still relatively recent, even in industrial countries. It may still be too soon to generalize, but it does not seem to be characterized as either a regular cycle or an inexorable downward trend. No universally robust predictor has been found. Those that have been considered include the relative

[1] Ce n'est pas seulement l'absence de méthodes efficaces de regulation des naissances qui a rendu une telle situation possible, mais aussi un contexte général dans lequel l'enfant "en trop" ne l'était pas tout à fait. . . . Il a pu se produire dans les années 60 un relâchement de la practique contraceptive, parce que les intentions des couples étaient devenues plus incertaines à cette époque.

generation size of men, the relative wages of women, the reliability of marriage, the level of religiosity, and the effectiveness of contraception. The latter may be the outcome of choices rather than a determinant.

6. Are State Policies Effective in Accelerating Fertility Decline?

Few issues have raised such heated debate in international circles as the question of whether population growth is so rapid that its effects should be curbed through deliberate antinatalist government policies. The effectiveness of state policies in curbing the rate of growth through promotion of family planning is a major concern of demographers. There is still a good deal of skepticism that policies can be effective in accelerating fertility decline due to the widespread belief that only socioeconomic transformation can succeed in providing couples with the motivation to reduce the number of children they bear. Crook (1977) points out that there was no state encouragement of the fertility decline in England and Wales, in the 19th century, quite the contrary. Yet almost all third world fertility declines involve some degree of state involvement. Mauldin and Berelson (1978) and later Lapham and Mauldin (1984) were among the first to attempt a systematic analysis of the impact of family planning program. These studies pointed to the association between the level of socioeconomic development and a measure of *program effort*, but the major conclusion was that strong programs do reduce fertility, a finding that reassured governments and donor agencies (Lee *et al*., 1995).

The 1994 International Conference on Population and Development, held in Cairo, marked a growing consensus among developed and developing countries, and among women's and environmentalist nongovernmental agencies, that they share an interest in advancing women's access to family planning. National policies and plans of action were agreed in Cairo, which had the objective of "incorporating population concerns in all relevant national development strategies" (United Nations, 1995, p. 67). There remains a need, however, to examine the factors that promote the effectiveness of policy and related public family planning services (Lee *et al*., 1995). The debate still hinges on whether family planning programs can effectively reduce *unwanted* fertility, in the absence of other broader social change.

Pritchett (1994) argues that the challenge is to reduce people's fertility desires, not to address the *unmet need* for family planning. He argues that the question is still whether family size preferences are determined by economic influences or social and cultural forces, concluding that family planning program effort, which provides knowledge and access to the means of control, is not a dominant factor determining differences in fertility. Rather, a low level of demand for children is both necessary and sufficient to reduce actual fertility. While pronatalist policies that reduce access to contraceptives will weaken the effect on fertility of any changes in preferences, the reverse may not work. Improvements in access to contraceptives are neither sufficient nor necessary for fertility reductions to occur, he contends, contrary to the study in Bangladesh. Some *active* family planning programs have gone on the offensive to recruit users using coercion. The Emergency in India in 1975 and some stages of the One Child Policy in China would be cases in point. Economic incentives (and sanctions) are then focused on the *motivators* and officials of the program (Crook, 1997). Curiously the much resented manner of the Emergency Family Planning Drive resulted in the overthrow of the Government of India, but not a sustained family-planning backlash.

New work to assess the demographic consequences of divergent population policies in a series of comparisons between pairs of countries similar in terms of crude socioeconomic indicators has recently been completed (Lee *et al*., 1995). This study's main conclusions were that when *policy elites*, responsible for putting family planning on the policy agenda, did not sustain momentum by building broadly based political coalitions, the family planning agenda was less well formulated and implemented. The study supports the view that government policies and programs can make a major difference, at least to the timing of fertility decline. In three of the four pairs, in countries with stronger programs the fertility decline started earlier; in the remaining pair, the pace of decline was sharper in the country with the stronger program. Moreover, countries with weaker family planning programs displayed more pronounced social differences in contraceptive use. In those with stronger programs, the main function of the program was to accelerate the spread of fertility control from the urban and educated elite to the rural and less educated. National policy-making by elites, and strong institutional and financial support for family planning programs are unlikely to have had direct relevance to fertility desires of individuals in these countries. These conclusions are in direct contrast to Pritchett's.

Both analyses appear influenced by ideology. Neither presents direct individual level evidence that might indicate which changes first, fertility preferences or access to the means of control. Thus, we return to the lament of Coleman and Schofield (1986) that demography suffers from lack of a theory that would integrate the influence of societal factors, such as social policies or economic development, with the individual-level behaviors of which national fertility levels are made up, a theory to integrate macro and micro.

In short, there are findings that government policies can make a difference to fertility, but also suggestions that they may not always be successful.

7. Are State Polices Effective in Preventing Fertility Falling Below Replacement?

Over the past century, there has been a gradual increase in political interest in population and family matters in developed countries. Common developments and diversity within them across countries are fully described by Gauthier (1996a) and in greater detail for Canada, the United States, Australia, and Great Britain by Heitlinger (1993). They include public regulation of abortion and contraceptives, cash benefits to families, policies facilitating the employment of women in general, policies facilitating the combination of paid work and parenthood, provisions for childcare and (although not specifically treated in these books) housing policy and immigration control. Few governments, other than the French, have explicitly stated pronatalist goals, although Nordic countries, especially Sweden, have a tradition of facilitating family life through liberal policy on voluntary parenthood, relatively generous social security, and a deliberate policy of promoting sex equality through recognizing a woman worker's right to parenthood (rather than a mother's right to employment). Since the 1970s the parental insurance scheme for the workforce has established right for fathers as well as mothers to have paid leave from employment in connection with their children, in an attempt to make gender equality and family life more compatible. The package of family friendly policies in Scandinavia also includes widespread public involvement in childcare. Elsewhere, policies affecting the family have been more likely to subsidize a traditional breadwinner–housewife division of labor between parents. The history of the development of this complex of policies suggests that they are in part the outcome of demographic change (Gauthier, 1996a). Their effectiveness at changing demographic trends is more debatable.

Gauthier (1996b) reviews evidence of effects on fertility (and family structure) of the impact of fiscal incentives on fertility. The evidence is for nothing more than a very modest effect (which was not altered when she included maternity provisions and child care provision in a study of twenty-two industrialized countries). The study by Ekert (1986) of eight European Community countries suggested that cash benefits of the order offered in France raised total fertility by 0.2 children. A similar result was confirmed by Blanchet and Ekert-Jaffé (1994) using data on 11 European countries, and by Ermisch (1988) on the basis of births of different order in England and Wales. His estimate was that a doubling of child-benefit would, eventually, raise completed family size by 0.15 children. This implies that should the government wish to be so un-British as to manipulate the birth rate, the costs of achieving significant change through cash benefits would be prohibitive.

Ermisch's analysis (and that of the study that preceded it, De Cooman *et al*., 1987) suggests that most of the reaction of births in England and Wales to changes in economic *signals* was in timing, particularly of the first birth, rather than completed family size. Postwar swings in period fertility have generally had a larger component of *tempo* rather than *quantum*, timing and spacing rather than completed cohort fertility. (Olsen, 1994; Hopflinger, 1994; Walker, 1995) While swings in birth timing may not make any difference to the family's ultimate size, they make a difference to the overall age structure, and make it difficult to plan age-specific services like schools, jobs for school leavers, and, eventually, pensions. It therefore remains important to understand and anticipate movements in births, even if there is little prospect of controlling them. Cigno (1991) suggests that governments are unintentionally fueling the deferment of childbearing through tax policies conducive to female labour force participation. This argument could be countered with the suggestion that female labor force participation need not be incompatible with childbearing, if policies are in place to facilitate their combination.

Such policies have been developed in Sweden. In the early 1990s, Sweden had one of the highest levels of fertility in Europe, and many commentators have been tempted to suggest that the package of family-friendly policies were responsible for raising the birth rate up to replacement levels. Two recent studies cast some doubt on such an interpretation. Both stress the importance of re-timing of Swedish births around historic and projected cohort fertility close to two children for all cohorts born this century. Hoem (1990) explains how a specific change in the parental leave regulations in 1986 created an inducement to space births more closely. There is evidence people did this, a response, perhaps unintended, to a policy change, but mainly one of tempo rather than quantum. Walker (1995) also concludes that there has been little effect of economic variables on life-cycle fertility between the 1950s and the 1990s, and that the impact of state policy was less important than broad trends of economic change in the labor market. He reaches this conclusion by estimating how the cost of childbearing (the *shadow price*) has changed over the years. He constructs series for the value of parental benefits, public child care, and child allowances that are compared with direct costs and opportunity costs in the labor market including foregone opportunities to accumulate human capital. The compression of wage differentials, another aspect of Swedish policy, during the 1970s and 1980s, is said to have reduced incentives for early childbearing and could also account for the shift in birth timing over the period (see also Cigno, 1991, Chapter 8), but pronatalist policies would have had to be even more generous than the standards set in Sweden to have much importance in the short or long term. The last resort for those who wish to defend Swedish family policy as a pronatalist instrument (rather than something that has merit in its own right) is that without parental leave and childcare apparatus the size of the Swedish family would have been smaller. This is an unfortunately unverifiable proposition. Drops in fertility in Sweden (and Eastern Germany) in the 1990s, may still prove to be temporary adjustments.

It remains to be seen whether pronatalist policies are likely to be effective at changing anything other than timing, and it is unlikely that we will see this put to the test. Hobcraft and Kiernan (1995) offer another reason to be skeptical of the state as engine of a reversal of the fertility decline—that it is unlikely that the public budget would ever devote substantial enough resources to make much difference. As Folbre (1994) would put it, the vested interests against investing in children as public goods, leave social reproduction increasingly the responsibility not just of private families, but the women left *holding the baby* as male responsibilities are eroded. She calls for a new definition of family policy enshrining an egalitarian restructuring of the genders.

Although state policies can certainly help to speed up the decline of high fertility, there is little evidence that they can successfully apply the brakes, should they want to, when the process falls below replacement level. Fertility and family life might be facilitated by a wide range of economic and social changes, but there is unlikely to be a substantial sustained response to measures aimed only at women and to the proximate determinants of their fertility.

CONCLUSIONS

The mastery of nature displayed in the control of fertility is itself one aspect of economic development. It is possible for this aspect to run ahead, behind, or alongside the mastery of other theaters of endeavor. When they run together, ideational and materialist explanations are compatible perspectives on the same process, the economist's signal is the sociologist's sanction. However, development is not always even, nor can one driving force always claim primacy. In some situations it may be possible to test the predictions of economic and social explanations against one another, but not all. Explanations that are observationally equivalent may have different implications for policy or prediction according to what the underlying mechanism is. It is thus difficult to reach simple conclusions about the social and economic determinants of fertility, other than to point to their complexity. Different forces on beliefs and behavior operate at individual, family, community and societal levels with different strengths, on different time scales and in different points in time and space. Furthermore there is feedback between the economic and the demographic, reflecting human kind's capacity to adapt and negotiate. The data one would ideally need to test such a model quantitatively defy description, let alone practicality in poor countries (David, 1998).

The lack of suitable longitudinal, multilevel, data (on paid and unpaid work, family history and structure, attitudes in the home and outside, and so on) may be one reason why social scientific knowledge of the determinants of fertility is not a clear consensus. It is easier to draw some conclusions about the beliefs of social scientists than about the beliefs and reproductive behavior of *real people*. The economic approach sees rational agents deciding to invest in child quantity, child quality or birth control according to changing price signals in the economic environment. The sociological view is that changing norms regarding reproduction lie behind fertility decline, accompanying, preceding and perhaps even promoting economic development. The *ideational* theory of fertility change has been criticized because it appears to allow no role for the cultural context in which diffusion of ideas takes place. *Institutional* theories of fertility contend that particular historic and political configurations shape social institutions in any society, and give rise to particular patterns of social organization. The sociologist's emphasis on the diffusion of ideas complements the economist's model in which agents are assumed to be equipped at no cost with perfect information that is assumed to permeate society (for an exception see Pollak, 1985).

Economic and sociologic analysts of fertility are too often in competition, but there is common ground. Notable attempts to synthesize across disciplines are offered by Folbre (1994), Hobcraft and Kiernan (1995), Pollak and Watkins (1993), and Pritchett (1994). For Folbre, constraints are gendered (men and women do not always face the same choices about social reproduction), but the salience of gender differences has varied over time and space.

Cigno's exposition of economic decision making in and about the family takes a more narrowly focused view. He confidently concludes that the decline in size of the family, and in its role as *an organizer of transactions between generations* are brought about as a response to *changes in the structure of incentives and disincentives associated with economic development*, rather than caused by changes in preferences or culture.

Indeed, some of what goes under the name of *culture* may be interpreted as the debris of past states of the economic environment—a consequence of the fact that, when the structure of incentives and disincentives changes, it takes time to realize it, time to decide how best to respond to it, and time to modify the pattern of one's life (Cigno, 1991, p. 195).

Most other economists writing about the family have been less sanguine about the completeness of the explanations to be offered by models of constrained optimization. Montgomery and Trussell (1986, p. 266) point out that the family still has to be set aside from other subjects of economic research because of "the pervasive roles played by individual physiology, by uncertainty, by irreversibility and by social norms." Irreversibility refers to the fact that responsible parents at least cannot trade in their children even if they turn out unsatisfactorily, like Gary Becker's Rotten Kid.

Readers may be more comfortable with the line taken by Pollak and Watkins (1993). They suggest that different cultural settings mediate how preferences and opportunities operate within groups differentiated by "language, ethnicity, region, and dress" (p. 488). They also suggest that opportunities and preferences are inseparably bonded, as the warp and weft of a fabric, and warn against pulling it apart. In declaring the need for a multidisciplinary approach they remark:

> The rational actor model is silent about the boundary between choice behavior and nonchoice behavior and about the determinants of nonchoice behavior. Perhaps 'culture' can fill this analytic vacuum by describing or explaining the location of the boundary separating nonchoice and choice behavior (p. 499).

The strength and diversity of social norms helps explain variation in behavior between and within societies. Secular and sacred values help to set targets and to sanction deviant behavior. But just as individuals may not be purely rational actors, neither are they merely passive agents of culture. They can shape and re-shape their culture, and may re-negotiate the norms, rules or values to their own benefit. The political and economic forces shape the cultural context of behavior, and individuals continually renegotiate their behaviors, rationalizing them within the context of social norms (Bledsoe *et al*., 1993).

Neither the view of individuals as passively adhering to cultural values, nor as rational utility maximizers is satisfactory. The notion that human motivation extends beyond economic incentives is useful not only for the understanding of fertility, but also relevant to other fields of activity, including the management of a paid workforce. No one set of explanations is sufficiently developed to provide an integrated view of why and how reproductive change occurs. There is a need for testable theories for how sociocultural and economic forces can be linked to variations in individual behavior. Both macro- and microlevel analysis is required to provide an understanding of how the political economy evolved and how the individual creates his definition of *proper* behavior (Greenhalgh, 1995; Kertzer, 1995).

The theories and evidence we have assembled paint the picture of the process of social reproduction as vitally connected with the process of production. They are not always harmoniously adjusted to one another. The gender division of labor and power sets the terms of cooperation and conflict between men and women in the reproductive and productive spheres. As their work so often spans reproductive and productive fields, it is obvious that women's lives need more than economics to explain them, but so do men's. Just as men and women have seldom been successful at cooperating on equal terms to achieve reproductive and productive goals, social scientists have not always found cooperative solutions to the challenge of understanding these important processes. We hope this chapter points toward a constructive renegotiation of the division of social scientific labor.

References

Appleton Simon, 1996. How does female education affect fertility? A structural model of the Côte d'Ivoire. *Oxford Bulletin of Economics and Statistics*, vol. 58(1), p. 139–166.

Axinn William G., 1992. Family organization and fertility limitation in Nepal, *Demography*, vol. 29(4), p. 487–502.

Bankole Akinrinola and Olaleye David O., 1995. Do marital partners have different reproductive preferences in Sub-Saharan Africa? *in:* P. Makinwa and An-Magritt Jensen, (eds.), *Women's position and demographic change in sub-Saharan Africa*, Liège: Scientific Study of Population, p. 147–168.

Basu Alaka M., 1992. *Culture, the status of women, and demographic behaviour: illustrated with the case of India*, Oxford, Clarendon Press. 265 p.

Becker Gary, 1981. *A treatise on the family.* (Revised and enlarged edition, 1991). Cambridge, MA, Harvard University Press. 424 p.

Blake Judith, 1972. Fertility control and the problem of voluntarism: Proceedings of the Twenty-Second Pugwash Conference on World Affairs, reprinted *Population and Development Review*, vol. 20, March 1993, p. 167–177.

Blanchet Didier and Ekert-jaffé Olivia, 1994. The demographic impact of family benefits: evidence from a micro-model and from macro-data. *in:* John Ermisch and Naohiro Ogawa (eds.), *The family, the market and the state in ageing societies*. London: Oxford University Press, p. 79–103.

Blank Rebecca, 1993. What should mainstream economists learn from feminist theory? *in:* Marianne A. Ferber and Julie A. Nelson (eds.), *Feminist economics today: beyond economic man*. Chicago: Chicago University Press. 209 p.

Bledsoe Carolyn, Hill Allan G., D'allesandro Umberto and Langerock Patrica, 1994. Constructing natural fertility: the use of western contraceptive technologies in rural Gambia, *Population and Development Review*, vol. 20(1), p. 81–114.

Bloom David, Freeman Richard and Korenman Sanders, 1987. The labour market consequences of generational crowding, *European Journal of Population*, vol. 3, p. 131–176.

Bongaarts John P., 1983. The proximate determinants of natural fertility, *in*: Rodolfo A. Bulatao, Ronald D. Lee, Paula E. Hollerbach and John Bongaarts (eds.), *Determinants of fertility in developing countries*, vol. 1, p. 103–138. New York, Academic Press.

Brass William and Kabir Mohammad, 1980. Regional variations in fertility and child mortality during the demographic transition in England and Wales, *in:* John Hobcraft and Phil Rees (eds.) *Regional demographic development*, London, p. 71–88.

Bulatao Rudolfo and Lee Ronald D. (eds.), 1983. *Determinants of fertility in developing countries*, New York: Academic Press, 2 vols.

Butz William P. and Ward Michael, 1979. The emergence of countercyclical U.S. fertility. *American Economic Review*, vol. 69(3), p. 318–328.

Cain Mead, 1982. Perspectives on family and fertility in developing countries, *Population Studies*, vol. 36, p. 159–175.

Cain Mead, 1985. On the relationship between land-holding and fertility, *Population Studies*, vol. 39(1), p. 5–16.

Cain Mead, 1986. The consequences of reproductive failure: dependence, mobility and mortality among the elderly of rural South Asia.*Population Studies*, vol. 40(3), p. 375–388

Cain Mead and McNicoll Geoffrey, 1988. Population growth and agrarian outcome, *in*: Ronald D. Lee, Brian Arthur, Allen C. Kelley, Gerry Rodgers, and T. N. Srinivan (eds.), *Population, Food, and Rural Development*, p. 101–117. Oxford, Clarendon Press, 215 p. (International Studies in Demography)

Caldwell John C., 1982. *Theory of fertility decline*. Academic Press, New York. 386 p.

Carvalho José Alberto, Paiva Paulo A. and Sawyer Donald R., 1981. The recent sharp decline in fertility in Brazil: economic boom, social inequality and baby bust, *Population Council Working Paper:* Latin America and Caribbean Regional Office. 62 p.

Cassen Robert, 1978. *India: population, economy, society*. London, Macmillan. 419 p.

Casterline John B., 1989. The state, social stratification and fertility transition, *in: International Population Conference, New Delhi*, vol. 1, Liège, IUSSP, p. 303–313.

Chamie Joseph, 1981. *Religion and fertility: Arab-Christian–Muslim differentials*. Cambridge, Cambridge University Press. 150 p.

Chesnais Jean-Claude, 1983. La notion de cycle en démographie: la fécondité post-transitionelle, est-elle cyclique? *Population*, vol. 38, 361–390.

Chisholm J. S., 1993. Death, hope, sex: life: history theory and the development of reproductive strategies, and invited commentaries, *Current Anthropology*, vol. 34(1), p. 1–24.

Cigno Alessandro, 1991. *The economics of the family*. Oxford, Clarendon Press.

Cleland John, 1985. Marital fertility decline in developing countries: theories and the evidence, *in:* Cleland, J. and Hobcraft, J. (eds.) *Reproductive change in developing countries: insights from the World Fertility Survey*, Oxford, Oxford University Press, p. 223–254.

Cleland John, 1993. Equity, security and fertility: a reaction to Thomas, *Population Studies*, vol. 47(2), p. 345–352.

Cleland John and Jeejebhoy Shireen, 1996. Maternal schooling and fertility: cvidence from censuses and surveys, *in:* Roger Jeffrey and Alaka M. Basu, 1996. *Girls' schooling, women's autonomy and fertility change in South Asia*, New Delhi, Sage Publications, p. 72–106.

Cleland John, Phillips James F., Amin S. and Kamal G. M., 1994. *The determinants of reproductive change in Bangladesh: success in a challenging environment*, Washington DC, The World Bank. 187 p.

Cleland John and Rodgriguez German, 1988. The effect of parental education on marital fertility in developing countries, *Population Studies*, vol. 42, p. 419–442.

Cleland John and Van Ginneken Jeroen K., 1988. Maternal education and child survival in developing countries: the search for pathways of influence, *Social Science and Medicine*, vol. 27(12), p. 1357–1368.

Cleland John and Wilson Christopher, 1987. Demand theories of the fertility transition: an iconoclastic view, *Population Studies*, vol. 41(1), p. 5–30.

Coale Ansley J., 1973. The demographic transition reconsidered. *in: International Population Conference. Liege*, 1973, vol. 1, International Union for the Scientific Study of Population, Liege, p. 53–72.

Coale Ansley and Watkins Susan (eds.), 1986. *The decline of fertility in Europe*, Princeton, Princeton University Press. 484 p.

Cochrane Susan H., 1979. *Fertility and education: what do we really know?*, Baltimore and London, The Johns Hopkins University Press. 175 p.

Coleman David and Schofield Roger, 1986. Introduction: the state of population theory, *in:* Coleman David and Schofield Roger (eds.) *The state of population theory*, Oxford, Basil Blackwell. 311 p.

Crook N., 1997. *Principles of population and development*. ed. Ian M. Timaeus. Oxford, Oxford University Press, 225 p.

David Patricia H., 1998. High fertility and short birth spacing: the poverty consequences of family-building patterns, *in:* Massimo Livi-Bacci and Gustavo DeSantis. (eds.) *Population and poverty in the developing world*, Oxford, Clarendon Press. p. 103–119.

Davies Hugh and Joshi Heather, 1994. The foregone earnings of Europe's mothers. *in:* Olivia Ékert-jaffé (ed.) *Levels of life and families*, Éditions I.N.E.D., Paris, p. 102–134.

Davis Kingsley, 1984.Wives and work: the sex role revolution and its consequences. *Population and Development Review*, vol. 10(3), p. 397–417.

De Cooman Eric, Ermisch John F. and Joshi Heather E., 1987. The next birth and the labour market: a dynamic econometric model of births in England and Wales. *Population Studies*, vol. 41(2), p. 237–268.

Demeny Paul, 1987. Re-linking fertility behaviour and economic security in old age: a pro-natalist reform. *Population and Development Review*, vol. 13(1), p. 128–132.

Dow Thomas E., Archer Linda, Khansiani Shaniyisa and Kerkevole John, 1994. Wealth flows and decline in rural Kenya. *Population and Development Review*, vol. 20, p. 343–354.

Duesenbery J. S., 1960. Comment on Becker's "An economic analysis of fertility," *in: Demographic and economic change in developed countries*. For National Bureau of Economic Research, Princeton, NJ, Princeton University Press, p. 231–234.

Easterlin Richard A., 1980. *Birth and fortune*: *the impact of numbers on personal welfare*. London, Grant Macintyre. 205 p.

Easterlin Richard A., Pollak Robert A. and Wachter, Michael L., 1980. Toward a more general economic model of fertility determination: endogenous preferences and natural fertility, *in:* Richard A. Easterlin (ed.), *Population and economic change in developing countries*. Chicago, University of Chicago Press, p. 81–149.

Ekert Olivia, 1986. Effets et Limites des aides financières aux familles: Une expérience et un modèle. *Population*, vol. 41(2), p. 327–348.

Ermisch John F. 1990. European women's employment and fertility again.'*Journal of Population Economics*, vol. 3(1), p. 1–18.

Ermisch John F., 1996. The Economic Environment for family formation, *in:* David Coleman (ed.), *Europe's population in the 1990s*. Oxford, Oxford University Press, p. 144–162.

Folbre Nancy, 1994. *Who pays for the kids? Gender and the structures of constraint*. London, Routledge. 335 p.

Freedman Ronald, Xiao Z., Li B. and Lavely William, 1988. Local area variations in reproductive behaviour in the People's Republic of China, 1973–1982, *Population Studies*, vol. 42, p. 39–57.

Friedlander Dov and Feldmann Carole, 1993. The modern shift to below-replacement fertility: has Israel's population joined the process? *Population Studies*, vol. 47(2), p. 295–306.

Gauthier Anne H., 1996a. *The state and the family: a comparative analysis of family policies in industrialized countries*. Oxford, Clarendon Press. 232 p.

Gauthier Anne H., 1996b. The measured and unmeasured effects of welfare benefits on families: implications for Europe's demographic trends, *in:* David Coleman (ed.), *Europe's population in the 1990s*. Oxford University Press, Oxford, p. 297–331.

Greenhalgh Susan, 1995. Anthropology theorizes reproduction: Integrating practice, political economic, and feminist perspectives, *in:* Susan Greenhalgh, (ed.), *Situating fertility: anthropology and demographic inquiry*, Cambridge, Cambridge University Press, p. 3–28.

Gustafsson Siv and Meisaari-polsa Tuija, 1991. Why does fertility in Sweden increase? *in: CEC Eurtastat Study on the Relationship between Female Activity and Fertility*, vol. 2, Brussels CEC DG V.

Hecht Jaqueline and Leridon Henri, 1993. Fertility policies: a limited influence?' *in:* Daniel Noin and Robert Woods (eds.), *The changing population of Europe*. Oxford, Blackwell, p. 62–75.

Heitlinger Alena, 1993. *Women's equality, demography, and public policies*, New York, St. Martin's Press. 383 p.

Hobcraft John and Kiernan Kathleen, 1995. *Becoming a parent in Europe* EAPS IUSSP European Population Conference Milano, vol. 1, Franco Angeli, Milan, p. 27–65.

Hoem Jan, 1990. Social policy and recent fertility change in Sweden, *Population and Development Review*, no. 4, p. 735–748.

Jejeebhoy Shireen J., 1995. *Women's education, autonomy, and reproductive behaviour: experience from developing countries*, Oxford, Clarendon Press. 306 p.

Joshi Heather E., 1985. *Motherhood and employment: change and continuity in post-war Britain. measuring socio demographic change*, Occasional Paper no. 34, London: OPCS, p. 70–87.

Joshi Heather E., 1990. Changing roles of women in the British Labour market and the family, *in:* Phyllis Deane, (ed.), *Frontiers of economic research*, Macmillan, Basingstoke, p. 101–128.

Joshi Heather E. and Davies Hugh B., 1994. The paid and unpaid roles of women: how should Social Security adapt? *in:* Sally Baldwin and Jane Falkingham (eds.), *Social Security: new challenges to the Beveridge model*, Hemel Hempstead, Harvester Wheatsheaf, p. 234–254.

Kaplan Hillard, 1994. Evolutionary and wealth flows theories of fertility: empirical tests and new models, *Population and Development Review*, vol. 20(4), p. 753–791.

Keilman Nico, 1996. Household, families and housing in *Demografia: Analisi e Sintesi*, Roma: Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," p. 25–50.

Kertzer David, 1995. Political-economic and cultural explanations of demographic behavior, *in:* Susan Greenhalgh (ed.), *Situating fertility: anthropology and demographic inquiry*, Cambridge, Cambridge University Press, p. 29–52.

Kreager Philip, 1986. Demographic regimes as cultural systems, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory*, Oxford, Basil Blackwell, p. 131–155.

Lapham R. J. and Mauldin Parker, 1985. Contraceptive prevalence: the influence of organized family planning programs, *Studies in Family Planning*, vol. 16(3), p. 117–137.

LeVine Robert A., LeVine Sarah E., Richman Amy *et al*., 1991. Women's schooling and child care in the demographic transition: a Mexican case study, *Population and Development Review*, vol. 17(3), p. 459–496.

Lee Kelley, Cleland John, Walt Gill and Lush Louisiana, 1995. *Population Policies and Programmes: Determinants and Consequences in Eight Developing Countries*. London, London School of Hygiene and Tropical Medicine and United Nations Population Fund, 95 p.

Leibenstein Harvey, 1974. An interpretation of the economic theory of fertility: promising path or blind alley? *Journal of Economic Literature*, vol. 12(June), p. 457–479.

Leridon Henri, 1978. Fécondité et structures démographiques: une hypothèse sur l'évolution de la fécondité depuis 1940. *Population*, vol. 2, p. 441–447.

Leridon Henri, 1985. La baisse de la fécondité depuis 1965: moins d'enfants désirés et moins de grossesses non désirées, *Population*, vol. 40(3), p. 507–526.

Lesthaeghe Ron, 1983. A century of demographic and cultural change in western Europe: An explanation of underlying dimensions, *Population and Development Review*, vol. 9(3), p. 411–435.

Lesthaeghe Ron and Surkyn Johan, 1988. Cultural dynamics and economic theories of fertility change, *Population and Development Review*, vol. 14, p. 1–45.

Lesthaeghe Ron and Wilson Christopher, 1986. Modes of production, secularization, and the pace of the fertility decline in Western Europe, 1870–1930, *in:* A. Coale and S. Watkins (eds.) *The decline of fertility in Europe*, Princeton, Princeton University Press, p. 261–292.

Lloyd Cynthia B., 1994. Investing in the next generation: the implications of high fertility at the family level, *Population Council Working Paper*, no. 63, New York, The Population Council. 59 p.

Martine George, 1996. Brazil's fertility decline, 1965–95: a fresh look at key factors, *Population and Development Review*, vol. 22(1), p. 47–76.

Mason K. and Taj A. M., 1987. Differences between women's and men's reproductive goals in developing countries, *Population and Development Review*, vol. 13(4), p. 611–638

Mauldin Parker and Berelson B., 1978. Conditions of fertility decline in developing countries, 1965–75, *Studies in Family Planning*, vol. 9(5), p. 90–147.

McNicoll Geoffrey, 1980. Institutional determinants of fertility change, *Population and Development Review*, vol. 6(3), p. 441–462.

McNicoll Geoffrey, 1994. Institutional analysis of fertility, *Population Council Working Paper*, vol. 62, New York, The Population Council. 41 p.

McNicoll Geoffrey and Cain Mead, 1990. Introduction and framework: institutional effects on rural economic and demographic change, *in:* Geoffry McNicoll and Mead Cain (eds.), *Rural devel-*

opment and population: institutions and policy, New York and Oxford, The Population Council and Oxford University Press, Suppl. to vol. 15, 1989, *Population and Development Review*, p. 3–44.

Monnier Alain and de Guibert-Lantoine Catherine, 1995. The demographic situation of Europe and the developed countries overseas: an annual report. *Population* (English selection), vol. 7, p. 187–202.

Montgomery Mark and Trussell James, 1986. Models of marital status and childbearing, *in:* Orley C. Ashenfelter and Richard G. P. Layard, (eds.), *Handbook of labour cconomics*, vol. 1 Amsterdam, Elsevier 205–271.

Murphy Mike, 1993. The Contraceptive pill and women's employment as factors in fertility change in Britain 1963–1980, *Population Studies*, vol. 47(2), p. 221–244.

Notestein F., 1945. Population: the long view, *in:* Theodore Schultz, (ed.), *Food for the world*, Chicago, University of Chicago Press, p. 36–57.

Olsen Randall J., 1994. Fertility and the size of the U.S. labour force, *Journal of Economic Literature*, vol. XXXII, p. 60–100.

Orubuloye I. O., 1995. Women's control over their sexuality: Implications for STDs and HIV/AIDS Transmission in Nigeria, *in:* P. Makinwa, and An-Magritt Jensen (eds.), *Women's position and demographic change in sub-Saharan Africa*, Liège: International Union for the Scientific Study of Population, p. 223–232.

Pollak Robert A., 1985. A transaction cost approach to families and households. *Journal of Economic Literature*, vol. 23, p. 581–608.

Pollak Robert J. and Watkins Susan, C., 1993. Cultural and economic approaches to fertility: proper marriage or misalliance? *Population and Development Review*, vol. 19(3), p. 185–187.

Preston Samuel. H., 1986. Changing values and falling birth-rates, *in:* Kingsley Davis, Mikhail S. Berstam and Rita Ricardo-Campbell (eds.), *Below replacement fertility in industrial societies: causes, consequences and policies*. Supplement to *Population and Development Review*, vol. 12, p. 176–195.

Pritchett Lant H., 1994. Desired fertility and the impact of population policies, *Population and Development Review*, vol. 20(1), p. 1–56. [Note also subsequent exchange of views in *Population and Development Review*, vol. 20(3), p. 611–630].

Promislow D. and Harvey P., 1990. Living fast and dying young: a comparative analysis of life-history variation among mammals, *Journal of the Zoological Society of London*, vol. 220, 417–437.

Qvortrup Jens, 1994. A new solidarity contract? The significance of a demographic balance for the welfare of both children and the elderly, *in:* Jens Qvortrup, Sgritta M. Bardy, and H. Wintersberger (eds.), *Childhood matters, social theory, practices and politics.* European Centre, Vienna. Avebury, Aldershot, p. 319–334.

Riley Nancy and Gardner Robert, 1997. *China's population: a review of the literature*, IUSSP Monograph. 131 p.

Sathar Zeba, 1996. Women's schooling and autonomy as factors in fertility change in Pakistan: some empirical evidence, *in:* Roger Jeffrey and Alaka M. Basu (eds.), *Girls' schooling, women's autonomy and fertility change in South Asia*, New Delhi, Sage Publications, p. 133–149.

Schultz T. Paul, 1981. *Economics of Population*. Reading, MA, Addison-Wesley Publishing.

Simons John, 1982. Reproductive behaviour as religious practice, *in:* Charlotte Höhn and Rainer Mackensen, (eds.), *Determinants of fertility trends: theories re-examined*, Proceedings of an International Union for the Scientific Study of Population seminar, Bad Homburg, April, 1980. Liège, Ordina, p. 131–145.

Simons John, 1986. Culture, economy and reproduction in contemporary Europe, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory*, Oxford, Basil Blackwell, p. 256–278.

Stigler Joseph and Becker Gary, 1977. De gustibus non est disputandum. *American Economic Review*, (March), p. 76–90.

Thane Pat, 1989. Old age, burden or benefit? *in:* Heather Joshi (ed.). *The changing population of Britain*. Oxford, Blackwell, p. 56–71.

Thomas Neil, 1991. Land, fertility and the population establishment, *Population Studies*, vol. 45(3), p. 379–398.

United Nations, 1995. Population and Development: Programme of action adopted at the International Conference on Population and Development, Cairo, 5–13 September 1994, New York: United Nations, Department for Economic and Social Information and Policy Analysis, vol. 1/ST/ESA/SER.A/149.

Vallin Jacques, 1992. *La population française*. Paris, La Découverte, 128 p. (Repères, no. 75).

Vallin Jacques and Lery Alain, 1978. Estimating the increase in fertility consecutive to the death of a young child, *in:* Samuel H. Preston (ed.), *The effects of infant and child mortality on fertility*, p. 69–90. New York, Academic Press, 262 p.

Van de Walle Etienne, 1992. Fertility transition, conscious choice and numeracy, *Demography*, vol. 29(4), p. 487–502.

Walker James R., 1995. The effect of public policies on recent Swedish fertility behavior *Journal of Population Economics* vol. 8(3), p. 223–251.

Watkins Susan, 1986. Conclusions, *in:* Ansley Coale and Susan Watkins (eds.), *The decline of fertility in Europe*, Princeton, Princeton University Press, p. 420–449.

Weinstein Maxine, Sun T. S., Chang, M. C. and Freedman Ronald, 1990. Household composition, extended kinship and reproduction in Taiwan: 1965–1985, *Population Studies*, vol. 44, 217–240.

Westoff Charles. F. and Ryder Norman B., 1977. *The contraceptive revolution*. Princeton, Princeton University Press. 388 p.

Willis Robert J., 1973. A new approach to the economic theory of fertility behaviour. *Journal of political Economy* March/April., vol. 81(2 Pt. 2), p. S14–64.

Wood Rebecca, Botting Beverley and Dunnell Karen, 1997. Trends in conceptions before and after the 1995 pill scare, *Population Trends*, vol. 89, p. 5–12.

Wright Robert E. 1989. The Easterlin hypothesis and European fertility rates. *Population and Development Review*, vol. 15, p. 107–122.

Wright Robert E. 1991. Cohort size and earnings in Great Britain, *Journal of Population Economics*, vol. 4(4), p. 295–305.

CHAPTER

38

The Geography of Fertility

Maps, Narratives, and Demographic Innovation

RON LESTHAEGHE AND KAREL NEELS

Interuniversity Program in Demography, Vrije Universiteit Brussel, Brussels, Belgium

Maps are not solely indispensable instruments for navigation on a plane or in space. Maps, and particularly historic ones, may also offer startling testimonies on how we have altered our environment, both urban and rural. Spatial representations of social science data furthermore offer clues about relationships between forms of collective behavior. Deductively, they can be used in hypothesis testing, or, inductively, for proposing broader theories. In other words, maps produce good narratives, and these in their turn are valuable for theory formation as well (van de Kaa, 1996).

In the sections that follow we will give two examples of such map-based narratives and their statistical elaboration. The first example deals with the history of the fertility transition in France, and subsequently with Coale's much broader formulation of preconditions for behavioral innovation (Coale, 1973; Lesthaeghe and Vanderhoeft, 1999). The second example illustrates with Belgian data how successive waves of demographic innovation, from adoption of contraception in the 1880s until the spread of premarital cohabitation in the 1980s, can follow stable diffusion patterns over a period of a century. The latter finding can be matched by counterexamples in which new behavior forms are spreading along different lines, but both types of spatial patterning beg the question of stable versus changing underlying causes and mechanisms. But before turning to the details of illustrations, a word needs to be said about the feature of *ecological fallacy*.

Maps representing social science data have a crucial drawback: They only tend to represent averages, not distributions, for spatial aggregates. Of course, outcomes of multivariate analyses, even if they use individual level data, showing means or relative odds for subclasses of covariates do exactly the same. In both instances, we are not producing Quetelet's single *homme moyen*, of which all real men would be imperfect copies, but a set of little average men, each representing a robot picture of the real ones that are brought together in the various categories of our covariates (Desrosières, 1993; Lesthaeghe, 2001). Hence, the covariate of *spatial proximity* is just as relevant as the others, such as age, gender, educational level, or social stratum. Nevertheless, statistical analyses performed on data for spatial aggregates have frequently been accused of falling into the trap of ecological fallacy. But there are two issues here, not just one. First, since individual variance within an area is dramatically reduced by just retaining the area's average, correlations between such averages are higher than correlations computed for individuals as data points. This is indeed a statistical artefact. But the *real* ecological fallacy only arises if relationships observed at the level of aggregates are being projected on individuals. An indicator reflecting an attribute of an aggregate (e.g., the proportions used in agriculture) cannot automatically lead to the pronouncement that individuals with this attribute (e.g., farmers) would exhibit the same statistical association (e.g., with fertility). In fact, coal-mining regions in Belgium in 1900 had lower fertility, but coal-miners themselves did not. In short, associations between characteristics of specific aggregates only have interpretations at this specific level of aggregation, and at no other, higher or lower. This holds *a fortiori* also for individual level data: If coal miners have higher fertility, then the same must not of necessity be true for coal-mining regions. Coal mining could trigger other forms of industrialization, which could raise aggregate incomes and so forth, and the whole industrialization package—of which the proportions in mining would be an excellent *indicator*—could produce lower, not higher, fertility. Attributes of regions and of individuals can have very different *meanings*.

Armed with this major caveat, we can now turn to our two examples and their relevance for theory.

I. A NARRATIVE OF THE HISTORIC FRENCH FERTILITY DECLINE

The control of fertility in France has attracted the attention of demographers and historians alike because it occurs very early (i.e., already in part during the second half of the 18th century) and precedes the great transformations of the 19th century associated with urbanization and industrialization. The history of the French fertility decline has been a major challenge to classic demographic transition theory because rising real incomes, investments in education, did not drive it nor did declines in child mortality.

1. The Geography of French Fertility at the Beginning of the 19th Century

Detailed descriptions of the geography of the fertility decline in France have already been presented by other authors (van de Walle, 1974; Wrigley, 1985; Bonneuil, 1997), but it is clear that by 1831 highly contrasting levels of marital fertility had come into existence. For instance, at that date, Coale's index (see Chapter 34) of marital fertility I_g indicates that the level had already dropped below 40% of the Hutterite standard of uncontrolled fertility in several Normand departments and in the Garonne valley, whereas it had remained in the vicinity of 80 percent of Hutterite fertility in Brittany, the southern part of the Massif Central and the Hautes-Alpes. The map of the index of marital fertility I_g as of 1831 and reconstructed by van de Walle is also shown in Figure 38–1 (map C) and clearly shows that there were essentially five zones of high fertility and late control:

The smallest area is the Département du Nord, which corresponds to parts of Flanders and Hainaut (former parts of the Austrian Low Countries);

To the east there is a larger zone made up of Lorraine, Vosges, Alsace and the Franche-Comté;

In the middle, there is an extended zone of high fertility linking departments north and south of the Massif Central (from Nièvre to Aveyron), and continuing across the Rhone valley into the departments of Ain, Isère and Hautes-Alpes. This also includes the Savoie area, which was then not yet a part of France.

The southern zone of high fertility is located along the Pyrenees and in the Landes, and also includes Corsica.

The western area, finally, covers the departments of Brittany and stretches as far south as the Vendée.

In an independent reconstruction of the index of overall fertility (i.e., without the partitioning in the shares of marital fertility, illegitimate fertility and marriage postponement), Bonneuil (1997) equally shows that there are essentially three zones of innovation of fertility control before 1830. These are:

The Normandy area, stretching further south to the Pays-de-Loire;

The departments of the Champagne area and northern Burgundy;

The southwestern area located in the Garonne valley and also comprising adjacent Charentes. It should be noted, however, that this southwestern corner had relatively low fertility to start with (i.e., before 1790), possibly as a result of a more marked pattern of birth spacing and more prolonged lactation (Lesthaeghe, 1992).

To sum up, van de Walle's map of the marital fertility index I_g as of 1831 still gives a good picture of the geographic demarcations of areas with an earlier rather than later pattern of fertility control.

In addition to innovation with respect to fertility control, France is also the country with a pronounced correlation between the levels of marital fertility and the degree of marriage postponement (van de Walle, 1974). More precisely, areas with lower marital fertility tend to have the earliest ages at first marriage, and areas with higher marital fertility are also the ones with the latest and least universal marriage pattern. This can easily be seen on Figure 38–1 (map D) by comparing Coale's index of early marriage (I_m) to his index of marital fertility (I_g). In the North, the Malthusian pattern of late and nonuniversal marriage (i.e., higher definitive celibacy at age 50 for women) still prevails

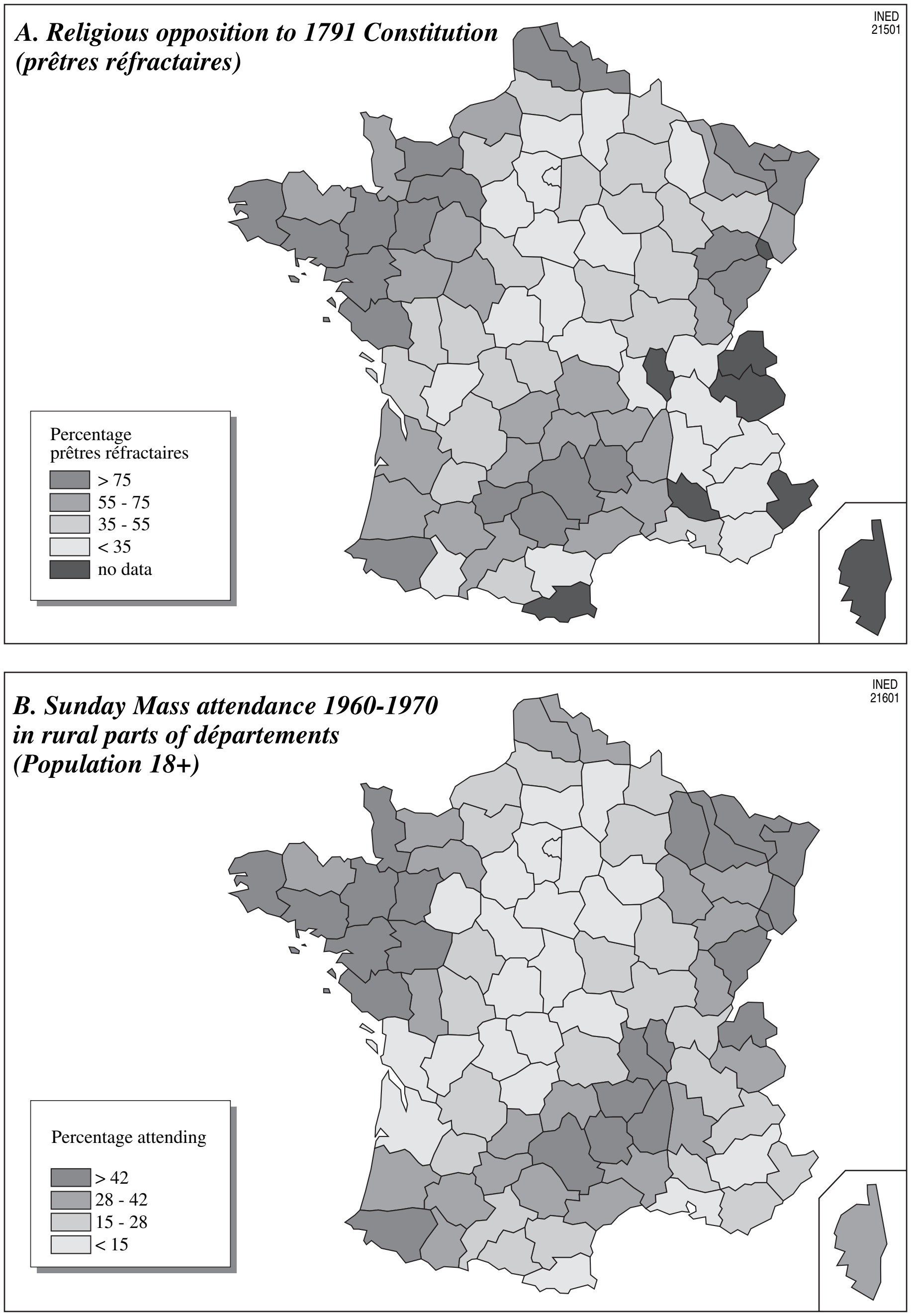
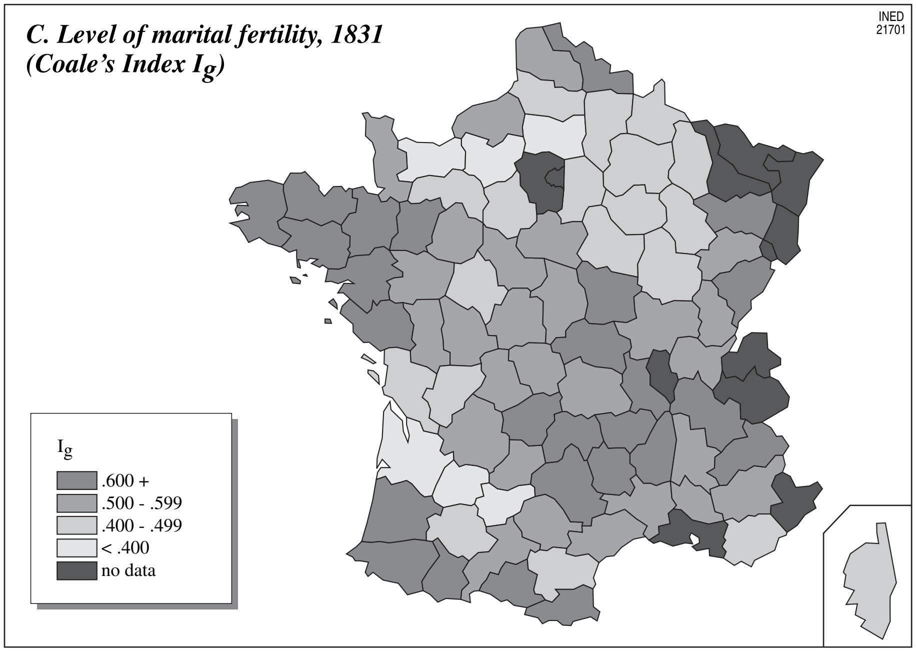


FIGURE 38–1 Geographic variations in secularization, compared with those in fertility level and in the marriage model. Source: **(A)** Tackett T., 1986. *La révolution, l'église, La France*. Paris, Editions du Cerf, 481 p. **(B)** R. Boulard, reprinted in Le Bras Hervé and Todd Emmanuel 1981. *L'invention de la France*. Paris, Librairie Générale de France, 511 p.

in 1831 in the departments of Nord and Pas-de-Calais. The eastern zones of high fertility from Lorraine to Franche-Comté are almost perfectly replicated on the map of late marriage. The southern half of the central area of high fertility and the Alpine area across the Rhone are also represented on the late marriage map, and the same holds for the western Pyrenees and Landes, and for the Brittany-Vendée western area. In short, with relatively few exceptions (parts of Normandy, northern half of the Massif Central), areas that had reduced marital fertility levels were equally the ones that had earlier marriage patterns and therefore deviated the most from the typical Malthusian system with postponed and nonuniversal marriage.



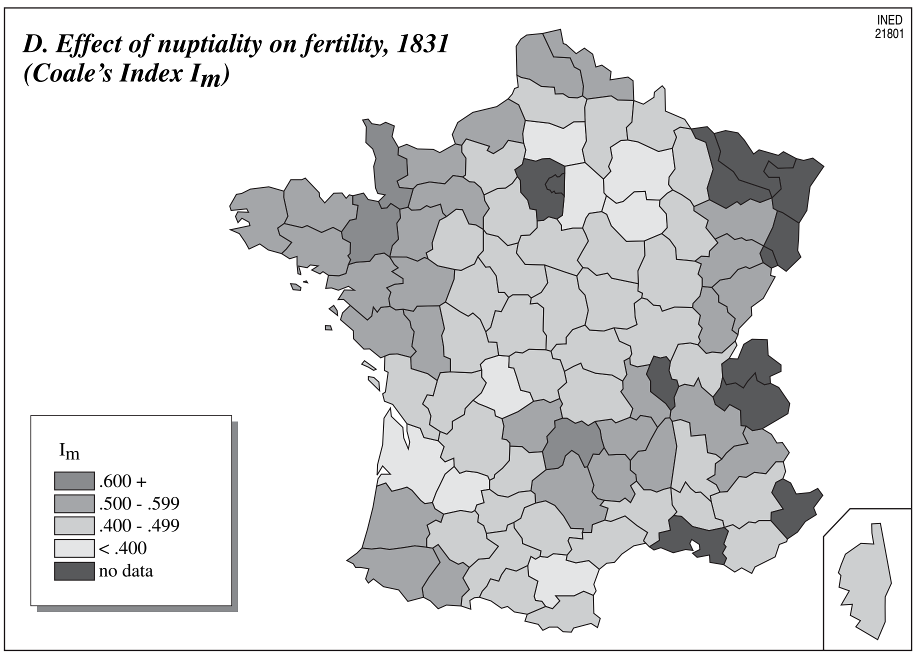


FIGURE 38–1 (*Continued*) Source: van de Walle Étienne, 1974. *The female population in France in the 19th century: a reconstruction of 82 départements*. Princeton, NJ, Princeton University Press, 483 p.

2. Explanatory Hypotheses

In this section we will treat four major hypotheses in accounting for the spatial patterns of marital fertility control and nuptiality. The first one deals with the production structure and with the economic motivations for reduced fertility centering on child utility. The second one links fertility to household structure, parental control, and patterns of inheritance. The third one is the secularization hypothesis and focuses on the moral legitimization of a new form of behavior. The last argument is of a political nature, and deals with the process of nation building and with patterns of regional and linguistic particularisms.

In earlier work dealing with the Western European fertility transitions (Lesthaeghe, 1983; Lesthaeghe and

Wilson, 1986) we have argued that familial forms of production such as small-holder or tenant-agriculture, whether or not in combination with cottage industries, are characterized by a high economic utility of children and by strong parental control over such family labor. Conversely, both the utility of children for household-based production and parental authority would weaken in systems with proletarization (i.e., with the growth of a wage earning class employed outside familial production sectors). Such wage earning populations can be used either in agriculture or in industry, but for the early French history of fertility control, only agricultural wage sectors are of importance.

Microeconomic theory has a similar point of departure. Weir (1982) argues that it is the relative value of children over the parents' life course that determines fertility. The value of family labor in its turn is a function of the economic situation of the family and of the prevailing labor market conditions in the community or wider area. Hence, Weir makes a distinction between the French areas with traditional peasant agriculture operating on a familial basis and redistributing any surplus labor via a domestic service outlet and areas with an early proletarization and a labor market recruiting wage earners for large agricultural states, typically owned by nobility, bourgeoisie, or the church.

According to these views, familial forms of production are conducive to the maintenance of high fertility, whereas the growth of agricultural wage labor fosters a fertility decline by virtue of reduced child utility over the parental lifetime.

The *mode of production* argument is of further significance for the nuptiality system as well. Peasant agriculture based on family labor is characterized by limited opportunities for an earlier establishment of new and economically independent households. Hence, according to the Malthusian principles of neolocal marriage (i.e., no coresidence of two or several married couples) and of economic independence, this means that those areas with small-holder and tenant farming ought to be characterized by late marriage. Furthermore, the circulation of surplus labor via domestic service (maids, manservants) also forcefully contributes to prolonged or even definitive celibacy (Hajnal, 1965, 1982). At this point, the interaction with the prevailing inheritance system has to be introduced. In regions with unequal division of property marriage postponement is typical for those who do not inherit or only get a smaller share, often in the form of a pension. Such persons tend to be pushed out of the system and enter other branches such as domestic service, army or navy, administration, or the clergy. They are also prime candidates for emigration.

One can also expect that families tied to small farms or to artisanal workshops are more heavily involved in arranging their children's marriage. A younger son from one family would be an ideal match for another family's heiress. The protection or enlargement of familial property is served by endogamy and even more by consanguineous marriage. By contrast, the predominance of wage labor in capitalist forms of extensive agriculture would not be associated with further marriage postponement or celibacy, since wage laborers have little to inherit and to wait for. Households can be set up independently more easily on the basis of wages earned outside the familial sector, and there is no pressure for arranged marriages and consanguinity.

The *mode of production* hypothesis has the advantage that it explains both the nuptiality and the fertility pattern: The duality between peasant agriculture and capitalist agriculture can be viewed as a common cause capable of explaining the observed correlation between nuptiality and marital fertility. In statistical terms, the strong negative correlation between Coale's indices I_g and I_m in France should largely vanish once a control is introduced for the prevailing form of agricultural production.

We now turn to the hypotheses focussing on patterns of parental and institutional authority. Todd (1983, 1988) introduces the parental control variable through a cross-tabulation of two other key variables—generational coresidence and form of inheritance system. He distinguishes between four types of family systems:

The nuclear residence pattern with equal division of property;
The nuclear residence pattern with unequal division;
The stem-family with a privileged son (primo- or ultimogeniture) who does not leave the family of origin upon marriage (whereas all others do);
The extended family in which coresidence of parents and married sons is maintained. Upon the death of the older generation, coresidence between married brothers remains possible, and inheritance tends to be egalitarian.

In Todd's patterning, the egalitarian nuclear family with equal division of property is characterized by weak parental control, and it would therefore also be a weak basis for authoritarian religious or political doctrines. By contrast, the stem family with its marked inequality among children and the extended control of the parental generation over the privileged son would be most supportive of strongly hierarchic religious structure. Todd's reasoning, although originating from variables other than those connected to the mode of production, bears some resemblance to our own line of argument. In our view, parental control is strongest in areas with familial forms of production by virtue of the restricted opportunities for adult children to set up their own independent household and because of the need to control marriages.

The argument of parental and religious authority is equally of relevance for the marital fertility decline since it provides a basis for the secularization hypothesis. If the act of consciously controlling fertility depends on the *moral legitimization* of the act as well as on the *calculus of advantage*, then areas with strong structural props in favor of institutional moral or religious patronage would lag behind in the fertility transition, whereas those without these props would be characterized by earlier secularization and by an earlier or faster fertility transition.

The last argument is of a political nature and deals with the process of nation building, and more particularly with regional or linguistic forms of particularism. It equally relates to patterns of spatial diffusion with particularisms of this nature acting as barriers to the spread of new forms of behavior. France obviously provides an example of early nation building with a central government gradually extending its unifying influence. As a result we should expect that characteristics of the core (here the larger Parisian Basin) would gradually be exported to the structurally and linguistically different periphery. Also Le Bras and Todd (1981) have stressed this feature and its corollary (i.e., the power of linguistic particularisms in France in defending the original structural characteristics of the periphery against the pressures stemming from the core). In addition, both authors point again at the importance of the various agricultural systems, but in contrast to Weir, much more attention is also being paid to other structural and cultural variables. Their position can be summarized as follows: Fertility control emerges in regions with early secularization, which in turn is fostered by the existence of a large rural proletariat and by nuclear family systems with equal division of property. Fertility control and secularization are both slowed down by the survival of peasant agriculture with inegalitarian inheritance and endogamy protecting the family capital from fragmentation, authoritarian family relationships, and linguistic particularism.

3. Empiric Evidence

The strong spatial resemblance between the fertility levels and marriage patterns of French departments and the pattern of secularization can be appreciated from the maps of Figure 38–1. Map A shows the religious divide as measured via the proportion of priests who refused to take the oath of allegiance to the secular government of the new Republic in 1791 (i.e., *prêtres réfractaires*). These priests remained loyal to the Pope, and their spatial distribution shows the extent of Catholic resistance to the Revolution. Tackett produced the original map in 1986. About 160 years later (i.e., in 1960–1970), Boulard's map of Sunday Mass attendance in the rural part of the departments (map B) exhibits a striking resemblance to Tackett's map of *prêtres réfractaires*: Five zones of greater fidelity to religion again appear. These five zones correspond once more with higher marital fertility in 1831 and later ages at first marriage as well.

A more elaborate statistical analysis is required to check our hypotheses. To this end, we have created a set of variables that characterize (a) the properties of the demographic regime (marital fertility, nuptiality, illegitimacy, endogamy); (b) the agrarian production system (peasant farmers versus wage labor, i.e., journaliers and métayers, domestic servants) and related inheritance systems; (c) the degree of secularization (prêtres *réfractaires*, ordinations, priests among army recruits, vacant vicarages, and Sunday Mass attendance); and (d) indicators of linguistic particularism and literacy. These indicators are defined in Table 38–1 for easy reference, and all but one pertain to the period between 1790 and 1900. Hence, they can be taken as typical for the 19th century.

In Table 38–2 we report the zero-order correlation coefficients between the level of marital fertility in 1831 (I_g) and the other indicators characterizing the agrarian modes of production, the cultural dimensions, and the nuptiality patterns. The structure is clear: high marital fertility and a late fertility transition are positively associated with two characteristics of peasant agriculture (servants in households, inegalitarian inheritance), with Catholic strength (refusal of revolutionary oath, priests among army recruits), linguistic particularism (no writing in French, survival of other languages) and with endogamy (cousin marriages). A low level of marital fertility and an early transition are associated with capitalist agriculture operating with wage labor, with high secularization (vacant vicarages), earlier marriage and higher illegitimate fertility.

The indicators used by Weir (1982) corroborate these findings in Table 38–3. The three indicators of the presence of familial peasant agriculture (fallow land, survival of the commons, and high percentage of owner-cultivators) are all three negatively correlated in a sample of 40 French villages with the amount of fertility decline witnessed among two sets of marriage cohorts (cohorts married between 1790 and 1819

TABLE 38–1 Definition and Sources of Indicators in Analyses of French Reproductive Regimes in 19th Century

A. Demographic indicators		
*I_g (1831, 1876)	Index of marital fertility relative to the Hutterite standard of uncontrolled fertility (Coale)	van de Walle, 1974: 226–467
*I_m (1831, 1876)	Index of proportions married (Coale)	ibidem
*I_h (1831)	Index of illegitimate fertility (Coale)	ibidem
*Age difference at marriage (1861–65)	Husband-wife difference in age at marriage	Le Bras & Todd, 1981: 440
*Endogamy (1911–13)	Rank order of départements according to the percentage cousin marriages (grouped from 0 = lowest to 11 = highest incidence)	Le Bras & Todd, 1981: 215
B. Agricultural system and inheritance pattern		
*Capitalist agriculture (1852)	Areas of large farms operating with *journaliers* and areas with *metayage* (=1) versus areas with smallholders and tenants (17% + fermage) (=0)	Todd, 1998: 84
*Inegalitarian inheritance (1900)	Inegalitarian system (=1) with primogeniture or other preferential treatment, versus egalitarian system (=0)	Todd, 1988: 32 (based on A. de Brandt, 1901)
C. Secularization		
*Refusal revolutionary oath (1791)	Refusal of revolutionary oath by clergy: 2 = more than 50% refused, 1 = 40%–50% refused, 0 = less than 40% refused	Todd, 1988: 64 (based on T. Tackett, 1986)
*Priests among army recruits (1825)	Priests per 10,000 recruits	Le Bras & Todd, 1981: 430
*Ordinations (1876)	Ordinations of priests per 100,000 population	ibidem
*Vacant vicarages	Proportion of vicarages that were vacant, i.e. without priests. Grouped values: 0 = less than 6%; 1 = 6%–15%; 2 = 15%–21%; 3 = more than 21%	ibidem: 379
*Sunday Mass attendance (1960–1970)	Percent of rural population attending Sunday Mass	ibidem: 430 (based on R. Boulard)
D. Literacy and language		
*No writing (1872)	Percentage of the population that can at most read French, but cannot write it	ibidem: 454
*Linguistic particularism (1863)	Persistence of patois or languages other than French. Grouped values: 0 = all speak French; 1 = traces of patois; 2 = large proportions or entire villages do not speak French; 4 = more than 50% non-francophone	ibidem: 279

compared to those married before 1770). Conversely, the indicators of the predominance of capitalist agriculture (artificial meadows, high percentage absentee landlords and high prices of arable land) are all three positively correlated with a faster decline of marital fertility.

The overall patterning can be brought out by means of a canonical correlation analysis. In this analysis we have brought together all demographic variables in one set (Y) and all the other indicators in a second set (X). From both sets, latent dimensions (canonical variables) are constructed in such a way that there is a maximal correlation between them. As the results in Table 38–4 indicate, the first canonical variate yields a correlation coefficient of 0.87 between the first Y-dimension and the first X-dimension (r_{x1y1}). For each of these dimensions, the best indicator can be listed by means of the correlations between dimension $Y1$ or $X1$, and their respective indicators. In this fashion, dimension $Y1$ is best identified by the indices of proportions married (I_m), the indices of marital fertility (I_g), the index of illegitimate fertility (I_h) and by endogamy. The signs reveal that $Y1$ captures a *slow* demographic transition: marital fertility and endogamy are high, proportions married and illegitimate fertility are low. These features are strongly correlated with the best indicators of the dimension $X1$: Strong adherence to Catholicism (five indicators), inegalitarian inheritance, and linguistic particularism. Conversely, negative correlates of dimension $X1$ are capitalist agriculture and vacant vicarages, which are associated with faster demographic innovation. Finally, the second canonical variate, which is uncorrelated with the previous one, tells us that there is a separate dimension indicated by a high difference in ages at marriage between spouses. This specific demographic feature is more exclusively correlated with areas of inegalitarian inheritance and with the survival of languages other than French.

An analysis of correlations brings out the general pattern of association, but it says little about synergistic or combinatory effects. A Boolean analysis highlights these interactive effects by exploring

TABLE 38–2 Correlation Coefficients Between Level of Marital Fertility (I_g) in 1831 and Selected Agricultural, Socioeconomic, and Cultural Variables; French Départements, 19th Century

Correlates (various dates)	Correlation coefficient with I_g
A. Modes of production	
Capitalist agriculture (versus smallholders and/or tenants), 1852	–0.46
Servants in households, 1856	+0.33
Inegalitarian inheritance, 1900	+0.40
B. Secularization	
Clergy refusing revolutionary oath, 1791	+0.46
Priests among army recruits, 1825	+0.31
Vacant vicarages, 1885	–0.52
C. Literacy and language	
No writing in French, 1872	+0.55
Persistence of patois or language other than French, 1863	+0.40
D. Nuptiality system	
Index of proportions married (I_m), 1831	–0.73
Index of illegitimacy (I_h), 1831	–0.30
Endogamy (cousin marriage), 1911–1913	+0.47

Source: see TABLE 38–1.

TABLE 38–3 Correlation Coefficients Between Change in Marital Fertility Among Cohorts Married Before 1770 and Marriage Cohorts of 1790–1819 and Indicators of Type of Agricultural Production (INED Sample of 40 Villages)

Indicators in 1852 (agricultural census) A: Familial peasant agriculture B: Capitalist agriculture with labor market	Correlation coefficient with marital fertility decline in two marriage cohorts (decline I_g)
A: percent fallow land	–0.31
Percent common land	–0.46
Percent owner cultivators	–0.48
B: Percent artificial meadows	+0.58
Percent absentee landlords	+0.46
Price per hectare of first quality arable land	+0.40

Source: D. Weir, 1982. *Fertility transition in rural France, 1740–1829*. Palo Alto, Dept. of Economics, Stanford University, 366 p. (PhD Economics).

TABLE 38–4 Canonical Correlation Results for Indicators of the Demographic Regime and for Socioeconomic or Cultural Indicators, 19th Century France, 90 Départements

First canonical variate:		Canonical correlation r_{x1y1} = 0.87 Wilks lambda: 0.05 Eigenvalue: 0.76	
Best indicators reproductive regime (**Y1**) (r > 0.40)		Best indicators socioecon/cultural dimension (**X1**) (r > 0.40)	
*I_m, index of proportions married 1831:	–0.86	*Sunday Mass attendance, 1960:	+0.90
*I_m, index of proportions married 1876:	–0.82	*Proportion not writing French, 1872:	+0.73
*I_g, level marit. fertility 1871:	+0.77	*Refusal revolutionary oath, 1792:	+0.70
*I_g, level marit. fertility 1831:	+0.75	*Ordinations per 100,000 pop., 1876:	+0.66
*I_h, level illegitimate fertility, 1831:	–0.65	*Capitalist agriculture, 1851:	–0.68
*Endogamy (cousin marriage) 1911–1913:	+0.46	*Vacant vicarages, 1855:	–0.61
		*Priests per 1000 recruits, 1825–1835:	+0.59
		*Inegalitarian inheritance, 1900:	+0.53
		*Linguistic particularism, 1863:	+0.50
Second canonical variate:		Canonical correlation r_{x2y2} = 0.74 Wilks lambda: 0.21 Eigenvalue: 0.54	
Best indicators reproductive regime (**Y2**) (r > 0.40)		Best indicators socioecon/cultural dimension (**X2**) (r > 0.40)	
*Husband–wife age difference at marriage 1861–1865:	+0.64	*Inegalitarian inheritance, 1900	+0.59
		*Linguistic particularism, 1863	+0.49

combinations of factors that are propitious in producing specific outcomes (Ragin, 1987). In the present example we have first dichotomized the independent variables as follows:

C: Capitalist agriculture with a substantial rural proletariat of *journaliers* and *métayers*
c: Peasant agriculture with a dominance of smallholders and tenants (1852)
P: Mainly partible inheritance
p: Mainly impartible or inegalitarian inheritance (1900)
F: Mainly French-speaking area
f: Matois or major non-Francophone groups (1863)
S: More secularized, with 50% or more of the clergy taking the oath of allegiance to the Revolution (1791)
s: Less secularized, with less than 50% taking the oath.

The dependent variable is equally dichotomized and indicates whether a department had marital fertility levels higher or lower than the mean in 1831 (mean $I_g = 0.56$). With four dichotomous predictors, one can inspect 16 different combinations ranging from cpfs to CPFS. The use of lower-case letters in cpfs means that none of the characteristics that are propitious to producing lower than average fertility are present, whereas the use of capitals in CPFS identifies the areas where all four of the conditions for a faster marital fertility transition are present. Among the possible 16 combinations, two have empty cells: CFps and CFSp, or simply CFp. In other words, we encounter a case of *limited diversity*. This is common in social science applications: systems have a degree of functional consistency and may therefore lack a number of *inconsistent* combinations. In this instance there were no departments with capitalist agriculture (C), and French speaking (F) and with impartible or inegalitarian inheritance (p): The latter seems to be inconsistent with capitalist agriculture or vanished during the course of history in the core regions of France.

Boolean minimization allows us to combine two Boolean expressions that differ in *one and only* one condition, yet produce the same outcome, into a new expression without that single contrasting condition. For instance if both CPfS and cPfS produce a faster fertility transition, the contrasting condition (here c versus C) can be dropped and the condition can be simplified to PfS. We have considered two types of outcomes for the 14 cells. First we wanted to establish which combinations of conditions are associated with less than a quarter of departments in each cell having I_g-values below the mean. In other words, what combinations lead to a strong majority of departments with a late or slow fertility transition? The resulting Boolean expression for strongly lagging is:

$$\text{LAG} = \text{Fcps} + \text{Cpfs} + \text{Pfc}$$

This expression states that Francophone (F) departments failed to experience a fast fertility decline by 1831 if this characteristic was combined with three counteracting conditions (i.e., c *and* p *and* s). Similarly, departments with only capitalist agriculture (C) also largely failed if the other three obstructing conditions (p *and* f *and* s) were simultaneously present. Finally, departments with egalitarian inheritance (P) also failed if they had a dominance of peasant agriculture and were not Francophone (c *and* f). If the last term were Pcfs, the rule would have been very simple: The effects of F, C and P, which on their own are propitious for low fertility, are neutralized if they are combined with the *joint* operation of counteracting factors. Stated differently, *if three of the four factors pushing in the direction of a faster fertility decline are missing, then the vast majority of such departments would lag behind in the fertility transition*.

The Boolean expression for the combinations leading to at least three quarters of the departments having lower than average levels of marital fertility in the 14 cells is:

$$\text{LEAD} = \text{CP(F + s)} + \text{S(Cpf + Fcp)}$$

The first term, CP(F + s), describes the prevailing conditions in 21 of the 24 departments that have a very early fertility decline. It indicates that the joint presence of capitalist agriculture and partible inheritance was necessary but not sufficient. These two conditions need to occur in tandem with either being Francophone or having low secularization. The second term, S(Cpf + Fcp), spells out the prevailing conditions in the remaining three departments (Creuze, Dordogne, Lot et Garonne). For these secularization is necessary but not sufficient since it has to be combined with another factor (either C or F) that is conducive to a low fertility level. Equally striking is that these three departments are located in the southwest and may have had the more marked pattern of birth spacing, leading to low pretransitional marital fertility already referred to for the Garonne area.

The main conclusions to be drawn from the Boolean analysis are as follows:

The sole distinction with respect to the type of agriculture does not constitute a sufficient condition for lower marital fertility in 1831. Other propitious factors need to be added to the combination.

When partible inheritance is added, the results are much stronger.

Cultural factors strongly interfere as brakes on the transition. This holds in particular for religious fidelity and linguistic particularism, as can be seen in the Boolean expression for *LAG*.

The present analysis is only based on regional cross-sectional data, and not on a more elaborate time–space analysis. Bonneuil (1997) offers such a more complete picture for the period 1831 to 1906, using Coale's index of overall fertility (I_f). This added time dimension leads this author to consider extra variables such as urbanization, the decline in mortality, the rise in education and migration. All of these become much more important after 1830. Yet, by that date, the basic map of the French marital fertility transition was already largely drawn, and as we have shown, it was shaped by more general systemic characteristics of both a structural and a cultural nature. Most of these are absent in Bonneuil's analysis for the subsequent period. In essence, these systemic characteristics have their roots in the diversity of agrarian production systems that developed in France prior to the 19th Century. These agrarian systems are associated with different nuptiality systems, differential use of nonfamilial labor (servants versus wage earners), contrasting inheritance patterns, strong timing differences in secularization, and patterns of marital fertility control. The diffusion of the latter innovation probably originated in more than one *center*, and it followed the lines of *weakest resistance* as spelled out by the covariates used here. Of equal importance is that this diffusion was hitting the obstacles formed by both linguistic and religious barriers. These barriers were not crossed so easily after 1830 either: In 1901 the map of marital fertility (van de Walle, 1974) still reflects the five zones of high fertility shown by the map for 1831, and the same holds in 1906 for Bonneuil's index of overall fertility, despite its contamination by differential marriage patterns.

II. NARRATIVE OF BELGIAN DEMOGRAPHIC INNOVATION

The Belgian example is an even more striking case of consistency in the spatial patterning of new forms of demographic behavior. We will try to document the following:

The features of the *second demographic transition*, such as the rise of divorce and of cohabitation after 1970, are strongly spatially correlated with the geography of the *first demographic transition* (i.e., the control of marital fertility and the weakening of the Malthusian marriage pattern after 1850);

Both are reflect a basic socioeconomic and cultural map that started to differentiate Belgian regions since the middle of the 19th century.

1. Background

As in France, there was older contrast between Belgian regions characterized by more intensive and more extensive farming. The former type was operated by family labor supplemented by unmarried servants, and it was mainly located on the sandy soils of Flanders. Since the beginning of the 18th century, such small-scale intensive farming was often combined with cottage industrial activities (especially textiles). The more extensive form is more common in the southern part of the country (i.e., in Wallonia) and uses wage labor as well. Cottage industries are rare. The demarcation is, however, not always clear-cut and does not exactly follow the linguistic border between the Dutch-speaking North and the French-speaking South. There are also exceptions within each of the linguistic areas. For instance, in Flanders more extensive agriculture is practiced on the rich alluvial clay soils of the Polders along the North Sea coast and of the Scheldt river estuary, whereas many Walloon farm laborers also operated their own small plots.

Nevertheless, political and religious differences developed along these lines, and most noticeably from the French period (1794–1815) onward. In fact, there was a first groundswell of secularization inspired by the ideals of the French Revolution, and carried by intellectuals and bourgeoisie, and later on (i.e., after Belgian independence [1830]) by the Liberal party. At the same time, Catholic opposition had also hardened, first against the French occupation (peasant revolts) and then against the Dutch king during the period of reunification of the Low Countries (1815–1830).

At the time of independence in 1830, two political pillars had formed, Liberal and Catholic, and until the 1880s these two were increasingly vying for popular support. Crucial for our purposes is the fact that the Liberal party remained strong, not only among the urban bourgeoisie, but also in the Walloon countryside. By contrast, the Catholic party remained firmly anchored in rural Flanders and particularly in the high population density belt characterized by intensive farming and cottage industries. For the period 1841 to 1847, for instance, we already have a good measurement of secularization for the entire territory via the percentage of annual marriages taking place in March and December (i.e., violating the Catholic ban on marriages during the two *closed periods* of Lent and

Advent). During the French period, there were clear signs of a weakening of this ban in various Walloon areas, but the map for 1841 to 1847 leaves no doubt: The secularization movement started from rural southern Wallonia and from the Liège area, and stops at the language border with Flanders by the 1860s. Contrary to common belief, the initial roots were not to be found in differential industrialization, since many strictly rural Walloon arrondissements showed levels of secularization that were as high as those witnessed in the emerging Walloon industrial belt (Lesthaeghe, 1991). In Flanders, there were only a few pockets with somewhat higher levels of secularization: the two towns of Antwerp and Ghent and to some degree the Polders strip along the coast.

The second wave of secularization is more closely associated with industrialization and urbanization, and its effect emerges essentially during the second half of the 19th century. It is also linked to the growth of the third political pillar (i.e., the socialist movement). At the same time, it is also matched by the Catholic revival and popular mobilization, as in other parts of Europe. New shades are then added to the map of secularization. By 1880, this map looks as follows:

There is a Walloon zone with the highest levels in the industrial belt, but also in all adjacent rural arrondissements (e.g., Thuin, Philippeville, Huy). The northern edge of this zone corresponds perfectly to the linguistic border with Flanders.

A second Walloon zone with somewhat lower levels dates back to the first secularization wave, and it is entirely made up of rural arrondissements, mainly located in the Ardennes region.

High levels are found in Brussels, the bilingual capital.

Moderate levels are found in a number of Flemish urban areas (Antwerp, Ghent) and in a few industrialized towns (e.g., Aalst).

Low levels prevail in the central part of Flanders (provinces of East Flanders and Brabant) and along the coast.

Very low levels are encountered in the western and eastern corners of Flanders, i.e., in the interior of the province of West Flanders and in the Antwerp Campine area and adjacent Limburg. These areas remain strongholds of Catholicism until the 1960s at least.

Much of this story is reflected in the statistical correlations presented in Tables 38–5 and 38–6. In Table 38–5 we have brought together the correlations between our various measurements for secularization, starting with the index of marriages during Lent and Advent (MLA) in 1841 to 1847, continuing with the parliamentary election results for 1919 (male universal suffrage) and 1958 (last *school war* between the Catholic pillar and the other two), and ending with absenteeism in the Sunday Mass attendance survey of 1964. As was expected, the spatial pattern of secularization measured through MLA in 1841 to 1847 is an early indicator that remains valid till the 1880s (correlation with MLA 1860 to 1865 = 0.91 and with MLA 1881 to 1884 = 0.83), but the added effect of the second wave of secularization, associated with the rise of the Socialist pillar, leads to a reduction of the spatial correlation (r with vote for secular parties in 1919 = +0.55). If, on the other hand, we start with MLA in the 1880s, the spatial pattern remains more intact, and if we take the election results of 1919 as a starting point, the subsequent maps of secularization remain almost identical till the 1960s. Table 38–6, finally illustrates that the rural character of arrondissements was less discriminating with

TABLE 38–5 Correlation Coefficients Between Various Indicators of Secularization, Belgian Arrondissements ($N = 41$)

	Marriage index during Lent and Advent			Socialist + Liberal + Com. vote		Absenteeism Sunday Mass 1964
	1841–1847	1860–1865	1881–1884	1919	1958	
Marriage index Lent + Advent (MLA)						
*1841–1847	—	+0.91	+0.83	+0.55	+0.36	+0.26
*1860–1865		—	+0.87	+0.73	+0.58	+0.50
*1881–1884			—	+0.83	+0.72	+0.65
Vote Socialist + Com. + Liberal (Sec vote)						
*1919				—	+0.91	+0.81
*1958					—	+0.91

Com. = Communist; MLA = marriages during Lent; Sec. = Secular.

TABLE 38–6 Correlation Coefficients Between Indicators of Secularization and Their Determinants, Belgian Arrondissements ($N = 41$)

	Marriage index during Lent and Advent (MLA)			Socialist + Liberal + Com. vote (Secvote)		Absenteeism Sunday Mass 1964
	1841–1847	1860–1865	1881–1884	1919	1958	
*Percent male active population in agriculture, 1890	–0.45	–0.59	–0.67	–0.81	–0.70	–0.72
*Percent male active population in agriculture + cottage industries, 1900	–0.26	–0.43	–0.59	–0.75	–0.75	–0.80
*Percent population aged 15–55 years, literate, 1880	+0.56	+0.46	+0.31	+0.13	–0.13	–0.14
*Percent Francophone, 1880	+0.63	+0.70	+0.76	+0.79	+0.61	+0.52

MLA, marriages during Lent and Advent.

respect to secularization levels until the 1860s, and these were better predicted by adult literacy (higher in Wallonia than in Flanders, also in Walloon rural areas) and by the proportion French speakers. The rural–urban distinction emerges more strongly after 1880, but the linguistic divide survives. The effect of differential adult literacy disappears of course during the 20th century as a consequence of universal primary education.

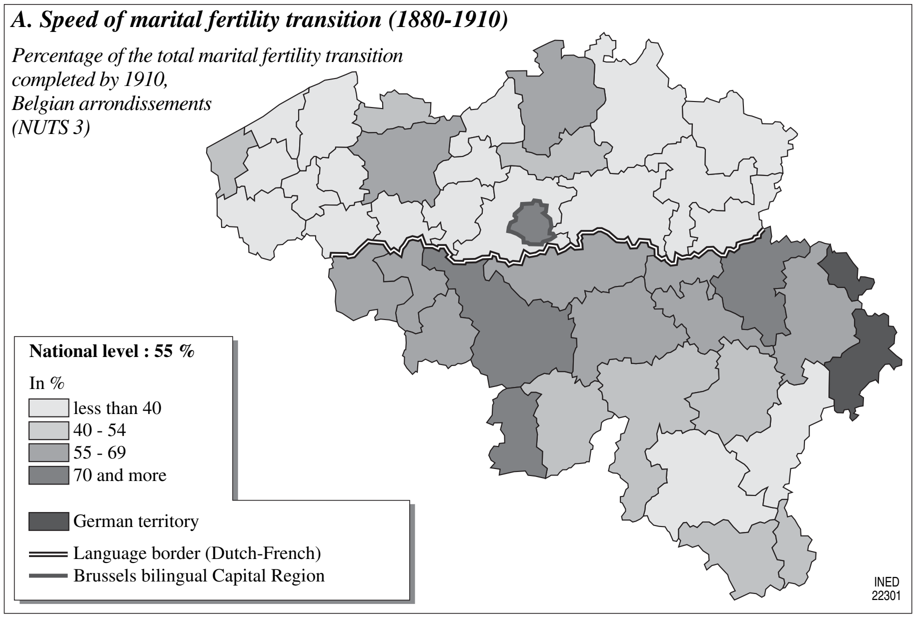
The overall outcome of the two-wave secularization process was a profound *pillarization* of Belgian society. Many aspects of public life (e.g., cooperative shops, labor unions, charities, pension funds, schools, leisure activities, media, and so forth) were organized by the three political families (Catholic, Liberal, and Socialist). In addition, the Flemish-Walloon contrast was further enhanced. Only after the 1960s would the political pillars weaken. The growth of other parties is related to this new process of *depillarization*, such as the Greens during the 1970s and right wing extremists (e.g., Vlaams Blok) during the 1990s. Both Liberal and Socialist parties had abandoned their anticlerical ideology, and at present, even the Christian democrats are debating whether the *C* should represent *Christian* or *Centrum*. But, as we will show, old and new forms of demographic behavior are still closely related to the older maps of secularization. It apparently takes longer for Belgian demography to *depillarize* than for the Belgian political scene.

2. Belgium's Spatial Aspects of the *First* and *Second* Demographic Transitions

A first impression of Belgian spatial pattern stability can be gleaned from the four maps presented in Figure 38–2. At the top we have juxtaposed the map of the speed of the marital fertility transition for the period 1880 to 1910 (map A) and the map of voting for the secularized parties in 1919 (map B). On both maps the central Walloon belt of industrial and adjacent rural arrondissements emerges very clearly. This was the area of early demographic *innovation* with respect to the adoption of birth control and the weakening of the Malthusian late-marriage pattern. This zone stops at the language border over its full east–west length. Fertility control has only spilled over this border in Brussels, which was already then a largely French-speaking city. It also started to develop in the urban arrondissements of Antwerp (to the north) and Ghent (to the west). The two zones of late adoption of fertility control and of Catholic loyalty also show up clearly at the eastern (Limburg + Campine) and western ends (Interior West-Flanders) of the Dutch-speaking region.

The last two maps in Figure 38–2 present two indicators of the second demographic transition (i.e., the rise of unmarried cohabitation during the 1970s and 80s [map D]) and the concomitant increase of the share of births out of wedlock in the 1980s (map C). The central Walloon zone shows up again in tandem with a broad ring of rural French-speaking arrondissements. Only in the southern corner of Wallonia is the rise of cohabitation more modest. The linguistic border is once more a major dividing line. In Flanders the growth poles are again the urban areas of Antwerp and Ghent, now followed by the urban area along the coast. The eastern and western corners of Flanders stand out by virtue of low levels of cohabitation and small shares of nonmarital births, in the same way as they were the last regions to practice fertility control a century earlier.

The story can also be told with a larger set of indicators measured over a period of about 150 years. These indicators are presented in Table 38–7, and include early and later measures of secularization, socioeconomic indicators in 1900 and 1970, and



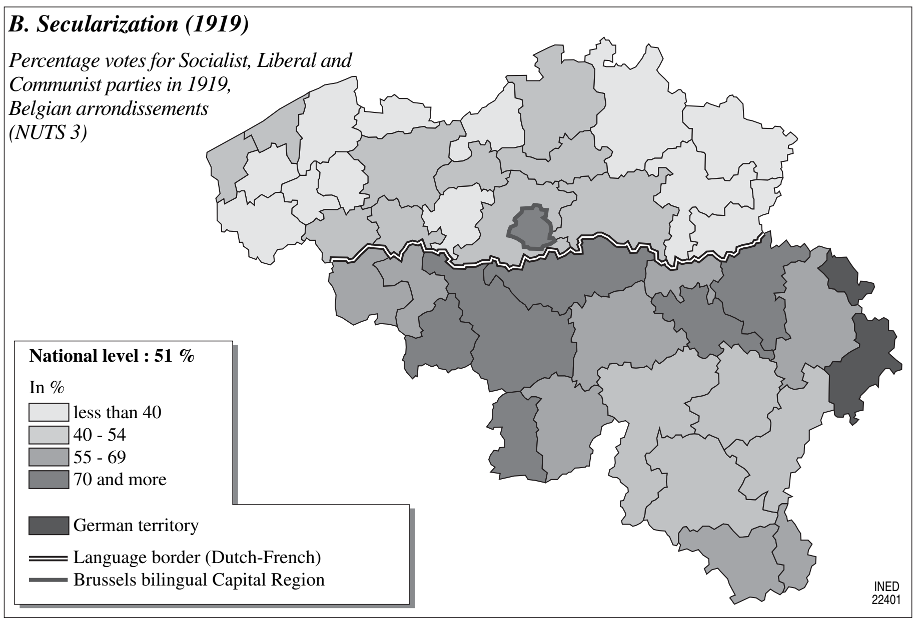


FIGURE 38–2 Pace of the marital fertility transition at the end of the 19th century **(A),** and level of secularization in Belgium **(B)**.

historic literacy differentials. They are paired to indicators of the *first* and *second* demographic transitions, including levels of marital fertility and nuptiality for the earlier period, and indicators of divorce, cohabitation, and nonmarital fertility for the later period. As in the example of France, a canonical correlation analysis has also been performed on these indicators, and the results are shown in Table 38–8.

As before, the demographic indicators are brought together in the *Y*-set and the socioeconomic and cultural factors in the *X*-set. The first canonical variate absorbs most of the information, and Y_1 and X_1 are very strongly correlated ($r = 0.98$). Virtually all demographic indicators load strongly on Y_1, starting with the speed of the marital fertility decline in 1880 to 1910 and with the recent level of nonmarital fertility of 1992 and the divorce incidence in 1967 to 1970. The current levels of cohabitation are next in line, together with the level of marital fertility more than 100 years earlier. The historic indicators of the weakening of the

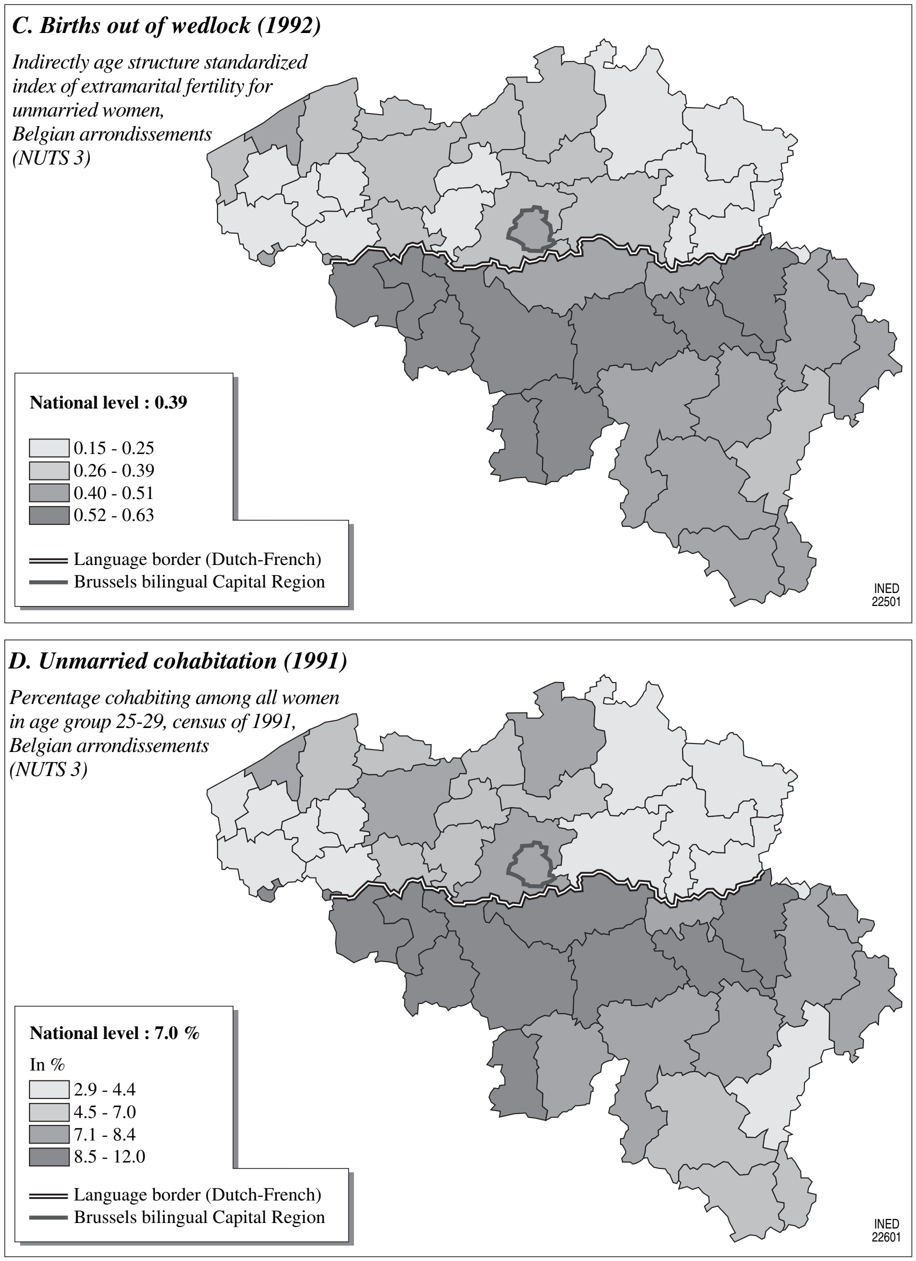


FIGURE 38–2 (*Continued*) Births out of wedlock in 1992 **(C)**, and unmarried cohabitation in 1991 **(D).**

Malthusian late marriage pattern are the weakest correlates of the dimension of demographic innovation, but still have loadings of +0.070. The corresponding X_1 dimension is almost perfectly identified by the voting for secularized parties in both 1919 and 1958 (loadings of 0.90 or more), closely followed by Sunday Mass absenteeism in 1964. This also reflects the socioeconomic structure as it existed in 1900 and indicated by the proportions then employed in agriculture and cottage industries. The 19th century measures of secularization are weaker indicators, but the index of marriages during Lent and Advent (MLA) of the 1860s still correlates with X_1 to a respectable degree (+0.69).

The only demographic indicator that identifies a second and uncorrelated canonical variate is the level of illegitimate fertility near the end of the 19th century. This map does not correspond as well to the secularization dimension, but is much more in line with the urban–rural divide and 19th century literacy levels. At that time, urban and especially industrial environ-

TABLE 38–7 Definitions of the Demographic, Socioeconomic, and Cultural Indicators Used in the Analysis for 41 Belgian Arrondissements

I_g	Coale's index (1963) of marital fertility, indirectly standardized for age and marital status composition and based on the Hutterite fertility standard; measured for 1880 and 1900.
Speed fertility transition	Percentage of the total marital fertility transition already completed before 1910; measured as $I_g(1910) - I_g(1880)/I_g(1880–0.200)$, where $I_g = 0.200$ is considered as the end point of the transition.
I_m	Coale's index of proportions married weighted by Hutterite fertility, 1880.
EM	Percentage ever-married women aged 20–24 years, 1880.
I_h	Coale's index of illegitimate fertility, 1880.
Divorce	Divorce rate per 10,000 married women, 1967–1970.
I_c	Index of nonmarital fertility comparable to Coale's I_h, except weighting by Belgian national age-specific fertility rates of 1989–1991 instead of Hutterite fertility, 1992.
Cohab	Percentage of all women aged 20–24 or 25–29 years currently in cohabitation, 1991.
MLA	Marriage index Lent and Advent (i.e., percent of total annual number of marriages in March and December divided by 2/12 [normal share for 2 months]); averaged over several years to neutralise effect of variable date for Easter. Measured for 1841–1847, 1860–1865, 1881–1884.
Sec vote	Percentage of votes for socialist, liberals and communists, parliamentary elections of 1919 and 1958.
Absent Mass	Percentage of adult population 15+ absent at Sunday Mass, 1964.
Agcot	Percentage of male active population employed in agriculture and cottage industries, 1900.
Agric	Percentage of male active population employed in agriculture, 1970.
Literacy	Percentage of population 15+ able to read + write, 1900.

Note: There are currently 43 arrondissements instead of 41. The old Brussels arrondissement has been reformed by bringing the Brussels Capital Region together with the Flemish arrondissement of Halle-Vilvoorde, whereas the newly formed French-speaking arrondissement of Mouscron (formerly a small section of the Flemish arrondissements of Ypres and Kortrijk) was dropped from the analysis. In other words, we have stayed as closely as possible to the historic division of the territory.

ments with a concentration of low literacy working class populations were the main settings of high illegitimacy. The present levels of nonmarital fertility are, as already indicated, spatially differentiated along the axis of the first canonical variate since they reflect much more the fertility levels of cohabiting couples, and are therefore typical indicators of the second wave of demographic innovation developing since the 1960s.

III. OTHER NARRATIVES: AN INVITATION FOR NEW RESEARCH

Obviously similar narratives can be produced for other European countries, and not all should exhibit similar degrees of spatial continuity over time as the examples of France and Belgium used here. For several countries a good start of the analysis is available, and many possess the necessary range of socioeconomic and cultural indicators. For instance, Livi-Bacci's work on Portugal (1971) and Italy (1977) clearly indicates that there are promising avenues to be followed for documenting the survival of older spatial patterns in the genesis of newer forms of demographic behavior. For Italy, the author is rather disappointed about the statistical explanatory power of classic socioeconomic variables in accounting for the historic marital fertility transition, but he finds that provincial fertility levels measured as early as 1911 are the best predictors of the regional voting results in favor of a liberalization of divorce laws in 1974 (Livi Bacci, 1977). Also the Swiss and Austrian narratives can be explored further because these countries too have long histories of spatial heterogeneity with respect to cultural, socioeconomic, and demographic variables. Both of them are therefore excellent testing grounds for the study of spatial continuity of first and second demographic transition variables.

There are of course counterexamples. The British history of the fertility decline is still an elusive case. Except for the initial Scottish–English divide, there are no clear geographic patterns of innovation and subsequent diffusion (Bocquet-Appel, 1997), nor are regional socioeconomic indicators providing convincing explanations (Teitelbaum, 1984). Social class distinctions rather than geographic ones may have been more important. Some Scandinavian countries may also produce a distinct class of narratives. For instance, unmarried cohabitation and concomitant extramarital fertility may have been features that were pushed to the northern frontiers of Norway and Sweden during a first phase, only to spread back south during the *second demographic transition*. Such a *revenge of the fringe*-hypothesis needs, however, more detailed doc-

TABLE 38–8 Canonical Correlation Results for Indicators of the *First* and *Second* Demographic Transitions and for Socioeconomic and Cultural Indicators, 19th and 20th Centuries, 41 Belgian Arrondissements

First canonical variate: canonical correlation $r_{x1y1} = 0.986$; Wilks lambda = 0.00; Best indicators demographic innovation Y_1 $r > 0.40$		Best indicators socio-econ. & cultural factors X_1 $r > 0.40$	
*Speed marit. fertility transition 1880–1910	+0.95	*Vote for secular parties, 1919	+0.95
*Level marital fertility, I_g 1900	–0.93	*Vote for secular parties, 1958	+0.90
*Level nonmarital fertility, I_c 1992	+0.92	*Absenteeism Sunday Mass, 1964	+0.83
*Divorce, 1967–1970	+0.91	*Percent male act. pop. agric. + cottage ind., 1900	–0.82
*Cohabitation women, aged 25–29, 1991	+0.86	*Marriages Lent + Advent, 1881–1884	+0.79
*Level marital fertility, I_g 1880	–0.82	*Marriages Lent + Advent, 1960–1965	+0.69
*Cohabitation women, aged 20–24 years, 1991	+0.78	*Marriages Lent + Advent, 1841–1847	+0.47
*Percent ever-married women, aged 20–24 years, 1880	+0.70		
*Index proportions married, I_m 1880	+0.70		
Second Canonical Variate: Canonical Correlation $r_{x2y2} = 0.927$; Wilks lambda = 0.01; Best indicators Y_2, $r > 0.40$		**Best indicators X_2, $r > 0.40$**	
*Index illegit. fertility, I_h 1900	+0.78	*% act. pop. in agriculture, 1970	–0.76
		*Literacy adult pop., 1900	–0.64
		*% male act. pop. agric + cottage ind., 1900	–0.42

umenting than is hitherto available. Furthermore, both Norway and Sweden have their more traditional Bible-belts, and to our knowledge, there has been no systematic investigation of this spatial cultural feature and its impact in shaping both first and second demographic transitions.

To sum up, many European countries provide settings of highly spatially differentiated cultural, socioeconomic and demographic developments that can be captured statistically since the 19th century. With a bit of imagination and luck, new indicators can be mined. The socioeconomic and demographic ones are easier to find, whereas the cultural ones may need more probing. However, a greater reward in the form of a more complete narrative may be down the line.

IV. NARRATIVES AND THEORIES OF DEMOGRAPHIC INNOVATION

The narratives produced in the Princeton Project of the 1970s on the history of fertility control and the modernization of the nuptiality systems in Europe inspired Coale (1973) to specify three preconditions for demographic innovation and adaptation. They are known as the *Ready, willing and able* specification. These terms refer to the following:

Ready: The new form of behavior must be economically advantageous and more so than the older form;

Willing: The new form must be culturally acceptable and not run counter to prevailing ethical or religious (or other ideational) convictions and values;

Able: There must be means, old or new, to implement the choice. This implies that these means must be known and are accessible.

In the narratives, readiness has been connected to socioeconomic structures and changes, to the evolution of real wages (Lesthaeghe, 1992, for France) and to economic aspirations (e.g., Dumont's *social capillarity*, 1880, or Denis' relative deprivation theory, 1899). At the micro-level, readiness also refers to the terms of economic calculus as for instance defined in Easterlin's (1978) demand function for children: utilities, disutilities, and relative means or income. Willingness, on the other hand, has been operationalized through the secularization indicators on the assumption that religious barriers would inhibit the adoption of new forms of demographic behavior, whereas secular outlooks foster the adaptation of behavior to new objective circumstance. The ability condition has not been operationalized here since we obviously do not have spatial information on the methods of contraception being used in historic times. In studies of much more recent fertility transitions, however, the ability condition could be operationalized via spatial patterns of knowledge and availability of modern methods of contraception.

At the micro-level, we can imagine that an individual or household i has a score on readiness (R_i), willingness (W_i), and ability (A_i), each ranging from low or unfavorable (0) to high or favorable (1), and with a zone of uncertainty in the middle (0.5). Coale's specification is a *bottleneck model* because the three conditions need to be met simultaneously. This means that the final decision score (S_i) is the *minimum* of the individuals' values for R_i, W_i and A_i. In other words:

$$S_i = \text{minimum}(R_i, W_i, A_i)$$

A particular person or household would only switch to the new behavior if his or her S_i is greater than 0.5 (i.e., when the minimum of the three component scores has crossed the boundary of indecision). Aggregating over all individuals or households implies that we are dealing with distributions for R, W, and A and for the minimum S (Lesthaeghe and Vanderhoeft, 1999). This, however, is a model for a single, homogeneous population. The spatial patterning requires a disaggregation according to distinct regional networks. A classic specification focussing on individual and network effects (Montgomery and Casterline, 1996) takes the following form:

$$Y_{i,t} = \beta_i X_{i,t} + \delta_i \sum \omega_{i,j} Z_{j,t-1} + e_{i,t}$$

The Greek letters indicate coefficients or weights, and the roman letters refer to variables. In this equation an outcome Y for individual or household i at time t is made up of the influence β_i of a set of *individual* characteristics $X_{i,t}$ and the influence δ_i attributed to the influence of a *network* composed of N persons or households. In this network, members j have characteristics Z_j, but the actor i gives differential *credibility* scores $\omega_{i,j}$ to each of them. These are then summed over all network members. The last term, $e_{i,t}$, is the error term.

Applied to Coale's conditions, we have to set up three such equations, one for R_i, W_i and A_i, respectively:

$$R_{i,t} = \beta_i X^R_{i,t} + \delta_i \sum \omega_{i,j} Z^R_{j,t-1} + e^R_{j,t}$$
$$W_{i,t} = \beta_i X^W_{i,t} + \delta_i \sum \omega_{i,j} Z^W_{j,t-1} + e^W_{j,t}$$
$$A_{i,t} = \beta_i X^A_{i,t} + \delta_i \sum \omega_{i,j} Z^A_{j,t-1} + e^A_{j,t}$$

The superscripts of the individual and network members' characteristics X and Z imply that these characteristics may vary or be of unequal relevance, depending on whether R, W, or A is considered. These characteristics can of course be objective attributes (e.g., wealth, income, and so forth), forms of behavior, or opinions and values. The individuals in the actor's network are likely to belong to the same kinship group, social class or have a shared location. The credibility weights are highly likely to vary according to social distance. The model above does not imply that R_i, W_i and A_i have to change at the same time: one of the three can be lagged and this would cause the minimum S_i to lag as well. Nor does the model imply a fixed sequence such as R→W→A (i.e., readiness increasing first, thereby causing a reevaluation of normative barriers, and then leading to a search for appropriate means). It may well be, as possibly in the case of the early French fertility decline, that normative barriers had weakened before (early secularization) and that means (e.g., coitus interruptus) were already well known before the economic motivations developed. In this instance, it would be the population distribution for R that would be the last to shift, and the bottleneck condition would not be defined by the distributions of W or A.

Let us now consider how the French and Belgian narratives fit with the model above. Familial forms of production (small-scale owner or tenant farming or cottage industry), inegalitarian inheritance (avoidance of fragmentation of land) and endogamous marriage strategies imply lower scores on readiness. Conversely, capitalist agriculture (C) and partible inheritance (P) imply higher scores on R. Early secularization (S)—whatever its initial cause—means higher scores on W. And linguistic particularism signals nonmembership of the core network. In this fashion, the Boolean expressions for LEAD and LAG describing the French marital fertility transition can be translated into the RWA model. Recall that the presence of one propitious factor (either P or C related to R, or S related to W, or F related to a core connection) but the absence of the other three are closely associated with a slow fertility transition. In other words, there was not one but a combination of bottlenecks (pc, ps, cs . . .) that produced the LAG-outcome. All such departments are low on secularization (s), and all of them have a missing element of the CP combination. Conversely, this CP-combination was the best predictor of a faster fertility transition and in the majority of LEAD departments, this was also combined with F (i.e., being part of the innovating core).

The Belgian fertility decline is characterized by two separate models as a result of the linguistic divide. In the South, the levels of secularization were already high before the start of the fertility decline so that W was no longer the bottleneck condition. The fertility transition can then be seen as advancing in tandem with the socioeconomic changes that influenced child utility and child costs. In the North, by contrast there were several urban centres of innovation with respect to fertility control, but this new form of behavior could only spread to the eastern and western corners of Flanders when moral and religious obstacles were weakening. In fact, the adoption of family limitation in these Catholic parts of Flanders may, at a later stage, have contributed to the gradual secularization of these areas as well (reversed causation).

The importance of region-linked networks in the Belgian example is further highlighted by the survival of subcultures that are either more permissive or more obstructive with respect to the typical features of the second demographic transition. Stable local networks then explain the replication of the diffusion model exhibited during the first demographic transition on both sides of the linguistic border: (a) diffusion from the Walloon core region towards the Ardennes in the southern corner of the country, and (b) diffusion from Brussels and urban areas in Flanders to the middle of Flanders first, and to the two Flemish corners later on. The connection with the secularization maps prior to 1970 furthermore indicates that the local subcultures kept their differentiation along the normative or the willingness dimension, even if the issues have changed (now: divorce, cohabitation, procreation within cohabitation). To sum up, countries characterized by strong and stable subcultures are also likely to produce narratives of long-term demographic continuity, similar to the Belgian experience.

CONCLUSION

The conclusion can be brief. In social science theories developed after the 1960s, it has become fashionable to attribute innovation of new behavior to the powers of individual agency. Individuals can cut all ties with tradition and adopt modes of action that are economically and functionally most suitable for a given objective situation. But, as our narratives have shown, demographic innovations over more than a century can originate in very much the same places, they can diffuse geographically in the same hierarchical fashion, and this can be steered along the same *lignes de force* governed by a *normative* dimension. Clearly, such a pattern can only emerge if (a) there is a continued and locally differentiated ethical and political socialization of successive generations, (b) if members of local and regional networks reinforce each others ways of thinking and of acting, and (c) if no outside forces (such as large migration streams) upset the stability of normative subcultures. In this sense, many Europeans may be more the product of their regional roots than that the primacy of individual agency has led them to believe.

Acknowledgments

The authors thank Didier Willaert for the production of the maps and Patrick Deboosere for the production of the recent Belgian cohabitation data by arrondissement.

References

Bocquet-Appel Jean-Pierre and Jakobi Lucienne, 1997. Diffusion spatiale de la contraception en Grande-Bretagne, à l'origine de la transition, *Population*, vol. 52(4), p. 977–1004.

Bonneuil Noël, 1997. *The transformation of the French demographic landscape, 1806–1906*. Oxford, Clarendon Press, 218 p.

Coale Ansley J., 1965. Factors associated with the development of low fertility: a historical summary, in United Nations (ed.), *United Nations World Population Conference 1965*, vol. 2, p. 205–209, New York, United Nations, 509 p.

Coale Ansley J., 1973. The demographic transition, *in:* IUSSP (ed.) *International Population Conference 1973*, vol. 1, p. 53–73. Liège, Derouaux, 494 p.

Denis H., 1899. *Le mouvement de la population et ses conditions économiques*. Brussels, Académie Royale des Sciences, des Lettres et des Beaux Arts (Mémoires couronnés et autres mémoires, vol. 59).

Desrosières A., 1993. *La politique des grands nombres—Histoire de la raison statistique*. Paris, La Découverte, 437 p.

Dumont A., 1880. *Dépopulation et civilisation. Étude démographique*. Paris, Éditions Lecrosnier & Babé, 520 p.

Easterlin R., 1978. The economics and sociology of fertility. A synthesis, *in:* C. Tilly (ed.) *Historical studies of changing fertility*, p. 57–134. Princeton, NJ, Princeton University Press, 390 p.

Hajnal J., 1965. European marriage patterns in perspective, *in:* David V. Glass and D. Eversley, *Population in History*, p. 101–143. Chicago & London, Aldine Publishing Co, 682 p.

Hajnal J., 1982. Two kinds of preindustrial household formation systems. *Population and Development Review*, vol. 8(3), p. 449–494.

Le Bras Hervé and Todd Emmanuel 1981. *L'invention de la France*. Paris, Librairie Générale de France, 511 p.

Lesthaeghe Ron, 1978. *The history of the Belgian fertility decline, 1800–1970*. Princeton, NJ, Princeton University Press, 350 p.

Lesthaeghe Ron, 1983. A century of demographic and cultural change in Western Europe. *Population and Development Review*, vol. 9(3), p. 411–436.

Lesthaeghe Ron, 1991. Moral control, secularization and reproduction in Belgium, 1600–1900, Société belge de démographie (Ed.) *Historiens et Populations*, p. 259–279, Louvain-la-Neuve, Editions Académia 836 p.

Lesthaeghe Ron, 1992. Beyond economic reductionism: the transformation of the reproductive regimes in France and Belgium in the 18th and 19th centuries, *in:* C. Goldscheider (ed.) *Fertility transitions, family structure and population policy*, p. 1–44. Boulder & San Francisco, Westview Press, 283 p.

Lesthaeghe Ron, 2001. Lambert-Adolphe-Jacques Quételet, 1796–1874, *in:* N. Smelser and P. Baltes (ed.) *International Encyclopedia of the Social and Behavioural Sciences*, p. 12673–12676. Oxford, Elsevier Science Ltd.

Lesthaeghe Ron and Vanderhoeft C., 1999. Une conceptualisation des transitions vers de nouvelles formes de comportements, *in:* Dominique Tabutin *et al*., *Théories, paradigmes et courants explicatifs en démographiei*, p. 279–306. Louvain-la-Neuve, Académia-Bruylant & l'Harmattan, 669 p.

Lesthaeghe Ron and Wilson C., 1986. Modes of production, secularization and the pace of fertility decline in Western Europe, 1870–1930, *in:* Ansley J. Coale and Suzan C. Watkins. *The decline of fertility in Europe*, p. 261–292. Princeton, NJ, Princeton University Press, 484 p.

Livi-Bacci Massimo, 1971. *A century of Portuguese fertility*. Princeton, NJ, Princeton University Press, 149 p.

Livi-Bacci Massimo, 1977. *A history of Italian fertility during the last two centuries*. Princeton, NJ, Princeton University Press, 311 p.

Montgomery M. and Casterline J., 1996. Social learning, social influence, and new models of fertility. *Population and Development Review* 22-supplement, p. 151–175.

Ragin C., 1987. *The comparative method*. Berkeley, CA, University of California Press, 185 p.

Tackett T., 1977. *Priests and parish in 18th century France*. Princeton, NJ, Princeton University Press, 350 p.

Tackett T., 1986. *La révolution, l'église, La France*. Paris, Editions du Cerf, 481 p.

Teitelbaum M., 1984. *The British fertility decline*. Princeton, NJ, Princeton University Press, 269 p.

Todd Emmanuel, 1983. *La troisième planète—Structures familiales et systèmes idéologiques*. Paris, Le Seuil, 251 p.

Todd Emmanuel, 1988. *La nouvelle France*. Paris, Le Seuil, 280 p.

van de Kaa Dirk, 1996. Anchored narratives: the story and findings of half a century of research in the determinants of fertility. *Population Studies*, vol. 50(2), p. 389–342.

van de Walle Étienne, 1974. *The female population in France in the 19th century: a reconstruction of 82 départements*. Princeton, NJ, Princeton University Press, 483 p.

Weir D., 1982. *Fertility transition in rural France, 1740–1829*. Palo Alto, Dept. of Economics, Stanford University, 366 p. (PhD Economics).

Wrigley E. A., 1985. The fall of marital fertility in 19th Century France: exemplar or exception? *European Journal of Population*, vol. 1(1), p. 31–60.

CHAPTER

39

Fertility: Theories, Frameworks, Models, Concepts

BART J. DE BRUIJN

Netherlands Interdisciplinary Demographic Institute (NIDI), Den Haag, Netherlands

Demography *is largely a social science and is becoming more so*. In the special issue of *Population Studies* to commemorate the journal's 50th anniversary, John Caldwell reached this conclusion after an assessment of demography's position in the scientific world (Caldwell, 1996). The unique character of demography may lie in the discipline's emphasis on quantification, analytical models, censuses, and large-scale sample surveys. It remains, nevertheless, a social science in that the central subject it deals with—population and population dynamics—is a social phenomenon: emerging from the aggregation of individual life events, to a great extent socially determined and with profound social consequences. The situation is aptly described by Schofield and Coleman (1986) in their introduction to *The State of Population Theory*. They characterized demography as a discipline with a hard mathematic core and a softer surrounding of an explanatory body of theory. Whereas this mathematic core has been called *a point of proper disciplinary pride* (McNicoll, 1989, p. 433), its development is not paralleled by an equally mature development of theory. The emerging picture of demography is that of an accurate, but relatively dull science; strong in accountancy, but relatively weak in conceptualization.

This unbalanced state of affairs is a matter of concern to many demographers. Schofield and Coleman, for instance, stated that

> Any subject which finds it necessary, or indeed possible, to consider its material divorced from an appropriate body of theory must be in trouble. This seems to be the case with demography at present (Schofield and Coleman 1986, p. 1).

During the development of the discipline of demography in the past decades, similar statements have repeatedly been made (e.g., Vance, 1959; Ford and De Jong, 1970; McNicoll, 1980, 1992; Greenhalgh, 1995b). Indeed, a one-sided sophistication of data collection and mathematic analysis can never substitute the role of theoretical fundament for the formulation of research questions and the interpretation of data. To reach a true understanding of demographic phenomena, we require theories, conceptual frameworks, and models that identify the causal mechanisms underlying the relations between relevant variables (Wunsch, 1995).

Demography traditionally focuses on three main subjects—fertility, mortality, and migration—and of these three, fertility has perhaps the most abundant and comprehensive theoretical foundation. However, even here complaints can be heard that although demography has yielded adequate descriptive instruments, the emerging vision is still insufficient; at least insufficient to be relevant for concrete situations (e.g., Freedman, 1987; Handwerker, 1986; McNicoll, 1992; Ryder, 1983; Willekens, 1990a). This may seriously impede development in areas in which demographic expertise is called on, such as the provision of sound population forecasts (Willekens, 1990b) or the underpinning for efficient and effective population policies (Andorka, 1989; Tsui *et al.*, 1992; World Bank, 1992). These complaints should not, however, disguise the fact that these inadequacies are not typical of demography alone, but represent the situation of social sciences in general. Furthermore, over the years a substantial number of theoretical orientations have entered the realm of fertility studies, leading to new and complementary insights.

I. THE STRUCTURE AND DEVELOPMENT OF FERTILITY THEORY

More than probably any other social science, demography is identified and demarcated by its subject matter—population and population change—rather than by conceptual premises of how to study it. The easily quantifiable nature of the major phenomena of interest bolsters up this empiricist bias. Other disciplines—economics, sociology, history, anthropology, biology, psychology and medical science—have invaded the relative vacuum of theoretical substance, justifying the label demography as an *interdiscipline* (Mayone Stycos, 1987). If this represents the status of demographic theory in general, it equally pertains to that of the theoretical body that deals with fertility, although several approaches are firmly grounded within the borders of demography itself. The resulting theoretical landscape of fertility is a colorful and mountainous patchwork of ideas, often without much coherence or substantial cross-fertilization. Leridon (1982) depicted the situation as a *cubist painting*, and in a recent overview of half a century of research into the determinants of fertility, van de Kaa (1996) recounts it as an evolving story consisting of a series of subnarratives from different disciplinary orientations. Each of these disciplinary approaches introduced a different perspective, a specific focus, methodology, level of analysis and assumptions about the mechanisms underlying reproductive behavior. In this respect, Hauser and Duncan's observation that demographic inquiry did not yield a principle coherent body of knowledge to explain the discipline's phenomena of interest is still valid today (Hauser and Duncan, 1959). A *grand unifying theory* remains far beyond the capacities of the discipline (Freedman, 1987; van de Kaa, 1996; Schofield and Coleman, 1986; Wunsch, 1995) and, for that matter, of any social science.

However, the inclination to adopt insights from neighboring sciences to study the determinants of fertility has not proven to be a guarantee to keep up with the developments at the frontiers of knowledge in these fields (Greenhalgh, 1996; McNicoll, 1992). The application of micro-economics theory has largely stuck to the new-home economics of 15 to 20 years ago and there seems to have been little development in psychological theorizing in demography since the emergence of the *value-of-children* concept in the 1970s and the early applications of the Fishbein-Ajzen type of attitude models. The notion of culture in demographic studies is notoriously reduced to austere proportions—if it is conceptualized at all—and stands a long way from the evolution of cultural understanding in the last decades (Greenhalgh, 1995b; Hammel, 1990; Handwerker 1986). New institutional approaches, social learning theory, and cognitive sciences in general have hardly touched the work of demographers or have only recently entered the field (Burch, 1980; McNicoll, 1992). Some theoretical traditions—like phenomenology, holistic anthropologic approaches—have never made it into the canonical streams of demography because they stand too far apart from the starting points of objectivity and quantification that are valued in demographic inquiry.

Nevertheless, theoretical developments in other social and behavioral sciences had their share in the shifts that occurred in the main orientations of fertility research. A prominent example is the application of economic decision models to fertility since the 1960s in the wake of the general expansion of the economic principles of consumer choice into other life domains. But other forces also operate on the direction of fertility research and its conceptualization. Thus, the availability of a huge amount of data after the series of the World Fertility Surveys (WFS) and the Demographic and Health Surveys (DHS) has been an important factor in the popularization of Bongaarts' model of proximate determinants of fertility. Partly also, demography followed the fashion of the day in scientific and policy circles: the prewar flirt with—and subsequent hasty abandonment of—eugenics, the attachment to the family planning movement, the nexus between population and environment and, especially since the run up to the 1994 International Conference on Population and Development in Cairo, the issue of reproductive health. Several authors have dealt in this respect on the influence of governments, pressure groups, scientific journals, funding agencies, and intellectual and institutional backgrounds (e.g., Caldwell, 1996; Caldwell and Caldwell, 1986; Demeny, 1988; Greenhalgh, 1996; McNicoll, 1992; Szreter, 1993).

In spite of the disciplinary variety in the approaches to fertility and the various orientation shifts over time, explaining the differences and changes in patterns and levels of fertility continues to be a central concern for demographers. In this respect, the theory of demographic transition offers a paradigmatic framework in which each of the theoretical contributions can be viewed as part of the total explanation of the generally observed trend from high to low fertility (Beaver, 1975; van de Kaa, 1996; Kirk, 1996). To organize a review of theories and models of fertility, one dimension that may therefore be used is the relevance of these theories for the explanation of fertility in the different stages identified in the perspective of demographic transition. Thus, the model of proximate determinants and its underlying concept of *natural fertility* bears particular relevance for the situation in (pretransition) historic populations and in many contemporary developing countries, whereas psychological choice models are fairly limited to the explanation of fertility in (post-transition) developed countries, and diffusion approaches seem to apply to the transition stage in between. Another, and perhaps even more important, criterion to classify theories and models of fertility is their emphasis on macro- or microlevel explanation. To provide a full understanding of fertility and fertility change, we must encompass both the structural determinants of the embedding context and the role of individual and intraindividual processes, and, consequently, the mechanisms that relate macro- and microlevels (de Bruijn, 1999; Greenhalgh, 1990; McNicoll, 1994; Wunsch, 1995).

The subsequent sections will address the major lines of thought in current demography as far as they are relevant to the study of fertility. We will leave aside strict mathematic demographic theories—stable population theory, the work of Lotka, model life tables, and so forth—and theoretical considerations in other disciplines that deal with the analysis of fertility—such as economics and anthropology—but remain outside mainstream thinking in demography.

1. Malthusian Thoughts and Population Theory

Ideas and observations with regard to fertility and population can be traced back to classic antiquity and ancient Chinese philosophers. Although these considerations foreshadowed the development of many principles of population, population growth, and development, and often addressed the implications for public policy, they remained largely speculative and at a low level of generalization (United Nations, 1973). The writings of Malthus at the end of the 18th and beginning of the 19th centuries are usually credited with being the first systematic account of the principles of population change. Malthus' body of thought was clearly rooted in the political, economic, and social issues of his time. The core of his first *Essay on the Principle of Population*, published in 1798, is the idea that the capacity of human populations to reproduce is (in principle) unlimited and proceeds with a geometric ratio, whereas the capacity to produce the means of subsistence is necessarily limited and increases at best in arithmetic fashion (Malthus, 1976). In subsequent versions of his *Essay*, Malthus further developed his theory and added various considerations.

The belief that population and means of subsistence increase with different ratios means that populations will grow where food production increases, but that eventually they will face a level where no more people can be sustained and any surplus population will die of starvation. Whereas Malthus maintained that populations always tended to a maximum increase, they were limited by the *positive checks* of mortality. These positive checks not only included famine and starvation, but also other *misery*, such as epidemics, wars, and plagues. In addition, he conceived of a number of *preventive checks*, which operated through people's voluntary acts to limit their number of children. The major principle Malthus saw in this respect—although it surprisingly only appeared in his later writings—was deferring marriage or refraining from marriage altogether. The other *moral restraint* he acknowledged—sexual abstinence within marriage—was considered ineffective because of the overpowering *passion between the sexes*. Malthus' world view (particularly as that of a clergyman) was violently opposed to other means of birth control, like abortion and contraception.

Since its inception, the Malthusian project has been often and severely criticized on empirical as well as ideologic grounds, and indeed it is still the subject of heated debate among scientists and policymakers (Rothchild, 1995). One of the main lines of attack concerns Malthus' assumption of the capacity to increase agricultural output. Preceded by many others, Boserup's orientation is a main contender of Malthusian theory. She advances that technologic progress might keep food production ahead of population increase (e.g., Boserup, 1981; see also J. L. Simon, 1977). In her view, population growth, and particularly increasing population density, is a main stimulus to innovative techniques in agriculture. Moreover, she reverses the Malthusian logic by suggesting that technologic progress only occurs under pressure of population growth. In other publications, Boserup elaborated on the intricate links between technology and fertility, asserting that modes of production have far-reaching effects on marriage patterns, gender relationships, and reproduction (Boserup, 1970, 1990).

Another shortcoming of Malthusian reasoning is its failure to distinguish the power of populations to increase from their tendency to do so. Many anthropologic studies (e.g., Bledsoe, 1990; Kreager, 1982, 1986; Howell, 1979, 1986) have shown that populations have a large variety of mechanisms at their disposal to keep population size in accordance with the carrying capacity of their environment, ranging from marriage patterns and migration to contraception and child fostering. However, this is an adaptation of the Malthusian scheme of balancing population and subsistence rather than a rejection, and in this respect Malthus' ideas remain an important background of fertility analysis.

This is also the case with regard to public policy and developmental issues. Despite the recognition that technologic innovations keep on pushing up the limits of population growth, the essence of the problem continues to permeate the debate, not only with regard to the food supply, but also in the broader perspective of sustainable development with regard to the exhaustion of non-renewable sources of energy (Meadows *et al*., 1972), global warming (Flavin, 1989), water scarcity (Falkenmark, 1990), soil erosion (Brown, 1989), global deforestation (Myers, 1990), and so forth.

2. The Theory of Demographic Transition

For most of this century, demographers and social scientists have been intrigued by the regularities of demographic change in many different settings all over the world. These regularities were considered so remarkable that their occurrence spurred the development of the major body of conceptualization available to demography. The apparent process of *demographic transition* proceeds in the course of modernization and economic development from a situation characterized by high mortality and high fertility to one where mortality and fertility are low, via a stage with declining death rates and declining birth rates lagging behind. This notion of demographic transition gained full momentum only after the seminal publications by Davis (1945) and Notestein (1945), although the full essence of the contingency between modernization and declining mortality and fertility, as well as the three-stage evolution had already been comprehensively formulated by Thompson in 1929. The major elements had also been addressed by Landry (1909, 1934) and in the ethnographic wealth of the work of Carr-Saunders (1922, 1964/1936).

The classic representation of the demographic transition, as for instance sketched by Notestein, claimed that mortality declined in the wake of the industrial revolution, which brought material changes in the sense of agricultural innovation, better communication, higher productivity, and improved health conditions. Fertility was much less responsive to such modernization and its decline depended to a large extent on the collapse—following mortality decline—of ideational and normative systems that supported high fertility.

In the past decades, what started out as a mere description or explanation of historic trends of mortality and fertility in Europe has become increasingly elaborated and has incorporated additional considerations, like different conceptualizations of modernization and the shift from socioeconomic to cultural–ideational and psychological determinants of fertility decline. The principles of historic demographic transitions were thought to be applicable to any contemporary situation in the sense that every nation, region, or population could be located on the evolutionary track of modernization and mortality and fertility decline. Other notions that transformed the concept from empirical observation to theoretical assumption posit that a substantial mortality decline invariably precedes a major decline in fertility, that the mortality decline is followed inevitably by reduced fertility, and that once a substantial fertility decline has been established, the process is irreversible and inescapable. In these respects, the notion of demographic transition has increasingly been considered as a theory with universal validity and predictive power. It is in this respect worthwhile mentioning that Dudley Kirk, another demographer who contributed to the initial formulation of the demographic transition concept, claims that neither Notestein nor Thompson initially thought of their ideas as a theory (Kirk, 1996).

In 1973, Ansley Coale stated that

> The power of the demographic transition concept . . . lies in the undeniable fact that with sufficient modernization fertility and mortality change in a predictable manner.

But the weakness of the concept, according to Coale, is

> The difficulty of defining a precise threshold of modernization that will reliably identify a population in which fertility is ready to fall (Coale, 1973, p. 64).

Thirteen years after Coale's remarks in what can be considered the last monument in the tradition of the demographic transition theory, Chesnais had to phrase the strengths and limitations of the theory in almost identical terms (Chesnais, 1986). Meanwhile, Coale and his associates from Princeton University had tried by means of a large-scale survey to identify the crucial variables that had determined the onset and pace of Europe's fertility transition. Their attempt failed in the sense that their study could not find any socioeconomic indicator of modernization that could unequivocally explain the occurrence of fertility decline in Europe (Watkins, 1986). Socioeconomic factors, which were emphasized by transition theory, appeared to be either spurious or inconsistent in the explanation of the timing of the decline or its tempo The simultaneous fertility decline in Hungary and England is a case in point, as in terms of socioeconomic indicators, Hungary lagged far behind England, at that time economically the most advanced nation in the world. With regard to contemporary societies, China, Kerala State, and Sri Lanka may also serve as illustrations, because there fertility is near or even below (Kerala) replacement level without meeting the assumed requirement of socioeconomic development. Bangladesh, one of the least-developed countries in the world, may be another good candidate, given the significant fertility reduction observed there. Evaluating the contemporary record on the onset and pace of the fertility transition, Bongaarts and Watkins (1996) found an enormous diversity in each of the socioeconomic indicators applied, confirming the conclusions of the Princeton study of historic Europe.

The presupposed sequence of mortality and fertility decline has also been called into question. In a historic perspective, France is the classic example of analysis contradicting the idea that mortality decline preceded fertility decline (E. van de Walle, 1978). Evidence from the Princeton study suggested that France was not an isolated case, and there is no statistical evidence for a general trend in the sequence of mortality and fertility decline (F. van de Walle, 1986). Sometimes fertility decline was found to follow a fall in infant mortality, sometimes it preceded it, and often they dropped simultaneously. Chesnais (1986), however, argues that some of these findings are based on methodologic inadequacies and in second instance results appear to be in accordance with the theory. Nevertheless, the exact causal relation between the decline in mortality and fertility remains difficult to establish (van de Kaa, 1996).

Although the theory of demographic transition suggests a historic perspective, the construct is largely devoid of time and change except for the intermediate stage of transition (Greenhalgh, 1995b). In the framework of transition theory, there is no history in either the pre- or the post-transition stage: In terms of development, time stands still. According to Greenhalgh (1994), many descriptions and analyses of fertility within the perspective of transition theory exist in historic vacuums and are not guided by the notion that the specific histories of the social environment can bear much relevance to their fertility patterns (McNicoll, 1994). The suggested homogeneity and immobility of *traditional* societies—historic or contemporary—is indisputably refuted by the empirical demographic record, which shows a large variety of fertility patterns and levels (Blake, 1985). Neither could transition theory adequately cope with the significant (although temporary) posttransition, postwar reversal of fertility trends in a number of Western countries, which resulted in the *baby boom* cohorts. To dispose of such phenomena as *temporary blips* (Kirk, 1996) is totally unsatisfactory and unacceptable given the huge impact they can—and do—have on society. Also, to find solace in an equilibrium or homeostatic framework, as Kirk suggests, to take account of the problem of where the fertility decline will eventually end, probably reflects more normative hope than a realistic forecast. The observation that in most European countries fertility dropped below the level of mortality—where transition theory assumed it to end—and a supposed historic inflection point in European society, inspired the conceptualization of a *second demographic transition* (Lesthaeghe and van de Kaa, 1986; Lesthaeghe and Verleye, 1992; van de Kaa, 1987).

Greenhalgh, among others, rightly claims that many of the assumptions of transition theory closely resemble those of evolutionary theories that formerly featured in anthropology and raise similar contempt with regard to overdue generalization, Eurocentrism, and supposed unilinearity of development (Greenhalgh, 1989, 1995b; Handwerker, 1986). Although the concept of demographic transition (or its fertility transition subset), if seen from sufficiently afar, stands better against the empirical record than comparable evolutionary thought in anthropology, closer scrutiny supports this criticism: There seem to be many roads to lower fertility, and onset and pace of the decline cannot be predicted anywhere near satisfactorily. This confirms the truth of Greenhalgh's statement "that the closer we get to understanding specific fertility declines, the further we move from a general theory of fertility transition" (Greenhalgh, 1990, p. 85). Therefore, the practical content of transition theory—either in terms of revealing the specific determinants of demographic change or in terms of providing tangible handles on population policies—is less than it initially promised. Szreter in this respect reflected that

> The principal [sic] virtue and function of the idea of demographic transition has always been in providing a graphic metaphor that summarily describes—and predicts—a long-term overall emergent pattern of change. As such it has enormous justification, motivational, and communicative value for agencies and institutions wishing to effect change. But, . . . a summary description of this metaphorical sort offers no necessary assistance or insight into the causal explanation of how such change occurred or occurs in any particular case (Szreter, 1993, p. 692).

What we need to explain and predict population development or design population policies is specific knowledge of particular settings and the mechanisms of social change and structure–agency interaction (Greenhalgh, 1988, 1990; McNicoll, 1992, 1994; Szreter, 1993; Teitelbaum, 1976; Willekens, 1990a, 1990b) and this cannot be adequately provided by the transition framework.

3. Biology and Proximate Determinants

A notion often associated with the transition from high to low fertility is the idea that in the posttransition stage, fertility is under complete control of couples and individuals, whereas in the pretransition stage it is to a large degree left to biologic principles, although constrained by socially constructed bounds. This notion was already conceived in the work of Landry, but the analysis of fertility in pretransition populations has immensely improved since Louis Henry's (1953) development of the concept of *natural fertility*. Henry defined this concept as fertility that existed in the absence of deliberated control through abortion or contraceptive practice, implying that reproductive behavior does not depend on the number of children already born to a couple. In these natural fertility situations, reproduction is determined by biologic principles, such as age at menarche, fecundability (the monthly probability of conception), time required for gestation, intrauterine mortality, and postpartum amenorrhea. In addition, fertility is determined by a number of social–behavioral factors, which are—at least from the point of view of the couples concerned—not intended to restrict childbearing. These factors might include marriage patterns (in particular as far as related to marital duration), spousal separation, (religious) rules for sexual abstinence in certain periods, and duration and intensity of breast-feeding, with its effects on the period of postpartum amenorrhea. The observed levels of natural fertility differ widely between societies especially because of the wide differentiation in these social mechanisms (Blake, 1985). It is, however, not always easy to establish whether behaviors are socially or individually determined and whether birth control considerations are involved. Thus, although people may want to abstain from sexual intercourse on the basis of normative rules, the efficacy of this may depend on motives of child health, birth spacing, and, for that matter, limitation of offspring (Caldwell *et al.*, 1982; Kakar, 1989; Knodel, 1983).

The concept of natural fertility has evolved over time (Henry, 1957, 1961; Leridon, 1977; Leridon and Menken, 1979) and has found its culmination point in the model of proximate determinants as developed by Bongaarts (Bongaarts, 1978; Bongaarts and Potter, 1983). Earlier, Kingsly Davis and Judith Blake provided a seminal contribution with the development of an analytical framework of intermediate determinants of fertility that affected either the exposure to intercourse or the exposure to conception or gestation and successful parturition. Divided over these three categories they identified eleven behavioral and biologic factors "through which, and only through which, any social, economic and environmental variable can influence fertility" (Davis and Blake, 1956, p. 214). Bongaarts further developed this framework by quantifying the effect of Davis and Blake's intermediate variables and collapsing them into eight, and later seven, proximate determinants of fertility. This resulted in a simple but powerful model for analyzing how fertility changes over time or differs from one group to another: Any level of fertility in a population can always be traced to variations in one or more of the following determinants:

1. The proportion of women of reproductive age that is married (as a measure of the proportion exposed to sexual intercourse)
2. The use and effectiveness of contraception
3. Induced abortion
4. Postpartum infecundability (as primarily determined by the duration and intensity of breast-feeding)
5. The frequency of intercourse (including the effect of temporary separation and abstinence practices)
6. The onset of permanent sterility (particularly as related to menopause)
7. Spontaneous intrauterine mortality

Each of these factors contributes to a reduction of the approximately 15 children a woman can have during her reproductive career. The empirical evidence showed that marriage, contraceptive practices, abortion, and postpartum infecundability have by far the strongest effect on levels and differentials of fertility (Bongaarts, 1993; Bongaarts and Potter, 1983). The model suggests therefore that the total fertility rate can be described as:

$$TFR = C(m) \times C(c) \times C(a) \times C(i) \times TF$$

where TFR is the total fertility rate, C(m) is the index of proportion married, C(c) is the index of noncontraception, C(a) is the index of induced abortion, C(i) is the index of lactational infecundity, and TF is the potential total fertility.

The framework's exceptional clarity and organizational power had an enormous impact on the research agenda of fertility studies—particularly for developing countries, but also for historic populations—and the WFS and DHS provided the necessary data to apply the model in a comparative perspective. The significance of the model is partly situated in the structuring of attention and efforts in the search for the ultimate determinants of fertility and fertility change.

Fertility itself is no longer the sole subject of research; in addition we need to look for the institutional and behavioral backgrounds of marriage, contraceptive use, breast-feeding, abstinence practices, and so forth or, to cite Freedman, we are faced with "the challenge of specifying the determinants of the proximate determinants" (Freedman, 1986, p. 30; Hull, 1983; Leridon, 1982). Whereas any such attempt was completely absent in the work of Bongaarts, Davis and Blake in fact used their framework of intermediate variables as a starting point to determine and analyze the institutional factors affecting fertility. Their explanation of fertility rested on the comparative analysis of social organization, which largely boiled down to an explanation in terms of family and kinship organization.

Whereas Bongaarts' model of proximate determinants, as well as the version of Davis and Blake, analyzes fertility at the level of populations and societies, several researchers tried to translate it to the individual level. Hobcraft and Little (1984), for instance, calculate fecundity and fertility as the outcome of the fecundity-reducing effects associated with the particular set of states that describe women's positions in their reproductive career (states related to pregnancy, absence from sexual relations, contraceptive use, post-pregnancy infecundity). Becker's model of adolescent fertility (S. Becker, 1993) specifies conditional probabilities of live birth, conception, and coitus on the basis of individual data. Hull (1983) and de Bruijn (1999) explicitly incorporate the proximate determinants in a decision-making approach. The value of this integration is that fertility is not seen as the product of one single decision, but as the possible combined effects of numerous decisions with regard to the—behavioral—proximate determinants such as marriage and divorce, contraceptive use, abortion, frequency and patterns of sexual intercourse, and breast-feeding practices. This reformulation represents a means to increase the relevance of the concept of individual choice for situations under conditions of natural fertility.

Although the popularization of the proximate determinants model is mainly due to its application to developing countries, it is perhaps a mistake to presume that its value may be discarded completely for developed countries (Easterlin and Crimmins, 1985). Although in Western countries fertility is considered to be under volitional control and childbirth to be a matter of demand rather than supply, it is worthwhile recognizing that here too fertility not only depends on behavioral factors, but also on biologic processes. Until recently the main fertility problem was the control of unwanted childbearing and the attainment of a *perfect contraceptive population* (Bumpass and Westoff 1970). In this respect Menken *et al.*, reflected that

> With great effort, fertility has been "turned off": People had come to believe that controlling fertility was the real problem and to expect that having children was easy (Menken *et al.*, 1986, p. 1393).

With respect to the ease of having children, they indicated the growing concern in both the popular and scientific literature with problems related to infecundity and infertility in Western countries, which have become increasingly evident with the generally observed rise in age at birth (see also te Velde, 1992).

4. Economic Theories of Fertility

The writings of Malthus left their mark in the field of classical and Marxist economics in the 19th and early 20th centuries, and later in the economic analysis of population and development in less-advanced countries, particularly in view of their unprecedented rates of population growth (United Nations, 1973). These macro-economic approaches never became standard material in demographic theorizing, unlike the micro-economic orientations that firmly entered the field in the 1960s.

Harvey Leibenstein (1957) may be called the progenitor of the view that the number of children is the result of individual decision making within an economic context of income and prices. Among others, Nerlove (1974), Willis (1973), Schultz (1981), and, most prominently Gary Becker (G. A. Becker, 1960, 1965, 1976, 1981, 1991; Becker and Lewis, 1973) developed the consumer choice theory into what became known as the new home economics of the Chicago school. This micro-economic approach not only involves the traditional variables of income and prices, but also the *quality of children* and budget constraints in terms of allocation of time and opportunity costs. Given these variables, households are assumed to produce a bundle of consumer commodities—including children—in accordance with the maximization of household utility. The model thereby links fertility decisions to other household decisions, including labor force participation and consumption. The notion of *child quality* became a key factor in Becker's work to account for the inverse relation between income and number of children as experienced in the fertility transition. The quality of children is assumed to be elastic with respect to income, whereas the quantity of children is not. This implies that the desired number of children may fall as income increases because the average cost per child may increase even faster.

The economic approach to fertility has been challenged on several grounds. Part of the criticism can be traced back to the concepts of choice and decision-maker that underlie micro-economic theories: strongly individualistic, decontextualized, static, relying on a narrow, substantive notion of rationality, and without a sufficient degree of (psychological) realism (de Bruijn, 1999). Obviously, representatives of rival disciplines embark upon such general criticism, but it is also voiced by a number of (behavioral and institutional) economists themselves (e.g., Simon, 1987; Lea *et al*., 1987; North, 1994). Among the economists working in the field of demography, Leibenstein and Arthur share some of these concerns. Arthur, in a critical review of Becker's *Treatise of the Family*, "call[s] for the use of rules, rights, agreements, hierarchies, organizational institutions—in short, structure" (Arthur, 1982, p. 395). These remarks touch upon the neglect by most economists of the social, cultural, and political environment of decision making. Leibenstein (1977, 1981, 1982) articulates that the concept of choice as used in economic choice theory is only selectively applicable in the study of reproductive behavior, and much of the fertility outcome must be seen as the result of routine and rule following procedures. With regard to the static nature of the new home economics, it can be observed that by and large economic analysis in demography does not allow for changes in preferences over lifetime as the result of learning and personal experience, and it assumes couples to have defined these preferences at the onset of marriage. Changes in behavior over time are in this perspective considered to be the result of variations in restrictions facing the decision-maker. However, a number of economists provide a more dynamic perspective by applying a life-cycle approach to fertility or accept the possibility of preference shifts (e.g., Moffit, 1984; Namboodiri, 1980, 1983; Rosenzweig and Wolpin, 1980; Siegers, 1987; Turchi 1991).

Although crucial to Becker's interpretation of fertility, the notion of child quality is not free from controversy, both in terms of conceptualization (e.g., Robinson, 1997) and with regard to the assumption that all children born in a family embody the same quality. The last assumption is refuted by the empirical findings that the value of children may differ by parity (Bulatao, 1981; Bulatao and Fawcett, 1981; Namboodiri, 1983) and by sex (e.g., Miller, 1981; Koenig and Foo, 1995; Nag, 1991; UNICEF, 1991). Criticism of the new home economics approach to fertility also refers to the assumption of a single joint household utility function. For one thing, it assumes altruism (with which Becker [1981] tries to deal) and harmony on behalf of both partners, which does not seem very plausible as both power and the distribution of costs and benefits of children may differ substantially between them (Caldwell and Caldwell, 1987; Lee and Bulatao, 1983; Fawcett, 1983; Simmons, 1985). Economic decision-making models are usually completely silent about the bargaining processes that settle possibly conflicting interests with regard to progeny. In several societies, even the absolute authority on fertility decision-making by the parents themselves may be seriously questioned (Khan and Singh, 1987; Koenig and Foo, 1992; Ryder, 1983). Furthermore, the definition of the household concept is notoriously difficult, not only with regard to relevant criteria, but also because of possible shifting and rotating membership of both parents and children. The widespread phenomenon of child fostering, especially in West Africa (e.g., Bledsoe, 1990, 1995; Page 1989), puts the standard economic view on direct and substitution costs of raising children in a different perspective. For these reasons, and because in many instances fertility is not a matter of demand but a matter of supply, the new home economics model is still relatively ill-suited for the analysis of fertility in developing countries (Simmons, 1985).

With regard to the neglect of supply mechanisms in the standard consumer choice theory and its new home economics version, a very important extension is provided by Richard Easterlin and collaborators (Easterlin, 1975, 1978a, 1978b; Easterlin and Crimmins, 1985; Easterlin *et al*., 1980). Easterlin's approach complemented the strictly demand-oriented model of new home economics with notions concerning the production side of fertility as derived from the intermediate variables framework of Davis and Blake and that of Bongaarts and the concept of natural fertility. In this synthesizing effort, he brings together the scientific paradigms of economics and sociology; but this effort clearly should also be considered as an attempt to bridge the gap between fertility analysis in contemporary Western situations and that under non-Western or historic conditions. This *Easterlin synthesis* has gained wide acceptance in demographic research and provided the basis of the agenda of the U.S. National Academy of Sciences (Bulatao and Lee, 1983). The model assumes that all determinants of fertility—public health, education, urbanization, family planning programs, and so forth—work through the categories of the demand for children (depending on household tastes for children and alternative goods, income, and costs and benefits of children), the supply of children (reflecting natural fertility determinants like exposure and frequency of intercourse, postpartum amenorrhea, spontaneous intrauterine mortality, and sterility) and the costs of fertility regulation (lumping together attitudes toward and access to fertility control methods and supplies, as well as the time and money required to obtain the birth control methods). Motivation to limit fertility only occurs if the supply of children exceeds their demand and the greater the excess of supply over demand, the greater this motivation. But the actual restriction of fertility also depends on the (economic and psychological) costs associated with various birth control methods. A modification in the specification of the demand function compared with the Chicago approach is that Easterlin's model incorporated variable and endogenous preferences, which allowed the explanation of difference in behavior on the basis of opportunities, but also in terms of differences in *tastes*. Furthermore the final dependent variable in Easterlin's model is not children, as much as children surviving to adulthood, thus making not only the supply of children endogenous to the explanatory framework, but child survival as well.

Easterlin contributed another important theoretical proposition to the economic perspective of fertility behavior. This contribution (Easterlin, 1978b, 1980) provided a dynamic element to the analysis of fertility and stands out as the sole fully developed cohort theory in demography. The *Easterlin hypothesis* concerns the adaptivity of fertility preferences (and subsequent fertility behavior) to the realization of a preconceived lifestyle. It asserts that an individual's desired lifestyle is molded by experience during the formative period in adolescence at the parental home. The degree of affluence promised by (male) labor market perspectives during the reproductive period determines the number of children that can be sustained while maintaining the standard of living that was experienced during the formative period. The labor-market or income-earning opportunities, in turn, are assumed to be negatively related to cohort size. This implies that the *tastes* for children or reproductive goals are not given, as assumed in traditional economic theory, but formed during the experience of the income effects caused by the entry of differently sized cohorts on the labor market.

Ní Bhrolcháin (1992) disputes the role of cohort effects in fertility. Although she recognizes the evidence of cohort mortality, which finds convincing foundation in epidemiologic and medical research (e.g., Barker, 1992; see also Caselli, 1990), she is unable to find similar evidence for fertility. Similarly, Wright (1989) has tested the Easterlin hypothesis for 16 European countries but has failed to find a statistically significant correlation between relative cohort size and fertility. From a review of empirical studies, Pampel and Peters (1995) conclude that the evidence for the Easterlin effect proves at best mixed and at worst completely wrong and that the degree of support varies across time periods, nations and level of measurement.

Apart from a number of technical difficulties involved in the measurement of the effect, there are also some conceptual inadequacies. For instance the theory focuses too much on men's roles in the labor market—whereas new home economics emphasizes the role of women—and there is some discussion about whether Easterlin's index of relative cohort size is a good indicator of relative income (Wright, 1989, p. 118). Furthermore, the theory's conceptualization of socialization is rather plain: The influence of personal experience operates from a distant and rather fixed past and neglects the influences of changing social environments apart from cohort size. Exclusive emphasis on the effects of the early socialization experience on a person's aspirations amounts to the neglect of the effect of peer group influences in new environments or experiences in adulthood in general (Freedman, 1987; Namboodiri, 1980). Moreover, there are other lifetime-specific experiences, such as education and labor force participation, which can have decisive impact on (later) considerations in decision making, for instance with regard to goal setting, dependency of others, or the value of children. Lastly, Easterlin's theory needs to specify better the mechanisms through which relative income influences motivation for fertility. Pampel and Peters (1995) suggest that if a number of additional conditions are met, the Easterlin effect still might emerge.

5. Psychological Approaches to Fertility

Compared with the impact of economics, sociology, anthropology, and biology, the contribution of psychology to demographic theory has been very limited (Burch, 1981; McNicoll, 1992b). The encounter between demography and psychology witnessed the application of two main theoretical approaches: the *value-of-children* approach and applications of psychological value-expectancy models.

The first line of thought can be traced back to a Maslowian perspective on motivation. Taking this as a starting point, Hoffman and Hoffman (1973) constructed a conceptual framework that depicted the way in which children could contribute to satisfying a number of material, social, and intrinsic needs. According to this framework, childbearing motivation depends on the evaluation of these satisfactions and the economic and noneconomic costs of children. The associated perceived value of children appears as an intermediate variable in the explanation of the relation between socioeconomic, cultural and gender aspects, and fertility behavior. The framework was elaborated and operationalized in the value-*of-children project* in the early and mid 1970s by James Fawcett and others (Fawcett, 1972, 1983; Fawcett and Arnold, 1973; Fawcett *et al*., 1972). Survey data were gathered in different countries, and subsequent analysis compared data for different countries and regions within countries at different levels of socioeconomic development. In this way the model tried to underpin the propositions of the theory of fertility transition, confirming some of them, but remaining undecided on others (Fawcett, 1983; Bulatao, 1982). It did suggest an evolution during socioeconomic development from economic and material considerations with regard to children, to more emotional rewards and psychological appreciation, which induced the introduction of the concept of the *transition in the value of children*. Whereas most applications of the model were carried out within the international comparative framework of the value of children project, several individual studies also relied on it (e.g., Niphuis-Nell, 1981). The value of children approach produced a substantial body of literature especially in the early 1980s, but has not been pursued much since then. Perhaps this is due to the fact that the research has provided relatively few generalizations about how background variables influence the perceptions of satisfactions and costs of children in order to affect fertility preferences and behavior (Fawcett, 1983).

Value-expectancy models like *expectancy x value* and, notably, the Fishbein–Ajzen model of reasoned action (Fishbein and Ajzen, 1975; Ajzen and Fishbein, 1980) were introduced into demography with regard to the field of fertility, contraceptive use, and female labor market participation (e.g., Bagozzi and Van Loo, 1991; den Bandt, 1982; Fishbein, 1972; Jaccard and Davidson, 1976; Moors *et al*., 1989; Rosenstiel *et al*., 1982; Wijsen, 1994). The Fishbein–Ajzen model states that the intention to perform certain behavior is a reliable indicator of the performance of that behavior. In turn, this intention can be assessed by measuring beliefs with regard to consequences of the behavior and the valuation of these consequences on one hand and perceptions of the opinions of others in combination with the importance attributed to these opinions on the other. An advantage of the model is that it takes into account to a certain extent the influence of the social environment by including a normative component in terms of the opinion of the important others.

To some extent, psychological approaches share the problems associated with micro-economic theories. These concern particularly the volitional character of individual behavior formation and the less than adequate incorporation of the institutional backgrounds determining that behavior. The theoretical framework underlying the value of children framework, and more explicitly the one underlying the psychological value-expectancy models, assumes deliberate choice and purposeful action, which do not always seem to be the most appropriate assumptions given the supply mechanisms of natural fertility, the lack of effective knowledge, the possibility of overriding social pressure in favor of childbearing, and the possibly restricted availability of and access to the means to control it. Ajzen and Fishbein themselves, in fact, recommended their model especially for situations where individuals have strong control over the factors associated with decision making, which is more applicable to developed than to less-developed countries. In this respect, the incorporation of the concept of self-efficacy (Ajzen, 1991; Bagozzi and Van Loo, 1991) may improve the explanatory power of psychological choice models. Other options to increase the relevance of decision making approaches to the situation in developing countries—although at the expense of quantification—relate to a broader definition of the concept of choice: either by extending decision-making analysis to the proximate determinants of fertility or by incorporating the *process* of choice in the sense of processes of attention and (routine and institutionalized) styles of decision making (de Bruijn, 1999; Bulatao, 1984; Shedlin and Hollerbach, 1981; Hull, 1983; Leibenstein, 1982). Such reformulations at least avoid the unjustified distinction between the concept of the person who populates pretransitional societies and who is cast in terms of passive recipients of culture and normative rules and guided by irrational beliefs, and the one who inhabits posttransitional societies and who is viewed as an active, independent decision maker with the capacity for rational thought and behavior (Blake, 1994; Greenhalgh, 1994).

6. Diffusion of Ideas and Technology

Over the years, the theory of demographic transition has incorporated a number of additional ideas to remedy some of its shortcomings. Many of these efforts were related to the failure to define socioeconomic development as the crucial variable of demographic change. The incorporation of some concept of culture—especially perceived as a principle involved in the spread of ideas—seemed to provide a promising alternative. In its wake, the concept of diffusion entered the theoretical edifice of demography (Cleland and Wilson, 1987; Retherford and Palmore, 1983; Watkins, 1986, 1987), or better, was reinstituted as a major explanatory strand.

Diffusion can be understood as the process by which innovations spread from one locale, social group, or individual to another (Retherford and Palmore, 1983). The spread of ideas, behaviors, and techniques has often been found to follow the grooves laid down by sociocultural forces, such as language, ethnicity, neighborhood, and workplace or channels of communication and exchange. Thus, Lesthaeghe's study of fertility decline in Belgium showed a clear demarcation of fertility patterns and levels along language borders for communities, which were otherwise—socioeconomically—very similar (Lesthaeghe, 1977). Kirk attributes the early transition in countries such as Hungary and Bulgaria to their location along the Danube as a prime artery of communication and commerce (Kirk, 1996). Various others point to migration avenues, network channels, and institutionalized lines of communication for the spread and explanation of family planning acceptance (e.g., Bongaarts and Watkins, 1996; Freedman, 1987; Cleland, 1987; Watkins, 1987, 1989). Entwisle *et al.* found evidence of the importance of conversational networks to directing and controlling the flow of information about contraceptive methods in rural communities in Thailand (Entwisle *et al.*, 1996). They associated the variety of method acceptance between communities and the typical predominance of one method within the communities with the structure of largely village-based social networks. The notion of diffusion also focuses attention on the different contextual levels—interpersonal, local, national, and global—involved in the transmission of information (Bongaarts and Watkins, 1996; Hammel, 1990; Montgomery and Casterline, 1996; Retherford and Palmore, 1983).

The reception of diffusion into fertility theorizing brings along its own conceptual and methodologic problems. With respect to diffusion, Greenhalgh cautions against a too reductionistic approach focusing almost entirely on communication about birth control, while neglecting the exchange of a broad scale of perceptions on other issues relevant for reproductive behavior (Greenhalgh, 1995b). To parts of the family planning movement the contribution of diffusion was even restricted to the spread of contraceptive technology (e.g., Ravenholt and Chao, 1974). Pollak and Watkins (1993), like Greenhalgh, argue that diffusion involves more than techniques and mere information. They refer to van de Walle who states that in early 19th century condoms were licit in extramarital liaisons, but not within marriage, and to Bledsoe who reaches a similar conclusion for several African countries. This pertains to the critique that the diffusion approach is problematic in its omission of the context of contraceptive communication, especially with respect to the requirement of its social legitimation and the spread of values (Lesthaeghe and Wilson, 1986; Handwerker, 1986). Related to this critique, many consider the diffusion concept merely a description without notable theoretical content (McNicoll, 1992; Pollak and Watkins, 1993); sometimes it entails just summary statistics that reflect the pattern of spatial and temporal spread of some phenomenon. Hammel's (1990) contribution can be mentioned as an important step to provide more substance to the notion of diffusion, but within demography, of course, Rogers (1962/1983, 1973) must also be referred to, with the work of Montgomery and Casterline (1996), who try to model the social structure of contraceptive diffusion. Still, a fundamental underpinning of the working of the diffusion mechanism—how information is conveyed in the social environment and how the messages are organized and interpreted, and finally lead to behavior—is in the lap of other disciplines such as (cognitive) anthropology, sociology, and, importantly, social learning theory (de Bruijn, 1999).

7. Culture, Structure, and Social Organization

The field of fertility theory covers a large number of interpretations and approaches that are not neatly classifiable as a disciplinary theory. In contrast to, for instance, economic, psychological, and proximate determinants approaches to fertility, they do not have an internally agreed set of variables, concepts, and theoretical propositions nor a strict methodology. Their common feature is the focus on the structural level of society, culture, institutions, or social organization—and often in combination—but otherwise it is a very heterogeneous compilation, its backgrounds largely in the more holistic disciplines of sociology and anthropology. They may involve aspects from theoretical approaches mentioned above, and many of them, in fact, can be interpreted as enrichments or modifications of the theory of demographic transition. Far from claiming to be exhaustive, a list of such approaches to fertility might include system–functionalist ideas derived from Adam Smith's *invisible hand* hypothesis (Wrigly, 1978) and other homeostatic interpretations (Howell, 1979, 1986; Kreager, 1982, 1986; Lesthaeghe 1980), *modes-of-production* paradigms (Boserup, 1970, 1990; Goody, 1976), Lesthaeghe's production–reproduction thesis (Lesthaeghe and Surkyn, 1988b; Lesthaeghe, 1989b), Cain's institutional approach of risk and insurance (1981), Caldwell's wealth flows theory (Caldwell, 1976, 1982), family-oriented analyses (Ariès, 1962; Cain, 1989; Freedman, 1975, 1987; Davis and Blake, 1956; Khan, 1987; Ryder, 1983), the notion of the second demographic transition (Lesthaeghe and van de Kaa, 1986; Lesthaeghe and Verleye, 1992; van de Kaa, 1987) and the recently emerging institutional approaches in the line of Greenhalgh (1988, 1990, 1994, 1995a) and McNicoll (1980, 1985, 1988, 1994).

In disappointment at failing to find the crucial determinants of fertility in socioeconomic indicators, some demographers have turned to culture (Cleland and Wilson, 1987). Research based on the results of the WFS (e.g., Cleland, 1985) and the Princeton study on the European transition experience (Anderson, 1986; Knodel and van de Walle, 1979; Watkins, 1986) signified the importance of cultural factors as the major independent determinants of fertility levels and the onset of fertility decline. These studies, however, did not elaborate on what was exactly meant by *culture*. Greenhalgh (1994) and Hammel (1990), therefore, accuse demographers of a widespread incompetence to conceptualize culture meaningfully. Apart from the work of a small number of researchers, culture has hardly gained any depth; it is usually only grasped in terms of language, ethnicity, or geographic region. Large-scale surveys on which traditional demographic research is based cannot grasp the meaning of culture, and so to many demographers, the concept connotes a *messy bag* (Lesthaeghe, 1989a), which may be assumed to contain all residual explanation. The problem of incorporating culture into theory, however, is not restricted to demographers alone: All social sciences consider culture a notoriously difficult concept, perhaps even more than the concept of social structure (Archer, 1996).

Culture is usually claimed to stand for the shared and intergenerationally transmitted beliefs and evaluations about the world and people's place in it. The role of culture in fertility change is presumed to be particularly located in this feature of transferring values and information within a culturally identifiable group (e.g., Lesthaeghe, 1977). But apart from this communicative feature, culture provides the normative and interpretive or meaning-giving rules with which people consider fertility and its proximate and ultimate determinants. The link between culture as an ideational or meanings system and social organization lies in the common order it provides for the definitions of social relationships and evaluation of individual behavior. Modes of production, intergenerational and gender relations, marriage systems, and so forth are reflected in culture. However, by providing social structure with a meaning, culture also legitimizes and, therefore, (re)produces society. This dualist manifestation has been acknowledged in demographic literature by several exponents of a cultural approach to fertility theory (e.g., Greenhalgh, 1989; Hammel, 1990; Lesthaeghe and Surkyn, 1988a). In the social, political, and anthropologic analysis of fertility, the family is the dominant institution: It is the locus of demand and supply of children, by and large it retains the function of socialization base, and often it constitutes the prime avenue to achieve things that are important in life, such as economic assistance, security, social interaction and status, information, and emotional and political support (Davis and Blake, 1956; Dyson and Moore, 1983; Freedman, 1987; McNicoll 1994).

Several broad lines of thought on the rather elusive concepts of culture and social structure can be distinguished in demography. Most of them are not acknowledged as separate theories, so the discussion here is arranged around the work of the major proponents and publications.

A first mode of thought with social structure as its point of departure is the holistic representation of society as an integrated system of arrangements and mechanisms that are principally geared to the reproduction and maintenance of that system. It can be viewed as a continuation of the Malthusian program, but based on modern research and insights. A milestone publication in this respect is Coleman and Schofield's *The State of Population Theory*. Kreager's work (1982, 1986) may be representative for this approach when he paints an anthropologic vision of demographic regimes as Durkheimian organic solidarities. Population process components—fertility, nuptiality, migration, and mortality—work in a coordinated fashion so as to adjust population size and distribution to the capacities of the environment.

Such studies concentrate on the colorful anthropologic uniqueness of societies, but sometimes tend to over-emphasize the cultural barriers to demographic change (Robinson, 1992; Lockwood, 1995). By their nature they are not infrequently rather static and descriptive, and fail to provide the links of encountered feedback mechanisms to the motivation of individuals to act in line with the long-run societal benefit. Lesthaeghe, for instance, rightly points out that there is no need to assume a latent aim of controlling population growth if other explanations, centering on direct goals, like survival of children or maintaining power relationships, are available to account for it (Lesthaeghe, 1980). On the macrolevel such individual pursuits of goals—importantly shaped by the structure and content of the social environment—may have the effect of maintaining a functional demographic balance in the long run (Blake, 1994).

Explaining demographic patterns only in terms of system needs easily ends in *ad hoc* or tautologic theorizing. The demonstration of the functional quality of certain collective behavior to a population system cannot demonstrate its necessity to be there, at that time and in exactly that form; there may very well be other, functionally equivalent, behavior patterns that could respond to the same motives (McNicoll, 1992; North, 1994). It is easy to interpret some institutional arrangement as valuable if it has found its established place in some situation, but the danger of ex post rationalization is acute: Conceivably under a different arrangement the situation would have been served even better. This reflection led Demeny to his gloomy remark that "the impression is overwhelming that history is a story of unrealised [sic] potentials that could have been within our grasp" (Demeny, 1986, p. 483).

The only line of thought that may claim the status of *demographic theory* and is as yet the only one that can conceptually rival the conventional transition theory (particularly in the setting of contemporary less-developed countries), is Caldwell's wealth flows theory (1976, 1982). Culture, here, is importantly represented by the meaning of kinship and family, traditionally a central issue in the anthropology tradition from which Caldwell originates. The theory states that the level of fertility is primarily imposed by the direction of the net wealth flows between parents and children, which include all the present and anticipated benefits over a lifetime. The outcome of this economic rationale is either maximum or zero fertility, but this is adjusted by the impact of personal, social, and physiological reasons. The commanding principle underlying the direction of intergenerational wealth flows is the social organization of the society, and specifically family structures. Caldwell argues that in all traditional societies the net wealth flow has been from younger to older generations, which means that economic motives promote high fertility. This flow will only be reversed if the economic and emotional primacy is withdrawn from the grip of broader family ties and is focused on the conjugal family. The nucleated family is, therefore, a *sine qua non* for low fertility and the transition from high to low fertility is a product of social change with economic implications, rather than economic change alone.

The primary force behind the transformation of the family is credited to universal education across a nation or cultural group. Mass education changes the values and costs of children within the family and introduces a Western family model into the society. Capitalizing on the general failure to identify critical socioeconomic variables for the onset of fertility decline, Caldwell argues that it is the process of Westernization, rather than economic modernization, which initiates the change toward low fertility (Ryder, 1983). Put to the test (e.g., Dow *et al*., 1994), Caldwell's theory received less support than might have been expected on conceptual grounds. It is not clear to what extent this is due to the severe operationalization and measurement problems—for instance the measurement of wealth flows—that are inherent in the theory's formulation (Schultz, 1983).

A third line of thought, like Caldwell's intergenerational perspective, involves the institution of the family, but with a different nuance. Here, family is primarily the outcome of large-scale historic socioeconomic and cultural processes, rather than the exclusive focus of demographic change. This tradition is most prominently recognized in the work by Lesthaeghe. The gist of his work is best articulated by the attention to the analysis of the first two of Coale's prerequisites for a fertility transition: First that the very deliberation about pros and cons of additional children to the family must be an acceptable form of behavior, and second that perceived social and economic circumstances must make reduced fertility seem an advantage to individual couples (Coale 1973, p. 65). Both preconditions are clearly anchored in the context of ideational systems and social organization, which is formulated by Stamm and Tsui as follows:

> The impact of family-planning technologies on reproductive parameters is a function of the systems of meaning which underlie the reproductive choices and preferences of the individuals comprising a society. Such systems of meaning define both what is and what is not subject to choice and the value of choice options (Stamm and Tsui, 1986, p. 159).

To arrive at reduced levels of fertility, there must be a favorable meaning-giving or ideational environment to direct the preferences and considerations that authorize the legitimacy of individual control over fertility as well as the desirability of smaller family sizes (Lesthaeghe and Wilson, 1986). Falling back on Maslow's notion that, along with development, intrinsic personal needs will become increasingly important, a universal emancipatory tendency of individualization may be assumed to have its effects on fertility behavior (Lesthaeghe, 1983; Lesthaeghe and Surkyn, 1988a). This may lead to a decline in fertility, but only if existing institutions that exert a pronatalist influence lose the legitimacy of their grip on individual decision making, and if socioeconomic conditions are such that the balance of subjective cost-benefit considerations is tipped toward smaller families. Although the complex associations between socioeconomic development and ideational change are explicitly discussed (e.g., Lesthaeghe,1989b; Lesthaeghe and Wilson, 1986), Lesthaeghe stresses the importance of autonomous ideational shifts toward liberal and, especially, secular values. In this respect Lesthaeghe's ideation with social–organization approach is a major modification of the classical notion of demographic transition.

This general scheme has been applied to the situation in sub-Saharan Africa (Lesthaeghe, 1989a, 1989b; Lesthaeghe and Eelensm 1989; Lesthaeghe and Surkyn, 1988b), to historic Western Europe (Lesthaeghe, 1983; Lesthaeghe and Wilson, 1986), and to recent demographic change in Western Europe (Lesthaeghe and Moors, 1992; Lesthaeghe and Surkyn, 1988a), cumulating in the concept of the *second demographic transition* (Lesthaeghe and van de Kaa, 1986; van de Kaa, 1987; Lesthaeghe and Verleye, 1992). Each time, the picture is painted differently, highlighting the specific relevant elements of the historic and institutional background. During Europe's fertility transition this included the nuclear family dominance, the evolution of a capitalistic mode of production, the waning of religious doctrines as guiding principles and general economic growth, fueling individual aspirations (Lesthaeghe, 1983; Lesthaeghe and Wilson, 1986). The second demographic transition can be distinguished primarily on the basis of a marked acceleration of the trend toward self-fulfillment and individual autonomy, bringing about new types of demographic behavior in the sense of new living arrangements and changed timing and prevalence of marriage and childbirth (Lesthaeghe and Verlye, 1992; van de Kaa, 1987). With regard to sub-Saharan Africa, the designation of social organizing principles relies heavily on the intellectual legacy of Boserup (1970) and Goody (1976) with respect to the structuring impact of modes of production on patterns of nuptiality, gender relationships, and progeny; but it also relies on classifications in terms of religious background (especially Islamization and the survival of traditional religions), female education and contraceptive use. The emerging picture reveals the complexity of the effects of interactions between socioeconomic development and social institutions on patterns of starting, spacing, and stopping childbearing (Lesthaeghe and Eelens, 1989).

A promising new line of interpretation is the institutional analysis of fertility. This approach, again, may be viewed as a reaction to the theory of demographic transition. Whereas the classic transition theory searched for general processes (including, at least in its original formulation by Notestein, the role of social institutions) and focused on macrolevel and socioeconomic aspects, the new institutional approach seeks situational and path-dependent specificity, and is sensitive to cultural interpretations and the interaction between structure and agency. The research and analyses of Lesthaeghe and Caldwell clearly largely fall into this interpretive framework. Well-known is also Cain's (1981, 1989) analysis of the value of children as a source of risk insurance in villages in India and Bangladesh, which suggests that the differences between the settings can be largely attributed to institutional elements like labor division between the sexes, patriarchy, legal status, and social security systems. The leading demographers in the field, however, are Susan Greenhalgh and Geoffrey McNicoll. Although Greenhalgh starts out from a political-anthropologic perspective and McNicoll particularly relies on a sociologic and institutional economics background, they are remarkably in unison in voicing the needs for and elaboration of institutional analysis.

Both largely reject the possibility of general schemes of fertility change. Compare, for instance, McNicoll's statement that:

> [F]rom a distance, the process of fertility transition that accompanies social and economic development shows many similarities across major world regions [. . .]. Yet at closer range fertility transitions are idiosyncratic. Their course is influenced by the institutional endowments each society has inherited through its particular historical experience (McNicoll, 1994, p. 2)

with that of Greenhalgh:

> There is no single demographic transition, caused by forces common to all places and all times. Rather, there are many demographic transitions, each driven by a combination of forces that are, to some unknown extent, institutionally, culturally, and temporally specific (Greenhalgh, 1990, p. 88).

A seminal article that contributed much to the attention for the institutional background of fertility and on the micro–macro link of fertility explanation was conceived by McNicoll (1980). Relying on Herbert Simon's concept of bounded rationality, he argued that the options for fertility behavior that are salient to the individual consist of only a selection of all options and that this salience depends on the structure of the information environment that is shaped by social institutions. In McNicoll's perspective, social institutions may be interpreted as the socially constructed (and sanctioned) rules that provide solutions to recurrent problems of individual action and interaction (McNicoll, 1985). This normative character of institutional rules may be complemented with their representational or meaning–giving dimension, which fits better with Greenhalgh's cultural interpretation. Both McNicoll (1994) and Greenhalgh (1995b) view institutions as social constructs that are constantly being made, remade and possibly dismissed in processes of negotiation and individual action. Although neither of them elaborate much on this issue, it may provide a tangible opening toward psychological and economic choice considerations, and thus a unique opportunity to narrow the gap between macro and micro approaches, structure and agency, and context and choice (de Bruijn, 1999).

The identification of interpenetrating local, regional and national institutions reflect the multi-level nature of context. Greenhalgh (1990, 1995) even goes on to the international arena, extending Watkins's remark that in 1870, the relevant community to which behavior pertained was largely local whereas in 1960 it was largely national (Watkins, 1989). Thus, aims of the international community with regard to women's rights and reproductive health as voiced at the 1994 Cairo Conference on Population and Development can be effectuated by supportive legislation at state level and women's organizations at lower levels, but can also be impeded by adverse family and gender systems or local labor market opportunities. Overarching institutions like religion or national family planning programs may be negotiated differently in a rural farming community than in the neighboring fishing community, because of the differences in the local economies (Niehof, 1985). Due to the sustained conjuncture of various institutions in specific social settings, the meaning of the individual institutions may change in the course of time, which is why, for instance Catholicism in Ireland has become different from that in Mexico or Sri Lanka (Handwerker, 1986). An institutional approach finds the understanding of fertility at least partly in the historic evolution of the specific amalgam of institutions: It views them as evolving processes that not only depend on current circumstances, but also, and crucially, on their past history, and, which moreover, evolve at every point in time, rather than only during a transition phase (Greenhalgh, 1995; McNicoll, 1994).

Most of the structural-cultural approaches may be understood as a reaction to the theory of demographic transition as a leading paradigm in demography. This reaction takes shape in various ways, from a reformulation of the central concepts of a transition (Caldwell), via a further specification of particular conditions under which transition is likely to occur (Lesthaeghe), to an outright rejection of a common force of fertility change (Greenhalgh), and further to the adoption of a situational analysis of the specific combination of relevant institutional forms (McNicoll). The theoretical profundity of the study of fertility declines essentially in the same order. Caldwell's approach is a straightforward theory, McNicoll's yields an analytic framework and, hopefully, a methodology or research agenda (McNicoll 1985, 1992).

CONCLUSION

The theoretical orientations presented here provide distinctive angles to view one and the same object of study. The consequence is a corresponding range of different answers to principal questions and requirements with respect to theory building and methodology. On its own, each theoretical approach does not provide a complete picture, but each advances its own propositions that contribute to understanding of fertility behavior. To a large degree, their differences are not complementary, but reflect different interests and assumptions, differences that might be irreducible unless human science in general succeeds in developing an encompassing meta-theory.

Given the state of fertility theory, how is it likely to evolve in the near future? This will depend on a number of factors, such as developments in other disciplines, newly emerging issues—childbearing at advanced ages, new fertilization technologies—and demands with respect to forecasting, policy development, and contributions to intervention programs. With respect to the latter, demographers have long been called upon to contribute their views, but have seemed to lack the appropriate conceptual competence to do so effectively. The demographic transition theory, and in a further distance Malthusian notions, provide a general background for family planning, but have not yet yielded the insight into individual behavior formation nor the specific knowledge about the context of that behavior that is necessary for the design of effective intervention programs. With regard to the interpretation of individual behavior, more contributions might have been expected from psychology, but psychological demography has largely stuck with the value-of-children approach and the 1970s based versions of choice models. It might be useful to consider recent developments in the area of cognitive psychology, and especially learning theory, for instance with regard to information education and communication activities, and, in general, the diffusion of knowledge and ideas. With respect to the situational analysis, demography has made prominent advances. The recently emerged institutional approaches offer good scope to provide an adequate representation of the social, economic, cultural, and political dimensions of the situation-specific context of fertility. Given advances in cognitive anthropology, sociology, and institutional economics, further elaboration and maturation of this field can be anticipated.

Finally, most theoretical approaches to fertility lack a dynamic perspective. Again, institutional analysis, relying on its historic and path-dependent interpretation, might improve this situation with respect to the social context of fertility. At the level of individual time, further development in the area of life course analysis is likely to occur. Although there is substantive literature in this respect (e.g., Birg *et al.*, 1991;

Coleman, 1983; Courgeau and Lelièvre, 1989; Willekens, 1991), it has not yet entered mainstream theoretical thinking in demography.

Acknowledgment

This chapter relies on earlier reviews of fertility theories by the author (de Bruijn 1993, 1999).

References

Ajzen Icek and Fishbein Martin, 1980. *Understanding attitudes and predicting social behaviour.* Englewood Cliffs (New York), Prentice Hall, 278 p.

Ajzen Icek, 1991. The theory of planned behaviour, *Organizational Behaviour and Human Decision Processes*, vol. 50, p. 179–211.

Anderson Barbara A., 1986. Regional and cultural factors in the decline of marital fertility in Western Europe, *in:* Ansley J. Coale and Suzan C. Watkins (eds.), *The decline of fertility in Europe*, p. 293–313. Princeton, Princeton University Press, 484 p.

Andorka Rudolf, 1989. Successes and failures in the field of population policies since 1984, *Population Bulletin of the United Nations*, no. 27, p. 30–41.

Archer Margaret S., 1996. *Culture and agency. The place of culture in social theory.* Cambridge, Cambridge University Press, 351 p.

Ariès Philippe, 1962. *Centuries of childhood: a social history of family life.* New York, Knopf, 447 p.

Arthur W. Brian, 1982. Book review of Gary S. Becker: A treatise on the family, *Population and Development Review*, vol. 8(2), p. 393–397.

Bagozzi Richard P. and Van Loo M. Frances F, 1991. A purposeful behaviour theory of work and family size decisions, *in:* Jacques J. Siegers, Jenny de Jong-Gierveld and Evert van Imhoff (eds.), *Female labour market behavior and fertility*, p. 101–129. Berlin, Springer Verlag, 301 p.

Barker David J.P., 1992. Fetal and infant origins of adult disease, *British Medical Journal*, vol. 301(6761), p. 1111.

Beaver Steven E., 1975. *Demographic transition theory reinterpreted. An application to recent trends in Latin America* Lexington, MA, Lexington Books, 177 p.

Becker Gary S. and Lewis H. Gregg, 1973. On the interaction between quantity and quality of children. *Journal of Political Economy*, vol. 82, p. 29–88.

Becker Gary S., 1976. *The economic approach to human behavior.* Chicago, University of Chicago Press, 314 p.

Becker Gary S., 1960. An economic analysis of fertility, *in:* Gary S. Becker (ed.), *Demographic and economic change in developed countries*, p. 209–240. Princeton (New Jersey), Princeton University Press, National Bureau of Economic Research, 356 p.

Becker Gary S., 1965. A theory of the allocation of time, *The Economic Journal*, vol. 75(299), p. 493–517.

Becker Gary S., 1981. *A treatise on the family*, Cambridge (Mass.), Harvard University Press, 288 p. (2nd ed., 1991, 424 p.).

Becker Stan, 1993. The determinants of adolescent fertility with special reference to biological variables, *in:* Ronald Gray, Henri Leridon and Alfred Spira (eds.), *Biomedical and demographic determinants of reproduction*, p. 21–49. Oxford, Clarendon Press, 482 p.

Birg Hernig, Flöthmann Jürgen and Reiter Iris, 1991. *Biographische Theorie der Demographischen Reproduktion* [Biographic theory of demographic reproduction]. Frankfort, Campus Verlag, 407 p.

Blake Judith, 1985. The fertility of transition: continuity or discontinuity with the past? *in:* IUSSP (ed.), *International Population Conference, Florence, 1985*, vol. 4, p. 393–405. Liège, IUSSP, 555 p.

Blake Judith, 1994. Fertility control and the problem of voluntarism. *Population and Development Review*, vol. 20(1), p. 167–177. (Presented at: Scientists and world affairs, Proceedings of the Twenty-Second Pugwash Conference on Science and World Affairs, September 7–12, 1972, London).

Bledsoe Caroline, 1990. The politics of children: fosterage and the social management of fertility among the Mende of Sierra Leone, *in:* Penn W. Handwerker (ed.), *Births and power: social change and the politics of reproduction*, p. 81–100. Boulder, Westview Press, 227 p.

Bledsoe Caroline, 1995. Marginal members: children of previous unions in Mende households in Sierra Leone, *in:* Susan Greenhalgh (ed.), *Situating fertility. Anthropology and demographic inquiry*, p. 130–153. Cambridge, Cambridge University Press, 304 p.

Bongaarts John and Potter Robert G., 1983. *Fertility, biology and behaviour. An analysis of the proximate determinants*. New York, Academic Press, 230 p.

Bongaarts John and Watkins Susan C., 1996. Social interactions and contemporary fertility transitions, *Population and Development Review*, vol. 22(4), p. 639–682.

Bongaarts John, 1978. A framework for analyzing the proximate determinants of fertility, *Population and Development Review*, vol. 4(1), p. 105–132.

Bongaarts John, 1993. The relative contributions of biological and behavioral factors in determining natural fertility: a demographer's perspective, *in:* Ronald Gray, Henri Leridon and Alfred Spira (eds.), *Biomedical and demographic determinants of reproduction*, p. 9–18. Oxford, Clarendon Press, 482 p.

Boserup Ester, 1970. *Women's role in economic development*. London, Allen and Unwin, 283 p.

Boserup Ester, 1981. *Population and technological change: a study of long terms trends.* Oxford, Basil Blackwell, 255 p.

Boserup Ester, 1990. Population, the status of women, and rural development, *in:* Geoffrey McNicoll and Mead Cain (eds.), *Rural development and population: institutions and policy*, p. 45–60. New York, Oxford University Press, 366 p. (Supplement to *Population and Development Review*, vol. 15).

Brown Lester, 1988. *The changing world food prospect: the nineties and beyond.* Washington DC, Worldwatch Institute, 58 p. (Worldwatch Paper, no. 85).

Bulatao Rodolfo A, 1981. Values and disvalues of children in successive childbearing decisions, *Demography*, vol. 18(1), p. 1–25.

Bulatao Rodolfo A, 1982. The transition in the value of children and the fertility transition, *in:* Charlotte Höhn and Rainer Mackensen (eds.), *Determinants of fertility trends: theories re-examined*, p. 95–122. Liège, Ordina Éditions, 310 p.

Bulatao Rodolfo A, 1984. Content and process in fertility decisions: a psychosocial perspective, *in:* United Nations (ed.), *Fertility and family. Proceedings of the Expert Group on fertility and family, New Delhi, 5–11 January 1983*, p. 159–199. New York, United Nations, Department of Economic and Social Affairs, 476 p. (Doc. ST/ESA/SER.R/53).

Bulatao Rodolfo A. and Fawcett James T., 1981. Dynamic perspectives in the study of fertility decision-making: successive decisions within a fertility career, *in:* IUSSP (ed.), *International population conference, Manilla, 1981*, vol. 1, p. 433–449. Liège, UIESP, 508 p.

Bulatao Rodolfo A. and Lee Ronald D. (eds.), 1983. *Determinants of fertility in developing countries: a summary of knowledge.* Washington D.C., National Academy Press, 2 vol., 642 p, 846 p.

Bumpass Larry and Westoff Charles F, 1970. The "perfect contraceptive" population. *Science*, vol. 169, p. 1177–1182.

Burch Thomas K., 1980. Decision-making theories in demography: an introduction, *in:* Thomas K. Burch (ed.), *Demographic behavior: interdisciplinary perspectives on decision-making*, p. 1–22. Boulder (Colorado), Westview Press, 242 p.

Burch Thomas K., 1981. Interactive decision making in the determination of residence patterns and family relations, *in:* IUSSP (ed.), *International population conference, Manilla 1981*, vol. 1, p. 451–461. Liège, UIESP, 508 p.

Cain Mead, 1981. Risk and insurance: perspectives on fertility and agrarian change in India and Bangladesh, *Population and Development Review*, vol. 7(3), p. 435–474.

Cain Mead, 1989. Family structure, women's status and fertility change, *in:* IUSSP (ed.), *International population conference. New Delhi 1989*, vol. 1, p. 181–188. Liège, UIESP, 452 p.

Caldwell John C. and Caldwell Pat, 1986. *Limiting population growth and the Ford Foundation contribution.* London, Frances Pinter, 194 p.

Caldwell John C. and Caldwell Pat, 1987. The cultural context of high fertility in Sub-Saharan Africa, *Population and Development Review*, vol. 13(3), p. 409–437.

Caldwell John C., 1976. Toward a restatement of demographic transition theory. *Population and Development Review*, vol. 2(2–3), p. 321–366.

Caldwell John C., 1982. *Theory of fertility decline.* London, Academic Press, 386 p.

Caldwell John C., 1996. Demography and social sciences. *Population Studies*, vol. 50(2), p. 305–333.

Caldwell John C., Reddy Palli H. and Caldwell Pat, 1982. The causes of demographic change in rural south India: a micro approach. *Population and Development Review*, vol. 8(4), p. 689–727.

Carr-Saunders Alexander M, 1922. *The population problem: a study in human evolution.* Oxford, Clarendon Press, 516 p.

Carr-Saunders, Alexander M, 1964. *World population: past growth and present trends.* London, Frank Cass & Co., 336 p.

Caselli Graziella, 1990. The influence of cohort effects on differentials and trends in mortality, *in:* Jacques Vallin, Stan D'Souza and Alberto Palloni (eds.), *Measurement and analysis of mortality: New approaches*, p. 229–249. Oxford, Clarendon Press, 405 p. (International Studies in Demography).

Chesnais Jean-Claude, 1986. *La transition démographique. Étapes, formes, implications économiques.* Paris, Presses Universitaires de France, INED, 582 p. (Travaux et Documents, Cahier no. 113).

Cleland John and Wilson Christopher, 1987. Demand transitions of the fertility transition: an iconoclastic view, *Population Studies*, vol. 41, p. 5–30.

Cleland John G., 1987. Socio-economic determinants of fertility: An assessment of recent findings and their implications, *in:* ESCAP (ed.), *Population policies and programmes: Current status and future directions*, p. 43–58. New York, United Nations, 216 p. (Asian Population Studies Series, no. 84).

Cleland John, 1985. Marital fertility decline in developing countries: theories and the evidence, *in:* John Cleland and John Hobcraft (eds.), *Reproductive change in developing countries. Insights from the World Fertility Survey*, p. 223–252. Oxford, Oxford University Press, 301 p.

Coale Ansley J., 1973. The demographic transition. *in:* IUSSP (ed.) *International Population Conference, Liége 1973*, vol. 1, p. 53–73. Liège, Derouaux, 494 p.

Coleman Samuel, 1983. The tempo of family formation, *in:* David W. Plath (ed.), *Work and Life Course in Japan*, p. 183–214. Albany, State University of New York Press, 267 p.

Courgeau Daniel and Lelièvre Eva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p.

Davis Kingsley et Blake Judith, 1956–56. Social structure and fertility: an analytical framework. *Economic Development and Cultural Change*, vol. 4, no. 4, p. 211–235.

Davis Kingsley, 1945. The world demographic transition, *Annals of the American Academy of Political and Social Science*, no. 273, p. 1–11.

de Bruijn Bart J., 1993. *Interdisciplinary backgrounds of fertility theory.* Amsterdam, Graduate School of Research in Demography, 65 p. (PDOD Paper no 16).

de Bruijn Bart J., 1999. *Foundations of demographic theory: choice, process, context*. Amsterdam, Thesis Publishers, 298 p.

Demeny Paul, 1986. Population and the invisible hand, *Demography*, vol. 23(4), p. 473–487.

Demeny Paul, 1988. Social science and population policy, *Population and Development Review*, vol. 14(3), p. 451–479.

den Bandt Marie-Louise, 1982. *Vrijwillig kinderloze vrouwen. Verkenningen rond een keuze.* [proefschrift] [Women who choose not to have children: investigation of a choice.]. Deventer, Van Loghem Slaterus, 378 p. (PhD Thesis).

Dow Thomas E. Jr., Archer Linda, Khasiani Shaniyisa and Kekovole John, 1994. Wealth flow and fertility decline in rural Kenya, 1981–92. *Population and Development Review*, vol. 20(2), p. 343–364.

Dyson Tim and Moore Mick, 1983. On kinship structure, female autonomy, and demographic behaviour in India, *Population and Development Review*, vol. 9(1), p. 35–60.

Easterlin Richard A, 1975. An economic framework for fertility analysis. *Studies in Family Planning*, vol. 6(3), p. 54–63.

Easterlin Richard A, 1978a. The economics and sociology of fertility: a synthesis, *in:* Charles Tilly (ed.), *Historical studies of changing fertility*, p. 57–133. Princeton (New Jersey), Princeton University Press, 390 p.

Easterlin Richard A, 1978b. New directions for the economics of fertility, *in:* J. Milton Yinger and Stephen J. Cutler (eds.), *Major social issues. A multidisciplinary view*, p. 310–317. New York, The Free Press, 575 p.

Easterlin Richard A, 1980. *Birth and fortune. The impact of numbers on personal welfare.* New York, Basic Books, 205 p.

Easterlin Richard A. and Crimmins Elein M, 1985. *The fertility revolution: a supply-demand analysis.* Chicago, University of Chicago Press, 209 p.

Easterlin Richard A., Pollak Robert A. and Wachter Michael L, 1980. Towards a more general economic model of fertility determination: endogenous preferences and natural fertility, *in:* Richard A. Easterlin (ed.), *Population and economic change in developing countries*, p. 81–94. Chicago, University of Chicago Press, 581 p.

Entwisle Barbara, Rindfuss Ronald R., Guilkey David K., Chamratrithirong Aphichat, Curran Sara R. and Sawangdee Yothin, 1996. Community and contraceptive choice in rural Thailand: a case study of Nang Rong, *Demography*, vol. 33(1), p. 1–11.

Falkenmark Molin, 1990. Rapid population growth and water scarcity: the predicament of tomorrow's Africa, *in:* Kingsley Davis and Mikhail S. Bernstam (eds.), ≪Resources, environment, and population. Present knowledge, future options≫, *Population and Development Review*, supplement to vol. 16, p. 81–94. New York, The Population Council & Oxford University Press, 425 p.

Fawcett James T. and Arnold Fred S., 1973. The value of children: theory and method. *Representative Research in Social Psychology*, vol. 4(1), p. 23–36.

Fawcett James T., 1972. *The satisfactions and costs of children: theories, concepts, methods.* Honolulu, East West Population Institute, East West Center, 522 p.

Fawcett James T., 1983. Perceptions of the value of children: satisfactions and costs, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries*, vol. 1, p. 429–458. New York, Academic Press, 642 p.

Fawcett James T., Albores S., Arnold Fred S., 1972. The value of children among ethnic groups in Hawaii: exploratory measurements, *in:* James T. Fawcett (ed.), *Psychological perspectives on population*, p. 234–259. New York, Basic Books, 522 p.

Fishbein Martin and Ajzen Icek, 1975. *Belief, attitude, intention and behaviour: an introduction to theory and research.* Reading, MA, Addison-Wesley, 578 p.

Fishbein Martin, 1972. Toward an understanding of family planning behaviour. *Journal of Applied Social Psychology*, vol. 2, p. 214–227.

Flavin Christopher, 1989. *Slowing global warming: a worldwide strategy.* Washington DC, Worldwatch Institute, 54 p. (Worldwatch Paper, no. 91).

Ford Thomas R. and De Jong Gordon F., 1970. *Social demography.* Englewood Ciffs (New Jersey), Prentice-Hall, 690 p.

Freedman Ronald, 1975. *The sociology of human fertility. An annotated bibliograpy.* New York, Irvington, 283 p.

Freedman Ronald, 1986. Theories of fertility decline, *in:* Kuttan Mahadevan (ed.), *Fertility and mortality: theory, methodology and empirical issues*, p. 30–36. New Delhi, Sage Publications, 351 p.

Freedman Ronald, 1987. The contribution of social science research to population policy and family planning program effectiveness, *Studies in Family Planning*, vol. 18(2), p. 57–82.

Goody Jack, 1976. *Production and reproduction: a comparative study of the domestic domain.* Cambridge, Cambridge University Press, 157 p.

Greenhalgh Susan (ed.), 1995a. *Situating fertility: anthropology and demographic inquiry.* Cambridge, Cambridge University Press, 304 p.

Greenhalgh Susan, 1988. Fertility as mobility: Sinic transitions, *Population and Development Review*, vol. 14(4), p. 629–674.

Greenhalgh Susan, 1989. New directions in fertility research: anthropological perspectives, *in:* IUSSP (ed.), *International population conference, New Delhi, 1989*, vol. 3, p. 437–449. Liège, IUSSP, 463 p.

Greenhalgh Susan, 1990. Toward a political economy of fertility: anthropological contributions, *Population and Development Review*, vol. 16(1), p. 85–106.

Greenhalgh Susan, 1994. *Anthropological contributions to fertility theory.* New York, The Population Council, Research Division, 49 p. (Working Paper no. 64).

Greenhalgh Susan, 1995b. Anthropology theorizes reproduction: integrating practice, political economic, and feminist perspectives, *in:* Susan Greenhalgh (ed.), *Situating fertility: anthropology and demographic inquiry*, p. 3–28. Cambridge, Cambridge University Press, 304 p.

Greenhalgh Susan, 1996. The social construction of population science: an intellectual, institutional, and political history of twentieth-century demography. *Comparative Studies in Society and History, vol.* 38(1), p. 26–66.

Hammel Eugène A., 1990. A theory of culture for demography, *Population and Development Review*, vol. 16(3), p. 455–485.

Handwerker W. Penn, 1986. Culture and reproduction: exploring micro/macro linkages, *in:* W. Penn Handwerker (ed.), *Culture and reproduction. An anthropological critique of demographic transition theory*, p. 1–28. Boulder, Westview Press, 389 p.

Hauser Philip M. and Duncan Otis D., 1959. The nature of demography, *in:* Philip M. Hauser and Otis D. Duncan (eds.), *The study of population: an inventory and appraisal*, p. 29–44. Chicago, University of Chicago Press, 864 p.

Henry Louis, 1953. Fondements théoriques des mesures de la fécondité naturelle, *Revue de l'Institut International de Statistique*, vol.21(3), p. 135–151.

Henry Louis, 1957. Fécondité et famille. Modèles mathématiques, *Population*, vol. 12(3), p. 413–444.

Henry Louis, 1961. Fécondité et famille. Modèles mathématiques II, *Population*, vol. 16(1), p. 27–48.

Hobcraft John and Little Roderik J.A., 1984. Fertility exposure analysis: a new method for assessing the contribution of proximate determinants to fertility differentials, *Population Studies*, vol. 38(1), p. 21–45.

Hoffman Lois W. and Hoffman Martin L., 1973. The value of children to parents, *in:* James T. Fawcett (ed.), *Psychological perspectives on population*, p. 19–76. New York, Basic Books, 522 p.

Howell Nancy, 1979. *Demography of the Dobe Kung.* New York, Academic Press, 389 p.

Howell Nancy, 1986. Feedbacks and buffers in relation to scarcity and abundance: studies of hunter-gatherer populations, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory. Forward from Malthus*, p. 156–187. Oxford, Basil Blackwell, 311 p.

Hull Terence H, 1983. Cultural influences on fertility decision styles, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries: A summary of knowledge*, p. 381–414. Washington D.C., National Academy Press, 846 p.

Jaccard James J. and Davidson Andrew R, 1976. The relationship of psychological, social, and economic variables to fertility-related decisions, *Demography*, vol. 13(3), p. 329–338.

Kakar Sadhio, 1989. *Intimate relations. Exploring Indian sexuality.* New Delhi, Penguin books India, 161 p.

Khan Falzur Rachid, 1987. Relationship between fertility behaviour and size, structure and functions of the family in Asia, *in:* ESCAP (ed.), *Population policies and programmes: Current status and future directions*, p. 67–78. New York, United Nations, 216 p. (Asian Population Studies Series no. 84).

Khan M. E. and Singh Ratanget, 1987. Woman and her role in the family decision-making process: a case study of Uttar Pradesh, India, *The Journal of Family Welfare*, vol. 33(4), p. 49–63.

Kirk Dudley, 1996. Demographic transition theory, *Population Studies*, vol. 50(2), p. 361–387.

Knodel John and van de Walle Étienne, 1979. Lessons from the past: policy implications of historical fertility studies, *Population and Development Review*, vol. 5(2), p. 217–246.

Knodel John, 1983. Natural fertility: age patterns, levels, and trends, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries: A summary of knowledge*, vol. 1, p. 61–102. Washington D.C., National Academy Press, 642 p.

Koenig Michael A. and Foo Gillian H. C., 1992. Patriarchy, women's status, and reproductive behaviour in rural north India. *Demography India*, vol. 21(2), p. 145–166.

Kreager Philip, 1982. Demography in situ. *Population and Development Review*, vol. 8(2), p. 237–266.

Kreager Philip, 1986. Demographic regimes as cultural systems, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory. Forward from Malthus*, p. 131–156. Oxford, Basil Blackwell, 311 p.

Landry Adolphe, 1909. Les trois théories principales de la population, *Scientia*, vol. 6(11–13), p. 3–29.

Landry Adolphe, 1934. *La révolution démographique.* Paris, Recueil Sirey, 231 p.

Lea Stephen E.G., Tarpy Roger M. and Webley Paul, 1987. *The individual in the economy: a textbook of economic psychology.* Cambridge, Cambridge University Press, 627 p.

Lee Ronald D. and Bulatao Rodolfo A., 1983. The demand for children: a critical essay, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries: A summary of knowledge*, p. 233–287. Washington D.C., National Academy Press, 642 p.

Leibenstein Harvey A., 1957. *Economic backwardness and economic growth: studies in the theory of economic development*. New York, John Wiley, 295 p.

Leibenstein Harvey, 1977. Beyond economic man: economics, politics, and the population problem, *Population and Development Review*, vol. 3(3), p. 183–199.

Leibenstein Harvey, 1981. Economic decision theory and human fertility behaviour: a speculative essay, *Population and Development Review*, vol. 7(3), p. 381–400.

Leibenstein Harvey, 1982. Relaxing the maximization assumption in the economic theory of fertility, *in:* Charlotte Höhn and Rainer Mackensen (eds.), *Determinants of fertility trends: Theories re-examined*, p. 35–48. Liège, Ordina Éditions, 310 p.

Leridon Henri and Menken Jane (eds.), 1979. *Natural fertility*. Liège, Ordina Éditions, 556 p.

Leridon Henri, 1977. *Human fertility. The basic components*. Chicago, The University of Chicago Press, 202 p.

Leridon Henri, 1982. Overview, *in:* Charlotte Höhn and Rainer Mackensen (eds.), *Determinants of fertility trends: Theories re-examined*, p. 297–305. Liège, Ordina Éditions, 310 p.

Lesthaeghe Ron J. and van de Kaa Dirk J., 1986. Twee demografische transities ? [Two demographic transitions?], *in:* Dirk J. van de Kaa and Ron Lesthaeghe (eds.), *Bevolking: groei en krimp*, p. 9–24. Deventer, Van Loghum Slaterus, 115 p.

Lesthaeghe Ron J. and Eelens Frank, 1989. The components of Sub-Saharan reproductive regimes and their social and cultural determinants: empirical evidence, *in:* Ron J. Lesthaeghe (ed.), *Reproduction and social organization in Sub-Saharan Africa*, p. 60–121. Berkeley, University of California Press, 556 p.

Lesthaeghe Ron J. and Moors Guy, 1992. Rationality, cohorts and values, *in:* Henk A. Becker (ed.), *Dynamics of cohort and generations research*, p. 165–202. Amsterdam, Thesis Publishers, 617 p. (Proceedings of a seminar held on December 12–14, 1991, at the University of Utrecht, The Netherlands).

Lesthaeghe Ron J. and Surkyn Johan, 1988a. Cultural dynamics and economic theories of fertility change. *Population and Development Review*, vol. 14(1), p. 1–45.

Lesthaeghe Ron J. and Surkyn Johan, 1988b. *Exchange, production and reproduction: Women in Sub-Saharan demographic regimes*. Brussels, Vrije Universiteit Brussel, 51 p. (IPD-Working Paper 1988–1).

Lesthaeghe Ron J. and Verleye Gino, 1992. De tweede demografische transitie; Conceptuele basis en recente evolutie. [The second demographic transition: conceptual basis and recent evolution], *in:* Nico van Nimwegen and Jenny de Jong Gierveld (eds.), *De demografische uitdaging: Nederland in Europa op weg naar de 21e eeuw*, p. 15–49. Houten, Bohn Stafleu, van Loghum, 147 p.

Lesthaeghe Ron J. and Wilson Christopher, 1986. Modes of production, secularization, and the pace of fertility decline in Western Europe, 1870–1930, *in:* Ansley J. Coale and Susan C. Watkins (eds.), *The decline of fertility in Europe*, p. 261–292. Princeton, NJ, Princeton University Press, 484 p.

Lesthaeghe Ron J., 1977. *The decline of Belgian fertility, 1800–1970*. Princeton (New Jersey), Princeton University Press, 259 p.

Lesthaeghe Ron J., 1980. On the social control of human reproduction, *Population and Development Review*, vol. 6(4), p. 527–548.

Lesthaeghe Ron J., 1983. A century of demographic and cultural change in Western Europe: An exploration of underlying dimensions. *Population and Development Review*, vol. 9(3), p. 411–435.

Lesthaeghe Ron J., 1989a. Introduction, *in:* Ron J. Lesthaeghe (ed.), *Reproduction and social organization in Sub-Saharan Africa*, p. 1–12. Berkeley, University of California Press, 556 p.

Lesthaeghe Ron J., 1989b. Production and reproduction in Sub-Saharan Africa: an overview of organizing principles, *in:* Ron J. Lesthaeghe (ed.), *Reproduction and social organization in Sub-Saharan Africa*, p. 13–59. Berkeley, University of California Press, 556 p.

Lockwood Matthew, 1995. Structure and behavior in the social demography of Africa. *Population and Development Review*, vol. 21(1), p. 1–32.

Malthus Thomas R., 1976. *An essay on the principle of population* (re-ed. by Philip Appleman). New York, Norton & Company, 260 p.

Mayone Stycos Joseph (ed.), 1987. Demography as an Interdiscipline, *Sociological Forum, Special issue: Demography as an Interdiscipline*, vol. 2(4).

McNicoll Geoffrey and Cain Mead, 1990. *Institutional effects on rural economic and demographic change*. New York, The Population Council, Research Division, 60 p. (Working Paper no. 14).

McNicoll Geoffrey, 1980. Institutional determinants of fertility change. *Population and Development Review*, vol. 6(3), p. 441–462.

McNicoll Geoffrey, 1985. The nature of institutional and community effects on demographic behaviour: a discussion, *in:* John B. Casterline (ed.), *The collection and analysis of community data*, p. 177–184. Voorburg, International Statistical Institute, 286 p.

McNicoll Geoffrey, 1988. On the local context of demographic change, *in:* John C. Caldwell, Alan G. Hill and Valérie J. Hull (eds.), *Micro-approaches to demographic research*, p. 10–24. London, Kegan Paul, 500 p.

McNicoll Geoffrey, 1992. The agenda of population studies: a commentary and complaint. *Population and Development Review*, vol. 18(3), p. 399–420.

McNicoll Geoffrey, 1994. *Institutional analysis of fertility*. New York, The Population Council, 42 p. (Population Council Working Paper no. 62).

Maslow, Abraham H., 1970. *Motivation and personality*. New York, Harper and Row, 369 p.

Meadows Donella H. *et al*., 1972. *The limits to growth. A report for the Club of Rome's project on the predicament of mankind*. New York, Universe Books, 205 p.

Menken Jane, Trussell James and Larsen Ulla, 1986. Age and infertility, *Science*, no. 233, p. 1389–1394.

Miller Barbara D, 1981. *The endangered sex. Neglect of female children in rural North India*. Ithaca, Cornell University Press, 201 p.

Moffitt Robert, 1984. Optimal life-cycle profiles of fertility and labor supply, *Research in Population Economics*, vol. 5, p. 29–50.

Montgomery Mark R. and Casterline John B, 1993. The diffusion of fertility control in Taiwan: evidence from pooled cross-section time-series models, *Population Studies*, vol. 47(3), p. 457–479.

Moors Hein, van Leusden Hans and Hogen Esch Irene, 1989. *Opvattingen over het Bevolkingsvraagstuk en Acceptatie van Beleid*. [Attitudes towards population questions and acceptability of policies]. The Hague, NIDI, 200 p. (NIDI Report no. 7).

Myers Norman, 1990. The world's forest and human population: The environmental interconnections, *in:* Kingsley Davis and Mikhael S. Bernstam (eds.), ≪Resources, environment, and population. Present knowledge, future options≫, *Population and Development Review*, Supplement to vol. 16, p. 237–251. New York, The Population Council and Oxford University Press, 423 p.

Nag Momi, 1991. *Sex preference in Bangladesh, India and Pakistan and its effect on fertility*. New York, The Population Council, Research division, 44 p. (Working Paper no. 27).

Namboodiri Narayanam Krishnam, 1980. A look at fertility model-building from different perspectives, *in:* Thomas K. Burch (ed.), *Demographic behavior: interdisciplinary perspectives on decision-making*, p. 71–90. Boulder (Colorado), Westview Press, 242 p.

Namboodiri Narayanam Krishnam, 1983. Sequential fertility decision making and the life course, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries*, vol. 1, p. 444–472. New York, Academic Press, 846 p.

Nerlove Marc, 1974. Household and economy: towards a new theory of population and economic growth, *Journal of Political Economy*, vol. 82(2), p. 200–218.

Ní Bhrolcháin Maire, 1992. Period paramount? A critique of the cohort approach to fertility, *Population and Development Review*, vol. 18(4), p. 599–629.

Ní Bhrolcháin Maire, 1993. Women's and men's life strategies in developed societies, *in:* IUSSP (ed.), *International population conference, Montreal, 1993*, vol. 2, p. 179–190. Liège, IUSSP, 493 p.

Niehof Anke, 1985. *Women and Fertility in Madura*. Leiden, University of Leiden, 373 p. (Ph.D. Thesis).

Niphuis-Nell Marry, 1981. *Motivatie voor ouderschap. Een onderzoek naar de invloed van attitudes op het proces van gezinsvorming.* [Motives for parenthood. An investigation of the influence of attitudes on family formation]. Deventer, Van Loghum Slaterus, 322 p. (Ph.D. Thesis).

North Douglas C., 1994. Economic performance through time, *American Economic Review*, vol. 84(3), p. 359–368.

Notestein Frank, 1945. Population: the long view, *in:* Theodore W. Schultz (ed.), *Food for the world*, p. 36–57. Chicago, Chicago University Press, 352 p.

Page Hilary J, 1989. Childrearing versus childbearing: coresidence of mother and child in Sub-Sahara Africa, *in:* Ron J. Lesthaeghe (ed.), *Reproduction and social organization in Sub-Saharan Africa*, p. 401–441. Berkeley, University of California Press, 556 p.

Pampel Fred and Peters H. Elisabeth, 1995. The Easterlin effect. *Annual Review of Sociology*, vol. 21, p. 163–194.

Pollak Robert A and Watkins Susan C., 1993. *Cultural and economic approaches to fertility: proper marriage or mésalliance*? *Population and Development Review*, vol. 19(3), p. 467–496.

Ravenholt R. T. and Chao John, 1974. Availability of family planning services the key to rapid fertility reduction, *Family Planning Perspectives*, vol. 6(4), p. 217–223.

Retherford Robert D. and Palmore James A., 1983. Diffusion processes affecting fertility regulation, *in:* Rodolfo A. Bulatao and Ronald D. Lee (eds.), *Determinants of fertility in developing countries*, vol. 2, p. 295–339. New York, Academic Press, 846 p.

Robinson Warren C., 1992. Kenya enters the fertility transition, *Population Studies*, vol. 46(3), p. 445–457.

Robinson Warren C., 1997. The economic theory of fertility over three decades, *Population Studies*, vol. 51(1), p. 63–74.

Rogers Everett M., 1962. *Diffusion of innovations*. New York, The Free Press of Glencoe, 367 p.

Rogers Everett M., 1973. *Communication strategies for family planning*. New York, The Free Press, 451 p.

Rosenstiel (von) L. Oppitz G. and Stengel M., 1982. Motivation of reproductive behaviour: a theoretical concept and its application, *in:* Charlotte Höhn and Rainer Mackensen (eds.), *Determinants of fertility trends: theories re-examined*, p. 79–93. Liege, IUSSP, 310 p.

Rosenzweig Marck R. and Wolpin Kenneth I., 1980. Life-cycle labor supply and fertility: causal inferences from household models, *Journal of political economy*, vol. 88(21), p. 328–348.

Rothchild Emma, 1995. Echoes of the Malthusian debate at the Population Summit, *Population and Development Review*, vol. 21(2), p. 351–359.

Ryder Norman B., 1983. Fertility and family structure, *Population Bulletin of the United Nations*, no. 15, p. 15–34.

Schofield Roger and Coleman David, 1986. The state of population theory, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory. forward from Malthus*, p. 1–13. Oxford, Basil Blackwell, 311 p.

Schultz T. Paul, 1981. *Economics of population.* Reading (MA), Addison-Wesley, 240 p.

Schultz T. Paul, 1983. Review of John C. Caldwell, Theory of fertility decline, *Population and Development Review*, vol. 9(1), p. 161–168.

Shedlin Michele G. and Hollerbach Paula E., 1981. Modern and traditional fertility regulation in a Mexican community: the process of decision making, *Studies in Family Planning*, vol. 12(6/7), p. 278–296.

Siegers Jacques J., 1987. Economische verklaringen van het kindertal. [Economic determinants of family size], *in:* Arie Kapteyn (ed.), *Demografische veranderingen en economische ontwikkelingen. Preadviezen van de Koninklijke Vereniging voor de Staatshuishoudkunde*, p. 89–124. Leiden, Stenfert Kroese, 303 p.

Simmons George B., 1985. Theories of fertility, *in:* Ghazi M. Farooq and George B. Simmons (eds.), *Fertility in developing countries. An economic perspective on research and policy issues*, p. 20–66. New York, St. Martin's Press, 533 p.

Simon Herbert A., 1987. Rationality in psychology and economics, *in:* Robin M. Hogarth and Melwin W. Reder (eds.), *Rational choice. The contrast between economics and psychology*, p. 25–40. Chicago, University of Chicago Press, 332 p.

Simon Julian L., 1977. *The economics of population growth.* Princeton, Princeton University Press, 555 p.

Stamm Liesa and Tsui Amy Ong, 1986. Cultural constraints on fertility transition in Tunisia: a case-analysis from the city of Ksar-Hellal, *in:* W. Penn Handwerker, (ed.), *Culture and reproduction. An anthropological critique of demographic transition theory*, p. 159–174. Boulder (Colorado), Westview Press, 389 p.

Szreter Simon, 1993. The idea of demographic transition and the study of fertility change: a critical intellectual history, *Population and Development Review*, vol. 19(4), p. 659–701.

te Velde E., 1992. De invloed van leeftijd op de vruchtbaarheid van de vrouw [Effects of age on women's fertility], *in:* Gijs Beets and Pauline Verloove-Vanhorick (eds.), *Een slimme meid regelt haar zwangerschap op tijd*, p. 21–27. Amsterdam, Swets & Zeitlinger, 141 p.

Teitelbaum Michael 1976. Relevance of demographic theory for developing countries, *Science*, no. 188, p. 420–425.

Thompson Warren S., 1929. Population, *American Journal of Sociology*, vol. 34(6), p. 959–975.

Tsui Amy Ong, Hermalin Albert I., Bertrand Jane T., Knowles J., Stover John and Stewart K. J., 1992. *Evaluating family planning program impact: New initiatives on a persisting question, Demography India*, vol. 22(2), p. 133–154.

Turchi Boone A., 1991. A unified theory of fertility and labour force participation, *in:* Jacques J. Siegers, Jenny de Jong-Gierveld and Evert van Imhoff (eds.), *Female labour market behaviour and fertility*, p. 237–262. Berlin, Springer Verlag, 301 p.

UNICEF, 1991. *Children and women in India. A situation analysis. 1990.* New Delhi, UNICEF, 216 p.

United Nations, 1973. *The determinants and consequences of population trends. New summary of findings on interaction of demographic, economic and social factors*, vol. 1, New York, United Nations, 661 p.

van de Kaa Dirk J., 1987. Europe's second demographic transition. *Population Bulletin*, vol. 42(1).

van de Kaa Dirk J., 1996. Anchored narratives: the story and findings of half a century of research into the determinants of fertility, *Population Studies*, vol. 50(2), p. 389–432.

van de Walle Étienne, 1978. Alone in Europe: The French fertility decline until 1850, *in:* Charles Tilly (ed.), *Historical studies of changing fertility*, p. 257–288. Princeton, NJ, Princeton University Press, 390 p.

van de Walle Francine, 1986. Infant mortality and the European demographic transition, *in:* Ansley J. Coale and Susan C. Watkins (eds.), *The decline of fertility in Europe*, p. 201–233. Princeton, Princeton University Press, 484 p.

Vance Rupert B., 1959. The development and status of American demography, *in:* Philippe M. Hauser and Otis D. Duncan (eds.), *The study of population: an inventory and appraisal*, p. 286–313. Chicago, University of Chicago Press, 864 p.

Watkins Susan C., 1986. Conclusions, *in:* Ansley J. Coale and Susan C. Watkins (eds.). *The decline of fertility in Europe*, p. 420–449. Princeton, Princeton University Press, 484 p.

Watkins Susan C., 1987. The fertility transition: Europe and the Third World compared, *Sociological Forum*, vol. 2(4), p. 645–673.

Watkins Susan C., 1989. Markets, states, nations and bedrooms in Western Europe, 1870–1960, *in:* William M. Mason, Barbara Entwisle, Susan Cotts Watkins, Marta Tienda, Karen Oppenheimer Mason and Ross M. Stolzenberg (eds.), *Micro-macro theoretical linkages in social demography: a thematic session*, p. 89–166. Michigan, Population Studies Center, University of Michigan (Research Report).

Wijsen Cécile, 1994. *Motherhood motivations in fertility decision making*. Amsterdam, Graduate School of Research in Demography (PDOD Working Paper, no. 27).

Willekens Frans J., 1990a. *Beweging in de demografie* [Paradigm shift in demography]. Groningen, University of Groningen (Inaugural Lecture).

Willekens Frans J., 1990b. Demographic forecasting; state-of-the art and research needs, *in:* Cornelius A. Hazeu and Gerard A. B. Frinking (eds.), *Emerging issues in demographic research*, p. 9–66. Amsterdam, Elsevier Science Publishers, 314 p.

Willekens Frans J., 1991. Understanding the interdependence between parallel careers, *in:* Jacques J. Siegers, Jenny de Jong-Gierveld and Evert van Imhoff (eds.), *Female labour market behavior and fertility*, p. 11–31. Berlin, Springer Verlag, 301 p.

Willis Robert J., 1973. A new approach to the economic theory of fertility behaviour, *Journal of Political Economy*, vol. 81(2 Pt. 2), p. 14–64.

World Bank, 1992. *Report on a donor workshop on effective family planning programs*, London, 4–6 February 1992. Washington, The World Bank, Population, Health, and Nutrition Division, Population and Human Resources Department.

Wright Robert E., 1989. The Easterlin hypothesis and European fertility rates, *Population and Development Review*, vol. 15(1), p. 107–122.

Wrigley Edward A., 1978. Fertility strategy for the individual and the group, *in:* Charles Tilly (ed.), *Historical studies of changing fertility*, p. 135–154. New Jersey, Princeton University Press, 390 p.

Wunsch Guillaume, 1995. "God has chosen to give the easy problems to the physicists" or why demographers need theory, *in:* EAPS and IUSSP (eds.), *Evolution or revolution in European population / Évolution ou révolution dans la population européenne. European Population Conference/Congrès européen de démographie, Milan, 1995*, vol. 1, Plenary sessions, p. 201–224. Milan, Fra Angeli, 266 p.

Index

B

D

G

H

J

K

L

N

O

P

Demography: Analysis and Synthesis

Demography: Analysis and Synthesis

Graziella Caselli, Jacques Vallin, and Guillaume Wunsch

with contributions by

Daniel Courgeau, Nico Keilman, Eva Lelièvre, James Vaupel,
Anatoli Yashin, and John Wilmoth

VOLUME II

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data
Caselli, Graziella.
[Dimographie. English]
Demography : analysis and synthesis : a treatise in population studies / Graziella Caselli, Jacques Vallin, and Guillaume Wunsch ; with contributions by Daniel Courgeau . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-12-765660-X (set : alk. paper)—ISBN 0-12-765661-8 (volume 1 : alk. paper)—ISBN 0-12-765662-6 (volume 2 : alk. paper)—ISBN 0-12-765663-4 (volume 3 : alk. paper)—ISBN 0-12-765664-2 (volume 4 : alk. paper) 1. Demography. I. Vallin, Jacques. II. Wunsch, Guillaume J. III. Title.
HB871.C37513 2005
304.6—dc22 2005019413

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 13: 978-0-12-765660-1
ISBN 10: 0-12-765660-X

For all information on all Elsevier Academic Press publications visit our Web site at www.books.elsevier.com.

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Contents
Volume II

Contents of Other Volumes

Volume I Contents

Volume III Contents

Volume IV Contents

Contributors

Amoakon Anoh, Ècole nationale supérirure de statistique et d'économie appliqué (ENSEA), Abidjan, Côte d'Ivoire.

Philippe Antoine, Institut pur la recherché et le developpement (IRD), Dakar, Senegal.

Isabelle Attané, Institut national d'études démographiques (INED), Paris, France.

Alexandre Avdeev, Université de Moscou, Moscow, Russia, and Université Marc Bloch, Strasbourg, France.

Maria Avdeeva, Departement de bibliographie et de l'information, Centre d'etudes demographiques, Faculte d'economie, Université de Moscou, Moscow, Russia.

Brigitte Baccaïni, Institut national d'études démographiques (INED), Paris, France.

Stefano Baldi, Permanent Mission of Italy to the United Nations, United Nations Plaza, New York, United States.

Magali Barbieri, Institut national d'études démographiques (INED), Paris, France.

Odo Barsotti, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Françoise Bartiaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Gijs Beets, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Giovanni Berlinguer, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Elza Berquo, Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil.

Francesco C. Billari, Istituto di Metodi Quantitativi, Università Bocconi, and Innocenzo Gasparini Institute for Economic Research, Milan, Italy.

Jean-Noël Biraben, Institut national d'études démographiques (INED), Paris, France.

Anna-Maria Birindelli, Dipartimento di Statstica, Università degli Studi Milano-Bicocca, Milan, Italy.

Alberto Bonaguidi, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Corrado Bonifazi, Instituto di Ricerche sulla Popolazione e le Politiche Sociali, Rome, Italy.

Marco Bottai, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Michel Bozon, Institut national d'études démographiques (INED), Paris, France.

Vittoria Buratta, Istituto Nazionale di Statistica (ISTAT), Rome, Italy.

Raimondo Cagiano de Azevedo, Facoltà di Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Gérard Calot, Institut national d'études démographiques (INED), Paris, France.†

Emmanuelle Cambois, Institut national d'études démographiques (INED), Paris, France.

Graziella Caselli, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Jean-Claude Chasteland, Institut national d'études démographiques (INED), Paris, France.

Jean-Claude Chesnais, Institut national d'études démographiques (INED), Paris, France.

Luciano Ciucci, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

†Deceased

Robert Cliquet, Population and Social Policy Consultants, Brussels, Belgium.

David Coleman, Department of Social Policy and Social Work, University of Oxford, United Kingdom.

Maria-Eugenia Cosio-Zavala, Universite de Paris X-Nanterre, Nanterre Cedex, France.

Daniel Courgeau, Institut national d'études démographiques (INED), Paris, France.

Jean Coussy, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Gianpiero Dalla Zuanna, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy.

Patricia David, Harvard School of Public Health, Boston, United States.

Lorenzo Del Panta, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna, Italy.

Bart de Bruijn, Netherland Interdisciplinary Demographic Institute (NIDI), Den Haag (La Haye), Pays Bas, Netherlands.

Jean-Michel Decroly, Laboratoire de Géographie Humaine, Université Libre de Brussels, Brussels, Belgium.

Ainhoa de Federico de la Rúa, Institut Federatif de Recherche sur les Economies et les Societes Industrielles (IRESI), Universite des Sciences et Technologies de Lille 1, Lille, France.

Arna Dellis, University of Hawaii, Manoa, Hawaii, United States.

Paul Demeny, Population Council, New York, United States.

Alessandra De Rose, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Paolo De Sandre, Departimento de Scienze Statistiche, Università di Padova, Padova, Italy.

Gustavo De Santis, Facoltà di Scienze Politiche, Università di Messina, Messina, Italy.

Martine Deville, Institut national d'études démographiques (INED), Paris, France.

Manon Domingues Dos Santos, Centre de recherché en economie et statistique (CREST), Malakoff, France.

Josianne Duchêne, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Viviana Egidi, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Vincent Fauveau, United Nations Population Fund, New York, United States.

Alexis Ferrand, Institut de Sociologies et d'anthropologie, Universite des Sciences et Technologies de Lille, Villeneuve d'Ascq, France.

Patrick Festy, Institut national d'études démographiques (INED), Paris, France.

Judith A. Fortney, Family Health International, Research Triangle Park, North Carolina, United States.

Luisa Frova, Instituto Nationale di Statistica (ISTAT), Rome, Italy.

Alexia Fürnkranz-Prskawetz, Vienna Institute of Demography, Vienna, Austria.

Hubert Gérard, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Giuseppe Gesano, Instituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy.

Piero Giorgi, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Valérie Golaz, Institut national d'études démographiques (INED), Paris, France.

Antonio Golini, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Rosa Gomez-Redondo, Universidad Nacional de Educacion a Distancia (UNED), Facultad de CCPP y Sociologia, Madrid, Spain.

Catherine Gourbin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Carl Haub, Population Reference Bureau, Washington, D.C., United States.

François Héran, Institut national d'études démographiques (INED), Paris, France.

Kenneth Hill, Department of Population and Family Health Sciences, Johns Hopkins University, Maryland, United States.

Serguey Ivanov, Population Division, United Nations, New York, United States.

Carol Jagger, Department of Epidemiology and Public Health, University of Leicester, United Kingdom.

Heather Joshi, Centre for Longitudinal Studies, Institute of Education, University of London, London, United Kingdom.

Nico Keilman, Department of Economics, University of Oslo, Oslo, Norway.

Shigemi Kono, Faculty of International Economics, Reitaku University, Chiba-ken, Japan.

Marlène Lamy, Institut de demographie de l'Universite de Paris (IDUP), Paris, France.

Jacques Légaré, Department of Demography, University of Montreal, Montreal, Canada.

Luc Legoux, Institut de démographie de l'université de Paris (IDUP), Paris, France.

Eva Lelièvre, Institut national d'études démographiques (INED), Paris, France.

Henri Leridon, Institut national d'études démographiques (INED), Paris, France.

Ron Lesthaeghe, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Thérèse Locoh, Institut national d'études démographiques (INED), Paris, France.

Michel Loriaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Enzo Lucchetti, Dipartimento di Biologia Evolutiva, Università degli Studi di Parma, Parma, Italy.

Dionisia Maffioli, Università degli Studi di Bari, Bari, Italy.

Godelieve Masuy-Stroobant, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

France Meslé, Institut national d'études démographiques (INED), Paris, France.

Alain Monnier, Institut national d'études démographiques (INED), Paris, France.

Karel Neels, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Annunziata Nobile, Dipartimento di Istituzioni politiche e Scienze sociali, Università degli Studi Roma Tre, Rome, Italy.

Alberto Palloni, Center for Demography and Ecology, University of Wisconsin, United States.

Sophie Pennec, Institut national d'études démographiques (INED), Paris, France.

Pierre Pestieau, Université de Liège, Liège, Belgium and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Victor Piché, Inter-university Centre for Demographic Studies, University of Montreal, Montreal, Canada.

Marc Pilon, Institut de recherché pour le developpement (IRD), Piagadpigpi, Burkina Faso.

Gilles Pison, Institut national d'études démographiques (INED), Paris, France.

Antonella Pinnelli, Department of Demographic Science, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Denise Pumain, Institut national d'études démographiques (INED), Paris, France.

S. Irudaya Rajan, Centre for Development Studies (CDS), Kerala, India.

Jean-Louis Rallu, Institut national d'études démographiques (INED), Paris, France.

Rosella Rettaroli, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna Italy.

Jean-Marie Robine, Démographie et Santé, Institut national de la sante et de la recherché medicale (INSERM), Montpellier, France.

Paul-André Rosental, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Silvana Salvini, Dipartimento di Statistiche, Università degli Studi di Firenze, Florence, Italy.

Antonio Santini, Dipartimento di Statistiche, Univrsità degli Studi di Firenze, Florence, Italy.

Francis Sartor, Departement d'epidemiologie-toxicologie, Institut scientifique de la Sante Publique, Brussels, Belgium.†

Bruno Schoumaker, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Catherine Sermet, Institut national d'études démographiques (INED), Paris, France.

Jolande Siebenga, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Patrick Simon, Institut national d'études démographiques (INED), Paris, France.

Lamberto Soliani, Dipartimento de Scienze Ambientali, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Salvatore Strozza, Dipartimento di Scienze Statische, Università di Napoli Federico II, Napoli, Italy.

Pierre Surault, Groupe d'études démographiques, economiques et socials (GEDES), Universite de Poitiers, Poitiers, France.

Dominique Tabutin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michael Teitelbaum, Alfred P. Sloan Foundation, New York, United States.

†Deceased

Marc Termote, National Institut of Scientific Research, University of Quebec, and Department of Demography, University of Montreal, Canada.

Laurent Toulemon, Institut national d'études démographiques (INED), Paris, France.

Tapani Valkonen, Département de Sociologie, Université de Helsinki, Helsinki, Finland.

Jacques Vallin, Institut national d'études démographiques (INED), Paris, France.

Celine Vandermeersch, Institut national d'études démographiques (INED), Paris, France.

Jean-Pascal van Ypersele, Institut d'astronomie et de geophysique G. Lemaitre, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

James Vaupel, Max Planck Institute for Demographic Research, Rostock, Germany.

Jacques Véron, Institut national d'études démographiques (INED), Paris, France.

Éric Vilquin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Paolo Vineis, Dipartimento di Scienze Biomediche e Oncologia Umana, Università di Torino, Turin, Italy.

Anatoly Vishnevsky, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Tania Vishnievskaia, Institut national d'études démographiques (INED), Paris, France.

Carolyne Wanja Njue, Population Council, Nairobi I, Kenya.

Christine Wattelar, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michel Willems, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

John Wilmoth, Department of Demography, University of California, Berkeley, California.

Guillaume Wunsch, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Anatoli Yashin, Max Planck Institute for Demographic Research, Rostock, Germany.

Sergei Zakharov, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Hania Zlotnik, Population Division, United Nations, New York, United States.

SECTION

I

DETERMINANTS OF MORTALITY

Introduction to Volume II

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Section I of Volume I examined the concept of population dynamics and its mechanisms, and Section II assessed the determinants of one of the three main factors of that dynamic: fertility. In Section I this volume, we consider the determinants of the second factor: mortality. Later sections look at the determinants of migration, the history and future of population settlement, the relations between population and society, the accuracy of population policies and theories, and the methods of data collection and complementary tools of analysis.

By definition, death is the end of a life course, and it often is the result of a shorter or longer process of failing vital energy. Determining the factors of mortality therefore requires examining the process that leads from health to illness and death. This is the subject of Part I (Health, Morbidity, and the Causes of Death), in which the process is considered only in terms of its medical connotations. However, the process resulting in death has a wide range of beginnings. The human organism has an inherently finite life span, and each individual is biologically predestined to die once his or her time has come, but the time of death may be precipitated or delayed by external events or by human behavior. Life and death are the result of a subtle balance between endogenous and exogenous factors, which are considered in Part II (Endogenous and Exogenous Mortality).

No individual is an island. Each person is part of a group whose characteristics predicate the biologic environment and influence behavior. These effects create a wide range of situations between populations and among the constituent groups and subgroups of each population. These differences are difficult to interpret. Evaluation of significant factors, such as gender, social status, and place of residence, is provided in Part III (The Intricacy of Differential Mortality Factors).

Diversity in populations is temporal as well as spatial. Morbidity and mortality rates have changed dramatically since humans first trod the Earth. The conditions of humankind's survival changed with the increasing ability to control the natural world. After millennia of hesitant progress, a sweeping revolution began in the 18th century. The attempts made to explain the dynamics of these changes and the diversity of situations are reviewed in Part IV (Theories and Frameworks of Mortality).

P A R T

I

HEALTH, MORBIDITY, AND CAUSES OF DEATH

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

In purely medical terms, death is usually a multifactorial event, the result of a shorter or longer process of failing vital energy leading from health through the various manifestations of morbidity—acute or chronic diseases, accidents, and other forms of violence—to the medical causes of death. The statistical and demographic approach to analyzing this process is the subject of Part I.

In Chapter 40, Catherine Gourbin and Guillaume Wunsch first discuss the pertinent concepts. Health, illness, and death are simultaneously biologic and cultural events, as well as individual and collective facts of life. Observational designs and data analysis must take these circumstances into account.

In Chapter 41, Catherine Sermet and Emmanuelle Cambois review the methods of collecting and analyzing data on morbidity, the state of impaired health that is curable but that may become a cause of death.

Chapters 42 and 43 address the collection and analysis of cause-of-death data. The first, by France Meslé, paints the general picture in which good records are systematically kept, and in the latter, Vincent Fauveau demonstrates the value of verbal autopsy methods in countries where vital registration is lacking or is deficient.

Cause-of-death analyses often erroneously assume that causes are independent. In Chapter 44, Guillaume Wunsch offers a causal approach that helps clarify cause-of-death interrelationships.

The next two chapters address two particular types of causes of death requiring specific study methods. Maternal mortality is examined in Chapter 45 by Judith Fortney, and the determinants of infant mortality are considered by Godelieve Masuy-Stroobant in Chapter 46.

Part I concludes with the presentation in Chapter 47 of a model by Viviana Egidi and Luisa Frova designed to estimate the unknown prevalence of a disease in a population on the basis of existing mortality data for that disease.

CHAPTER

40

Health, Illness, and Death

CATHERINE GOURBIN AND GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

To your health!

Apart from violent deaths due to accidents, homicide, or suicide, among other causes that may result from physical or mental ill health, most deaths are the result of a disease process and health impairment. But what is illness, and what is health? René Dubos defined health as "the circumstances in which an organism reacts by adaptation, at the same time preserving its individual integrity" (Dubos, 1973, p. 245). For Ivan Illich, "health is the ability that every individual has to cope with or assume responsibility for changing his environment" (Dufresne, 1985, p. 989). According to René Leriche, "health is life untroubled by one's body" (Leriche, 1936, p. 6.16.1), and for Jacques Dufresne, "health is not having to think about health" (Dufresne, 1985, p. 985).

I. HEALTH

Health is a concept that varies with time and place. Illness, as "the irrevocable decree of the Olympian gods," was thought to be a tragedy, and in the Judeo-Christian tradition, disease is considered a punishment (Arsenault, 1985). It is reasonable to wonder to what extent this conception still prevails, given the medical profession's reluctance in many Western countries to provide pain relief. In the Baoulé community of Ivory Coast, seers and augurers prescribe what is normal and abnormal—"evil possession to be treated and joyous possession to be preserved"—and these concepts are seen as norms with which to justify the existing social order (Laplantine, 1986, p. 37). In every age and in every society, the dominant models assign causes to diseases, whether ascribed to earth spirits or evil spirits, drinking, smoking, a fatty diet, viruses, or the gene pool.

The World Health Organization (WHO) defines health as "a state of complete physical, mental, and social well-being" and "not merely the absence of disease or infirmity" (WHO, 1958). Although it reflects the multidimensionality of the concept of health, this definition raises other issues. It misreads *health* as *well-being*, but health is only one component of well-being. It also refers to health mainly in terms of the individual's personal experiences, blurring the two dimensions of objective and subjective health. A diabetes sufferer may feel perfectly well as a result of irrefutable medical advances in the control of diabetes, but does that mean he is in good health? By the WHO definition, health is specific for the individual, making it hard to identify a common denominator to use as an aggregate health indicator (i.e., measure of the state of health of a population). For that measure, the society in which the individual lives must have consensus on what is meant by a *state of complete well-being*.

Another approach, exemplified by Christopher Boorse (1977), is to regard health as coextensive with the normal functioning of the individual, with *normal* being a statistical attribute. As pointed out in the literature on this topic, this definition does not accommodate the dynamic interaction between an organism and its environment, and it does not help the medical profession in addressing the demands of patients (with or without physical symptoms). Roberto Mordacci (1995) posited that there are two dimensions to health: adaptation and development. A healthy person can adapt to the physical or mental stressors of the environment, and health is one condition necessary for individual physical and psychological self-realization. Social health (reflecting the WHO definition) arguably refers to the network of social relations (e.g., family, friends, colleagues) that enables the individual to better withstand adversity.

Although the WHO defines health as more than the absence of disease, there is a clear connection between the two. Most people tend to think about health only when faced with illness. It may be, as Roberto Mordacci (1995) argues, that health is a precondition for being ill, but it remains a fact that only through illness (patient's own or that of others) do individuals compass their health. A distinction must be drawn between the *subjective* and *objective* components of health. An individual may feel unwell but not be suffering from any disease detectable by medical science, or a person may feel well but have a latent tumor, functional disorder, or organic condition with no external or perceived signs. In the former case, the individual is considered subjectively ill but objectively well, and the opposite descriptions obtain in the latter case. *Subjective health* is the status perceived by the individual, and *objective health* is the condition diagnosed by the expert (e.g., physician, healer, shaman). In these definitions, we find the components of health mentioned earlier: a properly functioning human organism as described by medical science and the individual's self-rating of well-being.

Most information on subjective health has come from cross-sectional or follow-up health surveys. The oldest was conducted in the 1950s in the United States. Annual health surveys have been done since 1956 in the United States, since 1964 in Finland, and since 1971 in Great Britain (Bowling, 2004). In the surveys, general health is assessed by questions such as *How would you describe your health?* These questions offer a range of answers, such as *very good, good, fairly good, average, poor, very poor*. Although common, this approach is not watertight. As sociologists have shown, using a single question to gauge a self-assessment is open to criticism. What is needed is a scale based on a range of propositions that can be synthesized into a single indicator as proposed, for example, by Likert or Guttman (Babbie, 2004). The question as asked tells nothing about the individual's reference norms. Are they peer-comparative, age-adjusted, lifestyle-referenced (e.g., smoking, diet, sport) standards, or are they referenced to already diagnosed health problems? Can health as a multidimensional concept be described by a single indicator even when multiple attributes are mentioned (Larson, 1994; Bergner, 1985; Boyle and Torrance, 1984)?

To resolve the difficulty of gauging health by a single question, medical sociology has developed a wide range of scales to assess the different components of health. Many measure the individual's self-assessment of health status, and some focus on mental well-being or happiness, defined as the individual's self-rated satisfaction with life. Other scales measure the extent of the social network and support or the individual's physical abilities. An excellent assessment of these scales has been provided by Ann Bowling (2004). These scales often are devised by health professionals, but the criteria articulated by individuals with no diagnosed illness are also significant. Most of the scales are based on closed questions, which may exclude other factors that some individuals may consider relevant. Another drawback is that the indicators used to measure different aspects of health do not necessarily carry equal weight within different cultures or social classes, at different life stages, for different diseases, or for different sexes (Bowling, 1995).

Individuals' subjective assessments of health are bound up with their life histories. Health is a personal concept, but it is also shaped by the individual's cultural environment and adaptability. *Healthy*, *unhealthy*, *abnormal*, and *normal* are concepts that vary with the beliefs and with the values of each society. Health is therefore a composite picture made up of the experiences, values, and information available to each individual and his or her understanding of the information. The medical knowledge that the individual acquires, particularly through the media or contacts with health practitioners, varies with time and place because of advances in medical knowledge, changing value systems, and improvements in information dissemination (Currer and Stacey, 1993). Of particular significance in this respect are the health education campaigns channeled by the medical literature and (mostly women's) mass-market magazines. These publications have revised individuals' subjective assessments of their health and changed their use of the care system, thereby altering the rates of diagnosed diseases. Health is a dynamic concept that differs between individuals and that changes for the same individual with age and circumstance, even when in good health. Health has become inseparable from the concept of well-being and physical and social adjustment to the environment. It includes the concept of improved well-being and quality of life, which may, for example, be expressed through sporting activity, a more balanced diet, or even cosmetic surgery insofar as it helps an individual become more socially integrated and have an improved sense of well-being.

In some Western societies, health has even become a *right*, like the right to education, security, free speech, or work. This concept was embodied in the WHO slogan—*Health for all by the year 2000*—which, however far-fetched, could be envisaged as an aim to be achieved in the remote future. With rights come responsibilities; if individuals have a right to health, they also have the duty to preserve their own and others' health. If health is everyone's business, the issue of individual responsibility for lifestyle choices then arises. An awakening to high-risk behaviors is beneficial in and of itself, but a society in which the right to health is a core precept and health costs are rising may qualify the principle of equal access to health care with that of individual responsibility, such that the cost of accidents sustained in freely chosen dangerous activities may not be picked up by the community. Eric Rakowski (1991) makes the distinction between *brute bad luck*, which is compensated by the social security system, and *option bad luck*, which is the result of a choice and the costs of which the victim (or private insurance) should bear. This kind of approach raises significant ethical questions. To what extent is option bad luck really chosen (Gilain, 1995)? Smoking is a matter of social circumstance as much as individual choice. The worst-off social groups are exposed to a high concentration of risk factors. Are they to be blamed for drug taking, smoking, or excessive drinking because of a harsh life, lack of education, failure to understand the health message, and the risks inherent in their conduct? In a resource allocation context, any health priority setting will come up against ethical problems. *Quality-adjusted life years* (QALYs), for example, attempt to evaluate treatment impacts on adding to life expectancy and quality of life, encompassed within a single measure. This prompts a host of questions. How should we compare the benefits of certain kinds of health care delivered in different care facilities according to different criteria to patients with different characteristics and different severities of illness? Are collective priorities compatible with essentially individual medical ethics? What considerations will apply: those of health professionals or those of healthy individuals? Does a year of healthy life have the same value for everyone? Should we be thinking in terms of health gain or level of health (Williams, 1996)?

Individuals' subjective assessments of their health is inseparable from their life histories. Broadly speaking, each person begins life with a *health potential*, which varies between individuals, improves or declines over the life course, and influences individuals' assessment of their own health. In Chapter 46, Godelieve Masuy-Stroobant describes the causes of heterogeneity in health at the youngest ages. She argues that this potential then increases over the life course in line with factors such as the place of children in each society, the role assigned to parents, vaccinations, lifestyle, a healthy diet, public health, rising living standards, rising educational levels of individuals and in society, and improved health infrastructure. However, these positive factors may be influenced by a series of individual or collective determinants, including an unhealthy lifestyle (e.g., alcohol abuse), an unhealthy environment, or malnutrition[1] due to drought or periods of economic uncertainty. Over time, the adverse factors will outweigh the beneficial ones, producing a gradual decline in health that leads to illness and death. Within this long-term decline, periods of improved health occur, only to be canceled out by new external or internal health stressors.

II. SICKNESS

He's sick, he tells us . . . I bet he is! Sick! So am I sick! What does "sick" mean? We're all sick! You'll be sick too, and not before long what's more! (Céline, 1952, p. 130)

The concept of disease has assumed many guises in time and through history. The *exogenous model* views disease as an accident due to the action of an actual or symbolic element extraneous to the patient. Depending on the time and place, the disease would be thought to originate in the malignity of a supernatural power or in a harmful agent perceived as natural and present in the environment. Disease presented incipient threats to the stability of the social order, which society must bring under control through modern or traditional medicine. It deviated from social norms whereby individuals had functions and duties to fulfill in society.[2] An extrinsic cause of disease had to be met by an equally extraneous cure. Depending on the culture, this might have been by thwarting the lot cast by fate or by administering appropriate medical treatment. The disease, being

[1] The United Nations (United Nations, 1992) estimates that 3% of the world's children are severely malnourished and that 34% of those younger than 5 years have average or moderate malnutrition or are underweight. Severe malnutrition affects children's physiologic and cognitive development.

[2] The purportedly neutral medical model in fact imbues the definition of disease with an ethical and moral dimension. For example, homosexuality was classed for a long time as a psychosis by the American Psychiatric Association (Renaud, 1985).

a foreign body, had to be expelled by ritual purification, bleeding, or antibiotic treatment. In this scheme of things, however, the disease, albeit extraneous to the patient, could nevertheless be caused by the patient's own failure to avoid it by breaking a taboo or adopting a risky behavior.

In the *endogenous model*, disease is caused by a general imbalance in the host organism. In this case, there would be, for example, an infectious agent, but it could develop only in conducive personal conditions. The disease would therefore stem partly from the individual himself. This model is specific to psychosomatic medicine disciplines, such as psychoanalysis and homeopathy. This twin-track model explains why there is no single definition of the morbid state or a single biomedical blueprint, but there are instead multiple concepts of disease and therefore many possible forms of medicine. This model explains why, even in Western societies, modern treatment based on the experimental scientific model is found rubbing shoulders with psychosomatic therapies and with lay healers. Likewise, in African societies, the treatment of disease symptoms may be left to the health clinic or hospital, and treatment of the cause would be the purview of the village shaman and healer (Zempleni, 1991).

For Claude Bernard, the father of modern medicine, health and disease are not two opposing *qualities* but the outcomes of *mere physiologic changes* that are simply stages along the health-disease continuum (Laplantine, 1986). The doctor approaches *health and illness* within the framework of an interpersonal relationship with the patient, whereas epidemiology and demography seek to identify *diseases* within a population group (Jenicek and Cléroux, 1982). Population morbidity research would therefore entail recognizing and evaluating the scale of health disorders within that population, identified from the number of diseased persons and their diseases (see Chapter 41). The demographic or epidemiologic approach is twofold: to measure the incidence of the disease in the population and to identify the determinants or risk factors involved in the occurrence of the disease. Public health epidemiology[3] is more concerned with individual risk factors such as lifestyle, whereas demography puts a greater focus on the impact of the economic and social context on behavioral variables.

Three components must be distinguished in analyzing morbidity. *Perceived morbidity*, also called *subjective morbidity*, enables a population's perceived needs and therefore health care demands to be estimated. However, it only partially reflects the degree of health care demand. Perceived morbidity can be determined through health surveys of a population or a sample population. *Diagnosed morbidity* is the extent of morbidity observed on use of medical provision (e.g., general practitioner, specialist, hospital) or determined in specific epidemiologic surveys designed to identify the scale and causes of a disease in a population, such as a prospective survey of cardiovascular disease. This is the type of morbidity found in hospital statistics or reports from sentinel networks of general practitioners. Both approaches to morbidity—through medical symptoms and by means of epidemiologic surveys—enable diagnosis of one or more disease processes and, if relevant, determination of their severity. The early stages of some diseases, such as Alzheimer's disease, are hard to diagnose because of their multiple, nonspecific symptoms. Alzheimer's disease is diagnosed on the basis of clinical criteria generated using standardized scales and protocols from which only a probability of the disease can be given. Only histopathologic criteria met by biopsy or during a postmortem examination make a certain diagnosis possible. *Objective morbidity* is the degree of morbidity identified by the methods of current medicine. It is measured by routine examinations of a population or subpopulation group (e.g., school students, company employees). The latter two aspects of morbidity initially were classified by the WHO as a cause-of-death classification, a category later extended to the diseases themselves (e.g., tooth decay, myocardial infarction). The WHO's *International Classification of Diseases* (ICD) has undergone adaptations for various medical disciplines, such as for oncology, ophthalmology, and stomatology. The ICD and other possible classifications are considered more extensively by France Meslé in Chapter 42.

As with health, different scales have been developed to measure quality of life during illness in general and especially to evaluate quality of life for those affected by specific diseases, because a cancer patient's quality of life, for instance, must be assessed differently from that of a mental patient. We therefore have tools for measuring general quality of life and various scales more specifically applicable to cancers, respiratory diseases, cardiovascular diseases, psychoses, neurologic disorders, rheumatic disorders, and other disease states. Ann Bowling (2001) has provided a detailed analysis of these scales, making it possible to gain insights into the effects and treatment of a disease.

Determining the incidence of diseases and their causes in a population necessarily involves measuring the phenomenon. In public health, the study of disease

[3] Mackenbach (1995) has provided a more extensive explanation of the distinction between clinical and public health epidemiology.

in a population can be characterized by its incidence and its prevalence. The *incidence* is the number of new (or recurrent) cases of a disease in a given period in a population, whereas the *point prevalence* is the total number of people affected by that disease in a given population at a point in time. As in demography, the following equation can be used in the absence of migratory flows:

$$\text{Prevalence } (t + n) = \text{Prevalence } (t) + \text{Entries } (t, t + n) - \text{Exits } (t, t + n)$$

In this equation, entries are determined by the incidence of new or recurrent cases and exits by the incidence of deaths or cures.

The incidence rate over a given period is calculated as the ratio of new cases observed in the period compared with the population at risk. The measure is similar to that of a death rate (see Volume I, Chapter 11). The prevalence rate is calculated as the point number of ill persons divided by the point population size. This gives a proportion of ill persons similar to that of the single population of marriageable age (see Volume I, Chapter 25).

A disease process may also be characterized by its *severity*, which can be measured by the length of sickness absence, the number of days of bed confinement or hospitalization, or even by severity scoring (Jenicek and Cléroux, 1982). Severity can also be measured by the *lethality* of the disease process, which is the ratio of the number of deaths from a disease to the number of people who contracted that disease over a given period. In certain circumstances, there is a relationship between mortality, incidence, and lethality (Jenicek and Cléroux, 1982). If entries to and exits from the ill population remain time constant, the ratio is determined by the following equation:

$$\text{Mortality} = \text{Incidence} \times \text{Lethality}$$

For example, if the annual incidence of a cancer is 80 cases per 100,000 people in the general population and the mortality is 40 cases per 100,000 people, lethality is calculated as follows:

$$\text{Lethality} = 40/80 = 0.5$$

Cancer will therefore be fatal for one-half of those affected. A health policy focused on reducing mortality may seek to reduce incidence, lethality, or both.

Often, data are lacking to calculate the morbidity indicators detailed previously. Two alternatives are then possible: measure mortality by cause or use whatever data may be available on care supply or demand. In the former case, life tables can be constructed by medical cause of death or life expectancies estimated in the absence of a given cause (see Chapter 42). The value of this approach depends greatly on the relation between the initial cause of death and the underlying disease process. Many diseases, such as osteoarthritis, have little impact on mortality, and an individual may live with a chronic disease for years before dying of it. The changing health landscape, increased life expectancy, and the emergence of chronic diseases have led to the definition of new indices that reflect quality of life more than quantity in terms of duration. At the same time, the WHO's development of the *International Classification of Impairments, Disabilities, and Handicaps* has helped to clarify the concepts. Into the basic sequence of etiology → pathology → manifestation were integrated the *consequences* of diseases that also predicate the demand for health care, generating the following proposed sequence: disease → impairment → disability → handicap. These concepts are defined with more precision by Jean-Marie Robine and Carol Jagger in Chapter 80 along with the newer indicators (e.g., health-adjusted life expectancy, disability-free life expectancy) developed from these ideas.

Another proxy of the measure of morbidity is based on health care supply and demand. For *supply*, health service infrastructure or macroeconomic indicators can be used. The first set of indicators comprises items such as the total number of physicians (e.g., general practitioners, specialists), ancillary medical personnel, dentists, and pharmacists per 1000 people in the general population; hospital provision (e.g., total and types of hospitals, total and types of beds); collective preventative medicine services (e.g., tuberculosis clinics, maternal and child welfare services, cancer screening centers, school and occupational health services); and the availability of home medical care services, nursing homes, home help services, and dependency treatment centers. The second set of indicators refers to items such as the share of gross national product spent on health and the percentage of the population covered by medical benefits insurance. Care *demand* indicators naturally focus on the use of this provision and therefore refer to total consultations of general practitioners or specialists, total hospital admissions and surgical operations, bed occupancy rates, total pharmaceutical consumption (distinguished by the targeted disease), total laboratory tests and clinical examinations performed, percentage of inoculated children, proportion of malnourished children, and type of malnutrition. Several sources provide an overall picture of the type of indicators proposed (OECD, 1996; WHO, 1986; Deliège, 1983). Evaluations of data on the use of private care provision or consumption of medicines not refunded by social security systems are other ways of assessing health care demand.

Factors other than morbidity also influence health care supply and demand (Béresniak and Duru, 1995). Health service funding affects mainly the private provision supply. The financing mechanism also affects demand, because extending health care coverage increases the use of medical services across the board. Supply and demand of health care also are affected by demographic factors; for example, specialist doctors are often concentrated in large towns and cities. A range of socioeconomic factors influences the market. The quantity and structure of supply are dictated by the level of economic development, whereas demand and its characteristics (e.g., uptake of preventative or curative care) vary significantly by the individual's age, sex, and social status (Grandfils *et al*., 1995). Demand and supply are not independent; demand rises as the number of doctors and hospitals and the level of diagnostic and treatment provision expand. There is a pull effect of supply growth on demand growth. For all these reasons, extreme caution must be exercised in using care demand and supply statistics as a proxy measure of morbidity, especially when the data for comparison is from countries with different health care and social security systems.

III. DEATH

All that is upright will be laid low.

(Mandenka proverb, Senegal oral tradition)

You die not because you are sick; you die because you are living. (Montaigne, Essays, Book III, Ch. XIII)

Under French and Belgian civil law, death results in the total loss of physical and legal personality; no other event results in such a total loss. The registrar draws up the death certificate, but as lawyers have pointed out, advances in medical science have complicated the problem of certifying death. Death has ceased to be a single event involving all the vital functions simultaneously, because current medical technologies allow individuals to be kept alive on prolonged life support who in other circumstances or times would have been pronounced dead (Philippart and Stroobants, 1984; Weill and Terré, 1978). The definition of death therefore varies with medical progress, and the same issue arises with euthanasia. According to the *Code of Medical Ethics* drawn up by the National Medical Council of Belgium (1975), the decision to turn off the life support systems of a person in irreversible coma can be made only in light of current medical knowledge (article 98).

The *Petite Encyclopédie Médicale Hamburger* (Leporrier, 1994) defines the end of life according to current medical thinking as brain death in the individual. This event results in a death certificate being issued certifying that irreversible damage has occurred to the central nervous system that is inconsistent with life. A patient whose lungs and heart are still functioning may therefore be clinically dead (Farber, 1988). Brain death may be diagnosed if four clinical conditions are met: total loss of consciousness and absence of any spontaneous activity apart from reflex responses, absence of reactivity of all cranial nerves, absence of spontaneous respiration, and a flat electroencephalographic pattern. All circumstances that may simulate death must be controlled for within a sufficiently long observation period (12 to 48 hours, depending on age) during which the signs of brain death remain unchanged.

The definition of death, like those of health and illness, is spatiotemporally variable. Likewise, humankind's representation of death has changed over the centuries. The conceptions of death and funeral rites partake of the entire sociocultural system. As Louis-Vincent Thomas points out, "Death is a biological occurrence but also a cultural fact for the perceptions that it creates about its nature and origins, the beliefs and representations it gives rise to, and the ways in which it is accepted, denied, or transcended. The funeral rites it brings about are a specifically human practice" (Thomas, 1985, p. 115). Drawing extensively on Philippe Ariès' work, it must be stressed that attitudes about death must be considered relative to the awareness of self and others and to the sense of individual or collective destiny. In the first millennium Christian West, death was prepared for within the family and represented the final rest accepted by people resigned to the fact of their own mortality. Death is also an organized public, albeit nondramatic, ceremony. In the Early Middle Ages (12th century), fear of Judgment Day transformed society's conception of death. It became a more harrowing experience, because the individual's final deeds would weigh more heavily in the scales. It had become what Philippe Ariès (1975) calls "the death of self," at least for the literate, powerful, and wealthy who saw the essence of themselves in their own death and discovered their individuality. In the 18th century, Romanticism changed the idea of death. The Western world dramatized and exalted it; the locus shifted to the death of the other, "the death of the significant you" (Ariès, 1975). Funeral rites had to reflect the grief of the survivors and the depth of their despair.

In the 20th century, death became a matter for shame; society and the family had to be spared the hideous anguish of it. We now die in hospitals, and death becomes a technical fact for the medical profession alone. Medical progress has helped change our conception of death, and we now expect science to postpone death indefinitely. Modern society, ensnared by its own technological progress, wonders about the fitting way to die, and the current debate on death is crystallizing around medical euthanasia. According to Louis-Vincent Thomas (1985),[4] this is likely to result in the loss of symbols, an extreme simplification of rites, desensitization symptomatic of desocialization (e.g., shortened ceremonies, handwritten letters of condolence, discreet mourning), and abnegation of death. However, in a survey conducted in France in 1994, approximately three-fourths of the respondents thought that people were more open about death than in earlier periods. That reaction was probably a consequence of the spread of acquired immunodeficiency syndrome (AIDS), which has claimed a variety of high-profile victims, but it also reflected the fact that death is a big seller for the film industry and other media (de Cacqueray, 1997).

Death in Western countries may be more an individual affair and less one of the family and society, but some writers such as David Armstrong (1987) or Lindsay Prior and Mick Bloor (1993; cited by Lupton, 1994) argue that since the mid-19th century, death has become a public affair. The introduction of death certificates has brought death into the realm of statistics, and the construction of life tables makes it a whole-population phenomenon. Concern has shifted from individual deaths to focus on the mean age at death or the sex- and cause-specific distribution of deaths. Death has become the subject of rational, mathematical representation. The *force* of mortality (i.e., instantaneous death rate) falls under the Gompertz-Makeham law (see Volume I, Chapters 8 and 11).

Each civilization in every era has developed its own strategies for dealing with death and its rituals for transcending this point of separation. In African traditional societies, death is the subject of public representations and ceremonies, because the community takes precedence over the individual. Death is part of life in these societies, and it is a daily presence because of the high incidence and the many songs, tales, and proverbs that feature it. Death is a moment in a process that encompasses life and continues after death. The life force leaves the body to continue in eternity. The dead are still therefore part of time and can be called on to resolve the problems of life. What remains is not a dead body, but an ancestor invested with new functions. The anguish of death is no less present in these traditional societies than in the Middle Ages or the modern Western world, but it is a socially regulated anguish, channeled through ritual and borne by belief: "Death is not freely vanquishable" (Thomas, 1975, p. 310). The fundamental difference between the Western and African traditions may be in the representation of ego:

> The African approach is participative and socially regulated; the West a heightened sense of personhood grappling with the godforsakenness of nothingness. The result is diametrically opposing attitudes to death and the dead: the one refusing the separation of life-death and living-dead; the other emphasizing both points of severance (Thomas, 1982, p. 251).

[4] This section owes much to the writings of Louis-Vincent Thomas (1975, 1982, and 1985 in particular), who was an authority in the field.

CONCLUSIONS

Health, illness, and death are at once biologic and cultural, individual and collective realities. Healthy and unwell individuals self-define themselves as such, but their health and their illness are also dictated by the norms of the society in which they live. Even death is a cultural fact according to whether it leads to the annihilation of ego or the survival of the ancestor in the afterlife. In the Western world, which is rational and ruled by science, the corpses of some wealthy Americans are lying cryogenically frozen, like the traveler in the *Space Odyssey*, in the unrealistic hope that the march of medicine will vanquish old age and disease. The ancient world, traditional Africa, and Western society each has its own way of answering Hamlet's question: "To be or not to be?"

Acknowledgment

This chapter was translated from the French version by Glenn D. Robertson.

References

Ariès Philippe, 1975. *Essais sur l'histoire de la mort en Occident du Moyen-Âge à nos jours*. Paris, Seuil, 223 p.

Armstrong David, 1987. Silence and truth in death and dying, *Social Science and Medicine*, vol. 24(8), p. 651–657.

Arsenault André, 1985. Mutations du discours thérapeutique, *in:* Jacques Dufresne, Fernand Dumont, and Yves Martin (eds.), *Traité d'anthropologie médicale*, p. 75–84. Silley, Quebec, Presses de l'Université du Quebec, 1245 p.

Babbie Earl, 2004. *The Practice of Social Research* (tenth edition). California, Wadsworth, 640 p.

Béresniak Ariel and Duru Gérard, 1995. *Economie de la santé*, 3rd ed. Paris, Masson, 182 p.

Bergner Marilyn, 1985. Measurement of health status, *Medical Care*, vol. 23(5), p. 696–704.

Boorse Christopher, 1977. Health as a theoretical concept, *Philosophy of Science*, vol. 44, p. 542–573.

Bowling Ann, 1995. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life, *Social Science and Medicine*, vol. 41(10), p. 1447–1462.

Bowling Ann, 2001. *Measuring disease*, 2nd ed. Buckingham, Open University Press, 420 p.

Bowling Ann, 2004. *Measuring health: a review of quality of life measurement scales*. Buckingham, Open University Press, 216 p.

Boyle Michael H. and Torrance George W., 1984. Developing multiattribute health indexes, *Medical Care*, vol. 22(11), p. 1045–1057.

Cacqueray Christian de, 1997. La mort est notre métier, *in:* Marie-Frédérique Bacqué (ed.), *Mourir aujourd'hui. Les nouveaux rites funéraires*, p. 21–47. Paris, Éditions Odile Jacob, 277 p.

Céline Louis-Ferdinand, 1952. *Voyage au bout de la nuit*. Paris, Gallimard, 505 p.

Conseil de l'ordre, 1975. *Code de déontologie médicale*. Brussels, Conseil de l'Ordre des médecins de Belgique, 50 p.

Currer Caroline and Stacey Meg (eds.), 1993. *Concepts of health, illness and disease: a comparative perspective*. Oxford, Berg, 315 p.

Deliège Denise, 1983. Les indicateurs de bien-être physique, mental et social, *World Health Statistics Quarterly/Rapport Trimestriel de Statistiques Sanitaires Mondiales*, vol. 36, p. 349–393.

Dubos René, 1973. *L'homme et l'adaptation au milieu*. Paris, Payot, 472 p.

Dufresne Jacques, 1985. La santé, *in:* Jacques Dufresne, Fernand Dumont, and Yves Martin (eds.), *Traité d'anthropologie médicale*, p. 985–1013. Silley, Quebec, Presses de l'Université du Quebec, 1245 p.

Farber Jean, 1988. *Déontologie médicale et législation professionnelle. Aspects juridiques et éthiques*. Brussels, Presses Universitaires de Bruxelles, 195 p.

Gilain Bruno, 1995. *Le financement équitable des soins de santé: approches récentes*. Louvain-la-Neuve, Chaire Hoover d'éthique économique et sociale, 71 p. (Document de travail no. 13).

Grandfils Nathalie, Le Fur Philippe, Mizrahi Andrée, and Mizrahi Arié, 1995. *Santé, soins et protection sociale en 1994*, Biblio no. 1103. Paris, CREDES, 154 p.

Jenicek Milos and Cléroux Robert, 1982. *Epidémiologie. Principes, techniques, applications*. St Hyacincthe (Québec), Edisem, 454 p.

Laplantine François, 1986. *Anthropologie de la maladie*. Paris, Éditions Payot, 411 p.

Larson James S., 1994. The weighting of an international health status index, *Social Indicators Research*, vol. 31, p. 265–275.

Leriche René, 1936. L'être humain. Paris, Comité de l'Encyclopédie française, Larousse. (Encyclopédie française, Tome 6).

Leporrier Michel, 1994. *La petite encyclopédie médicale Hamburger*, 18th ed. Paris, Flammarion Médecine-Sciences, p. 1747.

Lupton Deborah, 1994. *Medicine as culture: illness, disease and the body in Western societies*. London, Sage Publications, 182 p.

Mackenbach Johan P., 1995. Public health epidemiology, *Journal of Epidemiology and Community Health*, vol. 49, p. 333–334.

Montaigne Michel de, reprinted in 1998. Essai 13 De l'expérience, *in: Livre III des Essais de Michel de Montaigne. Présentation, établissement du texte, apparat critique et notes par André Tournon*, p. 425–505. Paris, Imprimerie Nationale Éditions, 683 p. (La Salamandre).

Mordacci Roberto, 1995. Health as an analogical concept, *The Journal of Medicine and Philosophy*, vol. 20, p. 475–497.

Oecd, 1996. *Health data on CD-ROM. Comparative analysis of health systems for Windows*. Paris, OECD (in collaboration with CREDES).

Philippart Fernande and Stroobants Marcelle, 1984. *État civil: décès*, 2nd ed, livre orange no. 5. Brugge, Éditions Vanden Broele, 64 p.

Prior Lindsay and Bloor Mick, 1993. Why people die: social representations of death and its causes, *Science as Culture*, vol. 3(3), p. 346–375.

Rakowski Eric, 1991. *Equal justice*. Oxford, Clarendon Press, 385 p.

Renaud Marc, 1985. De la sociologie médicale à la sociologie de la santé: Trente ans de recherche sur le malade et la maladie, *in:* Jacques Dufresne, Fernand Dumont, and Yves Martin (eds.), *Traité d'anthropologie médicale*, p. 281–291. Silley, Quebec, Presses de l'Université du Quebec, 1245 p.

Thomas Louis-Vincent, 1975. *Anthropologie de la mort*. Paris, Éditions Payot, 540 p.

Thomas Louis-Vincent, 1982. *La mort africaine. Idéologie funéraire en Afrique noire*. Paris, Éditions Payot, 272 p.

Thomas Louis-Vincent, 1985. *Rites de mort. Pour la paix des vivants*. Paris, Librairie Fayard, 294 p.

UNICEF, 1998. *The State of the World's Children*. Oxford, Oxford University Press, 134 p.

Weill Alex and Térré François, 1978. *Droit civil, les personnes, la famille, les incapacités*, 4th ed. Paris, Dalloz, 985 p.

WHO, 1958. *The first ten years. The Health Organization*. Geneva, World Health Organization, 12 p.

WHO, 1986. *Targets for health for all*. Copenhagen, World Health Organization, Regional Office for Europe, 201 p.

Williams Alan, 1996. QALYs and ethics: a health economist's perspective, *Social Science and Medicine*, vol. 43(11), p. 1795–1804.

Zempleni Andras, 1991. Corps blanc dans la famille noire: éthique et biomédecine en Afrique, *in:* Claude Debru (ed.), *Bioéthique et cultures*, p. 97–102. Paris, Librairie philosophique J. Vrin, 156 p.

C H A P T E R

41

Measuring the State of Health

CATHERINE SERMET AND EMMANUELLE CAMBOIS

Institut national d'études démographiques (INED), Paris, France

Measuring the states of health of a population leads to certain questions: What is health? What do we wish to measure, and for what reason? The responses to these interrogations subsequently will determine the nature of the mechanisms that can be devised.

Health status is a continuum from good health to death by way of disease or any other form of physical or mental change or decline. Definitions concerning states of health therefore depend on the definitions chosen to define good health. The World Health Organization (WHO) employs a positive idea of health ("state of complete well-being") that cannot be equated with "an absence of disease or infirmity" (WHO, 1946). Health is also "a state characterized by anatomic, physiologic, and psychologic integrity; ability to perform personally valued family, work, and community roles; ability to deal with physical, biologic, psychological, and social stress; a feeling of well-being; and freedom from the risk of disease and untimely death" (Last *et al*., 2001) or "an essential dimension in the quality of our lives . . . the opportunity to make choices and to gain satisfaction from living" (CDNHW, 1986). These definitions, which are based on the idea of positive health, are satisfactory on a conceptual level but are often considered barely operational because they are too general (Basch, 1990; Young, 1998). Unlike the positive definition of health, most assessments for measuring the health status are negative; they measure diseases and their consequences rather than health itself (Bergner, 1985). At the individual level, every person has a different view of health. Some think of it in terms of symptoms and diseases, whereas others define health in terms of interference with the normal activities of life, deviation from a norm, or an inability to respond to an environmental aggression. Confronted by a patient, the doctor attempts to make a diagnosis, estimate the need for care, and evaluate the results of treatment.

More collectively—and it is this aspect that we consider here—the approaches to the *health of the population* are equally varied. From the point of view of public health, interest is directed toward epidemic, acute, and chronic diseases; their causes and risk factors; and their distribution among the people. This approach attempts to anticipate diseases, reduce their consequences, and promote health. Those who control the health care system and its policies have other perspectives that reflect their preoccupation with the well-being of the population, the seriousness of handicaps, the costs of diseases, the evaluations of inequalities, and their definitions of priorities. Epidemiologists relate diseases to the individual and the environmental and social factors that are prone to change the frequency, distribution, and evolution of diseases and to define the biologic and environmental determinants of diseases. Demographers approach health matters in terms of the consequences on the duration of life (i.e., mortality rates, causes of death, and life expectancy) and knowledge of the social determinants or epidemiologic history of populations.

We therefore need to address two questions: What do we wish to know, and how should we apply this knowledge? Kue Yung (1998) encapsulated the principal reasons for inquiries into the health status of populations: "*to describe, to explain, to predict, to control*." Studies concerned with the health of populations enable the health status to be described, the causes of diseases to be explained, the individual and collective risks to be anticipated, and solutions for anticipating and controlling health problems to be proposed.

The WHO has listed eight ways of using morbidity statistics: control of communicable diseases; planning for development of preventive services; ascertainment of relation to social factors; planning for provision of adequate treatment services; estimation of economic importance of sickness; research into cause and pathogenesis; research on efficacy of preventive and therapeutic measures; and national and international study of distribution of diseases and impairments (Basch, 1990).

These various measurement targets and the uses that arise from them show that there cannot be a sole indicator for gauging a population's health status. Four large categories may nevertheless be distinguished: demographic data (e.g., age, sex), statistics on vital events (e.g., births, deaths), health statistics (e.g., morbidity, impairments, perceived health), and statistics relating to the health services (e.g., use of the services, indicators of medical supply) (Basch, 1990). To judge the realization of the objective proposed in *Health for All in the Year 2000*, the WHO asked all member countries to provide a series of indicators covering all the following categories: life expectancy at birth; infant mortality rate; proportion of persons aged 65 and older; percentage of gross national product devoted to health; and the number of cases of diseases such as acquired immunodeficiency syndrome (AIDS), poliomyelitis, and measles.

With these indicators, health statistics inform directly about the health status of populations, and the other three categories provide indirect information. These specific health indicators are evaluated in this chapter.

I. HOW TO MEASURE HEALTH STATUS

Until the beginning of the 1960s (Bergner, 1985), death rates formed the principal measure for the health of populations. With the lengthening of life expectancy, it became progressively evident that mortality no longer sufficed for measuring the changes that had come about in health and medical care. It was then that health surveys enabled the collection of numerous indicators of health status, environmental conditions, available health care, biochemical and physiologic measures, and days of inability to work.

In the 1970s, synthetic indicators appeared that were founded, for instance, on the combination of mortality rates and data on the incidence or prevalence of diseases or on distinguishing the years of life expectancy according to their quality. Some indicators of impairment or functional disability were also perfected, and they ranged from simple measures of physical limitations to the inclusion of a much wider field of physical, social, and psychological functions. Although they had been tested and validated before being widely used in many studies, very few had been routinely employed to measure the development of the state of population health over long periods. Indicators of mental health had been developed more quickly, probably because of the greater difficulties in providing precise diagnoses in this domain (Bergner, 1985).

Parallel to the development of general indicators, clinicians perfected their own evaluation tools for their patients. Directed toward diagnosis or surveillance, some of them had then achieved wider employment in more extensive population studies. There has been particular interest in measures of the quality of life, covering a considerable number of physical, social, and psychological dimensions and enabling many of them to calculate synthetic quality of life scores.

As suggested by Mildred Blaxter (1989), the various instruments for measuring the health status can be placed in three categories. The first set is determined according to a medical or biologic model of the disease, which defines the poor health status as a divergence from a physiologic or psychological norm. The second corresponds to a social and functional model in which the poor health status is defined as an inability to fulfill normal tasks or roles. In the third category, the measure of the subjective health status predominates, a model in which individual perception is paramount (Table 41–1). This idea of subjectivity merits clarification. Essentially, what is recorded in the subjective model is the deep feeling of the individual, his symptoms, and his physical or mental identification of health problems. In the medical model, as in the functional model, the account of the information may rest on the declaration of individuals and therefore be partly subjective, but the content of the information has a basis that can be objective, such as a diagnosis of disease or functional incapacity.

This classification system reflects the differences in the three English language terms employed when speaking of worsening of the health status (Blaxter, 1989). *Disease* represents the functional and structural perturbations of the human body that produce clinical signs and symptoms that are labeled as deviation from the norm (Young, 1998). This term generally refers to a biomedical nomenclature, and the diagnosis of diseases is within the province of physicians. It is essentially

TABLE 41–1 Commonly Employed Indicators for Summarizing States of Health

Medical or physiologic and psychiatric model	Social-interaction or functional model	Subjective or illness model
Clinical examination, physiologic or psychiatric screening for abnormality	Test of physical or psychological disability	Self-assessment of health
and/or Medical diagnosis of physical or psychiatric disease	*and/or* Ascertainment of functional status associated with the ill health (e.g., bedridden, not working)	*and/or* Reported experience of physical symptoms of ill health
and/or Self-report of the existence of medically defined diseases or abnormality	*and/or* Inability to perform normal tasks because of diseases, impairments, or illness	*and/or* Reported experience of psychosocial malaise

From Blaxter, 1989.

within the scope of Baxter's medical model. *Illness* represents the subjective experience of diseases, which Mildred Blaxter calls a *subjective model.* It is the perception of a change in the habitual functions or feelings of an individual. Disease can exist without the patient feeling ill, as in cases of hypertension. A person also can feel discomfort that cannot be expressed in diagnostic terms. *Sickness* is the societal response to diseases, and the term expresses the consequences of the latter on the relations of the individual with other people. This term is used more within the functional model, although this does not suffice to describe it completely.

II. THE MEDICAL MODEL: MEASURING MORBIDITY

1. Definitions

a. Morbidity

For *morbidity*, the epidemiologic dictionary offers the following definition (Last *et al.*, 1995): morbidity is "any departure, subjective or objective, from a state of physiological or psychological well-being." This approach covers the whole range of health, and it refers equally to medical diagnoses, functional disabilities, and subjective health.

The same source indicates that morbidity can be measured according to three units: the persons who were ill, the illnesses that theses persons experienced, and the duration of these illnesses. The term *morbidity surveys* makes explicit reference to the frequency of diseases within the population. This measure of morbidity is primarily a medical measure, because ideally, we want to count the diseases that are confirmed by a medical diagnosis and whose denomination necessitates highly encoded nomenclatures. In practice, the assessment is often based on the declarations of individuals, and it is this definition that we employ in this section.

b. Incidence and Prevalence

Incidence and prevalence are the two basic indicators for measuring the frequency of diseases in the population. Strictly speaking, the *prevalence* of a disease is the proportion observed at a given moment of people suffering from this disease. It represents the *stock* of diseases prevailing at a given moment, and it equals the number of sick people recorded in relation to the total population.

Incidence expresses the frequency of *new* cases of a given disease (i.e., persons newly attacked) during a given period within a certain population. It equals the number of *events* commencing during the period under consideration relative to the average population during the period.

When the population is stable, there exists a connection between the incidence (I) and the prevalence (P), which is expressed as follows (Bouyer *et al.*, 1995):

$$P_{t+1} = \frac{I_{t,t+1} \times d}{1 + (I_{t,t+1} \times d)}$$

In the previous equation, d is the average duration of the disease, t+1 is the date on which the prevalence is observed, and t,t+1 is the period of observation of the incidence. In the particular case when the incidence and prevalence are constant and when the prevalence is weak (<10%), the prevalence can be approximated as follows (Freeman, 1980; Bouyer, 1995):

$$P_{t+1} = I_{t,t+1} \times d$$

The prevalence of a disease is therefore a function of its duration, which sometimes makes the interpretation of its variations, in terms of the health status of the population, difficult. Eileen Crimmins (1996) stresses that if life expectancy lengthens as a result of a diminution in the incidence of diseases; it can effectively be maintained that the health status has improved. On the contrary, if it is longer because of an increased survival of the sick, it is not accompanied by an improvement in the health status.

The latter affirmation appears to be open to discussion. Improvement in medical care and therapeutic progress may prevent a certain number of deaths without curing the sick persons. Their survival is therefore prolonged, and prevalence increases. Can it be said that there is a decline in the health status of the population because the number of sick people has increased, or should it be maintained that there is an improvement, because the diseases that were hitherto mortal no longer systematically lead to death (Sermet, 1998)? The reply is subjective and depends on the quality of life experienced by people with the disease. To answer these questions, indicators were devised for healthy life expectancy—with or without chronic diseases, but also with or without disability or in good health as subjectively perceived. The WHO modeled this problem area and presented various possible developments in the health of the populations according to the evolution of the survival curves for the different states of health (WHO, 1984).

2. The Measure of Morbidity

The measure of morbidity of a population encounters two major difficulties: the definition of the disease and the type of morbidity measured. Definition of the disease immediately questions the very concept of normal and pathologic. Where does disease commence, and where does normality end? That which is normal for one population in a given culture is not inevitably so in another. What was normal 10 years ago might have become abnormal at the end of a decade. What is normal at the age of 60 years is not always so at 20 years. A normal clinical state may mask a complaint at a different clinical stage. When does a disease begin and end? When is a patient cured? In reality, there is a continuum between normal and pathologic states at the inception of a disease and at the end of it. Measuring morbidity bears a resemblance to measuring the surface of a field whose boundaries are unknown.

Recognition of a state as constituting a disease is far from being universal, and examples of cultural variations are numerous. For example, the hematuria that accompanies *Schistosoma* infection, a parasitic disease endemic in certain regions of Africa, is seen in certain cultures as a normal phenomenon, not requiring medical attention. In this situation, a record of morbidity based on the declarations of individuals cannot enable this symptom and the disease to be recognized as such. This example leads us to consider the question of the norm. We have become accustomed to consider the biomedical model developed in Western societies as the universal norm, but certain people consider that this model is only a reflection of our culture and that *reality* and *objectivity* are only social constructions (Young, 1998). For this reason, they are difficult to transpose from one society to another.

The role of cultural phenomena is even more important when it concerns the state of mental health. Many authorities emphasize the existence of syndromes that are closely linked to particular cultures. These *culture-bound syndromes* constitute entirely separate diseases in these societies, but they are not found in the models or classifications of other societies. The *mal ojo* of children in the Mediterranean and South American populations attributes sleep problems, crying, or diarrhea to the diabolic glance from a stranger, the evil eye. On the other hand, pathologies such as bulimia and anorexia are accepted in Western countries (Basch, 1990).

The temporal definition of the disease also is subject to discussion. Inasmuch as it is simple to establish the beginning and end of an acute disease such as influenza, it is just as difficult to determine the start and finish of a disease such as cancer. How should we date the commencement of the tumor when it might have been developing insidiously for years without being detected (not even clinically detectable), and how should we date its termination when metastases may occur several years after the treatment of the initial tumor?

The second difficulty to be confronted in the analysis of morbidity data depends on the definition adopted for the type of morbidity being studied. Morbidity based on the declarations of individuals represents only a partial view of the pathology because it is subject to voluntary or involuntary omissions, the possibility of considerable variations in individual perceptions of the disease, and sociocultural and psychological factors. Diagnosed morbidity requires contact with a doctor, and it covers only the pathology that motivated the person to seek treatment or that was revealed during a medical examination.

Three components of morbidity can be described according to the source of information (Fig. 41–1). Diagnosed morbidity, measured morbidity, and self-reported morbidity all exist within the framework of

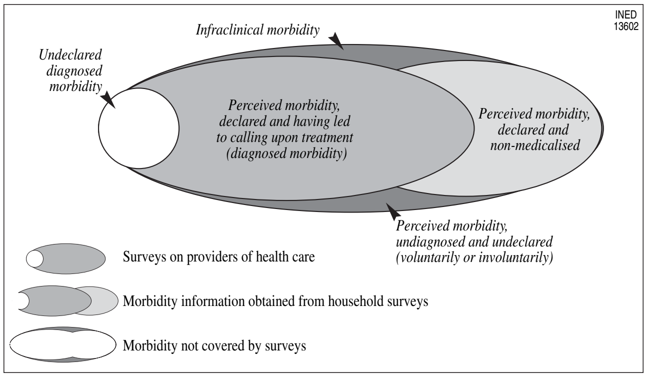


FIGURE 41–1 Types of morbidity. (From Sermet, 1994.)

the medical model described by Mildred Blaxter (1989).

Diagnosed morbidity corresponds to illnesses among individuals who have consulted doctors and been diagnosed and treated by the medical corps. A certain amount of data of this type is routinely recorded in many countries, such as notification of infectious diseases, registers of chronic diseases (particularly cancer), and hospital statistics, diseases, and accidents related to work. Only surveys conducted among doctors can provide pertinent data about diagnosed morbidity.

Measured morbidity can be objectified by clinical or paraclinical health examinations of the population. It forms a subsection of diagnosed morbidity.

Some large demographic surveys on health enhanced by a medical examination have been carried out on the whole or part of the sample. One of the first such surveys, the National Health Examination Survey (NHES), was begun in the United States in 1960. At first, it concerned only a few diseases and enrolled individuals between 18 and 79 years old. It was then extended to become the National Health and Nutrition Examination Survey (NHANES), which has been regularly repeated since 1974. Health examination surveys also exist in Canada (Canada Health Survey) and in various European countries, including Germany, Finland, Spain, and Great Britain. A European Commission study examined these surveys to evaluate the collected data (Koponen and Aromaa, 2001). The report showed that although costly and difficult to set up, this method of collecting data is increasingly used. A health examination will be integrated in the next French, Italian, Norwegian, and Portuguese health surveys.

One of the main difficulties in setting up such studies lies in the complexity at a national level, for example, of a standardized medical examination and of the limited number of diseases whose diagnosis depends on simple, noninvasive, and inexpensive examinations that can be applied to a large sample. Only a few illnesses, such as hypertension, or diseases whose diagnosis is biologic, such as diabetes, are easily measurable.

Frequently, national and regional surveys that include medical examinations confine themselves to a particular pathology, such as mental health in the Netherlands or diabetes in Greece. Surveys of this type have also been carried out on a European or international level; such is the case with the multicentric surveys of cardiovascular diseases (MONICA), osteoporosis (EVOS), or cancers (EPIC).

Self reported morbidity is information that can be obtained directly from individuals. It is also a partial and subjective view of the pathology, because no one can claim to know all the diseases with which he or she is afflicted, particularly those whose clinical presence is not yet detectable, and because of omissions and intentional nondeclarations. By definition, the data are subjective, because all people have their own conceptions of what is normal and pathologic and their own scale of values in ordering illnesses and naming them.

Self-report of morbidity is most frequently carried out through population surveys, commonly called *health surveys*. In France, those that are directed toward the general population include the Decennial Survey on Health and Medical Care of the National Institute of Statistics and Economic Studies (INSEE) and the Health Care and Health Insurance Survey of the Research and Information Institute for Health Economics (IRDES, formerly Centre de Recherhce, d'Etude et de Documentation en Economie de la Sante [CREDES]). Many countries in the world conduct surveys of this type. The European Commission has launched a program listing health interview surveys, which has enabled the compilation of about 40 national surveys of this type in Europe (Hupkens and Swinkels, 2001).

Many factors influence the declared level of diseases without affecting real morbidity. There are the individual factors, with the cultural and social characteristics of the individuals and their level of medical knowledge and their own health status. There also are factors associated with social organization, the environment, and medical development. The development of diagnostic methods, as well as those of the supply and quality of treatment, has enabled certain diseases to be detected more easily and to be identified more precisely. For example, the improvement in detection and the increased use of treatment for elderly people have greatly contributed to the prevalence of hypertension diagnosis in the United States. In France between 1980 and 1991, the number of lipid metabolism disorders reported by individuals was multiplied by three, indicating an increase in medical knowledge and improvement in diagnostic methods (Sermet, 1998). Societal changes also have amplified health-related expectations and requirements, as on the frontier between normal aging and geriatric pathology. Arthritis and dementia, formerly regarded as the signs of normal aging, are now designated diseases (Crimmins, 1996).

The characteristics of the illnesses themselves have influenced their declaration on surveys. Although it may be easy to recognize the presence of an acute disease whose symptoms may be evident and well delimited in time, what should be done in the case of diseases that manifest only sporadically? Sciatica, for example, is characterized by its successions of acute periods and more or less prolonged phases during which the anatomic damage persists but there are no clinical manifestations. The same can be said for other illnesses, such as myocardial infarction, transitory ischemic cerebral accidents, and seasonal allergic rhinitis. In this type of disease, individual attitudes vary considerably; some people declare the diseases, whereas others fail to do so when the diseases are not active.

As evidenced by the European inventory of health surveys, many methods are employed during the surveys to ask people about their diseases. There are questions of a general nature, such as that recommended by the European Office of the WHO within its harmonization program of measuring instruments for surveys by interview (Euro-HIS): “Do you have any long-standing illness or health problem?” Numerous variations of this question are to be found in the various surveys, and a question of this type has figured since 1990 in the United Kingdom census questionnaire. In other cases, to encourage those concerned to make declarations, lists of diseases or disorders are made available during the surveys. The methods employed in gathering the information are not without an effect on the declarations, such as the type of question (open or closed) and the recall period. Mental diseases constitute a particular case in the area of self-reported morbidity, because the disablement they cause and the taboos and misunderstandings of which they are the objects frequently lead to their being under-reported. Rating scales have been developed to enable mental diseases to be traced indirectly through the declarations of symptoms, such as the *Self-Rating Depression Scale* (Zung *et al.*, 1965), or by measuring their functional consequences, as in the *Mini-Mental State Examination* (Folstein *et al.*, 1975).

III. THE FUNCTIONAL MODEL: THE MEASURE OF DISABILITY

Because of the demographic evolutions at work and the increasing proportion of old and very old people, questions relating to the functional state of health have become decisive for public health. The functional model approaches health in terms of the consequences of diseases, accidents, or physiologic aging on the functioning of the individual in daily life. Consequently, the question that arises at the level of morbidity concerning the consequences of increased life expectancy applies equally to the level of functional health. As suggested by the WHO (WHO, 1985), a person must not only be “adding years to life, but life to years.” Authorities should appreciate the burden of disability and dependence that has evolved within the population. Some biomedical models have been developed to provide a conceptual framework for these investigations. They have enabled a precise meaning to be attributed to the terms *disability*, *handicap*, *dependency*, and *impairment* that previously were often confused. Models have placed these concepts within a dynamic process that includes various levels, at which the repercussions of the disease, accident, or aging are expressed in terms of the body’s functioning, daily activities, and social integration.

1. Concepts and Definitions

Saad Nagi (1965) developed the first conceptual works devoted to the disablement process. His aim was for a better definition of these concepts in the areas of functional health. From these activities and a revision in 1976, several important definitions emerged to describe the different levels of the disablement process. *Impairment* was defined as an anatomic, physiologic, intellectual, or emotional abnormality or loss (Nagi, 1976). Nagi then defined the *level of organismic performances*, for which he distinguished three dimensions: physical, emotional, and mental. *Disability* was defined as an inability or limitation in carrying out the activities and social roles related to work, the family, and an independent life. At the same time, the work of Philip Wood and Elisabeth Badley (1978) led to the development of the WHO *International Classification of Impairments, Disabilities, and Handicaps* (WHO, 1980). The WHO approach distinguishes the following levels: impairment, disability, and handicap.

a. Impairment

Impairment is defined (WHO, 1980) as "any loss or abnormality of psychological, physiological, or anatomical structure or function." The abnormalities are of a physical, mental, or sensorial nature, and according to the expression used by the WHO, they are the "exteriorized" form of the disease, malformation, or accident. Sometimes, the disease is assimilated to this exteriorized form, or in other words, to the impairment that it entails. Confusion then arises between the source and its consequences on the psychological, physiologic, and anatomic functions. Macular degeneration is a good example. The name of the disease is nothing other than the impairment it provokes, and it is difficult to clearly define the frontier between the impairment and its cause.

b. Disability

Disability corresponds to "any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being" (WHO, 1980). For the WHO, this limitation of activity is the "objectified" form of the disease, accident, or malformation. For example, macular degeneration can lead to a curtailing of activities, such as reading or driving, that are considered to be essential for the life of individuals, their work, social integration, and leisure activities.

In the first model proposed by Philip Wood and Elisabeth Badley (1978), a distinction was made within disability between functional limitations and activity restrictions. This distinction is related to that which Saad Nagi proposed in his model when defining the performance of the organism on one hand and activities on the other. Later studies based on numerous data collected in this domain have demonstrated that differentiating functional limitations from activity restrictions at the level of disability enables vital public health questions about taking care of the disability and dependence of people to be addressed (Robine *et al.*, 2000).

Functional limitation is situated at the level of the physical, sensorial, or mental body functions. It is the result of a disease, accident, aging, and consequent impairment on the functional abilities of the individual, independently of the environment or available technical aids. Functional limitation is expressed through difficulties in mobilizing body functions to, for example, walk, extend an arm, hear, or remember. In the example of macular degeneration, a functional limitation associated with this impairment would be difficulty in seeing. However, the localizations are not necessarily identical, and impairments linked to brain injuries, for example, may occasion motor, sensory, or mental functional limitations. Functional limitations, whether reversible or definitive, form intrinsic characteristics of individuals and determine their functional state. Environmental factors, personal resources, or technical aids, however, may modify the effect of functional limitations on the performances of the individual for daily activities.

Activity restrictions correspond to the individual's dysfunction in daily activities (e.g., personal care, household activities, work), and they depend on the person's functional status and environment. The WHO definition of disability is related to so-called complex activities. Macular degeneration that entails difficulties in seeing may compel the individual to limit activities such as driving a car or shopping.

In clarifying and ordering these two levels, a distinction can be made between people having functional limitations but who manage to compensate for them and those for whom the functional limitations result in restriction of the activities of daily life. This means that account may be taken of environmental influence and the personal resources of individuals that favor the adoption (or not) of a compensatory strategy. The literature on the subject presents functional limitation as a *preclinical stage* or predictive of activity restrictions (Fried *et al.*, 1996; Lawrence and Jette, 1996). Studies have given prominence to the determinative role of functional limitations in the disablement process (Fried *et al.*, 1996; Harwood, 1998; Lawrence and Jette, 1996; Tager, 1998). Mobility limitations, for example, have proved to be more predictive of activity restrictions than other categories of limitations (Lawrence and Jette, 1996; Manton *et al.*, 1998). They may also generate more recourse to the health services (Wolinsky and Johnson, 1991). Although fundamental, this distinction was not adopted by the WHO in its classification, and it reappeared only in the 1990s through the initiative of the Council of Europe (Council of Europe, 1990) and then on the occasion of later studies.

c. Handicap

Handicap for a given individual results from "an impairment or a disability that limits or prevents the fulfillment of a normal role (depending on age, sex and social and cultural factors)" (WHO, 1980). According to the WHO, handicap is a "socialized" form of disease, accident, or malformation. The WHO proposes six roles according to which handicap may be evaluated, among which are *minimum independence* (i.e., to be independent of someone's assistance for personal care activities), *independent mobility*, and *social* *integration*. The combination of several activities, such as washing oneself, feeding oneself, and going to the toilet, enables these roles to be carried out for the role of minimum independence. For this reason, the frontier between disability and handicap is blurred. Disability is usually measured by the activities that comprise these roles (e.g., personal care, professional activity) and as a consequence, while measuring disability, a handicap is at the same time being suggested.

In a classic scenario, a handicap results from a disability, which is itself provoked by an impairment. However, the WHO classification makes clear that in certain cases this sequence no longer exists or does not exist. In the realm of mental health or malformations, impairment may be the cause of a handicap without inducing a flagrant disability, but general behavior or stigmatization of the disease may form an obstacle to social integration.

The several reference models developed in the 1970s have generated discussion and debate. Newer models and definitions have been suggested for integrating the essential factors involved in the process (e.g., environment) or for defining more accurately the frontiers between the steps, which are sometimes difficult to discern (Robine *et al.*, 1997). In this context, the WHO classification, which served as a reference model for many years, has been revised to address the various criticisms of its original version. The main development to be found in the new version (*International Classification of Functioning, Disability, and Health*, [WHO, 2001]) has been better weighting for the effect of the contextual factors, both environmental and personal, on the functioning of individuals. It no longer rests directly on the principle of the disablement process, but it instead provides blocks of definitions that enable the functioning problems to be classified. As for the classification of diseases, the newer version provides a nomenclature for the problems of functioning:

Impairments are problems in body functions or structure such as significant deviation or loss. *Body functions* are the physiologic functions of body systems (including psychological functions). Body structures are the anatomic parts of the body, such as organs, limbs, and their components.

Activity limitations are difficulties an individual may have in executing *activities*, defined as execution of a task or action by an individual. *Participation limitations* are problems an individual may experience in involvement in life situations.

Environmental factors make up the physical, social, and *attitudinal* environment in which people live and conduct their lives.

The new version of the WHO classification does not explicitly mention the distinction between the functional limitations and activity restrictions. As with *disability* in the previous version, the current level of *activity limitations* covers both of these concepts, and what the WHO now calls *limitation of participation* designates the consequences of these activity limitations in the real life situation.

2. Measuring Instruments

a. Measuring the Various Concepts of the Disablement Process

As for morbidity, the various concepts of functional health may be understood through individual self-reports in interview surveys or by measurement during clinical examinations.

Impairment

Impairment is close to disease and constitutes the first consequence of it. The divide between disease and impairment is blurred, and for that reason, measurement of the latter is not easy. When health is measured through diagnosis, impairments may be approached on the basis of the diagnosed diseases or injuries by questioning about the commonly observed consequences of these diseases or injuries (e.g., altered visual systems, loss of memory). However, when the measurement is based on self-reported health, impairment is a level of alteration that is difficult to identify and evaluate. A number of surveys, such as the French Health Survey or the disability survey (i.e., Handicaps, Incapacités, Dépendance), include questions about disability and impairment, which is considered the source of the disability. These surveys list examples such as motor and auditive impairments, directing the person toward the level of body alteration. It is difficult, however, in this type of questioning to differentiate diseases, impairments, and functional limitations.

Disability

The aim is to measure the ability of individuals to carry out activities in terms of their state of health (i.e., diseases or impairments). For this purpose, scales were employed within the field of clinical research well before development models of disability came on the scene. In 1963, Sidney Katz and others set up the scale of elementary activities in daily life (i.e., activities of daily living [ADLs]), which was based on observing the development of independence in children: feeding themselves, dressing and undressing, showering or bathing, going from bed to armchair, using the toilet, and continence. This scale was later enhanced by Powell Lawton and Elaine Brody (1969) with so-called *instrumental* activities of daily life to measure the ability of individuals to live alone (i.e., instrumental activities of daily living [IADLs]): doing the housework, preparing meals, keeping the accounts, and performing other chores. In 1976, Saad Nagi introduced a scale for evaluating what he called *bodily performance* (i.e., functional limitations). The typical survey questions were about difficulties in activities such as seeing the printed characters of a newspaper, walking 500 meters, and going up and down stairs. The various scales were transformed into questionnaires and used in health interview surveys. Today, most of the health surveys conducted in developed countries include such question modules to evaluate disability and often target the elderly population (Cambois and Robine, 1994; Cambois *et al*., 2000; Wiener *et al*., 1990).

For organizing this field of research and providing the means of comparing the results of surveys, standardized instruments were proposed or recommended by bodies as the Organization for Economic Co-operation and Development (OECD) (McWhinnie, 1981) or the WHO (WHO and Statistics Netherlands, 1996). Certain instruments, particularly that recommended by the OECD, include simultaneously the ADLs, IADLs, and functional limitations such as those suggested by Saad Nagi (1976). Amalgamation of these different scales within one instrument makes the interpretation of the collected information difficult. What can be concluded in terms of the need for aid and care when classified together are a difficulty in *going up and down stairs* and a difficulty in *feeding oneself*? Disentangling the levels of functional limitations and activity restrictions enables prominence to be given to the disablement process; it allows consideration of the possibility that limitations are predictive of activity restriction and then consideration of the possible different needs for help and care associated with these two levels.

Measuring Functional Limitations The functioning of the individual is sought within different categories of functional limitations (e.g., physical, sensory, mental), based on the difficulties expressed (e.g., seeing, walking, remembering), the resort to a technical devices, or the reliance on personal assistance. In a survey based on interviews, the principle is to help the person to self-report possible limitations by suggesting to her or him some gestures or scenarios requiring various functions to figure out possible alterations, such as "bending the arm to shake hands" or "hearing what is said during a conversation." These questions constitute only suggested situations not indispensable in everyday life. Therefore, concerning the formulation of this type of questions, the individual must not be interrogated in terms of effective performance (Do you . . . ?), but should be asked in terms of ability (Can you . . . ?) (Glass, 1998). Different instruments have been suggested on the basis of Saad Nagi's proposals in 1976. A review of these instruments shows that functional limitations can be put into the following categories: physical functional limitations (e.g., balance, mobility, strength, endurance, agility), sensory limitations (e.g., vision, hearing), and mental limitations (e.g., memory, comprehension) (Coutton *et al*., 2000). Each of these categories can be understood on the basis of one or several questions: "Can you go up and down stairs without difficulty? Can you walk 500 meters without difficulty?" These functional limitations can also be tested during a medical examination by asking the individual to carry out certain gestures. This type of test is common in the practice of gerontology. Because of their specificity, mental functional limitations are most frequently tested in such a way, using instruments such as the *Mini-Mental State Examination* (Folstein *et al*., 1975) or small exercises involving reminding about the date of an important event, given the name of the present head of state, or counting backward.

In the interview surveys, the questions may be asked in two stages. The first is to find out the intrinsic functional level of the individual (i.e., excluding an adapted environment and without resort to specific aids or devices). These questions measure the intrinsic functional condition and enable a comparison of the functional state of the population in time or comparison of population groups. Second, by introducing the use of a device, it is possible to measure the effective functional state of individuals (i.e., their intrinsic state corrected by the aids available to them that may prevent further activity restrictions despite functional limitations).

Measuring Activity Restrictions Today, the reference activities for measuring activity restrictions are still those of the ADLs proposed by Sidney Katz (1963) and adopted in the recommendation by WHO Regional Office for Europe (1996). This indicator is generally employed for the older population, for whom it had been developed. Nevertheless, the permanent health survey in the United States, the National Health Interview Survey (NHIS), asks the population from the age of 5 years or older questions relating to activity restrictions of the ADL type. Activity restrictions may also be measured by the IADLs proposed by Powel Lawton to enlarge the ADL approach for assessing the chore daily activities, but what these activities really measure is now debated

(Robine and Jagger, 2000). Some studies suggest that the IADL assessment does more than evaluate the ability to live alone, because activities such as managing accounts or writing checks demand a certain degree of mental lucidity and could be useful to detect restriction due to cognitive functional limitations (Wolinsky and Johnson, 1991). Studies of these indicators have also shown the existence of a veritable hierarchy between the ADL and IADL assessments (Ferrucci *et al.*, 1998; Finch *et al.*, 1995; Lazaridis *et al.*, 1994; Spector and Fleishmen, 1998). A scale of severity of the restrictions can be constructed by ordering these activities. Difficulties in feeding oneself generally reveal the most severe level of disability, and difficulties in showering or bathing (without other difficulties) indicate the most moderate level among the ADLs, therefore a deterioration process can be observed along the scale. In the same way, the IADL assessment can indicate a less severe level of functional deterioration, except for using the telephone, which approaches the severity levels of the ADL type.

Activity restrictions are measured through the interview surveys by using a formulation based on the effective performances of the individual such as, "Do you feed yourself on your own without difficulty?" It is also possible to incorporate the use of a technical aid or the assistance of another person (Cambois *et al.*, 2000). Observation through performance tests, as described for functional limitations, does not furnish reliable information in this field (Kempen *et al.*, 1996; Reuben *et al.*, 1995); simulating activities such as feeding, preparing of a meal, or getting dressed provides more information about functional limitations than about the difficulties encountered in daily life to be independent in these activities in the person's own environment. These tests constitute laboratory conditions, which have very little to do with real life.

The soundness of instruments for measuring disability has been widely studied in relation, for example, to their sensitivity and reproducibility (Fried *et al.*, 1996; Kempen *et al.*, 1996; Reuben *et al.*, 1995; Tager *et al.*, 1998). The choice of items used and the formulation of the questions in measuring the functional limitations and the activity restrictions are important. The information can be refined by employing the quality criteria suggested by these studies.

Handicap or social participation is inferred from the activity restrictions, because the scales used are made up of activities considered to be necessary for ensuring social integration. Restrictions on activities for personal care may indicate the loss of minimum independence; restrictions on instrumental activities reveal difficulties in living alone at home. Handicap is frequently inferred from the restrictions on professional activities. This is most often measured through a synthetic question, and many surveys employ a single question for evaluating activity restrictions or handicap. However, the level of handicap may be studied more specifically. Within the sphere of the current European program, specific questions on social participation are being considered to estimate inequalities in the degree of social integration of persons with or without disability (e.g., work, leisure).

b. Synthetic Questions

Along with the scales, functional health can be evaluated through general questions about limitations in usual activities due to health problems. These questions are to be found in various forms in most health surveys (Perenboom *et al.*, 2000). They sometimes concern short-term limitations occurring during the 15 previous days or consider long-term problems by excluding transient problems. Although the scales presented previously are mostly employed for evaluating long-term disabilities, synthetic questions apply as much to long-term as to short-term problems. The information collected does not have the same use in both cases; the choice of questions depends on what the information is used to measure. The public health concern about care for disability and dependence tends to be evaluated through questions about long-term problems. In this context, a review of the indicators demonstrates that a period of more than 6 months enables transient problems to be screened out. Global questions about limitations in professional activity are also to be found in surveys, and in this case, the interrogations are related to the interruptions of activity because of a health problem or the modifications in the quantity and type of activity because of a health problem.

c. International Comparability

Relying on the international recommendations for disability measurement instruments, health surveys throughout the world have included such instruments in their questionnaires, adapting the formulation of the questions and the number of activities considered or sometimes adapting the type of activities considered according to their needs or to the place available in the surveys. The formulation of the questions and the choice of activities considered are important, and the variation from one survey to another on certain aspects makes the collected information difficult to compare. The European Union is developing research programs for evaluating the degree of comparability between these measures in Europe and concentrating on those that should be improved and recommended to obtain the most precise information. These studies have shown that it is difficult to implement recommendations such as those of the WHO and that the developed questions do not overall necessarily fulfill the quality criteria of such instruments. Proposals have been made in these European programs for promoting instruments that depend on current scientific knowledge (i.e., distinction between functional limitations and activity restriction, formulation of the questions, and hierarchy in the items) (Robine *et al*., 2000).

IV. THE SUBJECTIVE MODEL: PERCEIVED HEALTH AND QUALITY OF LIFE

1. Concepts and Definitions

Perceived health is a subjective measure that conveys the way in which individuals perceive their health status. It reflects the feeling, ideas, and beliefs held by the individuals concerning their health, and it does not necessarily represent the objective state of health. This perception varies according to the person. Two people with the same general state of health may have very different perceptions of it. As for the declaration of the diseases, perceived health is conditioned by the expectations of individuals concerning health, their general medical knowledge, and the information available to them concerning their objective state of health—diagnosed or undiagnosed disease and knowledge of the nature and severity of the complaint. Psychological problems also have an influence on the perception of the disease and may introduce symptoms that reduce the sense of well-being. Psychological problems may also provoke an exacerbation of signs that leads to a belief in a preexisting pathology (Connelly *et al*., 1989).

Perceived health can be measured by various types of instruments, and the data collected informs about its various dimensions, such as quality of life, well-being, and physical and mental symptoms. Synthetic instruments enable wider evaluations to be made.

2. Measuring Instruments

a. Global Approach of Perceived Health

One of the most frequently used instruments for assessing the subjective state of health is the question commonly entitled *perceived general health*. It involves assembling the general feelings of persons about their own state of health in relation to their expectations. It appears that one of the first questions of this type was asked in 1963: "How would you rate your health at the present time? (Excellent—Excellent for my age—Good—Good for my age—Fair—Fair for my age—Poor—Very Poor)" (Heyman and Jeffers, 1963).

A review of the instruments in current use and the investigations associated with them reveals the many formulations existing in the various surveys throughout the world (Sanderson *et al*., 2000). The terms used in the question and the number and type of response items proposed initiate a variety of replies and preclude comparisons between surveys. For the purpose of encouraging comparisons, the European Office of the WHO has suggested a standardized question (WHO and Statistics Netherlands, 1996): "How is your health in general?" It has five levels of response: very good, good, fair, bad, and very bad.

A particular method of response to this general type of question is the analogic visual scale. Individuals must locate their state of health on a thermometer-type vertical scale whose 0 point usually indicates death and the 100-point indicates the best possible state of health.

The interest of this question lies in its simplicity; it is easy to understand and administer, and it is easy to analyze and very informative. It includes all the physical, emotional, and social factors that enable individuals to judge their health status. Studies first revealed the correlation between replies to this type of question and mortality (Kaplan and Camacho, 1983; La Rue *et al*., 1979; Mossey and Shapiro, 1982). Many studies have subsequently confirmed this relationship, while showing a better correlation for men than for women (Idler and Benyamini, 1997). Perceived health is correlated with other health indicators, such as disability, chronic diseases, and symptoms (Kristensen *et al*., 1998) or even the consumption of medical care (Bergner, 1985).

b. Physical and Mental Symptoms

One of the methods used for measuring perceived health, associated or not with averred pathologies, consists of questioning individuals from lists of symptoms or requiring a more or less lengthy account of the symptoms using agenda-type techniques (Blaxter, 1989). Some surveys in Europe have collected symptoms according to one or the other method. The *General Household Survey* in 1977, the *British Health and Lifestyle Survey*, and the survey by the National Institute of Statistics and Economic Studies on living conditions in France in 1987 include questions on physical symptoms, such as fever, constipation, sour stomach, or others pains.

The last Decennial Health Survey in France (1991 to 1992) included a list of symptoms within the sphere of minor mental health problems, such as anxiety, sleep problems, fear of the future, loneliness, and nervousness. These types of symptoms relating to mental well-being have aroused considerable interest, and many indicators have been devised, ranging from simple lists to much more elaborate and validated rating scales. The *General Health Questionnaire* (GHQ) is a self-administered questionnaire comprising 28 questions about psychological symptoms and social insertion, enabling a total score and four subsidiary scores to be obtained (Goldberg and Hillier, 1979). As with the GHQ, there are numerous mental rating scales that inquire about symptoms of this type, and their intention is frequently to detect real pathologies, not just a feeling of discomfort. Some of these are disease specific, and they are about diseases that are very prevalent, incur a risk of prolonged hospitalization, or have a high mortality rate (Robine *et al.*, 2000).

c. Quality of Life

A complete review of health indicators must include the indicators of the quality of life. Their importance has grown in proportion to the increasing need to control the costs of health systems, and they have become invaluable for economic evaluations of costs in health systems. Paul Dolan (2000) describes four applications for these indicators of life quality: in randomized, controlled trials (e.g., to evaluate new therapies or technologies); in individual patient care to improve clinical practice; in describing, monitoring, and predicting overall levels of population health; and in informing resource allocation decision in health care. This latter category is not specific to the indicators of quality of life, but it is often proposed within this framework. These instruments also fulfill a frequently expressed need for synthetic indicators to incorporate diseases, functional state, mental state, and diverse measures of health status.

Various devices for measuring the quality of life have been developed. Alain Leplège (1999) highlights the existence of two types of indicators. The first of these, health-related quality of life indicators, is intended for measuring the effect of diseases on the quality of life. The assumption is that it is possible to separate what affects the quality of life from what does not. For the most part, these instruments combine four main dimensions:

1. The physical state of the subjects
2. Their somatic feelings
3. Their psychological state
4. Their social relations and relationship with their environment

These factors give considerable weight to functional abilities, according to the assumption that a person having difficulties in walking, for example, has a diminished quality of life compared with one who has no difficulties. A certain number of approaches combine life expectancy and health-related quality of life indicators. The *disability-adjusted life years* (DALYs) evaluate the loss associated with disability and premature death in terms of healthy life years. The *quality-adjusted life years* (QALYs) approach attributes a weighting of 0 to 1 for each period, corresponding to the quality of life associated with health status during that time. QALYs measure results in terms of increases in health, whereas the DALYs express them in terms of loss (Dolan, 2000). The *years of healthy life* (YHLs) measure life expectancy adjusted to quality.

The second type of indicator has a wider objective in that it measures the quality of life in general rather than that which is associated with health. As with health, the definitions of quality of life are numerous. The definition of the WHO adopted by Alain Leplège (1999) defines it "as an individual's perception of his place in existence, in the context of the culture and system of values within which he lives, and in relation to his objectives, expectations, norms and anxieties." Marilyn Bergner (1989) introduces the main areas in the quality of life as follows:

1. Symptoms
2. Functional status
3. Activities allied to the social role
4. Social functioning
5. Cognition
6. Sleep and rest
7. Energy and vitality
8. Emotional state or status
9. Perception of health
10. General satisfaction with life

In practice, most of the existing scales are instruments for measuring the health-related quality of life. They may, however, include aspects of the patients' lives that are only indirectly influenced by the state of health, and the frontier with instruments for measuring the quality of life is tenuous.

They may be divided into two groups, depending on the type of measure obtained: profiles and indices. The first group offers a multidimensional measure of life quality, whereas the second group provides a synthetic score (Teeling-Smith, 1988). In the latter case, the question of weighting the different items is essential.

Among the scales currently in use, some (e.g., *Quality of Life Index* [Spitzer *et al.*, 1981]) were initially constructed to measure the well-being of patients suffering from cancers or other serious diseases and receiving palliative treatment, and they can only be used for seriously ill patients (Nelson and Berwick, 1989). Others (e.g., *Nottingham Health Profile* [Hunt *et al*., 1985]) seek to evaluate the effect of diseases and treatments on health. They are not recommended for the general population, however, because they have poor discriminatory power (Leplège, 1999). The dimensions evoked vary according to scales, but physical and social fields are included in almost all of them. The number of questions may vary from about 20, as in the *Psychological General Well-Being*, to almost 140 in the *Sickness Impact Profile* (Bergner *et al*., 1981), and the length and complexity of the questionnaire does deter some people from using it on a routine basis (Nelson and Berwick, 1989). The number of conditions offered for each of these questions constitutes an important differentiation factor of the scales, and some (e.g., *Mos SF-36* [McDowell and Newell, 1996]) are reputed for having a greater sensitivity because of the greater number of response possibilities (Leplège, 1999). In certain cases, the causal link with health problems is explicit. If health problems are not the cause, it can be assumed that the person has no problems, and for certain people, this constitutes an important limit for the instruments (Strawbridge, 1998).

Specific scales of quality of life are supported by many international projects. In Europe, the *European Quality of Life Scale* (EuroQOL) (Brooks, 1996) was developed in the early 1990s. On an even broader level, the *World Health Organization Quality of Life Assessment* (the WHOQOL group, 1998), supported by the WHO, was destined to be used worldwide. This questionnaire provides an examination with six fields: physical health, the psychological domain, level of independence, social relations, environment, and spirituality.

CONCLUSIONS

All the indicators mentioned—morbidity, disability, and quality of life—are in one way or another accurate indicators of health states. Their diversity is a reflection of the multiplicity of possible approaches. The choice of indicator therefore depends solely on the aim of the measure. The indicator depends on the question asked. If the interest is focused on the need for the care of elderly people, it is important to assess disability. If the survey is devoted to the dissemination of diseases, measures of morbidity would be used. The diversity of the instruments enables differentiation and the possibility of covering different health dimensions to be able to reply in the most appropriate way to the questions asked.

Developments enabling reflection on health indicators tend to serve a double purpose in harmonizing and synthesizing the information. This desire to synthesize underlies the construction of numerous indicators. Is it therefore possible to imagine a single health indicator incorporating life expectancy, presence of diseases, functional state, and other health dimensions in such a way that it reflects the contributions of each element? Such an indicator would enable direct comparison of different population groups and follow-up of developments while also providing essential indications on each of the elements in its composition. However, this approach implies that it is possible to describe the continuum from health to disease to death in an unequivocal way and that every combination of factors leads always to a predetermined level of health (Bergner, 1985). Such an indicator still remains to be invented.

Harmonization is no easier to obtain. The proof of this lies in the profusion of scales for measuring disability or quality of life that have been produced in recent decades. International organizations are increasingly engaged, however, in procedures aimed at proposing common tools for the member states. It was in this way that the European Community set up a program for harmonizing the health information system in Europe (i.e., *Health Monitoring Programme*). The European Office of the WHO has been trying to promote the use of common measuring instruments for national health surveys, and from these activities, scales for measuring disability and instruments for measuring the quality of life have come about. Synthetic questions have aimed at providing a global approach to the state of health by covering the three health areas of perception, morbidity, and functional health.

The following three examples are questions proposed in Europe for measuring health:

1. How is your health in general? (Very good, Good, Fair, Bad, Very bad)
2. Do you suffer from any long standing illness or condition? (Yes/no)
3. For at least the past 6 months, have you been limited because of a health problem in activities people usually carry out? (Yes/no)

Acknowledgment

This chapter was translated from the French version by Ronald Wilkens.

References

Aïach Pierre and Cèbe Dominique, 1996. Évaluer la santé d'un groupe, une gageure, *La Recherche*, (284), p. 30–32.

Basch Paul F., 1990. *Textbook of international health*. New York, Oxford University Press, 423 p.

Bergner Marilyn, 1985. Measurement of health status, *Medical Care*, vol. 23 (5), p. 696–704.

Bergner Marilyn, 1989. Quality of life, health status, and clinical research, *Medical Care*, vol. 27(3), Suppl., p. S148–S156.

Bergner Marilyn, Bobbitt Ruth A., Carter William B., and Gilson Betty S., 1981. The Sickness Impact Profile: development and final revision of a health status measure, *Medical Care*, vol. 19(8), p. 787–805.

Blaxter Mildred, 1989. A comparison of measures of inequality in morbidity, *in*: John FOX (ed.), *Health inequalities in European countries*, p. 199–230. Aldershot, UK, Gower, 414 p.

Bouyer Jean, Hémon Denis, Cordier Sylvaine, Derriennic Francis, Stücker Isabelle, Stengel Bénédicte, and Clavel Jacqueline. *Épidémiologie, principes et méthodes quantitatives*. Paris, Les éditions de l'INSERM, xiii + 498 p.

Brooks Richard, 1996. EuroQol: the current state of play, *Health Policy*, vol. 37(1), p. 53–72.

Cambois Emmanuelle and Robine Jean-Marie, 1994. *Théorie et analyse des principales enquêtes santé portant sur l'incapacité dans les pays occidentaux*. Montpellier, Laboratoire d'épidémiologie et d'économie de la santé, 140 p. (Report to the MIRE).

Cambois Emmanuelle, Robine Jean-Marie and Jagger Carol, 2000. Difficulties in daily living beyond age 65 (ADL), p. 43–59, *in*: Jean-Marie Robine, Carol Jagger, and Viviana Egidi (eds.), *User's guide to health expectancies for the European Union: a coherent set of indicators*. Brussels, European Commission, 191 p.

Cdnhw, 1986. Achieving health for all: a framework for health promotion. Ottawa, Canadian Department of National Health and Welfare. (Report on the First International Conference of Health Promotion, Ottawa, 1986).

Connelly Julia E., Philbrick John T., Smith G. Richard. Jr, Kaiser Donald L., and Wymer Antoinette, 1989. Health perceptions of primary care patients and the influence on health care utilization, *Medical Care*, vol. 27(3), Suppl., p. S99–S109.

Council of Europe, 1990. Evaluation of the use of the International Classification of Impairments, Disabilities and Handicaps (ICIDH) in surveys and health-related statistics. Strasbourg, Conseil de l'Europe, 43 p.

Coutton Vincent, Van den Berg Jaap, and Robine Jean-Marie, 2000. Physical and sensorial functional limitations, *in*: Jean-Marie Robine, Carol Jagger, and Viviana Egidi (eds.), *User's guide to health expectancies for the European Union: a coherent set of indicators*, p. 25–43. Brussels, European Commission, 191 p.

Crimmins Eileen M., 1996. Mixed trends in population health among older adults, *Journal of Gerontology B Psychology and Social Science*, vol. 50, p. S223–S225.

Dolan Paul, 2000. The measurement of health related quality of life for use in resource allocation decisions in health care, *in*: Anthony J. Culyer and Joseph P. Newhouse (eds.), *Handbook of health economics*, p. 1723–1760. Amsterdam, Elsevier Science, 1910 p.

Ferrucci Luigi, Guralnik Jack M., Cecchi Francesca, Marchionni Niccolo, Salani Bernardo, Kasper Judith, Celli Romano, Giardini Sante, Heikkinen Eino, Jylhä Marja, and Baroni Alberto, 1998. Constant hierarchic patterns of physical functioning across seven populations in five countries. *Gerontologist*, vol. 38(3), p. 286–294.

Finch Michael, Kane Robert L., and Philip Ian, 1995. Developing a new metrics for ADLs, *Journal of American Geriatrics Society*, vol. 43, p. 877–884.

Folstein Marshal F., Folstein Susan E., and Mchugh Paul R., 1975. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician, *Journal of Psychiatric Research*, vol. 12(3), p. 189–198.

Freeman Jonathan and Hutchinson George B., 1980. Prevalence, incidence and duration, *American Journal of Epidemiology*, vol. 112(5), p. 707–723.

Fried Linda P., Bandeen-Roche Karen, Williamson Jeff D., Prasada-Rao Patricia, Chee Elisbeth, Tepper Sherri, and Rubin Gary S., 1996. Functional decline in older adults: expanding methods of ascertainment, *Journal of Gerontology: Medical Sciences*, vol. 51A(5), p. M206–M214.

Glass Thomas A., 1998. Conjugating the "tense" of function: discordance among hypothetical, experimental and enacted function in older adults, *The Gerontologist*, vol. 38(1), p. 101–112.

Goldberg David P. and Hillier V.F., 1979. A scaled version of the General Health Questionnaire, *Psychological Medicine*, vol. 9(1), p. 139–145.

Harwood Rowan H., Prince Martin, Mann Anthony, and Ebrahim Shah, 1998. Association between diagnoses, impairments, disability and handicap in a population of elderly people, *International Journal of Epidemiology*, vol. 27, p. 261–268.

Heyman Dorothy K. and Jeffers Frances C., 1963. Effect of time lapse on consistency of self-health and medical evaluations of elderly persons, *Journal of Gerontology*, vol. 18, p. 160–164.

Hunt Sonia M., Mcewen James, and Mckenna Stephen P., 1985. Measuring health status: a new tool for clinicians and epidemiologists, *Journal of the Royal College of general practitioners*, vol. 35(273), p. 185–188.

Hupkens Christianne and Swinkels Henk, 2001. *Health interview surveys in the European Union: overview of methods and contents*. The Hague, European Commission and Statistics Netherlands, 31 p.

Idler Ellen L. and Benyamini Yael, 1997. Self-rated health and mortality: a review of twenty-seven community studies, *Journal of Health & Social Behaviour*, vol. 38, p. 21–37.

Jammal Amal, Allard Robert, and Loslier Geneviève, 1988. *Dictionnaire d'épidémiologie*. St Hyacinthe, Quebec, Edisem, Maloine, 171 p.

Kaplan George A. and Camacho Terry, 1983. Perceived health and mortality: a nine-year follow-up of the Human Population Laboratory cohort, *American Journal of Epidemiology*, vol. 117, p. 292–304.

Katz Sidney, Ford Amasa B., Moskowitz Roland W., Jackson Beverly A., and Jaffe Marjorie W., 1963. Studies of illness in the aged. The Index of ADL: A standardized measure of biological and psychosocial function, *Journal of the American Medical Association*, vol. 185(12), p. 914–919.

Kempen Gertrudis I. J. M., Steverink Nardi, Ormel Johan, and Deeg Dorly J.H., 1996. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies, *Journal of Gerontology: Psychological Sciences*, vol. 51B(5), p. P254–P260.

Koponen Paivikki and Aromaa Arpo, 2001. *Health examination surveys (HES): review of literature and inventory of surveys in the EU/EFTA member states*. Helsinki, European Commission and National Public Health Institute (Finland), 97 p.

Kristensen Tage Sondergard, Smith-hansen Lars, Bjørner Jakob, Borg Vilhelm, and Skov Torsten, 1998. *Selvvurderet helbred og arbejdsmiljø*. Copenhagen, Arbejdsmiljøinstituttet og Insititut for Folkesundhedsvidenskab, 124 p.

Larue Arsenath, Bank Lew, Jarvik Lissy, and Hetland Monte, 1979. Health in old age: How do physicians' ratings and self-ratings compare? *Journal of Gerontology*, vol. 34, p. 687–691.

Last John M., Spasoff Robert A., Harris Susan S., and Thuriaux Michel C., 2001. *A dictionary of epidemiology*, 4th ed. New York, Oxford University Press, 180 p.

Lawrence Renée H. and Jette Alan M., 1996. Disentangling the disablement process, *Journal of Gerontology: Social Sciences*, vol. 51B(4), p. S173–S182.

Lawton Powell M. and Brody Elaine M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living, *Gerontologist*, vol. 9(3), p. 179–186.

Lazaridis Emmanuel N., Rudberg Mark A., Furner Sylvia E., and Cassel Christine K., 1994. Do activities of daily living have a hierarchical structure? An analysis using the Longitudinal Study of Aging, *Journal of Gerontology: Medical Sciences*, vol. 49A(4), p. M47–M51.

Leplège Alain, 1999. *Les mesures de la qualité de la vie*. Paris, PUF, 128 p.

Manton Kenneth G., Stallard Eric, and Corder Lary S., 1998. The dynamics of dimensions of age-related disability 1982 to 1994 in the US elderly population, *Journal of Gerontology: Biological Sciences*, vol. 53A(1), p. B59–B70.

Mcdowell Ian and Newell Claire, 1996. *Measuring health. A guide to rating scales and questionnaires*, 2nd ed. Oxford, Oxford University Press, 523 p.

Mcwhinnie John R., 1981. Disability assessment in population surveys: Results of the OECD common development effort, *Revue d'épidémiologie et de santé publique*, vol. 29, p. 413–419.

Mossey Jana M. and Shapiro Evelyn, 1982. Self-rated health: a predictor of mortality among the elderly, *American Journal of Public Health*, vol. 72, p. 800–808.

Nagi Saad Z., 1965. Some conceptual issues in disability and rehabilitation, *in*: Marvin B. Sussman (ed.), *Sociology and rehabilitation*, p. 100–113. Washington, DC, American Sociological Association, 265 p.

Nagi Saad Z., 1976. An epidemiology of disability among adults in the United States, *Milbank Memorial Fund Quarterly/Health Society*, vol. 54, p. 439–467.

Nelson Eugene C. and Berwick Donald M., 1989. The measurement of health status in clinical practice, *Medical Care*, vol. 27(3), Suppl., p. S77–S90.

Perenboom Rom, Van Oyen Herman, and Van Herten Loes, 2000. Limitations in usual activities, a global approach, *in*: Jean-Marie Robine, Carol Jagger, and Viviane Egidi (eds.), *User's guide to health expectancies for the European Union: a coherent set of indicators*, p. 59–69. Brussels, European Commission, 191 p.

Reuben David B., Valle Laura A., Hays Ron D., and Siu Albert L., 1995. Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered and performance-based measures, *Journal of American Geriatrics Society*, vol. 43, p. 17–23.

Robine Jean-Marie and Jagger Carol, 2000. Un modèle de l'incapacité pour le XXIe siècle: relier les concepts aux mesures, *Revue d'Épidémiologie et de Santé Publique*, vol. 48(3), Suppl., p. 3S42–C39.

Robine Jean-Marie, Jagger Carol, and Egidi Viviana (eds.), 2000. *User's guide to health expectancies for the European Union: a coherent set of indicators*. Brussels, European Commission, 191 p.

Robine Jean-Marie, Ravaud Jean-François, and Cambois Emmanuelle, 1997. General concepts of disablement, *in*: David Hammerman (ed.), *Osteoarthritis: a public health implication for the aging population*, p. 63–83. Baltimore, Johns Hopkins University Press, 251 p.

Sanderson Denise, Rasmussen Neils K., and Rosa Gispert Rosa, 2000. Perceived health, *in*: Jean-Marie Robine, Carol Jagger, and Viviana Egidi (eds.), *User's guide to health expectancies for the European Union: a coherent set of indicators*, p. 69–79. Brussels, European Commission, 191 p.

Sermet Catherine, 1994. De quoi souffre t'on? *Solidarité Santé*, 1(janv.–mars), p. 39–56.

Sermet Catherine, 1998. *Évolution de l'état de santé des personnes âgées en France, 1970–1996*. Paris, Les éditions de l'INSERM, CREDES, 334 p.

Spector William D. and Fleishmen John A., 1998. Combining activities of daily living with instrumental activities of daily living to measure functional disability, *Journal of Gerontology: Social Sciences*, vol. 53B(1), p. S46–S57.

Spitzer Walter O., Dobson Annette J., Hall Jane, Chesterman Esther, Levi John, Shepherd Richard, Battista Renaldo N., and Catchlove Barry R., 1981. Measuring the quality of life of cancer patients: a concise QL-index for use by physicians, *Journal of Chronic Diseases*, vol. 34(12), p. 585–597.

Strawbridge William J., 1998. Quality of life: what is it and can it be measured? *Growth Hormone and IGF Research*, vol. 8, Suppl. A, p. 59–62.

Tager Ira B., Swanson Ann, and Satariano William A., 1998. Reliability of physical performance and self-reported functional measures in an older population, *Journal of Gerontology: Medical Sciences*, vol. 53A(4), p. M295–M300.

Teeling-Smith George, 1988. *Measuring health: a practical approach*. Chichester, John Wiley & Sons, 268 p.

The Whoqol Group, 1998. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties, *Social Science and Medicine*, vol. 46(12), p. 1569–1585.

WHO, 1946. Constitution of the World Health Organisation, *Department of State Bulletin*, August 4, p. 211–219.

WHO, 1980. *International Classification of Impairments, Disabilities, and Handicaps: a manual of classification relating to the consequences of diseases*. Geneva, World Health Organization, 205 p. (revised edition in 1993).

WHO, 1984. *The uses of epidemiology in the study of the elderly: report of a WHO Scientific Group on the Epidemiology of Aging*. Geneva, World Health Organization, 84 p.

WHO Regional Office for Europe, 1985. *Targets for health for all*. Copenhagen, World Health Organization, 228 p. (European Health for All Series(1).

WHO, 1993. *International Classification of Impairments, Disabilities, and Handicaps*. Geneva, World Health Organization, 207 p.

WHO, 2001. *International classification of functioning, disability and health*. Geneva, World Health Organization, 299 p.

WHO and Statistics Netherlands, 1996. *Health interview surveys: towards international harmonization of methods and instruments*. Copenhagen, WHO Regional Office for Europe, 161 p. (WHO Regional Publications, European Series, no. 58).

Wiener Joshua M., Hanley Raymond J., Clark Robert, and Van Nostrand Joan F., 1990. Measuring the activity of daily living: comparison across national surveys, *Journal of Gerontology: Social Sciences*, vol. 45(6), p. S229–S237.

Wolinsky Fredric D. and Johnson Robert J., 1991. The use of health services by older adults, *Journal of Gerontology: Social Sciences*, vol. 46(6), p. S345–S357.

Wood Philip H.N. and Badley Elisabeth, 1978. An epidemiological appraisal of disablement, *in*: Albert E. Bennett (ed.), *Recent advances in community medicine*, p. 149–173. Edinburgh, Churchill Livingstone, 273 p.

Young T. Kue, 1998. *Population health, concepts and methods*. New York, Oxford, Oxford University Press, 315 p.

Zung William W., Richards Carolyn B., and Short Marvin J., 1965. Self-rating depression scale in an outpatient clinic. Further validation of the SDS, *Archives of General Psychiatry*, vol. 13(6), p. 508–515.

CHAPTER

42

Medical Causes of Death

FRANCE MESLÉ

Institut national d'études démographiques (INED), Paris, France

Whether it is a question of understanding the development of mortality or of explaining the differences between groups, there must be resort to the cause of death. Nevertheless, responses to the question "what did he die of?" are not straightforward. The same death that can be attributed to heart failure, myocardial infarction, angina pectoris, or hypertension also can be related to smoking, working conditions, stress, and diet. Must all these factors be taken into consideration, can some of them be eliminated, or can only one be retained? The first series of causes are the *medical causes* of death, and it is these to which this chapter is devoted. The second series is part of an entire socioeconomic and cultural context that is considered from different viewpoints in Chapters 48 to 52. The boundary between these two groups of causes is not certain; for example, it is not clear how to categorize alcoholism, nutritional problems, suicide, or accidents at work.

From the first attempts at collecting the causes of death to the very elaborate statistical systems employed in the developed countries, the main concern has always been to identify causes that could be prevented or treated. The search for classification criteria has led to the adoption of the *International Classification of Diseases* (ICD), which is used by most of the countries that collect cause-related mortality statistics. With the compilation of a body of strict rules for selecting the cause of death, physicians and statisticians have endeavored to improve temporal and spatial comparability. Attributing death to a single cause, however, remains simplistic in view of the complex succession of pathologic changes that finally lead to death. The study of *multiple causes* permits a better appreciation of the weighting of different pathologies but is still too limited.

The first part of this chapter retraces the path that led to the current system and stresses the difficulties shared by all classification research. After the information has been collected and classified, the analysis of levels and trends in causes of death entails the calculation of indicators that mostly employs classic demographic tools. In the second part, I describe the indicators that are most frequently used and compare their results.

I. COLLECTING, CLASSIFYING, AND CODING

The systems of health statistics in use in developed countries have been progressively constructed by two intersecting processes: setting up a system for collecting the causes of death and elaboration of a classification of these causes. A short history of this twofold process can elucidate the difficulties specific to this area of mortality statistics.

1. History of the Collection of the Causes of Death

As long as the means of combating diseases and death remained rudimentary, collecting the cause of death had little real interest for communities. The first experiments fell within the scope of the struggle against epidemics (Biraben, 1973). The records of causes on which John Graunt based his main observations were carried out after the 1592 plague epidemic in London. Following the example of Graunt, the first cause-related death tables appeared in Europe during the 18th century. In France, the first such table was published in 1767 by Jean Razoux, and it was based on the deaths recorded at the general hospital in Nîmes over 5 years. In 1776, in response to an investigation into the study of epidemics led by the Royal Medical Society, many physicians provided tables of mortality figures. In Denmark, after a plague epidemic, the cause of death was recorded in the town of Copenhagen from 1709 onward (Johansen, 1993). In Berlin, a weekly report on the causes of death was introduced in 1737 to follow the development of infectious diseases (Kintner, 1993).

In the 19th century, the first systematic cause-related mortality records were set up throughout Europe. In England and Wales, the 1837 *Registration Act* introduced, in addition to the registration of vital events (e.g., births, marriages, deaths), listing the cause of death (Fagot-Largeault, 1989). The British system was the first to be extended to the whole territory. In France, registration of the cause of death was first applied to Paris from the end of the 18th century, but it had to wait until 1855 for the system to be extended to towns with more than 10,000 inhabitants and until 1906 for the whole territory to be covered. In Scandinavian countries, the collection of the causes of death was organized at a national level in the middle of the 19th century.

By the early 20th century, the principle of a national registration of the causes of death was established in many European countries. Subsequent changes have essentially been of a qualitative nature. Medical certification of the cause of death has become the rule in Europe (Meslé, 1995a). Suitable arrangements for ensuring confidentiality most frequently accompany it, and these have contributed to an improvement in the quality of the statistics. The World Health Organization (WHO) proposes a standardized certificate, to which a number of countries have conformed more or less faithfully. This certificate enables the physician to declare several causes contributing to the death, and to better define the complex sequence of pathologic processes involved.

Although it has become standard practice in developed countries, statistics for the causes of death in most developing countries are lacking. To counter the shortcomings of the civil registration in these countries, survey methods have been developed to allow non-physicians to collect information from close relations regarding the symptoms and problems that preceded the death (an aspect that is developed by Vincent Fauveau in the next chapter). *Verbal autopsies* have contributed to a better understanding of the causes of death, particularly for children (Desgrées du Loû, 1996; Garenne and Fontaine, 1988, 1990; Zimicki, 1988, 1990). Such investigations cannot, however, replace regular statistics recording deaths by cause.

2. Classification Problems

Collecting the cause of death does not suffice for establishing a statistic. To exploit this information, it must be classified. The choice of classification criteria was at the heart of the discussions that finally produced the first version of the ICD. Successive revisions of this have not basically called into question its primary structure, but they have been responsible for some very unfortunate interruptions in the statistical series. The categorization now used by most of the countries that issue statistics on the causes of death remains organized around several lines of classification, but without any of them spanning it entirely.

a. Origin of the International Classification

The *Bills of Mortality*, on which John Graunt (1662) worked, presented the causes of death in alphabetical order. This type of classification was the one most frequently used in the first statistical collections. The first attempt at a systematic classification of diseases was in the 18th century. Anne Fagot-Largeault, in her work on *Les causes de la Mort* (1989), recorded several dozens of classifications published between 1718 and 1855. The best known was that of Boissier de La Croix, called *Sauvages* (1731), whose nosography in 10 categories was partly adopted by Linnaeus (1763) for the *Genera Morborum* (Vallin and Meslé, 1988).

The first attempt to devise a common classification for several countries appeared at the International Statistical Congress of 1853. Two physicians, William Farr and Marc d'Espine, responsible respectively for the health statistics of England and Wales and the canton of Geneva, were entrusted with "*establishing a uniform nomenclature of the causes of death applicable to all countries*" (Registrar general, 1856). Two years later, each of the two men presented his nomenclature based on different principles. The discussion that then opposed them is still a current one. Marc d'Espine (d'Espine and Farr, 1855) attempted to define a coherent classification according to the *nature of the diseases*. William Farr, more pragmatic, isolated the best-known diseases of the time, such as smallpox, scarlet fever, and tuberculosis, and classified the others according to their anatomic location. A compromise was adopted at the Congress of Paris in 1855, which tended to favor the Farr concept (IIS, 1956). Although revised three times (1874, 1880, and 1886), it was never intended to be used by any national statistical service with each country employing its own nomenclature.

In Paris, notably, a nomenclature, largely inspired by the classification proposed by Philippe Pinel (1797), was used for the cause-of-death statistics from 1802 onward. It was revised nine times during the 18th century, taking regular account of advances in medicine especially concerning progress in bacteriology, with the result that in 1886, the date of the last revision, it was quite modern in conception. When, in 1891, the International Statistical Institute decided to prepare a new international classification, the work was quite naturally entrusted to Jacques Bertillon, head of Statistical Work in the City of Paris. Taking up the nomenclature framework agreed to in Paris in 1886, he proposed, in 1893, a new international classification that was to enjoy rapid success (Bertillon, 1896).

This classification consisted of 14 divisions (Table 42–1), comprising a total of 203 categories (or items) whose precise contents were listed. Bertillon proposed two other, more abridged nomenclatures, one with 116 and the other with 44 items. Several countries and many towns immediately adopted the new classification. It was first revised in 1900 and then revised a total of 10 times during the 20th century. The number of countries using the ICD has not ceased to grow. All the countries with regular cause-related mortality statistics now use the ICD.

b. Successive Revisions

If the chapter headings of the 1893 ICD (Bertillon, 1896) are compared with those of the 10th revision adopted in 1989 (WHO, 1993) almost a century later, the changes appear to be unimportant (see Table 42–1). The number of chapters has increased from 14 to 21, but this results from an expansion of the first two chapters. The *general diseases* have been divided into *infectious and parasitic diseases*, *neoplasms*, *diseases of the blood and blood-forming organs and certain immunity disorders*, and *endocrine, nutritional and metabolic diseases*. Similarly, diseases of the nervous system and sense organs are now divided into four chapters, distinguishing

TABLE 42–1 Chapter Headings of the 1893 International Classification of Disease Compared with Those of the 1989 Tenth Revision

1893 Classification		1989 Classification[a]	
No.	Heading	No.	Heading
I	General diseases	I	Certain infectious and parasitic diseases
		II	Neoplasms
		III	Diseases of the blood and blood-forming organs and certain immunity disorders
		IV	Endocrine, nutritional and metabolic diseases
II	Diseases of the nervous system and sense organs	V	Mental and behavioral disorders
		VI	Diseases of the nervous system
		VII	Disorders of the eye and adnexa
		VIII	Diseases of the ear and mastoid process
III	Diseases of the circulatory system	IX	Diseases of the circulatory system
IV	Diseases of the respiratory system	X	Diseases of the respiratory system
V	Diseases of the digestive system	XI	Diseases of the digestive system
VI	Diseases of the genitourinary system and adnexa	XIV	Diseases of the genitourinary system
VII	Puerperal diseases	XV	Pregnancy, childbirth, and the puerperium
VIII	Diseases of the skin and adnexa	XII	Diseases of the skin and subcutaneous tissue
IX	Diseases of the locomotive organs	XIII	Diseases of the musculoskeletal system and connective tissue
X	Congenital malformations	XVII	Congenital malformations and chromosomal anomalies
XI	Diseases of infancy	XVI	Certain conditions originating in the perinatal period
XII	Diseases of old age	XVIII	Symptoms, signs, and anomalous clinical and laboratory results, not elsewhere classified
XIV	Ill-defined diseases		
XIII	Conditions produced by external causes	XIX	Injury, poisoning, and certain other consequences of external causes
		XX	External causes of morbidity and mortality
		XXI	Factors influencing health status and contact with health services

[a] Tenth revision of the International Classification of Diseases.

mental disorders from different localizations (e.g., nervous system, eye and ear).

The expansion of the chapter on *conditions produced by external causes* into two chapters on *injury, poisoning, and certain other consequences of external causes* and *external causes of morbidity and mortality* relates to another type of reasoning. This distinction, introduced in 1948, recognized the necessity with violent deaths for a double classification according to the nature of the lesion (e.g., wound, fracture) and according to the external cause (e.g., road accident, suicide). The only regrouping concerns the diseases of old age and those that are ill defined, which have constituted a single chapter since 1948. This attests to the increased precision of diagnosis and the desire of the medical community to assimilate senility to an ill-defined cause of death. The last chapter, *factors influencing health status and contact with health services*, was created by the last revision and has a particular status. Its items cannot be used as a principal cause of death.

Despite changes, the structure proposed by Bertillon has stood up well to 100 years of medical progress. It is true that because it depends essentially on the anatomic localization of diseases, it has a lesser risk of losing its relevance than a classification based on more evolutive criteria, such as etiology or avoidability.

This apparent stability of the structure masks the profound upheavals that the ICD experienced during the past century. The items have increased from 203 to more than 10,000. This enormous increase in precision reflected the progress made in the knowledge of diseases and the causes of death, which led those responsible for the nomenclature to refine the medical definitions, but it arose equally through the extension of the fields of use of the classification. When it was created, it was essentially a tool designed to serve the purpose of cause-related mortality statistics, it has been increasingly employed for all sorts of purposes in the realm of health statistics, leading to a not indispensable refinement in the study of cause-related mortality. In this respect, the title of the volumes of the 10th revision is revealing: *International Classification of Diseases and Related Health Problems* (WHO, 1993). The expression *causes of death* has entirely disappeared.

The multiplication of items with each revision would not be very troublesome if it concerned a simple phenomenon of expansion. In reality, the exchanges between the items are much more complex, making it impossible to follow a coherent series of causes of death through the various revisions. At the time of the first revisions, the authors meticulously justified the changes they introduced. This preoccupation has gradually disappeared coincidentally with the nomenclature becoming more complicated. Except in rare cases in which the body responsible for the cause-related mortality statistics has taken the trouble to operate a dual classification with the implementation of a new classification[1], it is necessary to undertake the painstaking task of comparing the contents of the items of the former classification with those of the new to identify the exchanges between items.

This comparison, term by term, of six transitions from one revision to the following one has been carried out by Institut national d'études demographiques (INED). It forms the first step in a method enabling the reconstitution of coherent series of deaths by cause (Vallin and Meslé, 1988, 1998; Mesléand Vallin, 1996). From the correspondence table established, it is possible in a second stage to define elementary associations between the items of the former and the new revision. These associations are constructed in such a way that the medical and statistical content are the same in the two revisions. In a third stage, they provide the necessary safety rail for calculating the coefficients of distribution by sex and age that permit the reclassification of deaths by sex and age from one revision to another.

The method applied to the available French data has enabled the 1925–1999 cause-related mortality series, classified according to the ninth revision of the ICD, to be obtained. For this purpose, it was necessary to deal with six transitions from the third to the ninth revision. Each of these presented particular problems, but the most acrobatic exercise was in the transition from the fifth revision (adopted in 1938) to the sixth revision (adopted in 1948). Despite an increase in precision, the first revisions of the ICD remained broadly influenced by the spirit of Bertillon and the European medical schools, especially that of France. After World War II, this changed. The revision of the International Classification was entrusted to the young WHO, in which the American school of medicine predominated. The nomenclature was overturned, and the number of items increased from 450 to 2000, producing the most detailed level ever offered. Only the structure in chapters remained relatively undisturbed.

c. Different Lines of Classification

As it appears in its most recent version, the ICD is still traversed by different lines of classification. At the most unrefined level, the division into chapters is not according to a single axis. Most chapters refer to an anatomic classification. This reflects a legacy from the

[1] The OPCS (Office of Population, Censuses and Surveys) in England undertook in 1979 a dual classification of a sample of deaths according to the intermediate lists of the eighth and ninth revisions of the ICD, which enabled the reconstruction of statistical series with constant definitions (Meslé and Vallin, 1993).

first Bertillon nomenclature and the influence of William Farr's classification favoring the classification of diseases according to their localization. Some chapters, however, come within the etiologic axis. This is the case in Chapter I for certain infectious and parasitic diseases, in Chapter II for neoplasms, and in Chapter XX for violent deaths. Other chapters refer to a time axis, such as diseases of early childhood, pregnancy, childbirth, or puerperium. The double classification axis introduced in 1948 for violent deaths has remained the exception. No other attempt has been made within the ICD to classify the different diseases according to more systematic criteria.

The diversity of classification criteria can be found at all classification levels, even within the basic items. Besides the detailed list of items and independent of the chapters, the ICD proposes different levels of groupings with a certain number of "intermediate" lists. These lists also are heterogeneous in respect of classification axes.

One of the essential aims in the study of cause-related mortality is to identify homogeneous pathologic groups whose developments may be influenced by common factors on which action may be undertaken. Many investigators have suggested more pertinent classifications than that of the simple anatomic reference. Many of these classifications focus on the concept of *avoidable mortality*. Causes of death may be divided into two groups, according to whether they are *avoidable* or *unavoidable*. This enables effective health policies to be determined, which are relevant to the first group.

The well-known distinction drawn by Jean Bourgeois-Pichat between endogenous and exogenous mortality (see Chapter 48) originates from this concept. He aims at identifying the causes of avoidable mortality that may be eliminated to reach as closely as possible the "hard core" of completely resistant causes. Working first on infant mortality, he relies more on age-specific mortality than on causes of death that are considered to be too imprecise at that age (Bourgeois-Pichat, 1951a, 1951b). In seeking to quantify a biologic mortality limit, he also distinguishes the deaths arising from exogenous causes (i.e., infectious diseases, respiratory diseases, and accidents) that it would be possible to eliminate from deaths arising from endogenous causes related to a weakening of the organism (Bourgeois-Pichat, 1952).

Employing the same idea, Erica Taucher proposed a division of avoidable deaths into several groups according to the type of intervention that could prevent them (Taucher, 1978). In this way, she distinguishes four categories of avoidable deaths (i.e., by vaccination or preventive medicine, by early diagnosis or medical treatment, by improvement in sanitary conditions, and by mixed actions) and three other categories (i.e., difficult avoidable deaths, those where the cause is ill defined, and other cases).

The concept of avoidable mortality enjoyed great popularity in the 1980s, and a European atlas of avoidable mortality was published by the European Community (Holland, 1988). It contained cause-specific mortality maps for a certain number of categories, assembling diseases that were recognized as avoidable within the framework of current medical knowledge. The limitations of such a classification are evident, however, in that revisions become continually necessary as medical knowledge advances. Although of practical use for establishing a health policy, the idea of avoidable mortality is ineffective in analyzing future developments.

The ideal solution would be to discover keys to a classification that would enable the cause of various diseases to be identified. This was the idea supported by Marc d'Espine at the International Statistical Congress in 1855, when he proposed distinguishing diseases solely according to their nature. A perusal of the classification that he had then established (Table 42–2) shows how utopian the idea was at a time when the nature of the pathologies was little understood. It is a pity, nevertheless, that the attempt declined in the mid-19th century has never subsequently been retried, except for the timid double classification of violent deaths.

After various attempts (Vallin and Nizard, 1978; Baloche and Nizard, 1973) in the area of reconstituting coherent cause-of-death series for France subsequent to 1925, we proposed the reclassification of deaths

TABLE 42–2 Nomenclature Proposed by Marc d'Espine at the 1855 International Statistical Congress

1st division:	stillborn
2nd division:	death from congenital malformation
3rd division:	death from old age or senile marasmus
4th division:	violent death or from an external accident
5th division:	death from morbid accident
6th division:	death from acute diseases, subdivided into three classes:
I.	Acute generalized or local inflammation
II.	Specific or general inflammation
III.	Inflammation allied to certain physiologic conditions
7th division:	death from chronic diseases, subdivided into two classes:
I.	Chronic inflammation or chronic diseases of local origin
II.	Chronic diseases resulting from various defects or chronic generalized diseases
8th division:	death from undetermined cause

according to an etiologic axis comprising seven major categories:

Infection
Malnutrition, intoxication (including drugs), allergy, and immunologic disorders
Accidental traumas and homicides
Neoplasms
Hereditary and congenital conditions
Degenerative disease
Suicide
Diseases of unknown origin and ill-defined or undeclared causes of death

The content of all the items in the eighth revision of the ICD was redistributed according to this axis. Most of the items only corresponded to a single etiologic category. Sometimes, however, it was necessary to distribute their contents among several causes. We then cross-referenced this etiologic dimension with an anatomic axis, which served as a safety device, having regard to the essentially anatomic character of the ICD. This type of exercise does not escape criticism that may be leveled at the fragile nature of such a classification and the arbitrary nature of certain etiologic groupings. The degenerative category, especially, is far from satisfactory. Above all, it hides our ignorance of the factors concerned with the origin of certain pathologic processes.

Nevertheless, the analysis of the evolution of mortality from this angle is not without interest. It enables, for example, more light to be cast on the real importance of infectious mortality, which is largely underestimated in the case of the ICD chapter on infectious diseases. Numerous items containing infectious diseases can be found in other chapters (e.g., influenza in diseases of the respiratory system, appendicitis in diseases of the digestive system), and the content of certain items even mixes infectious diseases with other conditions that relate to other causes. In the case of France, infectious mortality as estimated in 1925 was about 70% higher than that attributed to it in the single ICD Chapter I, and in 1978, it represented more than double of ICD Chapter I mortality (Vallin and Meslé, 1988).

It could be thought that in carrying out *ad hoc* regroupings of the items, the ever-increasing precision of the ICD enables the most pertinent classification axis to be selected. This is partly true, but it still remains for the authors of the ICD to rethink a more coherent and better-adapted classification for the cause-of-death statistics. One solution may be to develop a more systematic interlacing with an anatomic axis, unavoidable with an axis designed to define the etiologic process as precisely as possible.

3. Choice of Cause

A death results from the succession of a certain number of pathologic processes whose appearance or development may have been encouraged by other preexisting conditions. Nevertheless, most of the studies carried out on the causes of death make reference to a single cause. Identification of this so-called *initial* or *underlying cause* is decisive. With the changes in the structure according to age and cause of death that have come about over the past 50 years, this identification has become increasingly delicate, with the majority of deaths occurring among old people most frequently suffering from several chronic pathologies intertwined one with another. To ensure the coherence of statistics from one country to another, the WHO recommends a standard medical certificate of cause of death and a body of fairly strict rules for coding.

a. Rules of the World Health Organization

The death certificate has two parts (Fig. 42–1). The first is intended for a summary of the succession of the various processes that led to the death. The physician reporting the death must specify chronologically the various conditions that played a part in the occurrence of death. There is space, with three lines for three possible stages. On the first line (Ia), the physician states the condition immediately responsible for the death, entitled *direct cause*. The following line (Ib) is provided for the *intervening cause*, and on the third line (Ic), the *underlying cause* (i.e., primary or initial cause) is given. To verify the order of appearance of these various conditions, a column enables the interval between the appearance of each morbid process and the death to be recorded. The second part is reserved for the reporting of conditions that were not directly responsible for the death, but which could have contributed to it. The cause or causes appearing in this second part are the *contributory causes*. Most countries have adopted certificates quite similar to the international model.

In formulating strict coding rules, the WHO was aiming at a maximum guidance in the practice of national statistical institutes. Coding of the cause of death depends on a general rule, three selection rules, and six modifying rules (WHO, 1995).

The General rule stipulates that "if more than one condition is entered on the certificate, select the condition entered alone on the lowest used line of part I unless it is highly improbable that this condition could have given rise to all the conditions entered above it."

This principle is inapplicable if, on the last line of part I, several conditions are stated or a single condition that appears completely unlikely to have given rise to all the conditions mentioned in the previous

CAUSE OF DEATH		Approximate interval between onset and death
I		
Disease or condition directly leading to death *	*(a)* . . . . due to (or as a consequence of)	. . . .
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	*(b)* . . . . due to (or as a consequence of)	. . . .
	(c)	. . . .
II		
Other significant conditions contributing to the death, but not related to the disease or condition causing it	. . . .	. . . .
	. . . .	. . . .
* This does not mean the mode of dying, e.g., heart failure, asthenia, etc. It means the disease, injury, or complication which caused death.		

FIGURE 42–1 International form of a medical certificate of cause of death proposed by the World Health Organization.

lines. In this case, if it is impossible to obtain further information from the reporting physician, one of the first two rules of selection is to be applied.

Rule 1 stipulates the following: "If there is a reported sequence terminating in the condition first entered on the certificate, select the underlying cause of this sequence. If there is more than one such sequence, select the underlying cause of the first-mentioned sequence."

If no sequence is present, Rule 2 must be applied: "If there is no reported sequence terminating in the condition first entered on the certificate, select this first-mentioned condition."

Rule 3 may apply in addition to one of the above: "If the condition selected by the General Rule or Rules 1 or 2 can be considered a direct sequel of another reported condition, whether in Part I or Part II, select this primary condition."

In certain cases, however, the application of one of the above rules leads to the retention of an underlying cause that is of little interest from a statistical or epidemiologic point of view. The modification rules permit, for example, the choice of a different underlying cause when that which arises from the principal rules is senility, an ill-defined state, a nonserious condition, a cause allied to other conditions mentioned, a condition whose nature may be determined by other elements of the declaration, a condition comprising several stages or a late effect.

The WHO also gives a certain number of coding instructions for particular ICD items. However rigid the body of rules, the assessment of the sequences and links between various conditions remains very difficult. It may vary from one period to the next and from country to country. In addition, the obligation to select a single cause considerably reduces the information that could be analyzed. That is why attempts are increasingly being made to take account of the whole information provided by death certificates by analyzing the principal cause and all the conditions certified by the physician.

b. Analysis of Multiple Causes

The attractive path of analysis of multiple causes nevertheless encounters complex methodologic problems. The propensity of physicians to attribute the death to the concurrence of one or several causes varies considerably from one country to another or from one period to another, and this makes comparisons in space and time very difficult. In France, for example, the proportion of deaths for which the certificate contains more than one condition increased from 25% in 1955 to 67% in 1996 (Désesquelles and Meslé, 2000, 2004). Rather than a real complication in the pathologic phenomena, this must be seen as the result of a growing desire by certifying physicians to attempt a complete definition of the whole process leading finally to death.

The absence of comparability also arises from statistical practices that may evolve with the passage of time or vary between countries. In France, the National

Institute for Medical Research (INSERM), which until then had retained senility if it was mentioned on the death certificate as a contributory cause, decided in 1983 to no longer code this condition. The profile of multiple causes at mature ages was completely upset. The number of registered causes may also vary according to the statistical rules of each country (Meslé, 2000). In France, the statistics for the causes of death only include a maximum of four causes of death: an underlying cause, a direct cause, and two contributory causes. If the death certificate includes more, the coders must make a choice according to criteria that are not completely explicit. In the United States, the whole information (up to 20 conditions for a single death) contained on the death certificate is coded and may be analyzed. In practice, the custom in France is especially toward retaining a direct cause in addition to the principal one and nothing else. Death certificates only rarely contain contributory causes, and the INSERM in any case only includes two at the most. In the United States, the consideration of all the information to be found on the death certificate, and a more complete contribution by the physicians provides much more information. In any case, comparison of the two statistics based on such widely differing selection criteria could not be relevant.

Despite these difficulties, the analysis of multiple causes at a given time and in a given country may contain valuable information. Two principal paths are followed. On one hand, the number of times a condition is mentioned may be totalled, irrespective of its level of appearance. This type of approach has the advantage of highlighting the role played in mortality by certain pathologies (e.g., alcoholism, diabetes) that appear more rarely as a main cause. In a study conducted into the cause of death of people older than 60 years in France, it was shown that the proportion of deaths mentioning an endocrine disease or mental disorders was tripled when the registered causes as a whole were considered rather than the sole underlying cause (Désesquelles and Meslé, 2000, 2004).

A more precise interest may be taken in the most frequent associations in an attempt to identify the sequences of typical diseases in certain pathologic processes. In France, an association is unsurprisingly to be found between alcoholism and violent deaths, or more unexpectedly, between skin diseases and mental disorders. The problem with such analyses lies in the fact that significant associations can only be identified by using a very delicate level of detail. It is therefore very difficult to formulate a general picture, but it does nevertheless permit a very precise study of certain pathologies.

Analysis of the multiple causes of deaths is still in its infancy. Nevertheless, it will constitute an important challenge for the study of cause-related mortality in the future. The next step will entail the formulation of common rules for choosing and classifying the multiple causes, such as was done by the WHO for choosing the underlying cause of death. It is only in this way that it will be possible to really profit from this type of study.

II. ANALYTIC INSTRUMENTS

After collection and classification, the development of cause-related mortality may be analyzed by means of several indicators that generally assume the hypothesis of independence between the causes (discussed in Chapter 44). In reality, the hypothesis of independence between causes of death is most probably unsatisfactory. It constitutes, nevertheless, a useful simplification for carrying out a first description of cause-related mortality. The introduction of dependence between causes would entail the ability to dispose at the very least of a good recording of all the causes contributing to death, or better still, information on the cause of the disease.

1. Rates by Age and Standardized Rates

The use of rates (age-specific or standardized rates) is common in describing trends. Age and cause-specific rates are a simple way of measuring cause-related mortality with the advantage of being additive: the sum of death rates by cause equals the death rate from all causes. Table 42–3 gives the standardized male mortality rates in 1970 and 1992 in England and

TABLE 42–3 Standardized Mortality Rates per 100,000 Men for Seven Groups of Causes in England plus Wales and in Russia in 1970 and 1992

Causes	England and Wales		Russia	
	1970	1992	1970	1992
Infectious diseases	10	5	57	25
Neoplasms	328	322	313	361
Cardiovascular diseases	851	550	999	1098
Diseases of the respiratory system	299	142	228	134
Diseases of the digestive system	39	34	48	56
Other diseases	95	104	61	73
Violent deaths	62	46	245	300
All causes	1684	1203	1951	2047

Wales and in Russia for seven major groups of causes. The use of a standard population, in this case the European standard population proposed by the WHO, enables a comparison of cause-related mortality levels in time and space, independent of the differences in the age structure of the population. It is possible in this example to compare the cardiovascular mortality trends in England and Wales and in Russia. Starting at a slightly higher level, the Russian mortality from this cause increased between 1970 and 1992, whereas in England, it showed a net reduction, with the result that in 1992 the first was twice as high as the second.

Because this concerns an additive measure, it is possible also to measure the share of a cause of death in the total mortality. In the two countries, cardiovascular mortality accounted for 51% of the total mortality in 1970. In 1992, its share had reduced to 46% in England and Wales, whereas is reached almost 54% in Russia.

It is often forgotten, however, that the standardized mortality rate (see Chapter 15 of Volume I) is not an absolute measure of mortality. According to the standard population chosen, with the same age-specific death rates, the value of the standardized rates may vary and the relation between the two measures become reversed. This may be the case if comparison is desired for countries or causes in which the age structure of mortality is very different. Let us take the example of Russia and England and Wales once more. Figure 42–2 provides two versions of the annual evolution of the standardized rates of mortality in the two countries for three large groups of causes: neoplasms, cardiovascular diseases, and violent deaths.

The first was obtained through the WHO European standard population (see Fig. 42–2A). It shows a fairly classic hierarchy of causes of death in the two countries: cardiovascular diseases, then neoplasms, and finally violent deaths, at a very low level in England and Wales, but very close to neoplasms in Russia. The same divergence in the development of cardiovascular mortality is found as that described above, whereas the deaths from neoplasms are very close—slightly falling in England and rising in Russia, with the two curves intersecting in the mid-1980s.

Figure 42–2B presents, for the same causes and the same countries, the trends in the standardized rates calculated this time according to the world standard population that is also proposed by the WHO. At first sight, this new picture appears to be similar to the previous one. It differs from it, however, in a certain number of respects. The level of the standardized rates is much less high, irrespective of the cause. The younger age structure of the world population gives more weight to the ages where mortality is low. This difference between standardized rates according to the standard population adopted is all the more important when the cause of death especially affects advanced ages. In the case of Russia, employing the world standard population reduces the rates of mortality from cancer more than those for violent death; with this reference population, violent mortality occupies second place for almost the whole period.

Between Russia and England and Wales, at the level of these big regroupings of causes, the change in the reference population does not seriously change the commentary suggested by the reading of Figure

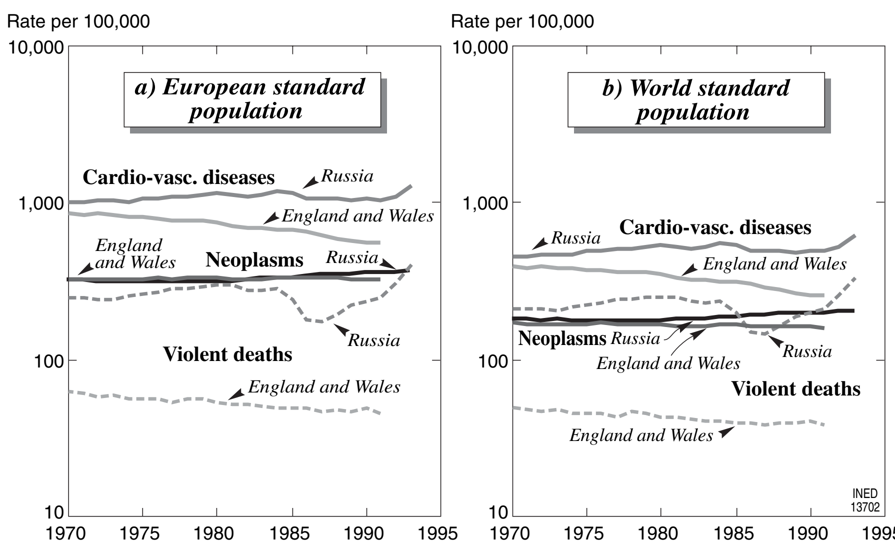


FIGURE 42–2 Annual trends of standardized mortality rates by neoplasms, cardiovascular diseases, and violent deaths in Russia and in England plus Wales according to the type of standard population employed.

42–2A. Nevertheless, a greater divergence is to be noted for neoplasms. The Russian cancer mortality, at the same level as the English one until the 70s, then rises above it very markedly. Cancers weigh more heavily on the young ages in Russia than in England and Wales. The use of a younger standard population penalizes the former.

2. Cause-Related Life Tables and Cause-Elimination Life Tables

Although, despite these reserves, the standardized rate remains a pertinent indicator of the trend and level of cause-related mortality, it does not enable cause of death and expectation of life to be directly related.

It is possible from the age- and cause-specific death rates to calculate the probabilities of dying by age for a given cause and to establish cause-related life tables (Monnier and Nizard, 1985; Baloche and Nizard, 1974; Schwartz and Lazar, 1961; Elveback, 1958; Berkson and Gage, 1950) (see Chapter 9 of Volume I). From these tables, the average or median age at death for a given cause can be calculated. The average age of death or expectation of life resulting from the effect of a single cause is not a very interesting indicator. In France, for example, life expectancy resulting from the action of cancer mortality changed, between 1925 and 1991, from 92.9 years to 87.7 years for men and from 91.5 years to 94.4 years for women. It certainly indicates the increase in male mortality and the decline in female mortality, but these figures do not tell us very much—what is the relevance of life expectancy related to the effect of a single operative cause? It would have to be possible to compare it to life expectancy in the absence of any cause other than the inevitability of death (i.e., a sort of maximum possible life expectancy), and this assumes the resolution of a very controversial question (see Chapter 48). The median age at death from a given cause informs more about changes in the age structure of cause-related mortality. Graziella Caselli (1993) showed how this median age at death had risen during the century for death from all causes. She cites, for example, the case of infectious diseases in Italy, for which the median age had increased from the 10–14 age category in 1930 to that of 75–79 years in 1985. At the same time, this median age varied little for accidents, from only 50–54 years to 55–59 years. This is an indicator that bears witness faithfully to the greater or lesser precociousness of a cause of death.

Nevertheless, if life expectancy resulting from a single cause is in itself not significant, and it is possible to ask what the expectation of life would be if one succeeded in completely suppressing certain causes of death. From the early 19th century, Emmanuel Duvillard had attempted to measure the increases in life expectancy anticipated from the elimination of smallpox, which it was then hoped to eradicate by means of a vaccine (Duvillard, 1806). Other authors have subsequently been interested in this question (Benjamin and Overton, 1981; Vincent, 1966; Greyville, 1954). It was also the idea supporting the work of Jean Bourgeois-Pichat when he differentiated endogenous and exogenous mortality to determine the anticipated gain from the complete suppression of mortality allied to external agents (see Chapter 48).

By calculating the difference in life expectancy from all causes and life expectancy obtained by removing one cause of death, we obtain a measure of the impact of a specific cause of death on life expectancy (Table 42–4).

At first sight, we have here a clear response to the question asked: what gain, in terms of life expectancy, can be anticipated by the removal of such and such a cause of death? According to the example proposed, in 1925, it was the eradication of infectious diseases that would have enabled the most effective gains in life expectancy (11.6 and 10.9 respectively according to sex). In 1978, suppression of this pathology did not contribute more than an additional year to life expectancy. On the other hand, the disappearance of deaths due to cancer enabled a gain of 3.9 years for men and 3.0 years for women, and that for degenerative disease (basically due to diseases of the circulatory system) was a gain of 6.4 or 6.7 years according to the sex.

The response is in reality not so simple. If the gains anticipated by the elimination of each cause are added

TABLE 42–4 Gain in Life Expectancy at Birth (Years) Anticipated by the Elimination of Certain Causes of Death in 1925 and 1978 in France

Cause of death	1925		1978	
	Men	Women	Men	Women
Infectious process	11.6	10.9	1.0	0.9
Neoplasms	1.1	1.6	3.9	3.0
Hereditary process	2.8	2.4	0.8	0.7
Degenerative disease	6.1	7.2	6.4	6.7
Malnutrition, intoxication	0.9	0.9	1.0	0.6
Diabetes	0.1	0.1	0.1	0.2
Accidents and homicides	1.4	0.8	1.8	1.0
Suicides	0.5	0.1	0.6	0.3
Total	24.5	24.0	15.6	13.4

From Vallin and Meslé, 1988.

up, the difference between the expected and maximum supposed life expectancy are not to be found. For the results presented earlier, the supposed maximum expectation of life was 101.5 years. In other words, to close the life tables, the life expectancy at 100 years was fixed at 1.5 years. The sum of expected increases in 1978 amounted to 15.6 and 13.4 years according to sex. If these gains in life expectancy in 1978 are added up, we obtain 85.4 years and 91.3 years, which is well short of the maximum of 101.5 years, which would logically be obtained if all the causes were removed.

This arises because the calculation takes account not only of the mortality due to the particular cause, but also of the mortality due to other causes. The results presented in Figure 42–4 were obtained by calculating the anticipated gain thanks to the elimination of one cause, all other causes remaining unchanged. If we now proceed successively to eliminate all the causes, the effect of each cause on life expectancy will depend on the rank order of elimination of the cause. The later this is eliminated, the greater will be the anticipated gain.

This gain is not directly related to the level of cause-related mortality. In 1978, women would have gained 3 years' life expectancy from the elimination of cancer, but in 1925, the difference was only 1.6. However, female mortality due to cancer was higher in 1925 than in 1978. There again, this paradox arises due to the fact that the elimination of this cause comes about in a certain context in which all the other causes of death intervene.

The simplicity of this indicator is only superficial. It must be handled with care, and does not, in any case, give a clear picture of the relations between life expectancy and causes of death.

3. Contribution of Each Cause to Differences in Life Expectancy

To have a measure of the weighting of each cause of death in life expectancy, several researchers have developed methods enabling differences of life expectancy between two populations to be allocated according to age or cause (Pressat, 1985; Arriaga, 1984; Andreev, 1982; Pollard, 1982, 1988, 1990) (see Chapter 16). Devised at approximately the same time, these various methods give very similar results (Toulemon, 1996).

They possess the enormous advantage of providing an additive decomposition of the life expectancy differences. According to the question we wish to answer, the basic elements of the decomposition can be regrouped over a certain interval of ages or according to the different causes of death.

It is therefore possible to quantify the contribution of the different causes of death in trends of life expectancy at birth (Table 42–5). Even if expectation of life at birth increases, certain causes of death may act negatively in hindering the progress. This is the case for the French example in Table 42–5. Over the last 4 decades, average duration of life has constantly increased, thanks primarily to the decline in mortality from infectious diseases and diseases of the respiratory system, then, more recently, to the fall in cardiovascular mortality (i.e., cerebrovascular diseases and other cardiovascular diseases). Between 1952 and 1991, male expectation of life has gained 8.5 years, for which infectious diseases and respiratory diseases account for 3.1 years and cardiovascular diseases as a whole for 2.8 years. Other causes, however, have hindered the increase in life expectancy. This is the case for cancers

TABLE 42–5 Contributions of Eight Major Groups of Causes to Variations in Life Expectancy in France during the Past 4 Decades

Period	All causes	Infections	Neoplasms	Cardiovascular diseases	Cerebrovascular diseases	Respiratory system diseases	Digestive system diseases	Other diseases	Violent deaths
Men									
1952–1962	2.58	1.09	–0.23	0.32	0.14	0.75	0.02	0.73	–0.20
1962–1972	1.54	0.40	–0.19	0.35	0.22	0.38	–0.05	0.74	–0.31
1972–1982	2.23	0.15	–0.15	0.40	0.41	0.37	0.27	0.45	0.33
1982–1991	2.18	–0.19	0.18	0.62	0.37	0.19	0.35	0.22	0.44
Women									
1952–1962	3.55	0.93	0.19	0.70	0.29	0.74	0.09	0.76	–0.12
1962–1972	2.46	0.22	0.25	0.64	0.34	0.55	–0.01	0.68	–0.21
1972–1982	2.66	0.12	0.25	0.60	0.51	0.36	0.24	0.40	0.18
1982–1991	2.25	0.02	0.20	0.60	0.53	0.13	0.24	0.27	0.26

From Meslé, 1995b.

(–0.39 years) and violent deaths over the first 2 decades.

The contribution of the causes of death to the differences in life expectancy between the sexes can also be calculated (Table 42–6 and Fig. 42–3). Table 42–6 may be read in two ways. Adding up the lines gives the participation of the different ages to these differences, and that of the columns the participation of the main causes. It is also possible to analyze which cause of death plays a major role at a certain age or even which is the age for which the contribution of a cause is most important.

With the difference in life expectancy between the sexes observed in France in 1990, it is violent deaths that play an overwhelming role from 15 to 29 years: 83% of the total contribution of this age group. Conversely, it is from 30 to 59 years that the contribution of violent deaths is greatest in absolute value: 0.71 per year as against 0.59 per year between 15 and 29 years, and much less for the other ages. The graphical representation of these contributions by age and cause is equally striking (see Fig. 42–3).

In the search for an indicator of the weighting of causes of death in life expectancy, decomposing the differences in average duration of life provides a tool that is interesting but remains relative, because it is always necessary to quantify the contribution of a cause of death within a difference between two life expectancies. An extension of this method (Vallin and Meslé, 1988) makes possible to measure the weighting of pathology in life expectancy at any given time, and for a given year to calculate the contribution of a cause to the difference between the life expectancy with all causes calculated for that year and the anticipated

TABLE 42–6 Contributions by Age Groups of the Major Causes of Death to Differences in Life Expectancy between the Sexes in France, 1990

Causes	All ages	0 yr	1–14 yr	15–29 yr	30–59 yr	60–74 yr	75 yr+
Infections	0.31	0.01	0.00	0.04	0.19	0.04	0.03
Neoplasms	2.85	0.00	0.01	0.01	0.82	1.29	0.73
Cardiovascular disease	1.61	0.00	0.00	0.01	0.46	0.69	0.45
Cerebrovascular disease	0.31	0.00	0.00	0.00	0.08	0.13	0.10
Respiratory diseases	0.56	0.01	0.00	0.00	0.07	0.19	0.28
Digestive diseases	0.43	0.00	0.00	0.01	0.18	0.17	0.07
Other diseases	0.54	0.13	0.01	0.04	0.16	0.12	0.09
Violent deaths	1.58	0.01	0.05	0.59	0.71	0.15	0.07
All causes	8.19	0.17	0.08	0.71	2.67	2.76	1.80

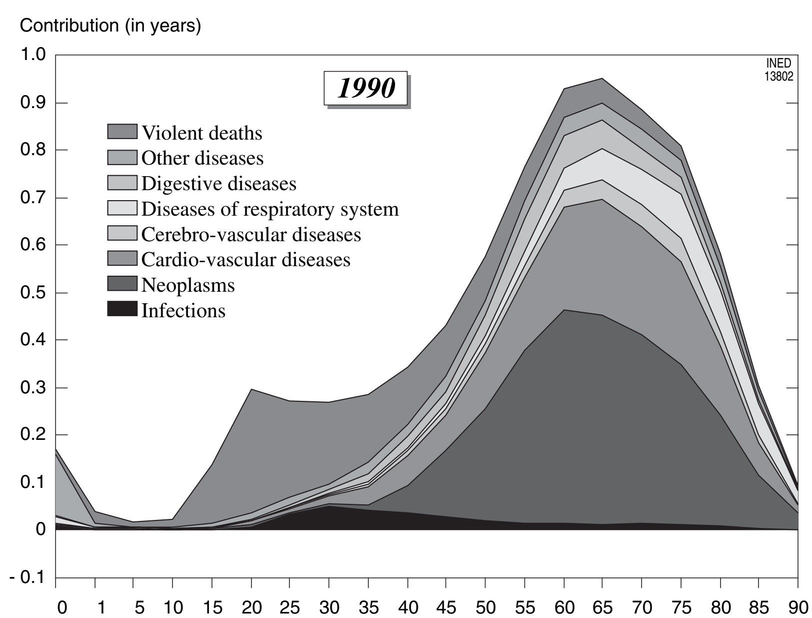


FIGURE 42–3 Contributions by age groups of the major causes of death to differences in life expectancy between the sexes in France, 1990. (From Meslé, 1995b.)

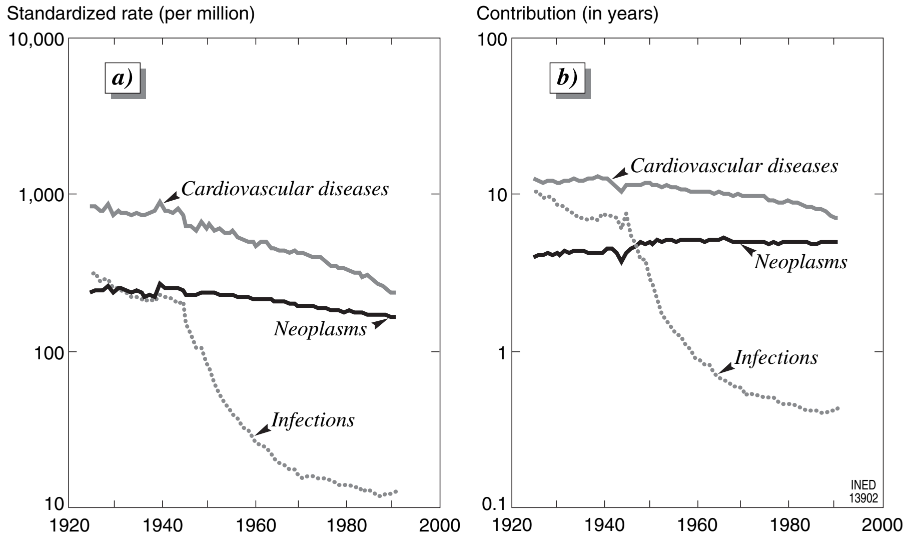


FIGURE 42–4 Comparison between the trend of the standardized mortality rate **(A)** and the trend of the contribution to the difference in life expectancy at birth and the maximum of 101.5 years **(B)** for three large groups of causes. Data are for females in France.

maximum life expectancy in the absence of all mortality before 100 years and with an average of 1.5 years of life to be lived after that age. The comparison between the trend of this contribution and that of the standardized rate is interesting (Fig. 42–4).

These two measures of the impact of infectious mortality develop in a very similar way—the lessening of the contribution of this pathology to the difference between life expectancy and 101.5 years faithfully follows the decline in the standardized rate of mortality. In the case of cardiovascular diseases, trends follow the same direction, but the decline in contribution is much slower than that of the rate. For certain periods, the movement is even contrary—during World War II, the standardized rate of cardiovascular mortality augmented, whereas the contribution of this cause to the limitation of life expectancy was diminishing. The paradox is even more evident for cancer mortality. Seen from the viewpoint of the standardized mortality rate, mortality was falling, whereas from the viewpoint of contributions, it was rising; at least until the end of the 70s.

The change in the standardized rate takes account of the evolution of the risk of death from a cause, independent of other causes. On the other hand, in the development of the contribution of pathology to the difference between life expectancy and a maximum of 101.5 years, changes in the structure of cause-related mortality also intervene. The contribution of cardiovascular diseases decreased during the war, because the increase in mortality due to this cause was relatively less important than that due to infections, and especially to violent deaths. Similarly, the increased contribution of cancers in a period where cancer mortality was diminishing is due to the fact that this reduction occurs at a much slower rate than the reduction in other conditions (infectious or cardiovascular diseases). This indicator therefore depends both on the level of mortality due to a cause and on the weighting of the cause in the total mortality.

4. Potential Years of Life Lost

Epidemiologists often use another measure of cause-related mortality: *the potential years of life lost* (Jenicek and Cléroux, 1982; Romeder and McWhinnie, 1978). This indicator quantifies the number of years that will not be lived because of premature deaths, occurring before a certain age (e.g., 70 years). The more a cause produces death at early ages, the more important will be the potential years of life lost.

The use of such an indicator presupposes, therefore, that an age is chosen at which death would subsequently be accepted as normal, whereas it would be considered premature if it happened earlier. In France, cancers were responsible in 1990 for a standardized rate of potential years of life lost between 1 and 70 years that was much higher than from other causes (Table 42–7). They were followed by cardiovascular diseases and by violent deaths. In this case, we have an indicator of the precociousness of a cause of death, rather than a real level of mortality due to this cause.

TABLE 42–7 Potential Years of Life Lost (Standardized Rates and Proportions) between 1 and 70 years and between 1 and 75 years in France, 1990, for Both Sexes

Groups of causes	PYLL, 1 to 70 yr		PYLL, 1 to 75 yr	
	Rate (per 100,000)	Proportion (per 100)	Rate (per 100,000)	Proportion (per 100)
Infectious diseases	655	3.0	797	2.8
Neoplasms	9008	41.7	11933	41.2
Cardiovascular diseases	3374	15.6	5070	17.5
Cerebrovascular diseases	958	4.4	1497	5.2
Diseases of respiratory system	780	3.6	1196	4.1
Diseases of digestive system	1453	6.7	1835	6.3
Other diseases	1609	7.4	2222	7.7
Violent deaths	3812	17.6	4385	15.2
All causes	21649	100	28935	100

PYLL, potential years of life lost.

The main problem arising from this indicator lies in the arbitrary nature of the fixed age limit. Table 42–7 compares the results obtained when the age limit is increased from 70 to 75. The integration of a supplementary group of ages quite clearly increases the standardized rate of potential years of life lost. The relative increase is not the same for all causes, and this means that the distribution of these lost years of life is modified. In the current example, the proportion of neoplasms varies little, but that of cardiovascular diseases increases appreciably, and the contribution of violent deaths diminishes.

To limit the arbitrariness of the choice of age limit, certain authors have proposed adopting expectation of life as the reference age (Estève *et al.*, 1993). In this way there is an estimate of the number of years lost in relation to the average number of years theoretically remaining to be lived. As the study authors themselves stress, the comparison in the time and space of the results obtained is not of great significance, because the calculations were based on different life expectancies.

CONCLUSIONS

Whatever the imperfections of the methods for measuring and the difficulties of ensuring comparability of the data in time and space, a thorough analysis of the causes of death has become indispensable for understanding the changes in mortality and for predicting its future developments. Throughout the 20th century, identification of the cause of death has continually improved due to progress in medical knowledge and the development of increasingly standardized tools for collecting and classifying. With ever-increasing concentration of deaths in older age groups, improved understanding about the morbidity processes leading to death will require perfecting newer analytic methods that take account of the multiplicity of causes and their interactions.

Acknowledgment

This chapter was translated from the French version by Ronald Wilkens.

References

Andreev Evgueny, 1982. Метод компонент в анализе продолжительности жизни [Analysis of life expectancy by means of its components], *Vestnik Statistiki*, March (3), p. 42–47.

Arriaga Eduardo, 1984. Measuring and explaining the change in life expectancies. *Demography*, vol. 21(1), p. 83–96.

Baloche Chantal and Nizard Alfred, 1973. Évolution de la mortalité par type de cause de décès en France 1950–1967, essai d'une nouvelle mesure de la mortalité, *in*: IUSSP (ed.), *International population conference, Liège, 1973*, vol. 3, p. 185–218. Liège, IUSSP, OrdinaÉditions, 470 p.

Benjamin Bernard and Overton Elisabeth, 1981. Prospects for mortality decline in England and Wales, *Population Trends* vol. 23, p. 22–28.

Berkson Joseph and Gage Robert, 1950. Calculation of survival rates for cancer, *in: Proceedings of the staff meeting of the Mayo Clinic, Rochester, 11 May*, p. 270–286. Rochester, MN, Mayo Clinic, 720 p.

Bertillon Jacques, 1896. De la nomenclature des maladies (causes de décès, causes d'incapacité de travail) adoptée par le service de statistique de la ville de Paris, *in: Annuaire statistique de la Ville de Paris*, p. 87–142. Paris, Ville de Paris, 826 p.

Biraben Jean-Noël, 1973. Essai sur la statistique des causes de décès en France sous la Révolution et l'Empire, *in: Hommageà Marcel Reinhard sur la population française au XVIII*e *et au XIX*e *siècles*, p. 59–70 Paris, Société de démographie historique, 599 p.

Bourgeois-Pichat Jean, 1951a. La mesure de la mortalité infantile. I. Principes et méthodes. *Population*, vol. 6(2), p. 233–248.

Bourgeois-Pichat Jean, 1951b. La mesure de la mortalité infantile: II. Les causes de décès. *Population*, vol. 6(3), p. 459–480.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme. *Population*, vol. 7(3), p. 381–394.

Caselli Graziella, 1993. L'évolutionà long terme de la mortalitéen Europe, p. 111–164, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European demography. II. Population dynamics*. Paris, John Libbey, INED, 526 p. (Congress and Colloquia, no. 9.)

Désesquelles Aline and Meslé France, 2000. *First insight into the patterns and trends of old-age mortality using French multiple cause-of-death statistics.* Paris, INED, 25 p. (Paper presented at the REVES Conference no. 12, Healthy life expectancy, linking policy to science, Los Angeles, 20–22 March, 2000.)

Désesquelles Aline and Meslé France, 2004, Intérêt de l'analyse des causes multiples dans l'étude de la mortalité aux grands âges: l'example français, *Cahiers québécois de démographie*, vol. 33, no. 1, p. 83–116.

Desgrées du Loû Annabel, 1996. *Sauver les enfants. Le rôle des vaccinations: une enquête longitudinale en milieu rural à Bandafassi, au Sénégal.* Paris, CEPED, 262 p. (Les études du CEPED, no. 12.)

Duvillard (De Duran) Emmanuel E., 1806. *Analyse et tableaux de l'influence de la petite vérole sur la mortalité à chaque âge, et de celle qu'un préservatif tel que la vaccine peut avoir sur la population et la longévité*. Paris, Imprimerie Nationale, 210 p.

Elveback Louis, 1958. Estimation of survivorship in chronic disease: the "actuarial" method, *Journal of the American Statistical Association*, vol. 53, p. 420–440.

Espine Marc (d') and Farr William, 1855. *I. Projet de classification des causes de mort physiologiques, accidentelles et morbides à l'usage de la statistique mortuaire de tous les pays, préparé pour le congrès international de statistique de Paris, par M. le Dr Marc d'Espine, . . . II. Projet de classification proposé par M. Will. Farr*. Paris, impr. de Vve Bouchard-Huzard, 42 p. (Paper presented at the Second Congress of the International Statistical Institute, Paris 1855.)

Estève Jacques, Benhamou Ellen, and Raymond Luc, 1993. *Méthodes statistiques enépidémiologie descriptive.* Paris, Éditions de l'INSERM, 307 p.

Fagot-Largeault Anne, 1989. *Les causes de la mort. Histoire naturelle et facteurs de risque.* Paris, Librairie philosophique Vrin, 420 p.

Garenne Michel and Fontaine Olivier, 1988. Enquête sur les causes probables de décès en milieu rural au Sénégal, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 123–142. Paris, INED, PUF, 458 p. (Travaux et Documents, Cahier no.119.)

Garenne Michel and Fontaine Olivier, 1990. Assessing probable cause of death using a standardized questionnaire: a study in rural Senegal, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analyses of mortality: new approaches*, p. 123–142. Oxford, Clarendon Press, 416 p. (International Studies in Demography.)

Graunt John, 1662. *Natural and political observations made upon the bills of mortality*, London. Translated into French: *Observations naturelles et politiques [. . .] sur les bulletins de mortalité*. Paris, INED, 176 p., 1977.

Greville T. N. E., 1954. On the formula for the L-function in a special mortality table eliminating a given cause of death, *Transactions of the Society of Actuaries*, vol. 6(1).

Holland Walter W. (ed.), 1988. *European community atlas of "avoidable" death.* Oxford, Oxford University Press, 356 p.

Iis, 1956. *Second International Statistical Congress, 1855. Compte-rendu général des travaux du Congrès international de statistique dans ses séances tenues à Paris*. Huzard, Institut international de statistique, Imprimerie de Mme veuve Bouchard, 542 p.

Jenicek Milos and Cléroux Robert, 1982. *Épidémiologie. Principes, techniques, applications.* St. Hyacinthe, Quebec, Edisem, 454 p.

Johansen Hans C., 1993. *The development of reporting systems for causes of deaths in Denmark.* Indiana University, Population Institute for Research and Training, 14 p. (Paper presented at the conference, The history of registration of cause of death, Bloomington, Indiana, 1993, CD no. 9.)

Kintner Hallie J., 1993. *Recording the epidemiologic transition in Germany, 1816 to 1934.* Indiana University, Population Institute for Research and Training, 24 p. (Paper presented at the conference, The history of registration of cause of death, Bloomington, Indiana, 1993, CD no. 12.)

Linnaeus (von) Carl A., 1763. *Genera Morborum [..] Upsaliae apud Christ. Erh. Steinert. 1763.* Stockholm, Lars Salvius, Upsaliae, 40 p.

Manton Kenneth G. and Stallard Eric, 1984. *Recent trends in mortality analysis*. Orlando, FL, Academic Press, 342 p. (Studies in Population.)

Meslé France, 1995a. L'enregistrement des causes de décès en Europe, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 401–431. Louvain-la-Neuve, Academia/L'Harmattan, 671 p.

Meslé France, 1995b. La mortalitéen France: le recul se poursuit. *Population*, vol. 50(3), p. 743–776.

Meslé France, 2000. *Causes de décès aux grands âge: validité et comparabilité*. Paris, INED, 22 p. (Paper presented at the IUSSP seminar, Longevity, Individual Life Duration and Growth of the Oldest-Old, Montpellier, 23–25 October, 2000.)

Meslé France and Vallin Jacques, 1988. Les composantes de la mortalité cardio-vasculaire en France depuis 1925: résultats d'une reconstitution historique. *Population*, vol. 43(2), p. 391–425.

Meslé France and Vallin Jacques, 1993. Causes de décès: de la 8e à la 9e révision, deux cas différents, la France et l'Angleterre, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European demography. II. Population dynamics*, p. 421–445. Paris, John Libbey, INED, 526 p.

Meslé France and Vallin Jacques, 1996. Reconstructing long-term series of causes of death, *Historical methods*, vol. 29(2), p. 72–87.

Monnier Alain and Nizard Alfred, 1985. Les causes de décès, *in:* Roland Pressat (ed.), *Manuel d'analyse de la mortalité*, p. 99–122. Paris, INED, OMS, 164 p.

Pinel Philippe, 1797 (an VI). *Nosographie philosophique ou la méthode de l'analyse appliquéeà la médecine.* Paris, Maradan, 2 vol., 307 p. + 403 p. (Revised edition, Paris, Hachette, 1975.)

Pollard John H., 1982. The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries*, vol. 109, part II, p. 225–240.

Pollard John H., 1988. Causes de décès et espérance de vie: quelques comparaisons internationales, *in:* Jacques Vallin, Stan d'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 291–313. Paris, INED, PUF, 458 p. (Travaux et documents, Cahier no. 119.)

Pollard John H., 1990. Cause of death and expectation of life: some international comparisons, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analyses of mortality: new approaches*, p. 269–292. Oxford, Clarendon Press, 416 p. (International Studies in Demography.)

Pressat Roland, 1985. Contribution des écarts de mortalité par âge à la différence des vies moyennes. *Population*, vol. 40(4–5), p. 766–770.

Razoux Jean, 1767. *Tables nosologiques et météorologiques très étendues dressées à l'Hôtel-Dieu de Nîmes depuis le 1er juin 1757 jusqu'au 1er janvier 1762.* Basle, J. Im-Hof et fils, 359 p.

Registrar General, 1856. *Sixteenth annual report 1856.* London, Registrar General of England and Wales, 73 p.

Romeder Jean-Marie and McWhinnie John R., 1978. Le développement des années potentielles de vie perdues comme indicateur de mortalité prématurée. *Revue d'épidémiologie et de santé publique*, vol. 26(1), p. 97–115.

Sauvages (Boissier de Lacroix) François, 1731. *Nouvelles classes de maladies qui dans un ordre semblable à celui des botanistes comprennent les genres et les espèces de toutes les maladies avec leurs signes et leurs indications.* Avignon, Bibliothèque universitaire de Montpellier. (EC83, 8e, Boissier de Lacroix: classes nouvelles.)

Schwartz Daniel and Lazar Philippe, 1961. Taux de mortalité pour une cause donnée de décès en tenant compte des autres causes de décès ou de disparition, *Revue de l'Institut International de Statistique*, vol. 29(3).

Taucher Erica, 1978. *Chile: mortalidad desde 1955 a 1975. tendencias y causas.* Santiago du Chili, CELADE, 83 p. (Series A, no. 162.)

Toulemon Laurent, 1996. *La part des âges et des causes de décès dans les différences d'espérance de vie.* Unpublished. Paris, INED, 10 p.

Valkovics Emil, 1984. L'évolution récente de la mortalité dans les pays de l'Est: essai d'explication à partir de l'exemple hongrois. *Espace, Populations, Sociétés*, no. 3, p. 141–168.

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978.* Paris, INED, PUF, 608 p. (Travaux et Documents, Cahier no. 115 + 7 volumes annexes.)

Vallin Jacques and Meslé France, 1998. Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925 à 1993, *in:* Gérard Pavillon (ed.), *Enjeux des classifications internationales en santé*, p. 113–156. Paris, Éditions INSERM, 220 p. (Questions en santé publique.)

Vallin Jacques and Nizard Alfred, 1978. Les causes de décès en France. I.Pour une typologie simple et homogène: application à la période 1968–1974, *Population*, vol. 33(3), p. 547–608.

Vincent Paul, 1966. Effets sur la mortalité de la disparition éventuelle de certaines causes de décès, *in: Documents officiels de la Conférence démographique européenne. Strasbourg, 1966*, p. 1–18. Strasbourg, Conseil de l'Europe, pag. mult. (4 volumes).

WHO, 1993. *International statistical classification of diseases and connected health problems*, 10th revision, vol. 1. Geneva, World Health Organization, 1335 p.

WHO, 1995. *International statistical classification of diseases and connected health problems*, 10th revision, vol. 2. Geneva, World Health Organization, 167 p.

Zimicki Susan, 1988. L'enregistrement des causes de décès par des non-médecins: deux expériences au Bangladesh, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 101–122. Paris, INED, PUF, 458 p. (Travaux et Documents, Cahier no. 119.)

Zimicki Susan, 1990. Approaches to assessment of the cause structure of mortality: a case-study from Bangladesh , *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analyses of mortality: new approaches*, p. 99–122. Oxford, Clarendon Press, 416 p. (International Studies in Demography.)

CHAPTER

43

The Assessment of Causes of Death in Developing Countries

VINCENT FAUVEAU

United Nations Population Fund, New York, United States

The importance of collecting and analyzing causes of death for demography and public health has been much emphasized (Lopez, 1990; Vallin and Meslé, 1988; WHO, 1987, 1994). The concept of mortality by cause has become predominant in the analysis of the evolution of general mortality and the mortality of particular groups (e.g., infants, children, women of child-bearing age), in comparing the mortality of different groups (e.g., rural–urban, boys–girls, rich–poor) or evaluating the effect of public health interventions (e.g., vaccinations, mass treatment, hygiene). Health ministries of every country now have a medical statistics department responsible for publishing the main causes of death, often relying on the famous World Health Organization (WHO) International Classification of Diseases (ICD), which was adopted for the first time in 1893, or on variants adapted to local conditions.

If most developed countries fulfill the main condition for this analysis (i.e., good quality of collected data) because of their rigorous medical tradition and the generalization of deaths in medical environments, the same is not true for developing countries. In situations where most deaths occur outside of a medical context, it has become necessary to find alternative solutions to the problem of establishing causes, either in experimental and limited conditions (on sample groups or in certain pilot zones) or preferably as a routine practice. In this latter case, the first issue is to ensure that all deaths in the population sample are recorded, despite the absence of a reliable civil registration system. To consider only deaths occurring in a hospital environment does not guarantee representation and is therefore not suitable for any satisfactory statistical analysis from a public health point of view.

This chapter provides the historical background of the origins of alternative methods of establishing causes of death in developing countries, before introducing a detailed description of the practice of *verbal autopsy*, its applications and its limitations. Other possible techniques are also mentioned. The chapter ends with a discussion of the conditions of extension for the use of these alternative methods of collecting causes of death.

I. BACKGROUND

Arab doctors were the first to perform autopsies in an attempt to understand causes of death. In medieval Europe, autopsies were rarely performed, and if so, they were done illegally. It was not until the beginning of modern medicine at the end of the 18th century that this technique was officially accepted, which is now well codified.

As France Meslé described in Chapter 42, medical certification and its corollary, the classification of the causes of death, became widespread in France during the 19th century, eventually leading to the development of the international terminology that has been several times revised by WHO (Vallin and Meslé, 1988; WHO, 1977).

Regarding the establishment of causes of death in countries with incomplete statistics, in 1956, Yves Biraud published the first article recommending the use of information supplied by a deceased person's family to estimate the cause of death. In 1978, WHO started publishing lists of causes of death for the use of nonphysicians working in developing countries.

The first demographic studies aimed at collecting information on causes of death in the general population were conducted in Khanna, Northern India, between 1955 and 1960 (Wyon and Gordon, 1971) and in Companiganj, Bangladesh, between 1975 and 1978 (Chowdhury and Khan, 1980). In the same period, a team from Johns Hopkins University were starting work on a longitudinal study of a health program in rural Punjab in Narangwal, India. Among those researchers, Arnold Kielmann and colleagues (1983) were the first to coin the term *verbal autopsy*, a questionable expression but one which is now universally used.

Shortly after and in parallel, other researchers worked on the same subjects in large longitudinal population study laboratories, which were run by demographers and physicians in Matlab, Bangladesh (Zimicki *et al*., 1985; D'Souza, 1981), Niakhar, Senegal (Garenne and Fontaine, 1988, 1990), Kenya (Omandi-Odhiambo *et al*., 1984), and Gambia (Lamb *et al*., 1984). Verbal autopsy methods were gradually theorized, discussed and refined, particularly in Matlab (Fauveau *et al*., 1994; Zimicki, 1988, 1990).

The first international seminar organized to share different experiences of verbal autopsies took place in Baltimore in March 1989 at the Johns Hopkins University School of Public Health (Gray *et al*., 1990). During this meeting, the issue of multiple causes and that of putting verbal autopsies into algorithms, particularly those concerning infant mortality, were discussed for the first time. In 1994, a second seminar was organized by the WHO at the London School of Hygiene and Tropical Medicine, focusing on the use of the verbal autopsy for the causes of female deaths and particularly for maternal deaths. Once more, the debates were focused on the analysis of multiple causes, the design of algorithms, and on validation problems (WHO, 1995).

Since then, the verbal autopsy method has been widely used, particularly on samples from some demographic and health surveys, and numerous publications report its use in several developing countries. The Department of Reproductive Health and Research of the WHO included the methodology of verbal autopsy in their publication on maternal deaths, *Beyond the Numbers* (WHO, 2003).

II. THE VERBAL AUTOPSY METHOD

Verbal autopsy can be defined as "a technique of using the information acquired from a recently deceased person's entourage to reconstruct the events and the symptoms that preceded the death so as to deduct a medically acceptable cause or causes of death" (Fauveau, 1995). This definition (every word counts) contains the limitations of the method, particularly concerning the validity or the official recognition and the medically acceptable aspect. Its advantage remains its capacity to compensate for the absence of a medical certification in regions where that is not available. Of the six successive stages of the method and their potential variations described hereafter, the most difficult is the first, the stage that affects the statistical validity of the entire procedure.

Stage 1: Gathering and Recording Death Information

To begin the procedure, all deaths occurring during the considered period in the relevant age group (or if necessary, in all age groups) must be counted and recorded. The information should include the date and place of death, sex, age (date of birth if known), and name and address of the deceased person. This very basic but precise task, which does not require any medical knowledge, is generally undertaken by a representative of the demographic surveillance system or of civil registration. Using only deaths occurring in a hospital environment must be avoided as in most cases, these deaths constitute a biased group, and they are certainly not representative of all deaths.

Stage 2: Interview on the Cause of Death

The information gathered on the deaths is transmitted to a health agent (or health worker) who visits the family or friends of the deceased, ideally within a period of 1 to 3 months, to interview them about the circumstances of death in the local language. This interview, which must not under any circumstances be perceived as a police interrogation, consists of two parts. First, a free narration of the illness is recorded with emphasis on the chronologic order of events. Second, the important elements are revisited for details of the patient's observed symptoms or at least those symptoms perceived and described by the family. It is recommended that this interview is conducted in the presence of two or three people who were present during the progression of the illness to be able to optimize the details by cross-checking. The interviewer must be very knowledgeable in the local pathology and in the locally used names and explanations, but she or he must not be responsible for the diagnosis.

Stage 3: Interpretation by a Physician or an Experienced Health Worker

The reported facts are then interpreted by a public health doctor, an experienced nurse, or a community health worker experienced in the interpretation of local symptoms to make a diagnosis of the cause of death: first the *principal cause*, which was responsible for the original illness or the sequence leading to the death; then the *immediate cause*, which directly caused death; and finally, the *associated causes*, ailments that were not responsible for the patient's death but which contributed toward the chain of events leading up to death. The health professional notes his or her conclusions and then compares the causes with existing lists consisting of the categories of causes that are being investigated and prepared by the public health authority. In the most rigorous variant of this process, the reported facts are given to three health professionals who independently write their conclusions. A fourth person then compares the diagnoses by applying the rule of simple majority. If all three readers have reached the same conclusion, it is accepted; if only two have come to the same conclusion, a second reading takes place before the conclusion is accepted; and if none of the three readers is in agreement with the others, the case is filed in the *impossible to attribute* category.

Stage 4: Returning to the Field for Incomplete Cases

When the readers indicate that they have been unable to make a diagnosis in the absence of precise information, they must list on the form the missing information required to proceed. The visiting health agent (or field worker) must return to see the deceased person's family to try to collect the necessary information before resubmitting the completed forms into the system.

Stage 5: Coding and Classifying

There are many classification and coding systems. After a system has been selected, it is important to adhere to it in the long term to be able to follow the evolution of the causes of death over the years or compare the development of certain causes over a certain period. WHO's detailed ICD list is not directly useable for verbal autopsies because it is too complex and detailed, and it is too often based on precise medical diagnoses, relying greatly on anatomic-pathologic information. However, it is useful to refer to its main categories by allocating the first number of a three-digit code according to large categories (e.g., respiratory diseases, diarrheal diseases, accidents). It is also often useful to add certain codes according to a combination of pathologies or frequently observed local causes. For example, for a death attributed to a complication of measles, three different codes can be used: one for death due to an immediate neurologic complication of measles (i.e., during eruption), another for a death due to a late respiratory complication, and a third for a death due to a late digestive complication (i.e., an invasive type of diarrhea). During the analysis, this cause of death could therefore be classified as neurologic, respiratory, digestive, or infectious (measles) according to the chosen orientation. It could also be useful to add a prefix or a suffix to indicate the diagnostics confirmed by a paraclinical examination (e.g., laboratory, radiology), even if these cases are rare. Special procedures are recommended for certain categories of causes, representing three cases in particular.

a. Deaths in the Neonatal Period

Deaths in the neonatal period are great in number but difficult to attribute using information collected by verbal autopsy and particularly so in regions where supernatural explanations persist (e.g., magic, linked to the "evil eye"). To better process the causes of these deaths, particular attention must be paid to the chronology of events with regard to the birth. For example, on the basis that it is biologically nearly impossible to die of neonatal tetanus before the fourth day of life and that these deaths are rare after the fourth week of life, the diagnosis of neonatal tetanus can be refined by differentiating it from neurologic complications due to anoxia, obstetric trauma, or other neonatal infections. By trying to determine the length of the child at birth rather than the weight (which is rarely measured), knowledge of deaths linked to prematurity and intrauterine growth retardation can also be improved (Fauveau *et al*., 1990c). If a special questionnaire is used for all neonatal deaths, it would be of no use for any other deaths (Srivastava *et al*., 2001).

b. Deaths from Malnutrition

During an interview on infant death, particular attention should be paid to symptoms of malnutrition and their chronology. Except for certain cases of acute malnutrition, these symptoms appear progressively and are not immediately noticed by the mother. In general, neither the traditional growth chart nor measurement of weight can be relied on. It is often remarks made by family members who only see the child occasionally that draw attention to the condition. Moreover, coding and classification of malnutrition will be different according to whether it is to be presented as one cause of death amongst others or as an underlying cause, which is more plausible and more useful from a public health point of view (Fauveau *et al.*, 1990a).

c. Deaths of Women of Child-Bearing Age

Regarding maternal mortality, as in the previous case, and especially in those regions where the attitudes and roles concerning female health problems are very distinct (in particular when men have little access to information concerning obstetrics and gynecology), it is highly advisable to entrust the verbal autopsy questionnaire to a female health worker who is respected in the local community and who is knowledgeable in local practices and beliefs. To detect the specific role of the complications of an induced abortion, a sensitive questionnaire should be used to determine the presence and regularity of menstrual periods, the respondent's (or her family's) awareness of pregnancy in the early stages, her attitude and comments regarding her wish to carry the pregnancy to term, contraceptive practices and methods used, and social, economic, or marital problems (Fauveau *et al.*, 1988, 1989) (see Chapter 45).

Stage 6: Analysis and Publication

An initial analysis can be produced simply with a spreadsheet (EPI-INFO type) which is capable of calculating the exact age of the person at the time of death from the dates of birth and death (when known). The results are presented in tables organized by age group, sex, and categories of causes of death. When several years of observation are available, they can be grouped in the same table, thereby neutralizing the effect of annual variations (especially if the number of cases is very low); otherwise, the annual results can be presented to analyze the trends. Obviously, aggregating the causes of death for a whole population, including all age groups, should be avoided because of the differences in cause of death according to age and the variations of age structure. The classic table includes the absolute number of deaths by each category of cause in one column and the percentage of each category with regard to all deaths in the group in another column. In this way, the *proportional mortality rate* is obtained. This is the most frequently used indicator in publications concerning these types of data. However, if denominators are available, represented by the mid-year population or the population of the middle of the period under scrutiny, a third column of *cause-specific mortality rates* per 1000 or 100,000 persons at risk by age group and sex can be introduced. Table 43–1 gives an example of both indicators.

Probabilities of dying for 1000 live births can also be constructed, particularly for infant and child mortality. It is interesting to compare causes of death between years (or periods before and after a program), between regions (e.g., a control zone with a case zone), and between sexes using appropriate statistical tests (i.e., comparison of rates or proportions).

III. USES OF THE VERBAL AUTOPSY

A first use of the verbal autopsy method is to show the comparative importance of the main causes of death in a given population and in a given age group, either in a relative way by ranking them, or in an absolute way using cause-specific mortality rates. This application can be made at a local level in a population laboratory (Hoj *et al.*, 1999; Marsh *et al.*, 1993; Chowdhury *et al.*, 1991; Fauveau *et al.*, 1991d; Fauveau *et al.*, 1989; Omandi-Odhiambo *et al.*, 1984) or at a national level from a representative sample, such as a demographic and health survey (MSP, 1990) or by following the evolution of specific causes.

The verbal autopsy technique also is used in guiding the decision-making process that precedes the implementation of a specific intervention or set of interventions, such as the targeting of nutritional programs (Fauveau *et al.*, 1990a) or the choice of an optimal age for vaccination (Fauveau *et al.*, 1991a).

A third category of use is in evaluating the effect of specific programs on the corresponding cause-specific mortality. One of the classic objectives of public health is to evaluate the effect of interventions (preventative and curative) by a comparative examination of mortality rates. In developing countries, analyses were usually limited to crude death rates or restricted to certain age-groups and sex, such as infant or childhood mortality and mortality of women of child-bearing age (Lamb *et al.*, 1984). However, under pressure from international organizations and in connection with the establishing of large-scale vertical programs to fight epidemics or endemic diseases, it quickly became apparent that it was necessary to refine the presentation of mortality rates by subdividing them into cause-specific rates. In the same way, analysts can combine the cause-specific mortality rates with those by age and sex to further refine the evaluation. This difficult exercise was used to evaluate measles and tetanus vac-

TABLE 43–1 Proportional Mortality Rates and Cause-Specific Mortality Rates by Sex for Ages 1 to 4 Years According to Verbal Autopsy in Matlab, Bangladesh, 1986 to 1987

Causes of death	Boys (n = 24,219)			Girls (n = 22,652)			Difference between rates
	N	CSMR (per 100)	PMR (%)	N	CSMR (per 1000)	PMR (%)	
Acute watery diarrhea	12	0.5	4.7	22	1.0	5.2	0.5
Acute invasive diarrhea	21	0.9	8.3	35	1.5	8.3	0.7[a]
Post-measles diarrhea	14	0.6	5.5	27	1.2	6.4	0.6[a]
Persistent diarrhea	12	0.5	4.7	31	1.4	7.3	0.9[b]
Total diarrhea	59	2.4	23.2	115	5.1	27.2	2.6[b]
SM with chronic diarrhea	52	2.1	20.5	136	6.0	32.2	3.9
SM with respiratory infections	3	0.1	1.2	8	0.4	1.9	0.2
Other SM	16	0.7	6.3	23	1.0	5.4	0.4
Total SM	71	2.9	28.0	167	7.4	39.5	4.4[b]
Acute lower respiratory infection	34	1.4	13.4	27	1.2	6.4	–0.2
Vaccine Preventable Diseases[c]	12	0.5	4.7	13	0.6	3.1	0.1
Fever, septicemia	5	0.2	2.0	0	0.4	2.4	0.2
Other infectious diseases	4	0.2	1.6	10	0.4	2.4	0.3
Total other infectious diseases	55	2.3	21.7	60	2.6	14.2	0.4
Drowning	50	2.1	19.7	51	2.3	12.1	0.2
Other accidents	4	0.2	1.6	4	0.2	0.9	0.0
Total accidents	54	2.2	21.3	55	2.4	13.0	0.2
Other causes	3	0.1	1.2	9	0.4	2.1	0.3
Impossible to determine	12	0.5	4.7	17	0.8	4.0	0.3
Total deaths	254	10.5	100.0	423	18.7	100.0	8.2[b]

CSMR, cause-specific mortality rate; PMR, proportional mortality rate; SM, severe malnutrition.
[a] $P < 0.05$.
[b] $P < 0.01$.
[c] Measles, whooping cough, and tetanus.

cination programs (Koenig *et al.*, 1991; Aaby *et al.*, 1988), programs to reduce maternal mortality (Fauveau *et al.*, 1991c), programs to reduce mortality from acute respiratory infection (Fauveau *et al.*, 1992a; Bang *et al.*, 1990), and programs to reduce mortality caused by acute diarrhea (Fauveau *et al.*, 1992b) or malaria (WHO, 1991). Verbal autopsy has also been used in therapeutic tests on tropical pathology (Pacque-Margolis *et al.*, 1990).

A fourth category of the use of verbal autopsy applies to the selection of strategies or *packages of health interventions* with specific objectives, such as to reduce the excess mortality of girls compared with boys in certain communities (Fauveau *et al.*, 1991b) or to optimize the cost-effectiveness of a health program (Attanayake *et al.*, 1993).

Verbal autopsy techniques also can be used to make international comparisons. For instance, in the case of the International Study Group on Diarrhoeal Mortality (Victora *et al.*, 1993), in a coordinated multicenter study, a common methodology was applied to several sites around the world and comparative results were then produced.

IV. PROBLEMS AND LIMITATIONS OF THE VERBAL AUTOPSY

Several issues relative to the methodology appear in publications reporting the use of the verbal autopsy: coverage of deaths, the quality of the interview and the aptitude of the interviewers, coding and classification, processing of undetermined causes and multiple causes, use of algorithms, validity and validation, cost and continuity, and particular cases of diseases with nonspecific symptoms.

1. Coverage of Deaths

To maximize the value of interpretations of the results of verbal autopsy analysis, it is necessary to avoid selection bias and to ensure that *all* deaths in the relevant age group occurring in the community have been recorded and have been subject to the necessary investigations. For example, a selection of deaths recorded by a hospital or a community clinic would be of no value in producing usable statistics in public health in most developing countries. However, the verbal autopsy method has been used for deaths in hospitals in validation studies (Chandramohan *et al.*, 1998; Snow *et al.*, 1992). The same problem applies to verbal autopsies conducted on samples at the time of demographic and health surveys if the samples are not representatively weighted.

2. Quality Control of the Interviews

In the case of an issue as sensitive as that of the death of a relative, the interview should take place as soon as possible after the event, which means that it will often take place during the period of mourning. It requires subtlety and tact to be comprehensive and precise while respecting local customs and cultural sensitivity. The need for the interviewer to be respectable and respected in the local community, to be preferably female when a woman's death is under review, to speak the local language, and to guide the progression of the interview without interpreting or speeding events has been emphasized. It would not be appropriate for the interviewer to be the same health worker who had treated the deceased person before death. It is essential to pay attention to the chronology of events. A supervisor should revisit a certain number of interviews at random to check their validity.

3. Algorithms

Taking into account the diversity of agents involved in collecting information in this process and in a concern for standardization, use can be made of decision trees or algorithms (Quigley *et al.*, 1996; WHO, 1995; Chandramohan *et al.*, 1994; Snow *et al.*, 1992). This procedure is useful to standardize interviews, but it does require careful development and thorough validation because an erroneous branch could be used that would lead to an absurd conclusion. Standardized questionnaires and simplified algorithms have been used, particularly since 1992, for the Demographic and Health Surveys (notably phase III, Macro International, 1992). An example is given in Figure 43–1.

4. Coding and Classification

WHO's ICD cannot be applied in its initial state to verbal autopsies because it is too complex and based on anatomic-pathologic factors. A simplified classification is therefore needed (Chandramohan *et al.*, 2001). Classifications are too frequently based on symptoms because of the nature of the collected information, and systems based on causes would be of better use in public health care. Marie-Hélène Bouvier-Colle and colleagues (1976) and Jacques Vallin and Alfred Nizard (1978) proposed a more coherent classification based on causes to be applied to developed countries (Meslé, 1999; Vallin and Meslé, 1988; Meslé and Vallin, 1984). However, no standardized simplified classification has been adapted for use with verbal autopsy.

5. Multiple Causes

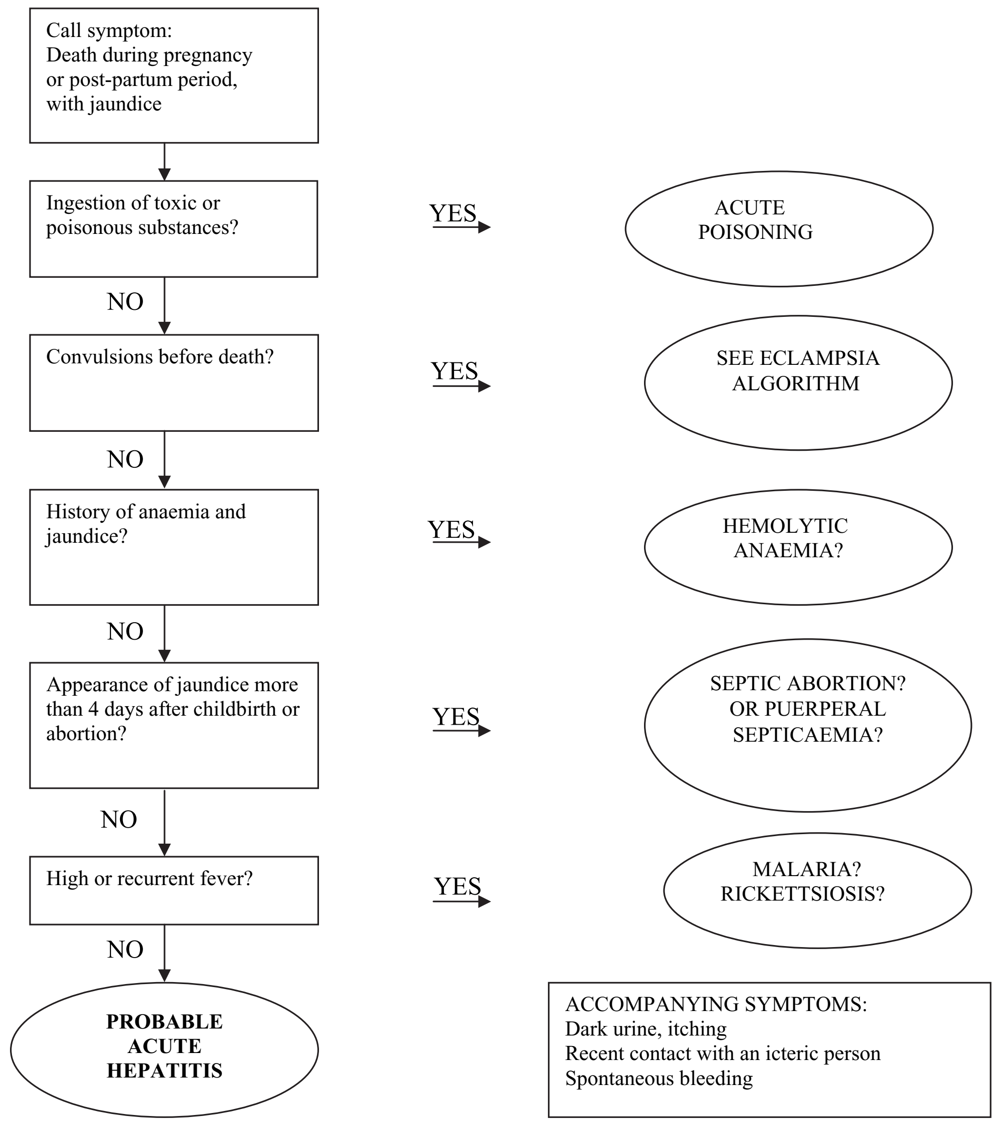
If, as for the ICD, the practice has been spread to code a *principal* cause (which initiated the sequence of events leading to death, such as abortion), an *immediate* cause (which directly caused the death, such as genital hemorrhage), and an *associated* cause (which contributed to the death, such as anemia), the standardization of directives to attribute these codes remains incomplete and the comparability of the studies subject to caution. Several large categories of causes present difficult coding problems: trauma or violent causes, diabetes, alcoholism, malnutrition, and acquired immunodeficiency syndrome (AIDS).

Moreover, it has often been recommended in public health to consider the sociobehavioral causes of death to shed more light on avoidable factors by approaches that are not exclusively medical, such as "contraceptive failure," "illegal abortion," or "refusal of medical assistance" (Fauveau and Blanchet, 1989). The famous classification of "the three delays" has been proposed as an explanation for maternal mortality: a delay in the decision to refer the parturient in difficulty, a delay due to transport time, and a delay in treatment once the patient has arrived at the referral centre. It is clear that this presentation leads to a proposal for preventive action of a behavioral nature rather than medical intervention.

6. Undetermined Causes

The recognized proportion of causes impossible to determine by verbal autopsy, without counting those for which the information is not available, varies significantly according to different studies. High proportions (>20%) raise doubts about the interpretation of the results provided, but recognition of the impossi-

FIGURE 43–1 Example of an Algorithm for the Logical Progress of the Diagnosis by Verbal Autopsy of Puerperal Hepatitis



From WHO, 1995.

bility of attributing causes generally guarantees greater credibility of the presented results.

7. Validation

The issue of objective validation of the results of verbal autopsies was raised from the first tests of the method (Mobley *et al*., 1996; Kalter *et al*., 1990; Dutta *et al*., 1988), only to run into the problem of the gold standard. In theory, validation can be made only by comparing verbal autopsy results with results from medical certificates made by qualified doctors using complete medical case records. Snow and colleagues (1992) tested this comparison by administering a verbal autopsy on a sample of 303 Kenyan children's deaths, the diagnoses of which had been established in a hospital. They obtained a good specificity (between 80% and 95%) for measles, neonatal tetanus, and trauma but an insufficient sensitivity (less than 50%) for malaria, anemia, acute respiratory infections, gastroenteritis, and non-epidemic meningitis. Quigley and colleagues (1996), using the same series, showed that for the diagnosis of malaria, a sensitivity greater than 70% could be obtained by using an algorithm more adapted to the local symptoms. Validation studies were applied to neonatal deaths (Marsh *et al*., 2003).

8. Particular Cases of Certain Pathologies

Simple cases in which the presence of a pathognomonic sign indicates an indisputable cause without any ambiguity are rare. Cases with nonspecific symptoms, for which the analyst must be satisfied with a group of indicators (e.g., malaria, AIDS) are more common.

a. Malaria

Numerous researchers have underlined the difficulty of diagnosing malaria in endemic regions because of the weak specificity of symptoms and clinical aspects and the rarity of biologic confirmation (Todd *et al*., 1994; Snow *et al*., 1992). The classic combination of fever, headaches, shivers, and seasonal anemia remain most common, but this profile is rarely complete. Validation studies have shown sensitivity below 50%. Clinical aspects of pernicious malaria are often impossible to distinguish from those of meningitis, except in the case of a typical epidemic.

b. Acquired Immunodeficiency Syndrome

In some countries or population groups, AIDS has emerged as a major, if not principal, cause of death. Knowing that AIDS sufferers generally die of interrelated infections or neoplasms and often of multiple complications, it has become necessary to develop a set of criteria that can lead to this diagnosis or at least indicate a strong likelihood (Doctor and Weinreb, 2003; Okongo *et al*., 1998; Kamali *et al*., 1996; Dowell *et al*., 1993). Among the principal criteria are belonging to a group most at risk (e.g., young adult, sex worker) and the presence of characteristic symptoms (e.g., cough or recurring diarrhea, weight loss, buccal mycosis, appearance of particular tumors). The diagnosis is even more difficult in the case of the AIDS-tuberculosis combination and AIDS in children.

V. OTHER TECHNIQUES

The objective of this chapter is not to detail the innovative techniques intended to determine and record all deaths occurring in a population. However, an alternative method of investigating the causes of death, the *medical audit*, should be mentioned. This method was developed to improve patient care in hospitals, particularly for perinatal (Edouard, 1985) and maternal mortality (Anonymous, 1992). The method consists of closely examining all elements of patient care from arrival at the hospital or even from the decision to transfer the patient to hospital. Factors such as decision-making delays, errors in orientation, errors in diagnosis, treatment errors, and lack of surveillance are analyzed by an external investigator or, more frequently, by the hospital or clinic staff, always with a view toward improving care by insisting on extra training to prevent a future recurrence. Details of these newer techniques applied to maternal deaths are found in the WHO's *Beyond the Numbers* (WHO, 2003).

CONCLUSIONS AND PROSPECTS

The increase in publications by public health organizations concerned with developing countries and using alternative methods to establish causes of death, notably the verbal autopsy, indicates a growing interest in these methods. Most authors express reservations about the absolute validity of the results, but they recognize that in the absence of an universal, and universally reliable, medical certification, these methods have acquired legitimacy. Several authors stress their weak sensitivity, especially in the context of endemic diseases with nonspecific symptoms. Nonetheless, numerous organizations, health ministries, and research institutes are still unable to respect the criteria for reliability in the use of these methods and run the risk of lessening their credibility.

Under what conditions can the extension of use of the verbal autopsy be recommended?

1. *The demographic surveillance system should be independent from the service provider system*. The neutrality and objectivity of determining causes of death are not guaranteed if we ask the physicians and nurses concerned by these deaths (which are perceived as failures of the system) to provide explanations for them. However, maintaining two parallel systems at the same time does require a large budget.
2. *The quality of the procedures* depends on training, initial and continuous, of the agents involved and on the constant and constructive supervision of these agents. Combined with the logistic needs, maintaining quality involves high costs.
3. *Consistency, sustainability, and continuity* of quality, essential conditions for studying chronologic series, involve high maintenance costs that are often incompatible with the resources of a developing country.
4. *Validation* by comparison with reliable hospital data, when available, should be an integral part of the

procedure. However, hospital data too often include only the diagnosis made on admission, which is not reviewed on release after (or before) death. This can lead to wrong conclusions.

5. International consensus does not exist in *the standardization of procedures*, making regional or international comparisons difficult. Several decision trees and algorithms have been published, showing that the best combinations of sensitivity, specificity and replicability can only be applied to a small number of distinct pathologies (e.g., tetanus, measles, meningitis, epidemics of cholera). Sometimes, even algorithms have been shown to be a source of confusion.

6. *Standardization of classifications* suffers from a lack of consensus, even though it seems more feasible. Only the WHO can propose appropriate lists. The processing of principal, immediate, and underlying associated causes, as well as multiple causes, remains subject to debate. A certain amount of caution should also be applied when processing causes that are impossible to determine.

Acknowledgment

This chapter was translated from the French version by O. D. Chimere-Dan.

References

Aaby Peter, Hansen Henrik L., Tharup Jesper, Sodemann Morten, Knudsen Kim, Jensen Thoger G., Kristiansen Hans, Poulsen Anja, Jakobsen Marianne, Da Silva Maria C., and Whittle Hilton, 1988. Trial of high-dose Edmonston Zagreb measles vaccine in Guinea-Bissau: protective efficacy, *Lancet*, vol. 2(8615), p. 809–814.

Anonymous, 1992. Setting up a medical audit, *Safe Motherhood Newsletter*, no. 9, p. 8.

Attanayake Nimal, Fauveau Vincent, and Chakraborty Jyotsnamoy, 1993. Cost-effectiveness of Matlab MCH-FP project in Bangladesh, *Health Policy and Planning*, vol. 8(4), p. 327–338.

Bang Abhay T., Bang Rani A., Morankar V., Sontakke P., Tale O., Solanki J., Kelzarkar P., Dudhbade A., Jengathe D., and Wargantiwar R., 1992. Diagnosis of causes of childhood deaths in developing countries by verbal autopsy: suggested criteria. The Search Team, *Bulletin of the World Health Organization*, vol. 70(4), p. 499–507.

Bang Abhay T., Bang Rani A., Tale O., Sontakke P., Solanki J., Warguntiwar R., and Kelzarkar P., 1990. Reduction in pneumonia mortality and total childhood mortality by means of community-based intervention trial in Gadchiroli, India, *Lancet*, vol. 336, p. 201–206.

Biraud Yves, 1956. Méthodes pour l'enregistrement par des non médecins des causes élémentaires de décès dans les zones sous-développées. Geneva, WHO, 65 p. (Unpublished document no. HS/60.)

Biraud Yves, 1961. Une méthode d'enregistrement des causes approximatives de décès dans les régions sous-développées, *in:* IUSSP (ed.), *General population conference, New York, 1961*, vol. 2, p. 348–354. London, IUSSP, 597 p.

Bouvier-Colle Marie-Hélène and Rotschild Adrienne, 1976. Causes médicales de décès selon les processus étiopathogéniques and localisations anatomiques. Essai d'un nouveau mode de regroupement des rubriques de la International Classification of Diseases (8th revision), *Bulletin of the World Health Organization*, vol. 54, p. 495–506.

Chandramohan Daniel, Maude Gillian, Rodrigues Laura, and Hayes Richard, 1994. Verbal autopsies for adult deaths: issues in their development and validation, *International Journal of Epidemiology*, vol. 23(2), p. 213–222.

Chandramohan Daniel, Rodrigues Laura, Maude Gillian, and Hayes Richard, 1998. The validity of verbal autopsies for assessing the causes of institutional maternal deaths, *Studies in Family Planning*, vol. 29(4), p. 414–422.

Chandramohan Daniel, Setel Philip, and Quigley Maria, 2001. Misclassification error in verbal autopsy: can it be adjusted? *International Journal of Epidemiology*, vol. 30(3), p. 509–514.

Chowdhury Hafizur, Fauveau Vincent, Yunus Mohamed, Zaman Khalid, and Briend André, 1991. Is acute watery diarrhea an important cause of morbidity and mortality among rural Bangladeshi children? *Transactions of the Royal Society of Hygiene and Tropical Medicine*, vol. 85, p. 128–130.

Chowdhury Shafiq and Khan Ah, 1980. *Causes of death, Companiganj, Bangladesh, 1975–1978*. Dacca, Christian Commission for Development in Bangladesh, 80 p.

Cohen Joel E., 1988. Estimating the Effects of Successful Malaria Control Programs on Mortality, *Population Bulletin of the United Nations*, no. 25, p. 6–26.

Doctor Henry and Weinreb Alexander, 2003. Estimation of AIDS adult mortality by verbal autopsy in rural Malawi. *AIDS*, vol. 17, p. 2509–2513.

Dowell Scott F., Davis Homer L., Holt Elizabet A., Ruff Andrea J., Kissinger Patricia J., Bijoux Judith, Boulos Reginald, Boulos Carlo, and Halsey Neal A., 1993. The utility of verbal autopsies for identifying HIV1-related deaths in Haitian children, *AIDS*, vol. 7(9), p. 1255–1259.

D'souza Stan, 1981. *A population laboratory for studying disease processes and mortality: the demographic surveillance system, Matlab, Comilla, Bangladesh*. Dhaka, ICDDR, B, 29 p. (Special Publication, no. 13.)

Dutta N, Mand M, and Kumar Vinod, 1988. Validation of causes of infant death in the community by verbal autopsy, *Indian Journal of Pediatrics*, vol. 55(4), p. 599–604.

Edouard Lindsay, 1985. Epidémiologie de la mortalité périnatale, *Rapport Trimestriel de Statistiques Sanitaires Mondiales*, vol. 38(3), p. 289–301.

Ewbank Douglas, 1984. Uses of mortality data for evaluating the success of specific health and development programmes, *in:* United Nations (ed.), *Databases for mortality measurement*, p. 18–30. New York, United Nations, 164 p. (Population Studies, no. 84.)

Fauveau Vincent, 1995. *La mortalité par cause dans la planification and l'évaluation des programmes de santé maternelle and infantile*. Nancy, Université de Nancy 1, 164 p. (Doctoral thesis on public and social health).

Fauveau Vincent and Blanchet Thérèse, 1989. Deaths from injuries and induced abortion among rural Bangladeshi women, *Social Science and Medicine*, vol. 29(9), p. 1121–1127.

Fauveau Vincent, Briend André, Chakraborty Jyotsnamoy, and Sarder AbdulMajid, 1990a. The contribution of severe malnutrition to childhood mortality in rural Bangladesh: implications for targeting nutritional interventions, *Food and Nutrition Bulletin*, vol. 12(3), p. 215–219.

Fauveau Vincent, Chakraborty Jyotsnamoy, Sarder AbdulMajid, Khan Mehrab, and Koenig Michael, 1991a. Measles among under-9-month-olds in rural Bangladesh: its significance for age at immunization, *Bulletin of the World Health Organization*, vol. 69(1), p. 67–72.

Fauveau Vincent, Koenig Michael, Chakraborty Jyotsnamoy, and Chowdhury Anwarullah I., 1988. Causes of maternal deaths in rural Bangladesh, *Bulletin of the World Health Organization*, vol. 66(5), p. 643–651.

Fauveau Vincent, Koenig Michael, and Wojtyniak Bogdan, 1991b. Excess female deaths among Bangladeshi children: an examination of cause-specific mortality and morbidity, *International Journal of Epidemiology*, vol. 20(3), p. 729–735

Fauveau Vincent, Stewart Katherine, Chakraborty Jyotsnamoy, and Khan Shameem, 1992a. Impact on mortality of a community-based programme to control acute lower respiratory tract infections, *Bulletin of the World Health Organization*, vol. 70(1), p. 109–116.

Fauveau Vincent, Stewart Katherine, Khan Shameem, and Chakraborty Jyotsnamoy, 1991c. Effect on mortality of community-based maternity care programme in rural Bangladesh, *Lancet*, vol. 338, p. 1183–1186.

Fauveau Vincent, Wojtyniak Bogdan, Chowdhury Hafizur Rahman, and Sarder AbdulMajid, 1994. Assessment of cause of death in the Matlab demographic surveillance system, *in:* Vincent Fauveau (ed.), *Matlab: Women, Children, and Health*, p. 65–86. Dhaka, ICDDRB, 462 p. (Special Publication, no. 35.)

Fauveau Vincent, Wojtyniak Bogdan, Koenig Michael, and Chakraborty Jyotsnamoy, 1989b. Epidemiology and cause of death among women in rural Bangladesh, *International Journal of Epidemiology*, vol. 18(1), p. 139–145.

Fauveau Vincent, Wojtyniak Bogdan, Mostafa Golam, Sarder AbdulMajid, and Chakraborty Jyotsnamoy, 1990c. Perinatal mortality in rural Bangladesh: a community-based study, *International Journal of Epidemiology*, vol. 19(3), p. 606–612.

Fauveau Vincent, Yunus Mohamed, Shafiqul Islam, Briend André, and Bennish Michael, 1992b. Does ORT reduce diarrheal mortality? *Health Policy and Planning*, vol. 7(3), p. 243–250.

Fauveau Vincent, Yunus Mohamed, Zaman Khalid, Chakraborty Jyotsnamoy, and Sarder AbdulMajid, 1991d. Diarrhea mortality in rural Bangladeshi children, *Journal of Tropical Pediatrics*, vol. 37, p. 31–36.

Garenne Michel and Fontaine Olivier, 1988. Enquête sur les causes probables de décès en milieu rural au Sénégal, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 123–142. Paris, INED, PUF, 458 p. (Travaux et documents, Cahier no. 119.)

Garenne Michel and Fontaine Olivier, 1990. Assessing probable cause of death using a standardized questionnaire: a study in rural Senegal, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analyses of mortality: new approaches*, p. 123–142. Oxford, Clarendon Press, 416 p. (International Studies in Demography.)

Gray Ronald, Smith Gordon, and Barss Peter, 1990. *The use of verbal autopsy methods to determine selected causes of death in children*. Liège, IUSSP, Baltimore, The Johns Hopkins University School of Hygiene and Public Health, 46 p. (JHUSHPH Occasional paper no. 10 and IUSSP document no. 30.)

Hoj L, Stensballe J, and Aaby Peter, 1999. Maternal mortality in Guinea-Bissau: the use of verbal autopsy in a multi-ethnic population, *International Journal of Epidemiology*, vol. 28, p. 70–76.

Kalter Henry, Gray Ronald, Black Robert, and Gultiano Socorro, 1990. Validation of a postmortem interview method to ascertain selected causes of death in children, *International Journal of Epidemiology*, vol. 19, p. 380–386.

Kamali Anatoli, Wagner Hans-Ulrik, Nakiyingi Jessica, Sabiiti Imelda, Kengeya-Kayondo Jane F., and Mulder Daan W., 1996. Verbal autopsy as a tool for diagnosing HIV-related adult deaths in rural Uganda, *International Journal of Epidemiology*, vol. 25(3), p. 679–684.

Kielmann Arnold, Desweemer Cecile, Parker Robert, and Taylor Carl, 1983. Analysis of morbidity and mortality, *in: Child and maternal health services in rural India, the Narangwal experiment*, vol. 1. *Integrated nutrition and health care*, p. 172–214. Baltimore, Johns Hopkins University Press, 340 p.

Koenig Michael A, Fauveau Vincent, and Wojtyniak Bogdan, 1991. Mortality reduction from health interventions: the case of immunization in Bangladesh, *Population and Development Review*, vol. 17(1), p. 87–114.

Lamb W.H., Foord Frances A., Lamb Colette M.B., and Whitehead Roger G., 1984. Changes in maternal and child mortality rates in three isolated Gambian villages over ten years, *Lancet*, vol. 2, p. 912–914.

Lopez Alan D., 1990. Causes of death: an assessment of global patterns of mortality around 1985, *World Health Statistics Quarterly*, vol. 43, p. 91–104.

Macro International Inc., 1999. *Interviewer's manual for use with model B questionnaire for low prevalence countries*. Calverton, MD, 120 p. (Demographic and Health Survey III, Basic Document.)

Marsh D., Majid N., Rasmusen Z., Mateen K., and Khan Ah. A., 1993. Cause-specific child mortality in a mountainous community in Pakistan by verbal autopsy, *Journal of the Pakistan Medical Association*, vol. 43(11), p. 226–229.

Marsh D.R., Sadrudin S., Fikree Faryial, Krishnan C., and Darmstadt Gary, 2003. Validation of verbal autopsy to determine the cause of 137 neonatal deaths in Karachi, Pakistan, *Paediatric and Perinatal Epidemiology*, vol. 17(2), p. 132–142.

Meslé France, 1999. Classifying causes of death according to an aetiological axis, *Population Studies*, vol. 53, p. 97–105.

Meslé France and Vallin Jacques, 1984. The problem of studying mortality patterns by cause over a long period of time: an example from France, 1925 to 1978, *in:* Jacques Vallin, John Pollard, and Larry Heligman (eds.), *Methodologies for collection and analysis of mortality data*, p.449–492. Liège, IUSSP, Ordina Editions, 530 p.

Mobley Cyntia C., Boerma J. Ties, Titus Stephen, Lohrke Britte, Shangula K., and Blake Robert E., 1996. Validation study of a verbal autopsy method for causes of childhood mortality in Namibia, *Journal of Tropical Pediatrics*, vol. 42(6), p. 365–369.

Msp, 1990. *Étude nationale sur les causes and les circonstances de la mortalité infanto-juvénile (ECCD), 1988–1989*. Rabat, Ministry for Public Health, Morocco, 88 p.

Okongo Martin, Morgan Dilys, Mayanja Billy, Ross Amanda, and Whithworth James, 1998. Causes of death in a rural population-based HIV-1 natural history cohort in Uganda, *International Journal of Epidemiology*, vol. 27(4), p. 698–702.

Omondi-Odhiambo A.M., Voorhoeve A.M., and Van Ginneken Jeroen K., 1984. Age-specific infant and childhood mortality and causes of death, *in:* Jeroen van Ginneken and Alexandre S. Muller (eds.), *Maternal and child health in rural Kenya*. *An epidemiological study*, p. 213–222. Beckenham, Croom Helm, 373 p.

Pacque-Margolis Sara, Pacque Michel, Dukuly Zwannah, Boateng John, and Taylor Hugh R., 1990. Application of the verbal autopsy during a clinical trial, *Social Science and Medicine*, vol. 31(5), p. 585–591.

Quigley Maria A., Armstrong Johanna, and Snow Robert, 1996. Algorithms for verbal autopsies: a validation study in Kenyan children, *Bulletin of the World Health Organization*, vol. 74(2), p. 147–154.

Ronsmans Carine, Vanneste Anne-Marie, Chakraborty Jyotsnamoy, and Van Ginneken Jeroe, 1998. A comparison of three verbal autopsy methods to ascertain levels and causes of maternal deaths in Matlab, Bangladesh, *International Journal of Epidemiology*, vol. 27(4), p. 660–666.

Srivastava S.P., Kumar Anjani, and Ojha Arvind Kumar, 2003. Verbal autopsies determined causes of neonatal deaths. *Indian Pediatrics*, vol. 38, p. 1022–1025.

Snow Robert W., Armstrong Johanna R.M., Forster Dayo, Winstaley M.T., Marsh V.M., Newton C.R., Waruiru C., Mwangi I., Winstanley P.A., and Marsh R., 1992. Childhood deaths in Africa: uses and limitations of verbal autopsies, *Lancet*, vol. 340(8815), p. 351–355.

Todd J.E., De Francisco Andres, O'dempsey T.J.D., and Greenwood Brian M., 1994. The limitations of verbal autopsy in a malaria-endemic region, *Annals of Tropical Paediatrics*, vol. 14(1), p. 31–36.

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978*. Paris, INED, PUF, 608 p. (Travaux et Documents, Cahier no. 115.)

Vallin Jacques and Nizard Alfred, 1978. Les causes de décès en France. I. Pour une typologie simple et homogène: application à la période 1968–1974, *Population*, vol. 33(3), p. 547–608.

Victora Cesar G., Huttly R.A. Sharon, Fuchs S.C., Barros F.C., Garenne Michel, Leroy Odile, Fontaine Olivier, Beau J.B., Fauveau Vincent, Chowdhury Hafizur R., Yunus Mohamed, Chakraborty Jyotsnamoy, Sarder AbdulMajid, Kapoor S.K., Bhan M.K., Math L.B., and Martines J.C., 1993. International differences in clinical patterns of diarrheal deaths: a comparison of children from Brazil, Senegal, Bangladesh and India, *Journal of Diarrhoeal Diseases Research*, vol. 11(1), p. 25–29.

WHO, 1977. *Manuel de la Classification statistique internationale des maladies, traumatismes and causes de décès*, vol. 1. Geneva, World Health Organization, 782 p.

WHO, 1978. *Lay reporting of health information*. Geneva, World Health Organization, 27 p.

WHO, 1987. *Studying maternal mortality in developing countries: rates and causes. A guidebook*. Geneva, World Health Organization, 63 p.

WHO, 1991. *Directives pour le développement de protocoles pour les études destinées à évaluer l'effet des moustiquaires imprégnées d'insecticide sur la mortalité*. Geneva, World Health Organization, 165 p.

WHO, 1994. Measurement of overall and cause-specific mortality in infants and children: memorandum from a WHO/UNICEF meeting, *Bulletin of the World Health Organization*, vol. 72(5), p. 707–713.

WHO, 1995. *Verbal autopsies for maternal deaths*. Geneva, World Health Organization, 213 p. (Report of a WHO Workshop, London School of Hygiene and Tropical Medicine, 10–13 January, 1994.)

WHO, 2003. *Beyond the numbers: reviewing maternal deaths and complications to make pregnancy safer*. Geneva, World Health Organization, 141 p., http://www.who.int/reproductive-health/publications/btn/btn.pdf

Wyon John B. and Gordon John E., 1971. *The Khanna study, population problems in the rural Pendjab*. Cambridge, MA, Harvard University Press, 437 p.

Zimicki Susan, 1988. L'enregistrement des causes de décès par des non-médecins: deux expériences au Bangladesh, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 123–142. Paris, INED, PUF, 458 p. (Travaux and Documents, Cahier no. 119.)

Zimicki Susan, 1990. Approaches to assessment of the cause structure of mortality: a case-study from Bangladesh, *in:* Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analyses of mortality: new approaches*, p. 99–122. Oxford, Clarendon Press, 416 p. (International Studies in Demography.)

Zimicki Susan, Nahar Luftun, Sarder AbdulMajid, and D'souza Stan, 1985. *Demographic surveillance system—Matlab*, vol. 13. *Cause of death reporting in Matlab: source book of cause-specific mortality rates 1975–81*. Dacca, ICDDRB, 103 p. (Scientific report no. 63.)

C H A P T E R

44

Dependence and Independence of Causes of Death

GUILLAUME WUNSCH
Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

The problem of competing risks and the assumptions of *independence* and of *continuity* are discussed in Chapter 9. In this chapter, much of the same ground is covered in relation to the possible dependencies between causes of death or, more specifically, between the risks of dying of these causes. Consider a group of persons with ischemic heart disease followed over time to determine age at death or, more generally, their *failure-time*. In this situation, other causes of death compete with heart disease in addition to loss attributed to follow-up because of, for instance, out-migration. All these events censor the observation, but one cause of censoring—heart disease—is itself the topic of the study. The question is not in this case what would have been the result if there had been no censoring effects but what would have been the probabilities of dying of heart disease if there had been no competing events and, in particular, no other causes of death. The question is especially acute when computing probabilities by cause of death. Because everybody dies from one or another cause of death, some deaths (e.g., heart disease) have not been observed because persons ill from this disease have died of another disease (e.g., stroke) or an accidental cause (e.g., car crash).

I. PROBLEMS WITH THE CLASSIC APPROACH TO PROBABILITIES OF DYING

The probability of dying from one cause (e.g., ischemic heart disease) is usually estimated demographically by using the formulas of Elveback (1958) or Berkson and Gage (1950) (see Chapters 9 and 42), assuming independence between causes of death. This approach to evaluating probabilities of dying by cause of death is not without problems. First, because most deaths in countries with low mortality are due to chronic diseases, with often a long lag of time between the onset of a disease (e.g., heart disease, cancer) and death, it is often incorrect to assume, for example, that if someone had not died of heart disease between ages x and x + 5, he or she could have died of cancer. This assumption becomes less valid as the age interval decreases, because the chance of getting ill and dying of the competing disease decreases with the length of the period. Because a person has to be ill with cancer to die of the disease, either the person could not possibly die of cancer during the age interval considered if she or he did not have cancer or (s)he already suffered from both diseases. The single decrement approach would not be correct in the first situation, and the methods used commonly in this approach (e.g., Berkson or Elveback formulas for estimating the net probability) would therefore not be plausible. Multistate morbidity-mortality tables are required instead (see Chapter 47).

II. INDEPENDENCE OR DEPENDENCE AMONG CAUSES OF DEATH

Consider the assumption of *independence*. Is the risk of death from ischemic heart disease independent from, for example, the risk of death from cancer? Smoking has a negative impact on the incidence and lethality of lung and breast cancers and heart disease. For smokers, the probabilities of dying of heart disease or cancer are therefore not independent, because these effects have a common cause (i.e., smoking). Probabilities of dying from ischemic heart disease should therefore be computed among the nonsmokers, because in this case cancer and heart diseases cannot be associated through smoking behavior, although other known or unknown common causes can still play havoc with the results. The common causes of death should be controlled for when computing probabilities of dying by medical cause of death. The same can be said for the probabilities of becoming ill. For example, it is necessary to distinguish breast cancer incidences among female active smokers, passive smokers, and those not exposed to smoking. In this case, for the population of Geneva between 1992 and 1993 (Morabia, 1996), the following incidence rates were obtained for 100,000 women: 276 for active smokers, 252 for passive smokers, and 114 for the nonexposed subjects. These differences point to a possible impact of smoking (active or passive) on the incidence of breast cancer.

In the previous example, smoking behavior was a common cause of cancer and heart disease. Other common causes inducing dependencies between medical causes of death can be of a genetic nature. It has been shown, for example, that variations in the *APOE* gene are associated with variations in the risks for ischemic heart disease and Alzheimer's disease (Ewbank, 1997). These medical causes of death or disease processes are not independent in a causal sense, because they have a common genetic origin. When speaking of causes, it must be understood that what is meant is not a *sufficient* cause. There are few sufficient causes in the social sciences. A weaker causal relationship is assumed in our view of a causal structure, "such that the truth of the antecedent statement would at least give some support to the truth of the consequent" (Karhausen, 1996) or, in other words, that the occurrence of the cause increases the probability of the effect.

What *types* of competing events can be distinguished? As discussed in Chapter 8, an event such as migration competes with death for the censoring of the individual life line in the geographic region concerned. In a closed population, migration would not be problematic.[1] In an open population, migration perturbs the observation of mortality. Risks of dying can then be corrected for the occurrence of migration, using the Berkson method, for example, and assuming independence between mortality and migration in the absence of better information (Leridon and Toulemon, 1997). In the case of death, diseases may compete among themselves as the cause. An interesting discussion in the case of cancer is provided by Estève and colleagues (1993). According to these authors, probabilities of dying of cancer and those for other causes could be positively correlated, because being ill from cancer or being treated for cancer can increase the probability of dying of other causes. The opposite situation also can happen according to Estève and colleagues (1993), because cancer patients are medically monitored and adopt less risky lifestyles. In this case, probabilities of dying of other causes may be lower than in the general population.

What causal structures can be assumed among causes of death? Suppose an individual suffers from two lethal disease processes or causes of death, X and Y. The occurrence of death due to X prevents the occurrence of death due to Y, and vice versa, but the occurrence of one or the other would lead to life's end. X or Y are, in this case, *disjunctive causes* of death, although there can be *interaction effects* between X and Y. In this latter case, if death due to X occurs before death due to Y, death would not be due solely to X but also to the XY interaction effect between both disease processes. The same holds for Y happening before X. As an example of *interaction*, smoking and carrying the *APOE4* allele are separate risk factors for Alzheimer's disease, but it seems that smoking has a beneficial impact for those who have the *APOE4* allele and smoke (Breteler *et al*., 1999).

In another situation, both causes, K and L, lead to death in *conjunction*. Death occurs because of both K and L (e.g., acquired immunodeficiency syndrome and tuberculosis), and attributing death solely to K or to L would not be relevant.

In still another case, it is possible to have a situation in which M causes N. If someone who has terminal cancer (M) commits suicide by poisoning (N), cancer should be considered the true cause of death, and cancer would then be the initial or *underlying cause* of death and suicide the *immediate cause*, because taking the poison was intermediate in this case between cancer and death.

The single decrement approach is not a very satisfactory model in most of these circumstances, and demographers should make more use, in a causal perspective, of the medical information available on the lethal disease processes they are studying, taking into account among others the underlying, immediate, and associated causes of death as stated on the death certificate (Manton and Stallard, 1984). They should also develop multistate morbidity-mortality models or stochastic compartment models if the necessary data are available (see Chapter 47). For example, Manton and Stallard (1984) compute multiple decrement life tables for chronic diseases based on the underlying cause and on all medical conditions mentioned on the death certificate (i.e., the underlying, intermediate, immediate,

[1] Some countries have been closed in the past for political or geographic reasons. A small population size can lead to inbreeding, and genetic consequences can include health problems and death (see Chapter 79).

and associated causes of death); the definitions are provided in Chapter 42. They obtain the so-called pattern-of-failure life tables better suited to multiple pathologies than the classic tables by underlying cause alone, because the former take into account patterns of dependent conditions and the latter does not. According to these life tables, Table 44–1 compares for the United States (1969) the proportion of white males eventually dying of various chronic diseases (i.e., cancers, ischemic heart disease, stroke, diabetes mellitus, and generalized arteriosclerosis) with the proportions derived from the underlying cause life tables.

Pattern-of-failure life tables lead to higher proportions of deaths than the underlying cause life tables, because the former are based on the additional mentions of chronic disease in the multiple-cause data, whereas the latter take only the underlying cause into consideration. The difference between the two approaches is small for cancers but large for generalized arteriosclerosis.

Mortality by cause can also be studied using statistical methods, such as event-history analysis (see Chapter 23), based on individual life courses. The advantage of this approach is to take account of population heterogeneity and of the past and present individual risk factors. However, these methods require longitudinal individual data that are rarely available in the field of mortality for the whole population. Although statistical methods based on individual data are used in the field of fertility research (see Chapter 23) because of the wide availability of retrospective surveys, they are still not of common usage in mortality studies requiring costly and cumbersome prospective longitudinal surveys (Alderson, 1983), or individual data linkage between census results and vital registration data (Wunsch *et al*., 1996). As Courgeau and Lelièvre (see Chapter 23) rightly point out, although these methods are used in the field of clinical tests and epidemiology, they have not had great success in demography due to the nature of the data required.

TABLE 44–1 Proportions of White Males Dying from a Specific Chronic Disease According to the Underlying Cause and Pattern-Of-Failure Life Tables (United States, 1969)

Chronic disease	Underlying cause life table (%)	Pattern-of-failure life table (%)
Cancer	16.49	19.30
Ischemic heart disease	39.25	45.71
Stroke	9.54	15.12
Diabetes mellitus	1.47	5.74
Generalized arteriosclerosis	1.73	15.39

Manton and Stallard, 1984.

As an example, Zhu an colleagues (2000) used a Cox *proportional-hazard model* (Collett, 1997) to examine the possible relation between cognitive impairment and mild dementia or stroke, controlling for various confounders such as age, sex, education, and systolic blood pressure. The data were obtained from a population-based, prospective, longitudinal survey conducted in Kungsholmen (Stockholm); subjects were 75 years old or older at the baseline date (October 1, 1987), and the data were linked to the Stockholm inpatient hospital register. Results of this research demonstrate a significant positive relation between cognitive diseases and stroke, probably because of a common cause such as small cerebral lesions not detected at the clinical stage. The study covered 1551 individuals followed during a 3-year period. Because of the high cost involved, such a study could not be carried out on the whole population of Sweden. Demographers can nevertheless take advantage of this approach in the mortality field if they are looking for an answer to a specific question that does not require following a large sample of the whole population of the country considered.

Continuing reflection on the possible relations between causes of death, it was stated in Chapter 8 that a condition of *continuity* was required when evaluating fertility rates or frequencies in the absence of competing risks. Continuity supposes no causal relation between competing processes, such as fertility not changing the probability of dying. The problem is more complex in the field of mortality. Suppose there are only two causes of death or lethal disease processes, A and B, and no common causes. The condition of independence due to the absence of common causes is therefore satisfied. A person may be affected with A, with B, or with A and B, or the person may be in good health during the time interval considered. Being ill with A, however, may change the lethality of process B or vice versa, as in the cancer case given previously. Suppose the analyst is interested in mortality by cause A; cause B is therefore the disturbance or competing risk. In this case, for individuals suffering from both A and B, suppose that having contracted illness B modifies the probability of survival due to A (i.e., introduces a *discontinuity*). For example, the presence of syphilis modifies the risk of developing arterial aneurism. In this situation, both diseases are dependent, not because of a common cause but because of a causal relationship between them.[2]

[2] From a causal viewpoint, dependence between two (or more) risks can result from common causes or from a causal relation between the risks themselves. It can be shown that *statistical independence* of the cause-specific ages at death is a sufficient condition for *risk independence* (Manton and Stallard, 1984, chap. 4).

Probabilities or rates should take account of this factor and compute one probability of dying due to A conditional on the fact that the person is ill with A alone and compute a second probability of dying of A conditional on the fact that the person is ill with both A and B, with the presence of B possibly modifying the lethality of A. The population in good health or suffering only from B cannot die of A during the time interval considered because they are not ill with A. We may conclude that demographic analysis requires a thorough understanding of the processes involved and that, in this case, methods such as causal analysis (see Chapter 130), multistate increment-decrement models (see Chapter 47), and event-history modeling (see Chapter 23), would be better suited to the study of complex phenomena such as mortality by cause of death than the single decrement approach.

CONCLUSIONS

Because populations are *heterogeneous*, it must be remembered (see Chapter 10) that individuals with the higher risks of dying are on average excluded earlier from the population than those with lower risks. The *frailty* of the population therefore changes over time due to this selection effect in favor of those in better health. Without appropriate information, it is therefore not possible to distinguish the situation in which the age-specific heterogeneous probabilities of dying are constant for each individual from the situation in which probabilities decrease by age for a homogeneous group of individuals, because both cases lead to declining average probabilities over time. In the field of mortality by cause, this means that those who become ill with the most deadly diseases die first, leaving behind those with less severe conditions. This selection effect can explain, for example, the possible crossover in age-specific death rates between groups of different mortality levels and the fact that those who survive until old age can still be in quite good health on average.

References

Alderson Michael, 1983. *An introduction to epidemiology*, chap. 6. London, Macmillan, 335 p.

Berkson Joseph and Gage Robert, 1950. Calculation of survival rates for cancer, *in*: *Proceedings of the Staff meeting of Mayo Clinic, Rochester*, 11 May, p. 270–286. Rochester (Minnesota), Mayo Clinic, 720 p.

Breteler Monique M.B., Bots Michiel L., Ott Alewijn, and Hofman Albert, 1999. *Vascular disease and vascular risk factors and dementia*, *in*: Richard Mayeux and Yves Christen (eds.), *Epidemiology of Alzheimer's disease: from gene to prevention*. Berlin, Springer, 136 p.

Collett Dave, 1997. *Modelling survival data in medical research*. London, Chapman & Hall, 347 p.

Elveback Louis, 1958. Estimation of survivorship in chronic diseases: the "actuarial" method, *Journal of the American Statistical Association*, vol. 53, p. 420–440.

Estève Jacques, Benhammou Ellen, and Raymond Luc, 1993. *Méthodes statistiques en épidémiologie descriptive*. Paris, INSERM, 307 p.

Ewbank Douglas, 1997. The genetic make-up of population and its implications for mortality by cause of death: links between Alzheimer's and ischemic heart disease, *in*: *Symposium on health and mortality*, New York, Population Division, United Nations, in collaboration with Population and Family Study Centre, p. 219–236.

Karhausen Lucien R., 1996. The logic of causation in epidemiology. *Scandinavian Journal of Social Medicine*, vol. 24(1), p. 8–13.

Leridon Henri and Toulemon Laurent, 1997. *Démographie*. Paris, Economica, 442 p.

Manton Kenneth G. and Stallard Eric, 1984. *Recent trends in mortality analysis*. Orlando, Academic Press, 342 p.

Morabia Alfredo, 1996. *Epidémiologie causale*. Geneva, Editions Médecine et Hygiène, 174 p.

Wunsch Guillaume, Duchêne Josianne, Thiltgès Evelyne, and Salhi Mohammed, 1996. Socio-economic differences in mortality: a life course approach, *European Journal of Population*, vol. 12(2), p. 167–185.

Zhu Li, Fratiglioni Laura, Guo Zhigang, Winbald Bengt, and Vitanen Matti, 2000. Incidence of stroke in relation to cognitive function and dementia in the Kungsholmen Project, *Neurology*, vol. 54(1), p. 2103–2107.

C H A P T E R

45

Measurement and Levels of Maternal Mortality

JUDITH A. FORTNEY

Family Health International, Research Triangle Park, North Carolina, United States

Only 50 years ago, maternal mortality in Europe was as high as it is in parts of the developing world today, and it had a significant demographic and dynastic impact. Maternal mortality has changed the course of history and has resulted in untold suffering for the mothers themselves and their bereaved husbands and children. Princess Charlotte's death in 1817, after 48 hours of mismanaged labor (Dewhurst, 1980), began the chain of events that led to Queen Victoria's place on the British throne and her amazing impact on European history (Potts and Potts, 1995).

In a less spectacular way, the relentless loss—one per minute—of women in childbirth affects families throughout the developing world. Today, 99% of maternal deaths occur in developing countries, and the disparity between the developed and developing world is the greatest of any public health indicator. Estimates put the loss (in 1990) at 585,000 deaths per year (WHO and UNICEF, 1996). Apart from these maternal deaths, only 25% of their babies born alive survived their first year (Koenig *et al*., 1988). Although data are scarce, older children (especially girls) whose mothers die often terminate their education to care for the younger siblings, and many street children are motherless (Piraino, 1991).

After discussing current definitions of maternal mortality, this chapter describes measurement issues, means of estimating maternal mortality, and levels of maternal mortality in the world today and in the past.

I. DEFINITIONS

The definition of a maternal death is surprisingly contentious. Although the World Health Organization's (WHO) *International Classification of Disease*, 10th revision (ICD-10), definition of maternal mortality (Table 45–1), and acknowledgment of "pregnancy-related death" is an improvement over the ICD-9 version, some researchers think it does not go far enough. There are two contentious issues. Maternal mortality, like infant mortality, is partly defined by time. Infant mortality is universally accepted to be deaths of infants younger than 1 year of age from any cause. In the case of maternal mortality, however, both the period and the significance of cause are disputed.

Traditionally, a death is defined as maternal when it occurs during pregnancy or within 42 days of termination of pregnancy, although there are variations on this period. In the ICD-10, WHO acknowledges the need for an extended period and refers to "late maternal deaths," which occur after 42 days and up to 1 year after delivery. The call for extending the relevant period comes mainly from the developed countries where modern medicine can delay death beyond the traditional 42-day period. This is not the case in developing countries. Research in Bangladesh, using a 90-day definition, found that only 5% of maternal deaths in that country occurred 43 to 90 days after delivery (Koenig *et al*., 1988). Excluding the deaths after 42 days

TABLE 45–1 World Health Organization's ICD-10 Definitions of Maternal Mortality

Maternal death: The death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes.

Pregnancy-related death: The death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the cause of death.

ICD-10, International Classification of Disease, 10th revision.

TABLE 45–2 Variations in the Period for Defining Maternal Mortality

Organization	Period
WHO, ICD-10	42 days
FIGO	42 days (follows ICD-10)
CDC	12 months
ACOG	Indefinite if cause is maternal
RCOG	42 days (follows ICD-10)

ACOG, American College of Obstetrics and Gynecology; CDC, U.S. Centers for Disease Control and Prevention; FIGO, Fédération Internationale de Gynaecologie et Obstétriques; ICD-10, International Classification of Disease, 10th revision; RCOG, Royal College of Obstetrics and Gynaecology (UK).

reduced the maternal mortality ratio from 550 to 520 per 100,000 live births.

Researchers must take care when making comparisons that the data are comparable with regard to the period used in the definition, as well as other variations. Occasionally researchers use unconventional periods; for example, Bardet and colleagues (1981) use 60 days. Even international organizations may use different periods (Table 45–2). This is unsatisfactory because it discourages comparability.

There is another problem in comparative research in addition to the period used for calculations. Traditionally, maternal deaths have been divided into direct obstetric, indirect obstetric, and accidental, and incidental types (i.e., apparently not caused by the pregnancy or delivery, such as accidents, homicides, suicides), with the last category excluded from the maternal mortality ratio statistic. Because it is sometimes difficult to distinguish between the last two categories in classifying a death, the maternal mortality ratio may be underestimated because of misclassifications into the accidental or incidental category.

There is growing recognition that what the ICD-10 definition refers to as "accidental or incidental" causes may not be incidental to the pregnancy. Vincent Fauveau and Thérèse Blanchet (1989) working in Matlab, Bangladesh, report that 20% of deaths of pregnant unmarried women were caused by suicide compared with 5% for married women. Kleiner and Greston (1984) and Frautschi and colleagues (1994), in reviewing the literature, found that suicide precipitated by pregnancy is not an uncommon occurrence. Several investigators (Gazmararian *et al*., 1995) report that domestic violence in the Unites States, regardless of whether it results in death, may be triggered by pregnancy, especially if the pregnancy was unintended. Some pregnant, unmarried women are killed by their families because of the pregnancy.

Some accidents may be pregnancy related. Fildes and his colleagues (1992) refer to trauma as the leading cause of maternal death in the United States. In the state of Maryland, approximately 10% of pregnant women are hospitalized for accidental injury during pregnancy (Greenblatt and Dannenberg, unpublished observations). Reviewing the literature on accidental injury during pregnancy, Sherer and Schenker (1989) ask, "Is the pregnant woman accident prone? Ever changing physical dimensions, increasing abdominal girth, continuous weight gain combined with physiologically compensatory lordosis create a change in the pregnant woman's center of gravity. Affected are both her maneuverability as well as her ability to react to approaching dangers. Neufeld *et al*. state that as pregnancy progresses, the incidence of minor trauma increases and by the third trimester will occur more frequently than at any other time during female adulthood." This question is still debated. Some studies show the opposite relationship, with pregnant women being less accident prone than others (Ronsmans *et al*., 2001).

Pregnancy affects immune status and therefore susceptibility to infections. Several infectious diseases are more severe during pregnancy and are more likely to cause death than in nonpregnant women. The most important of these (in terms of numbers) are malaria, tuberculosis, hepatitis, and measles. A study in Kenya reports that there is an 11-fold decrease in recovery from malaria infection among pregnant women (Brabin, 1983). Kwast and her colleagues (1986) report that 15.5% of maternal deaths in Addis Ababa, Ethiopia were caused by hepatitis. For measles, there is an excess female mortality, which peaks during the reproductive years of 15 to 44 and lasts until age 50 years. In Senegal, for example, excess mortality during the reproductive years is 43% (Garenne, 1994).

Breast cancer has a worse prognosis when diagnosed during pregnancy because it is usually in a more advanced state at the time of diagnosis. Whether this is because the hormonal environment causes it to grow more rapidly during pregnancy or because diagnosis is delayed because of other changes in the breast due to the pregnancy is unknown (Petrek, 1996). Regardless of the cause, pregnancy contributes a higher level of mortality from breast cancer.

For the reasons previously cited, many of the leading researchers in maternal mortality prefer a definition that includes *all* deaths of women who are pregnant or within 1 year of termination of pregnancy, regardless of the cause of death. Although this definition will doubtless include deaths that are not attributable to pregnancy or delivery, the resulting overcount is less problematic than the previous undercount, and it has the advantages of being conceptually simple, having a precedent in infant mortality, and being far simpler to implement than the present ICD definition.

II. MEASURES OF MATERNAL MORTALITY

Several indicators of maternal mortality are used in the literature, and investigators should be aware of their differences in comparative studies.

1. Maternal Mortality Ratio

The most frequently used measure of maternal mortality is the following ratio:

$$(\text{number of maternal deaths} \,/\, \text{number of live births}) \times 100{,}000$$

Although 10,000 or 1000 live births is sometimes used, 100,000 is increasingly the preferred denominator, especially because ratios in the developed world drop below 10 per 100,000 live births. For example, 8 maternal deaths per 100,000 live births is more easily interpreted by nontechnicians than 0.8 per 10,000 or 0.08 per 1000. Use of the 1000 live births denominator is largely a holdover from times and places with much higher maternal mortality rates. Although developing country researchers often (for good reasons) still use the 1000 live births denominator, the 100,000 denominator is recommended in the interests of comparability.

Why use live births as denominator? It has been suggested (Thuriaux and Lamothe, 1984) that the number of all pregnancies is a more correct denominator. This suggestion has not received serious attention because of the impossibility of implementing it. Few countries are able to count or even recognize all pregnancies.

Any period can be used as a reference as long as it is identical for both numerator and denominator. Periods longer than 1 year are often desirable because larger numbers of deaths make for more stable estimates.

The term *maternal mortality ratio* began to be used in the mid-1980s. Before that (and still by some writers), it had been called the *maternal mortality rate*. It is not, however, a true rate because the numerator is not a subset of the denominator. The misuse of the word *rate* can be confusing to the beginning investigator of maternal mortality.

2. Maternal Mortality Rate

The *maternal mortality rate* for a defined period is calculated as follows:

$$(\text{number of maternal deaths} \,/\, \text{number of women aged 15 to 45 years}) \times 100{,}000$$

In this case, the numerator (i.e., maternal deaths) is a subset of the denominator (i.e., women between 15 and 45 years old), and it is therefore a true rate.

Despite its usefulness and conceptual importance, the maternal mortality rate is not widely used. Although the ratio measures the risk to a woman *once she becomes pregnant*, the rate measures the risk that a woman *will become pregnant* and *that she will die after she has become pregnant*. It is strongly affected by the level of fertility in the region of interest. The ratio is influenced by the quality and availability of health services, but the rate is influenced by the quality and availability of family planning as well as health services.

3. Lifetime Risk

The lifetime risk measures the risk that a woman entering adulthood will die at any point from maternal causes. The more frequently she becomes pregnant, the more frequently she is exposed to the risk. The lifetime risk is written as "1 in N" risk of experiencing a maternal death, and a smaller N represents a greater risk. It is actually 1 minus the probability of surviving all pregnancies.

For example, using age and parity-specific mortality data from Matlab in Bangladesh (Chen *et al.*, 1974) and assuming a TFR of 6, with one birth before age 20, two births at age 20 to 29, two births at age 30 to 39, and one birth at age 40 to 49, the following probabilities of survival by age and parity are obtained (Table 45–3) multiplying the survival probabilities:

$$0.9912 \times 0.9964 \times 0.9964 \times 0.9975 \times 0.9975 \times 0.9920 = 0.9713$$

This is the probability of surviving six births at those ages in that place and time. The probability of

TABLE 45–3 Probabilities of Female Survival by Age and Parity: Statistics from Matlab, Bangladesh, in the Early 1970s

Age	Probability of surviving (1 − probability of dying) at each specified age and parity			
	1	2–3	4–5	6+
10–19	0.9912	—	—	—
20–29	—	0.9964	—	—
30–39	—	—	0.9975	—
40–49	—	—	—	0.9920

Adapted from Chen *et al.*, 1974.

maternal death given that reproductive pattern is calculated as follows:

$$1 - 0.9713 = 0.0287$$

This is equal to about 1 in 35 (1/0.0287 = 34.8). In practice, however, the calculation can be made without access to such detailed data (i.e., parity-specific risks are rarely available). The maternal mortality ratio found in the study shown above was 570 per 100,000 live births or 0.0057 for each individual birth, or $0.0057 \times 6 = 0.0342$ (i.e., about 1 in 29, a close approximation to 1 in 35).

4. Proportional Mortality

Proportional mortality refers to the proportion of all deaths of women of reproductive age that are maternal deaths (i.e., that occur during pregnancy or within 42 days [or 12 months] of the termination of pregnancy). In developing countries, it is typically 20% to 25% and the leading cause of death in this age group, whereas in developed countries, it is usually less than 1%.

III. MEANS OF ESTIMATING MATERNAL MORTALITY

Adult mortality generally is not well measured in the developing world. Maternal mortality, a subset of adult mortality, is also not measured well.

There are several ways of estimating maternal mortality and, for a variety of reasons, none is entirely satisfactory (WHO, 1987). Even where maternal mortality is very high, deaths are relatively rare in any one community. For example, with a maternal mortality ratio of 500 per 100,000 live births, a crude birth rate of 30 per 1000 population, and a community population of 5000, the number of births per year is 150, and the number of maternal deaths per year is 0.75, or an average of three deaths every 4 years. Surveys designed to identify maternal deaths must cover a very large population base or extend over many years (or both) to identify enough deaths from which to estimate any measure with some degree of stability.

1. Vital Registration

Maternal deaths can be identified from several sources. In principle, vital registration is the ideal source of information on maternal death. However, experience has shown that underestimation is usual (Dye *et al*., 1992; Bouvier-Colle *et al*., 1991; Benedetti *et al*., 1985; Speckhard *et al*., 1985). An extensive review of deaths of women of reproductive age in France (Bouvier-Colle *et al*., 1991) resulted in an increase in the maternal mortality ratio from 10 to 22 per 100,000 live births. In the United States, several states have added pregnancy status to the death certificate or link birth and death certificates. These efforts have increased the maternal mortality ratio in those states by 30% to 153% (Berg *et al*., 1996).

Few developing countries have systems of vital registration. Even where vital registration is of good quality, there is underestimation of maternal mortality, most often because the person certifying the cause of death fails to mention the pregnancy status. The risk of this occurring increases as the time after termination of pregnancy is extended. For example, a woman who experiences eclampsia (i.e., convulsion resulting from hypertension induced by pregnancy) develops kidney failure as a result of the eclampsia; she dies weeks (or months) later from kidney failure, and her death takes place on the medical rather than the obstetric ward. The physician who is treating her is concerned only with the kidney failure (because that is what requires treatment), not its underlying cause, and does not mention eclampsia or pregnancy on the death certificate.

In another example, a woman with obstructed labor (i.e., because the baby is in the wrong position, the baby is too big, or the pelvis is too small) has a cesarean section and dies because of an anesthetic accident. The physician who certifies the death reports it as an anesthetic or surgical accident without mention of cesarean section or pregnancy.

A physician certifying the death of a woman from postpartum hemorrhage may write hypovolemic shock on the death certificate. Research from developed and developing countries has repeatedly shown that the contribution of pregnancy to a death is omitted from death certificates. Few developing countries have functional systems of vital registration, and even where registration is virtually complete, the cause of death reporting is usually incomplete for maternal deaths. For example, in Egypt, the deaths of women who died of postpartum hemorrhage are often reported on their death certificates (death certification and recording of adult deaths is virtually complete in Egypt) as hypovolemic shock (Grubb *et al*., 1988). Although technically correct, postpartum hemorrhage is more informative and permits designation as maternal death, which the former designation does not. In most developed countries, unexplained deaths are usually investigated by autopsy, but in developing countries, autopsies are most often reserved for situations in which violence is known or thought to have occurred (even where the cause of death is obvious, such as a motor vehicle accident).

2. Hospital Data

Hospital data are commonly used in developing countries where other data are lacking. They are, however, virtually impossible to interpret when some births or other pregnancy terminations take place outside of hospitals. Deaths from spontaneous or induced abortions or ectopic pregnancies are most likely to occur outside of hospitals, and in most developing countries, many births occur at home. It can be safely assumed that hospital births are not representative of all births in this situation. However, are they more representative of complicated deliveries, births to wealthier women, or births to urban women? In any case, it is not possible to infer anything about the population in general. Although hospital death rates are indicative of conditions in hospitals and can be useful, they should not be referred to as maternal mortality ratios or rates.

3. Population-Based Surveys

Several population-based surveys are considered to be of reasonably high quality (Boerma and Mati, 1989; Bhatia, 1988; Fortney, 1986). In these surveys, the families of women who died are interviewed to determine the circumstances surrounding the death and to determine the cause of death. This has been called the *verbal autopsy* method (WHO and LSHTM, 1995) (see Chapter 43). This approach has yielded useful information about the circumstances surrounding maternal deaths and for verifying that the deaths were maternal. Deaths of young adults are relatively easier to diagnose because they are more likely (than deaths of children or older people, for example) to be caused by accidents or acute conditions, and the deceased is likely to have verbalized her symptoms. Nevertheless, for some proportion of deaths, it will be impossible to assign a cause.

Identifying the deaths being investigated can be done by several means: vital registration, reporting by community leaders, and reporting by multiple informants (Boerma and Mati, 1989; Measham *et al*., 1981). If all deaths of women of reproductive age are investigated, they are sometimes called *reproductive-age mortality surveys* (RAMOS), a term coined by James Shelton of United States Agency for International Development (USAID) in 1979.

One limitation of population-based surveys is that deaths are often unreported or misreported if they are an object of shame (e.g., after an illegal abortion, if the mother was not married); are unrecognized (e.g., ectopic pregnancy); are forgotten (e.g., when the mother has low status or has no surviving or nearby relatives); occur to women who live in remote areas; or involve illegal activity (e.g., abortions, homicides, suicides). Other limitations include the large population that needs to be covered (and therefore the high cost) and the need for sensitive and expert interviewing, although families are often willing to discuss the circumstances despite their grief.

4. Sisterhood Surveys

The *sisterhood method* (Graham *et al*., 1989) of estimating maternal mortality was developed because of the difficulty and expense in getting such data in other ways. It involves asking just four questions, which can be added relatively easily to an ongoing survey of an appropriate population, thereby reducing costs. Because of its relative ease of data collection it has been widely adopted and is routinely added to many surveys, including the Demographic and Health Surveys (DHS). The four questions cover the number of sisters the respondent has (had), the number of sisters who reached reproductive age, the number who are still alive, and the number who died a maternal death.

An important limitation of the method is that it cannot be used to measure short-term change because it estimates mortality for an extended period (10 to 12 years) preceding the survey. Work in modeling estimates of maternal mortality (Stanton *et al*., 1995) has suggested that the sisterhood method consistently underestimates, by about one-third, mortality due to maternal causes. This claim is supported by work in Bangladesh (Shahidullah, 1995a, 1995b) and Burkina Faso (Garenne and Friedberg, 1997; Garenne *et al*., 1997).

5. Modeling

The numerous difficulties in estimating maternal mortality and identifying maternal deaths have led to increased interest in indirect estimation through modeling. Modification of the 1990 global estimate is a result of modeling, and the numbers provided in the next section are based on this model. There has been only one published report of a model to estimate maternal mortality (Stanton and Hill, 1996). This model was based on indicators known to influence maternal mortality.

An earlier model by the same authors was abandoned because it was based on life tables. Although a model based on actual life tables would be acceptable, for most developing countries, actual life tables have not been calculated, and only model life tables exist. All countries using the same model life table would have the same estimated level of maternal mortality, regardless of variations in availability and quality of health services. Model life tables are strongly influenced by levels of infant mortality. This results in maternal mortality being estimated from infant mortality, and they are believed to be only moderately associated.

Garenne and colleagues (1997a, 1997b) used a simulation model with real data from Burkina Faso to evaluate the direct and indirect methods of sisterhood estimation. The use of models to estimate maternal mortality remains controversial. The WHO and UNICEF 1996 estimates were hotly contested by many countries whose estimates were increased. Nevertheless, other methods also have their limitations, and an increased role for models in the future is both likely and desirable.

IV. LEVELS OF MATERNAL MORTALITY

Table 45–4 shows levels of maternal mortality for selected countries. All these estimates are taken from the WHO and UNICEF (1996) estimates for 1990 based on a model. The disparities between the developed and the developing world are large, 300-fold when comparing the highest (Sierre Leone: 1800) with the lowest (Canada or Norway: 6; two small countries, Iceland and Luxembourg, have an estimated ratio of 0). The developed world as a whole has a maternal mortality ratio of 27, compared with 480 in the less developed regions, almost a 20-fold difference.

What do these numbers mean in terms of demographic impact? In Afghanistan, with a maternal mor-

TABLE 45–4 Point Estimates of Maternal Mortality Ratios for Selected Countries in Six Regions of the World, 1990

Europe (36)		North America (11)		Central and South America (140, 200)	
Bulgaria	27	Canada	6	Bolivia	650
Denmark	9	USA	12	Brazil	220
France	15	Mexico	110	Colombia	100
Germany	22			Costa Rica	55
Iceland	0			Guatemala	200
Italy	12			Haiti	1000
Norway	6			Peru	280
Sweden	7			Uruguay	85
UK	9			Venezuela	120
Africa (870)		**Asia (390[a])**		**Middle East (330)**	
Angola	1500	Australia	9	Afghanistan	1700
Burkina Faso	930	Bangladesh	850	Egypt	170
Cameroon	550	Cambodia	900	Iran	120
Côte d'Ivoire	810	Hong Kong	7	Iraq	310
Ethiopia	1400	India	570	Israel	7
Gambia	1100	Indonesia	650	Kuwait	29
Ghana	740	Japan	18	Libya	220
Kenya	650	Nepal	1500	Pakistan	340
Nigeria	1000	Sri Lanka	140	Saudi Arabia	130
South Africa	230	Thailand	200	Syria	180

Source: [a] Excluding Australia and New Zealand.
From Revised 1990 Estimates of Maternal Mortality: A New Approach by WHO and UNICEF. WHO FRH/MSM/96.11, Geneva, 1996.

TABLE 45–5 Maternal Mortality Ratios in Europe for the 16th to 19th Centuries

Period	European ruling families[a]	Seven Swedish parishes[b]	Thirteen English parishes[c]	France[d]
1500–1599	1940	—	—	—
1600–1649	—	—	1530	—
1650–1699	1949	—	1880	—
1700–1749	2020	—	1360	1140
1750–1799	—	—	880	970
1800–1849	1880	546	—	850
1850–1879	1470	690	—	—
1880–1889	—	537	—	—
1890–1900	—	595	427	—

[a] Data from Peller, 1965.
[b] Data from Högberg and Broströ, 1985.
[c] Data from Wrigley and Schofield, 1983; adjusted for nonmaternal causes and unobserved pregnancies.
[d] Data from Guitierrez and Houdaile, 1983. Actual periods are 1700–1749, 1750–1789, and 1790–1829.

tality ratio of 1700 and a total fertility rate (TFR) of 6.9, a woman has a 1 in 8 or 9 chance of dying a maternal death—a significant demographic impact. In Sri Lanka, where the maternal mortality ratio is 140 and the TFR is 2.6, a woman's likelihood of dying a maternal death is only 1 in 275—a relatively small demographic impact.

1. Decline in the Developed World

In historic times, maternal mortality in the developed world was at least as high as it is in the developing world today. Tables 45–5 and 45–6 show data from some European countries for which historical data are available. Table 45–5 presents maternal mortality ratios, and Table 45–6 shows what is assumed to be close to a maternal mortality rate. The existence of relatively complete records for England, Sweden, France, and for the aristocracy in several countries makes estimates possible (Högberg *et al*., 1986; Wrigley and Schofield, 1983; Peller, 1965).

Elsewhere in Europe, the ratio in Geneva was 1470 in the 1600s, 1511 in the 1700s, and 1333 in the 1800s (Perrenoud, 1981). In the United States, the ratio in the year 1900 was about 600, which was quite similar to that in Europe at about the same time (Högberg and Broström, 1985). Between 1975 and 1982 in the United States, a religious group that refused all medical care had a maternal mortality ratio of 874 (Kaunitz *et al*., 1986).

Högberg and Broström (1985) attribute the decline in the number of maternal deaths in the 1700s in Sweden to a decline in fertility rather than an improvement in obstetrics. This trend is partly evident also in

TABLE 45–6 Average Maternal Mortality per 100,000 Fertile Women from the Ruling Families of Europe Who Were Married in the Specified Years

Age	1500–1699	1700–1799	1800–1849	1850–1899
20–24	3100	5540	3400	2610
25–29	3450	2560	2470	1750
30–34	2450	1320	1250	1720
35–39	2740	1750	1330	1180
40–44	1770	—	—	—
45–49	300	390	—	—

Adapted from Peller, 1965.

Table 45–6, which shows rates (not ratios) from European ruling families.

Several events caused the decline of maternal mortality in the developed world beginning in the 19th century. In interpreting the data in this section, it is necessary to understand that most (roughly 80%) of maternal mortality can be attributed to the following five causes:

Infection
Hemorrhage
Eclampsia (i.e., convulsion resulting from hypertension induced by the pregnancy)
Obstructed labor (i.e., because the baby is in the wrong position, the baby is too big, or the pelvis is too small)
Abortion

The first major breakthrough in obstetrics came with discovery of the contagiousness of puerperal or childbed fever. This appears to have been discovered by Alexander Gordon in Scotland around 1790 and independently in the 1840s by Oliver Wendell Holmes in the United States and Ignaz Semmelweiss in Austria. The discovery was resisted, perhaps as late as the 1930s, by some obstetricians (Loudon, 1992), but the idea was adopted by many others. Högberg and his colleagues (1986) show that the incidence of puerperal sepsis in Sweden dropped from 5263 per 100,00 deliveries in 1881 to 220 in 1900. Obstetric forceps were used by some obstetricians from about 1740 on (Loudon, 1992), but they were a carefully guarded secret and available only to wealthy patients. Although they represented obstetric progress, they had little impact on maternal mortality in general. The development in the 1840s of anesthesia was not sufficient to make cesarean section widely available, but toward the end of the century, surgical techniques also improved, averting many deaths from obstructed labor (Loudon, 1992). Sulfonamides became available in the 1930s, as did penicillin in 1945, providing the means to treat puerperal sepsis. In Britain, more than 80% of the decline in total maternal mortality between 1934 and 1940 resulted from the decline in (non–abortion-related) sepsis (Loudon, 1992).

A rapid decline in maternal mortality occurred in the developed world between 1934 and 1950. In Britain, these 16 years saw a drop from 440 deaths per 100,000 live births to 87. Loudon attributes this dramatic decline to the following factors:

Introduction of sulfonamides and penicillin

A decline in the virulence of *Streptococcus* (which causes puerperal sepsis)

Introduction of ergometrine, blood transfusion on a large scale, and the growth of obstetric flying squads, all of which played a major role in reducing deaths from obstetric hemorrhage

Improvement in the standard of living and the nutrition of the population (reducing, for example, the prevalence of rickets which cause pelvic deformity)

Improvements in obstetric training

More than most issues of public health importance, the reduction in maternal mortality was a function of medical intervention rather than prevention; the important exception was the belated and reluctant introduction of infection control, mainly by handwashing (Fig. 45–1). With the exception of rickets, the role of nutrition is difficult to document and far from clear. Historical fluctuations in maternal mortality could have been attributed to variation in food availability or to variations in obstetric customs. Food availability is unlikely to have affected ruling families, however, and great fluctuations are also evident in this population.

The role of transport in reducing maternal mortality in the developed world has yet to be studied, but the improvement of roads and the development of mechanical transport undoubtedly improved

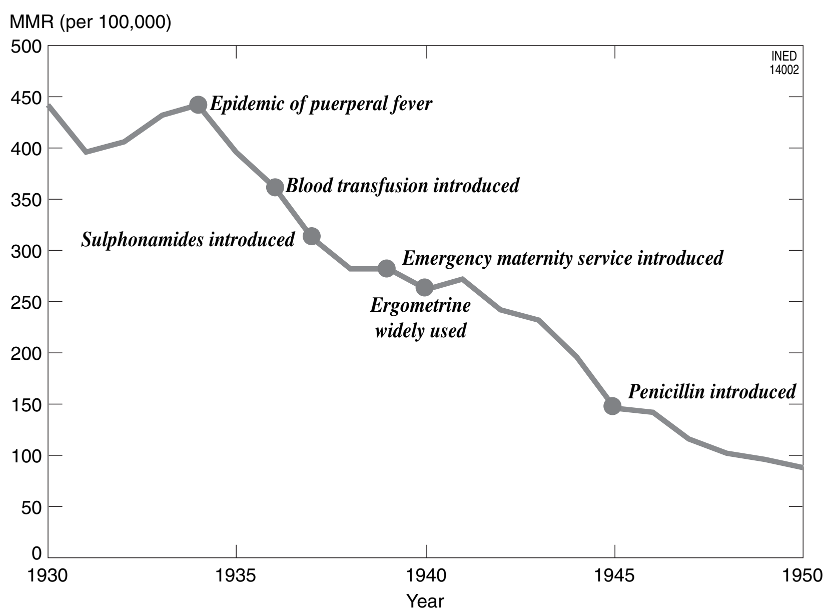


FIGURE 45–1 Maternal mortality ratios (MMRs) in England and Wales from 1930 to 1950 associated with advances in medicine.

women's access to medical care when it was necessary. The midwife on her bicycle and the doctor making house calls by car were common long before most women delivered in the hospital.

2. Decline in the Developing World

The situation of maternal mortality in the developing world is fundamentally different from that in the developed world in the early 19th century. In 1800, there were no antibiotics, no barbiturates, no anesthesia, and no blood transfusions. All of these are now technically feasible, not technically difficult, and not especially expensive, even in the developing world. Why then do women still die in childbirth?

The reasons are many and complex, and they go far beyond medical technology. In the developed world, medical technology became available as economies were burgeoning; charities often countered poverty; educational opportunities expanded; and transport improved. Although some of these changes are occurring in the developing world, it is often at a slower rate and with more frequent setbacks than during the periods of economic expansion in the West. Desertification, war, anticolonialism, government inefficiency, corruption, interrupted transport, structural adjustments, and the brain drain all contribute to reducing availability of health care. Where developing countries (e.g., Sri Lanka, Costa Rica) have made public health a national priority, maternal mortality is relatively low.

The last 15 years of the 20th century saw great strides in understanding the factors that contribute to high maternal mortality in developing countries. Although the developed world expresses its interest in helping the developing world address the issue and many developing countries have stated their desire to address it, neither has yet made the necessary political and financial commitment to achieve the goal. Although medical and technologic advances of the past century have given us the means to reduce persistent high levels of maternal mortality in the world's less developed regions, the medical infrastructure is not always in place, services are not always geographically and financially accessible, and the training, accountability, and commitment of providers are sometimes limited.

References

Bardet Jean-Pierre, Lynch Katherine, Mineau Géraldine, Hainsworth Malcom, and Skolnick Mark, 1981. La mortalité maternelle autrefois: une étude comparée (de la France de l'ouest à Utah). *Annales de démographie historique*, p. 31–48.

Benedetti Jacqueline, Starzyck Patricia, and Frost Floyd, 1985. Maternal deaths in Washington State. *Obstetrics and Gynecology*, vol. 66, p. 99–101.

Berg Cynthia J., Atrash Hani K., Koonin Lisa M., and Tucker Myra, 1996. Pregnancy-related mortality in the United States, 1987–1990. *Obstetrics and Gynecology*, vol. 88, p. 161–167.

Bhatia Jagdish C., 1988. *A study of maternal mortality in Anantapur District, Andhra Pradesh, India*. Bangalore, Indian Institute of Management, 24 p.

Boerma J. Ties and Mati Joseph K.G., 1989. Identifying maternal mortality through networking: results from coastal Kenya. *Studies in Family Planning*, vol. 20, p. 245–253.

Bouvier-Colle Marie-Hélène, Varnoux Noëlle, and Hatton Françoise, 1991. Reasons for the underreporting of maternal mortality in France, as indicated by a survey of all deaths among women of childbearing age. *International Journal of Epidemiology*, vol. 20, p. 717–721.

Brabin Bernard J, 1983. An analysis of malaria in pregnancy in Africa. *Bulletin of the World Health Organisation*, vol. 61, p. 1005–1016.

Chen Lincoln, Gesche Melita C., Ahmed Shamsa, Chowdhury Anwarullah I.A., and Mosley W. Henry, 1974. Maternal mortality in rural Bangladesh. *Studies in Family Planning*, vol. 5(11), p. 334–341.

Dewhurst John, 1980. *Royal confinements: a gynaecological history of Britain's royal family*. New York, St. Martin's Press, 205 p.

Dye Timothy, Gordon Howard, Held Berel, Tolliver Nancy J., and Holmes Alan P., 1992. Retrospective maternal mortality case ascertainment in West Virginia, 1985 to 1989. *American Journal of Obstetrics and Gynecology*, vol. 167(1), p. 72–76.

Fauveau Vincent, Blanchet Thérèse, 1989. Deaths from injuries and induced abortion among rural Bangladeshi women. *Social Science and Medicine*, vol. 29(9), p. 1121–1127.

Fildes John, Reed Laura, Jones Nancy, Martin Marcel, and Barrett John, 1992. Trauma: the leading cause of maternal death. *Journal of Trauma*, vol. 32(5), p. 643–645.

Fortney Judith A., 1987. The importance of family planning in reducing maternal mortality. *Studies in Family Planning*, vol. 18, p. 109–114.

Fortney Judith A., Susanti Inne, Saleh Sanneya, Gadalla Saad, Rogers Susan M., and Potts D. Malcolm, 1986. Reproductive mortality in two developing countries. *American Journal of Public Health*, vol. 76, p. 132–138.

Frautschi Schuyler, Cerulli Annamaria, and Maine Deborah, 1994. Suicide during pregnancy and its neglect as a component of maternal mortality. *International Journal of Gynaecology and Obstetrics*, vol. 47(3), p. 275–284.

Garenne Michel, 1994. Sex differences in measles mortality: a world review. *International Journal of Epidemiology*, vol. 23, p. 632–642.

Garenne Michel, Friedberg Patrice, 1997. Accuracy of indirect estimates of maternal mortality. *Studies in Family Planning*, vol. 28, p. 132–142.

Garenne Michel, Sauerborn Rainer, Nougtara Adrien, Borchert Matthias, Benzler Justus, and Diesfel Jochen, 1997. Estimating maternal mortality in Burkina Faso. *Studies in Family Planning*, vol. 28, p. 54–61.

Gazmararian Julie A., Adams Melissa M., Saltzman Linda E., Johnson Christopher H., Bruce F. Carol, Marks James S., Zahniser Christine, and Prams Working Group, 1995. The relationship between pregnancy intendedness and physical violence in mothers of newborns. *Obstetrics and Gynecology*, vol. 85, p. 1031–1038. (Synthesized in *Family Planning Perspectives*, vol. 27, p. 222–223.)

Graham Wendy, Brass William, and Snow Robert W., 1989. Estimating maternal mortality: the sisterhood method. *Studies in Family Planning*, vol. 20, p. 125–135.

Greenblatt Jesse F., Dannenberg Andrew L., and Johnson Christopher J., 1997. Incidence of hospitalizations for injuries among pregnant women in Maryland, 1979–1990. *American Journal of Preventive Medicine*, 13(5), 374–379.

Grubb Gary S., Fortney Judith A., Saleh Sanneya, Gadalla Saad, El-Baz Ahmad, Feldblum Paul J., and Rogers Susan M., 1988. A comparison of two cause-of-death classification systems for deaths among women of reproductive age in Menoufia, Egypt. *International Journal of Epidemiology*, vol. 17, p. 201–207.

Guitierrez Hector, Houdaille Jacques, 1983. La mortalité maternelle en France au XVIII^{e} siècle. *Population*, vol. 38(6), p. 975–993.

Högberg Ulf, Broström Göran, 1985. The demography of maternal mortality Seven Swedish parishes in the 19th century. *International Journal of Gynaecology and Obstetrics*, vol. 23, p. 489–497.

Högberg Ulf, Wall Stig, and Broström Göran, 1986. The impact of early medical technology on maternal mortality in late 19th century Sweden. *International Journal of Gynaecology and Obstetrics*, vol. 24, p. 251–261.

Kaunitz Andrew M., Spence Craig, Danielson T.S., Rochat Roger W., and Grimes David A., 1984. Perinatal and maternal mortality in a religious group avoiding obstetric care. *American Journal of Obstetrics and Gynecology*, vol. 150(7), p. 826–831.

Kleiner George J. and Greston Wilma M. (eds.), 1984. *Suicide in pregnancy*. Boston, John Wright, 213 p.

Koenig Michael A., Fauveau Vincent, Chowdhury Anwarullah I., Chakraborty Jyotsnamoy, and Khan Mehrab A., 1988. Maternal mortality in Matlab, Bangladesh: 1976–85. *Studies in Family Planning*, vol. 19(2), p. 69–80.

Kwast Barbara E., Rochat Roger W., and Kidane-Mariam Widad, 1986. Maternal mortality in Addis Ababa, Ethiopia. *Studies in Family Planning*, vol. 17(6), p. 288–301.

Loudon Irvin, 1992. *Death in childbirth: an international study of maternal care and maternal mortality, 1800–1950*, p. 160, 258. Oxford, UK, Clarendon Press, 622 p.

Measham A.R., Obaidullah M., Rosenberg M.J., Rochat Roger W., Khan A.R., and Jabeen S., 1981. Complications from induced abortion in Bangladesh related to types of practitioner and methods, and impact of mortality, *Lancet*, 1(1), 199–202.

Peetrek J., 1996. Breast cancer and pregnancy, *in*: J.R. Harris, M.E. Lippman, M. Morrow, and S. Hellman (eds.), *Diseases of the breast*. Philadelphia, Lippincott-Raven.

Peller Sigismund, 1965. Births and deaths among europe's ruling families since 1500, *in*: David V. Glass and David E.C. Eversley (eds.), *Population in history: essays in historical demography*, p. 87–100. Chicago, Aldine Publishing Company, 682 p.

Perrenoud Alfred, 1981. Surmortalité féminine et condition de la femme (XVIIe–XIXe siècles). Une vérification empirique. *Annales de démographie historique*, p. 89–104.

Piraino M., 1991. *Pathways of child survivors of maternal mortality in Africa. an exploratory report on Kenya, Namibia, Tanzania, Zambia, and Zimbabwe*. New York, Defence for Children International—USA.

Potts D.M. and Potts William T.W., 1995. *Queen Victoria's gene*. Gloucestershire, Alan Sutton Publishing, 100 p.

Ronsmans Carine, Khlat Myriam, Kodio B., Ba M., Bernis (De) L., and Étard Jean-François, 2001. Evidence for a "healthy pregnant women effect" in Niakhar, Senegal? *International Journal of Epidemiology*, vol. 30, p. 467–474.

Shahidullah Mohammed, 1995a. The sisterhood method of estimating maternal mortality: the Matlab experience. *Studies in Family Planning*, vol. 16, p. 101–106.

Shahidullah Mohammed, 1995b. A comparison of sisterhood information on causes of maternal death with the registration causes of maternal mortality in Matlab, Bangladesh. *International Journal of Epidemiology*, vol. 24, p. 937–942.

Sherer D.M. and Schenker J.G., 1989. Accidental injury during pregnancy. *Obstetrics and Gynecology Survey*, vol. 44, p. 330–338.

Speckhard M.E., Comas-Urritia A.C., Rigau-Perez J.G., and Adamsons M., 1985. Intensive surveillance of pregnancy-related deaths, Puerto Rico, 1978–1979, *Boletin de Asociacion Medical de Porto Rico*, vol. 77, p. 508–513.

Thuriaux Marie-C. and Lamotte Jean-Marie, 1984. Maternal mortality in developing countries: a note on the choice of denominator, *International Journal of Epidemiology*, vol. 13, p. 246–247.

WHO , 1987. *Studying maternal mortality in developing countries: rates and causes*. *A guidebook*. Geneva, World Health Organization, 63 p. (Document WHO/FHE/87.7.)

WHO and Lshtm (London School of Hygiene and Tropical Medicine), 1995. *Verbal autopsies for maternal deaths*. Geneva, World Health Organization. (Report of a WHO Workshop, London, 10–13 January 1994.; WHO/FHE/MSM/95.15. Geneva.)

WHO and Unicef, 1996. *Revised 1990 estimates of maternal mortality: a new approach by WHO and UNICEF*. Geneva, World Health Organization, 16 p. (Document WHO/FRH/MSM/96.11.)

Willmott-Dobbie B.M., 1982. An attempt to estimate the true rate of maternal mortality, sixteenth to eighteenth centuries, *Medical History*, vol. 26, p. 79–90.

C H A P T E R

46

The Determinants of Infant Health and Mortality

GODELIEVE MASUY-STROOBANT

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Starting from a level of potential health capital at birth, which can be roughly estimated through indicators such as birth weight, gestational age, or the presence of a congenital handicap, an infant's capacity to maintain or restore this potential, if necessary, depends on a whole series of factors on which the newborn has almost no means of action. These are referred to by demographers and epidemiologists as *determinants* of perinatal, infant, or child health and mortality.

I. DETERMINANTS OF WHICH INFANT MORTALITY?

Infant mortality, which is defined as the probability for a live-born child to die before its first birthday, is a multifaceted phenomenon that needs to be placed in its historical, geographic, and social contexts. This particularly frail year in the life of a human being is, at the individual level, part of a process that from conception to death could potentially last more than 100 years. Being born alive presupposes having escaped fetal death, and the study of the first year of life is a recognition of its specific nature, resulting primarily from the very high mortality risks observed therein compared with succeeding ages. In very-low-mortality countries, equivalent mortality risks are only reached at 60 years of age. In situations of high mortality, only old people are exposed to risks higher than those for small children, and going beyond the stepping stone of the first birthday is but one stage in the race for survival because child mortality (from 1 to 4 years) is still far from negligible (see Table 46–3), whereas it is mainly due to accidents in low infant mortality countries.

In a historical context, the rapid and continuous decline in infant mortality in countries where it has reached current in low levels was accompanied by a concentration of deaths at even younger ages. Between 40% and 50% of the deaths that occur during the first week of life are mostly children born preterm (less than 37 weeks' gestation) or with low birth weight (less than 2500 g or even 1500 g). It is therefore not surprising that research is increasingly focusing on the cause of late fetal deaths, which are deaths immediately preceding the early neonatal period, of which it is conceptually and chronologically the extension (see Chapter 32). In these situations, the concept of *perinatal mortality* (grouping late fetal deaths or stillbirths and early neonatal deaths) tends to supersede that of infant mortality, even if there is regained interest in the post-neonatal period in the context of prevention of the sudden infant death syndrome. In situations of high mortality levels, attention is focused on the already born child, and the concept of infant mortality is completed by that of child mortality (1 to 4 years), because mortality risks remain high until the age of 5.

Today, we are faced with two mortality systems that differ fundamentally in their levels, timing (age at death), structure of medical causes, measurement, search for determinants, and identification of priorities for action. Through these differences, a certain number of regularities are observed, such as the well-known phenomenon of social inequality, which persists irrespective of mortality levels, or the series of measures applied toward their elimination.

At the global level, the most recent estimates produced based on civil registration for developed countries, on retrospective surveys such as the Demographic and Health Surveys (DHS), or generated by indirect methods from census or survey questions on the survival status of children for developing countries, show that infant mortality varies from 195‰ in Sierra Leone to 4‰ in Japan (Lévy, 1999). The levels by large regions are shown in Table 46–1. Europe was still characterized by differences in risks of such magnitude in the early 20th century, with an infant mortality rate of 80‰ in Norway from 1901 to 1905 (Masuy-Stroobant, 1997) and perhaps more than 300‰ in Russia (Zakharov, 1996) at the same period.

As much as available data permits, the high mortality levels of the past show some similarities with what is observed in many poor countries today. Mortality remains high beyond the first year of life, and the causes of death were mostly of infectious origin. It is paradoxical that we have civil registration data for medical causes of infant mortality in 19th century Europe, that, although containing a high proportion of ill-defined causes of death, give a good idea of the situation at the national level. In Belgium between 1895 and 1900, enteritis and diarrhea were leading causes, followed by infectious diseases (e.g., measles, smallpox, whooping cough, scarlet fever, diphtheria) and then by acute respiratory diseases (Masuy-Stroobant, 1983, p. 32). Concerning early neonatal mortality, Iceland showed very high levels of mortality due to tetanus in the 18th century (Vasey, 1997).

TABLE 46–1 Infant Mortality per 1000 Live Births in the World in 1999

Region	Rate (‰)[a]	Region	Rate (‰)[a]
Africa	**88**	**America**	**28**
North	51	North	7
West	86	Central	34
East	105	South	35
Central	104	Caribbean	41
Southern	55		
Asia	**56**	**Federation of Russia**	**17**
West	54	**Europe**	8
South Central	74	North	6
South East	46	West	5
East	29	East	13
		South	7
Oceania	**24**	**World**	**57**

[a] Most of these figures are the results of estimates made by the Population Reference Bureau based on available statistics in different countries. In low-mortality countries, they come from civil registers, and in high-mortality countries, they come from retrospective surveys.

From Lévy (1999).

For developing countries and especially for countries with very high infant mortality, these sources of data do not exist, and it is often necessary to rely on intensive surveillance systems, which are generally not representative of the countries in which they function, or on highly selective hospital data concerning predominantly urban populations. In these cases, the major causes of death in childhood (0 to 5 years) are mainly infectious or parasitic: "They are mostly severe respiratory conditions, diarrhoea and enteritis, malaria, measles and tetanus" (Hill and Hill, 1988, p. 75). Undoubtedly, we must include today the devastation caused by the acquired immunodeficiency syndrome (AIDS) epidemic that directly, by transmission from mother to child, or indirectly, by the greater risk of becoming an orphan, increases the risk of mortality in childhood, especially in urban areas of Central and Eastern Africa (Rossi-Espagnet *et al*., 1991).

At the other extreme, the main causes of infant deaths in low-mortality countries are perinatal conditions, among which those related to immaturity are most common, followed by congenital anomalies and the sudden infant death syndrome, which in certain countries belongs to the larger category of "ill-defined symptoms and morbid states". If perinatal causes are concentrated in the first month of life and cot death appears after the first month, infant deaths due to congenital anomalies, especially frequent during the first month, may also occur later. These three large groups represented 73% of infant deaths in Belgium and in Great Britain in 1994.

The fact that measles, enteritis, diarrhea, and other infectious diseases, considered preventable today by vaccination or easily cured and rarely lethal in low-mortality countries, are such terrible killers in poor countries no doubt evokes the nutritional hypothesis proposed by Thomas McKeown (1976) to explain the general mortality decline in Europe since the middle of the 19th century. According to him, the reduction in lethality of disease was brought about by an improved resistance of the host (in this case, children) because of improved living conditions in general and better nutrition in particular. Other authors invoke the advantages of public health measures, which by the end of the 19th century provided cities with sewage systems and drinkable water distribution services, when domestic and personal hygiene practices became more popular. As much as it is possible to understand and explain *a posteriori* events that belong to the past, these two hypotheses have been confirmed by facts.

Infant feeding however, shows some specificity, and if the medical corps had already noticed a higher weight gain among breast-fed newborn babies compared to bottle-fed babies (Bouchut, 1878), artificial feeding of the newborn could not yet be carried out safely. Pasteur's discoveries were applied on a large scale only at the end of the 19th century by methods of industrial pasteurization of milk and by the diffusion of Pasteur's recommended hygiene rules and pediatric care (through well-baby clinics) and by the sterilization of milk and feeding bottles at home, hygiene, and care of babies. These rules are still practiced today, and their spread and the measurement of their effects by regular medical check-up of children (e.g., weight, height, monitoring of changes in feeding habits, general vaccination) are the basis of the European organization of preventive medicine, which has since spread all over the world.

The observed past interaction between malnutrition and infectious diseases is quite similar to morbidity and mortality by infectious diseases among children in the Third World today. It is estimated that malnutrition is associated with 60% of deaths in childhood for developing countries (Blanc, 1991). Measles among malnourished children is frequently associated with other infectious diseases such as pneumonia or diarrhea. Children who survive this severe episode remain weakened for a long time, and various studies show a delayed effect of measles on the children's survival, particularly in sub-Saharan Africa (Clements *et al.*, 1992). Maternal malnutrition is also partly responsible for the high proportion of low-birth-weight children reported in some estimations (Kramer, 1987). These children, already weakened at the onset, will show lower resistance to subsequent infectious diseases.

Pasteur and the improvement in infant nutrition are not the only factors responsible for the steady decline in infant mortality, which for developed countries constitutes one of the most important achievements of the 20th century. This decline was accompanied by a general economic and social development that seems to have been particularly favorable to women and children; the generalized access to education and the development of effective contraception gave rise to an unprecedented improvement in women's status in these societies. Equally important was the considerable development in medical care during pregnancy and delivery and the fight against infectious diseases by means of vaccination or antibiotics.

Medically assisted procreation and the care of increasingly preterm and low-weight babies have since the mid-1980s gradually reshaped the context of health and infant mortality in rich countries. An increasing number of children, who otherwise would not even have been registered because of their extreme immaturity, can be now be cared for and saved. There is an increase in low-birth-weight births (less than 2500 g) and of very low birth weights (less than 1500 g) in most developed countries with the access to intensive neonatal care techniques. This phenomenon may cause an apparent stagnation in infant mortality levels, but the increasing possibility saving babies of very low weight does not exclude the risk of handicap for those who survive (Topp *et al.*, 1997). The question for these countries is that of the consequences of this decrease in mortality on the health of children, especially children being born with a very low birth weight.

II. A COMPLEX PHENOMENON TO BE ANALYZED AT DIFFERENT LEVELS

The determinants of infant mortality, which are generally different from medical causes of death, can be analyzed, explained, or acted on at different levels. The coexistence of very different mortality systems is generally interpreted at the macro-societal level in terms of the level of economic and social development. The association between national or regional levels of infant mortality and level of development, measured according to various indicators, is so systematic that infant mortality is generally considered to be one of the most sensitive indicators of a country's level of development. At the individual level, the extent of social inequality in infant mortality could also serve as an indicator of equity in the distribution of resources within a country.

The determinants of infant mortality have been traditionally identified by simple comparative analyses of available data. The term *determinants* was then attributed each time a variable shows a recurrent significant association with infant mortality in various situations.

At the national level, political choices made by governments are likely to improve female literacy and education of young girls, increase access (physical and financial) to health care, and improve living conditions (e.g., through reclamation of swampy land, installation of sewage systems, extension of drinkable water distribution, collection and treatment of waste). They may also promote access to employment with a view to ensuring better living standards for all.

At the community level, the quality of water for domestic use or the availability of drinking water; access to electricity and to a sewage system; a healthy environment; adequate housing; and climatic conditions have been major determinants of the mortality decline in Europe until the early 20th century and are still important factors in poor, high-mortality countries. Access to health services (preventive and curative) at the local level is also a part of this group of determinants. Generally, the urban or rural nature of the area determines access to such collective services.

The household's composition where a child is born and the importance given to each member in the division of tasks or the distribution of available resources (e.g., food, health care, education) may be promising research themes, as well as the status of women and girls in particular and the relation of children to adults in general. Beyond the analysis of the bargaining power between generations or between spouses, an analyst has to consider the help and the affective or psychological support the household members, relatives, and friends can provide in sharing childcare.

At the individual level, the position of the infant's family in the social hierarchy affects in various ways access to resources needed to maintain good health and coping strategies for everyday life conditions. This position is often measured by the father's income and occupation or the mother's level of education. The legal nature of the union, the nationality, and the ethnic group are other social status criteria.

The mother's biologic characteristics, such as her own weight at birth, her pregestation weight, and weight gain during pregnancy, are other nutritional indicators that enable us to predict a child's birth weight and, consequently, its frailty.

1. Preventative Behavior

The fertility scheme (i.e., age at child birth, number and spacing of births, and in particular, the last birth interval) adopted by the mother is part of the numerous recommended preventive behaviors to carry her pregnancy to term and maintain her child in good health. The identification of risk factors in her past pregnancies history, such as miscarriages, preterm births, or infant deaths, is part of the standard anamnesis during pregnancy because of their predictive power on the outcome of the current pregnancy.

Prevention can start before conception by the search for hereditary defects likely to be transmitted to the child. It may continue during pregnancy by early prenatal care and the search for signs of acquired immunity by the mother (e.g., tetanus, rubella, toxoplasmosis, sexually transmitted diseases, testing for human immunodeficiency virus) against a certain number of infectious or parasitic diseases (malaria) and the existence of maternal pathologies (e.g., diabetes, heart disease) that could endanger the life of the mother, the fetus, or the newborn.

The early detection of congenital anomalies can lead to the decision to undergo a so-called therapeutic abortion, thereby reducing the infant mortality level but mostly reducing the prevalence of handicapped children in the population. Beyond this medical type of prevention, there is also the adoption of daily preventive behavior, which can be summed up as living a healthy life: not overtiring oneself, avoiding stress and heavy or tedious work, having a balanced diet, and when necessary, using vitamin supplements to manage preexisting deficiency (i.e., anemia) or deficiency caused by the pregnancy itself, avoiding alcohol but especially smoking or taking drugs. Childbirth should, if possible, take place in a hospital equipped to cope with dystocic pregnancies and high-risk births.

As soon as the child is born, prevention should be continued by providing the baby with adequate nutrition (especially in the case of artificial feeding and during the delicate weaning period), ensuring the newborn's body hygiene, remove sources of infection or accidents, and give it the attention and love enabling it to grow well. From a medical point of view, vaccinations should be done according to a recommended schedule, and attention is paid to the baby's growth by regular visits to well-baby clinics or health centers. It is also necessary to have the capacity to identify early symptoms of illness to seek adequate care in due time (e.g., water diet or use of oral rehydration salts in the case of diarrhea, evaluation of the severity of episodes of fever).

2. Psychological Factors

Already identified in 1955 by J. Morris and J. Heady (1955), the concept of mother's capacity, which indicates an active rather than passive psychological profile when facing difficulties, is a research theme that is being developed (Kronenfeld, 1997).

3. The Child's Characteristics

Besides birth weight, pregnancy duration, and the presence of anomalies, the male mortality disadvantage is the rule, at least until the first birthday, which in high-mortality situations may give way to a female disadvantage that can last until the ages of 14 to 15 years (Tabutin and Willems, 1996). These trends, duly reported at the end of the 19th century in Europe, where they gradually disappeared during the early 20th century, are still observed in several high-mortality countries today. Twin births represent an increased risk because they are often preterm births. The spread of medically assisted procreation in low-mortality countries has triggered an unprecedented increase in multiple and very multiple (triplets and more) births in the past 2 decades (Masuy-Stroobant, 1994).

This already long list is far from exhaustive, but it reflects the importance of research on the cause of death in infancy, the complexity of its causality, and its analysis because of the interdependence of the different levels of explanation.

The often mentioned theme of maternal education (see Chapter 56 on theories and explanatory frameworks on child mortality) can be analyzed at several levels. Its mode of operation and the status it confers to a woman depend on the political context.It no doubt affects the choice and place of residence and lifestyle, and the educational level, if high enough, gives rise to better living standards, which gives people access to more comfortable, better furnished, and easier to maintain housing, as well as to financial access to adequate health care. The status it confers to a woman and mother in the household empowers her to better negotiate the distribution of resources in favor of better care for her child. It is also generally associated with behavior more attuned to anticipation, investment in the future, and prevention, and it gives a woman the cognitive means and the rationality to find solutions to the difficulties she might encounter.

Most of the determinants discussed here rarely exist in isolation, and they may have an effect on child health or mortality only in the presence of other determinants, making it difficult or impossible to isolate a totally independent effect. This is the case, for example, for smoking, which is the object of numerous studies. Smoking often appears to exert a direct effect on birth weight by producing intrauterine growth retardation. Studies have shown that the smoker's diet during pregnancy is of less good nutritional quality than that of nonsmokers, nutritional deficiency in itself contributing also to fetal growth retardation (Trygg *et al*., 1995). The social and behavioral profile of female smokers also is significantly different from that of nonsmokers in many ways. How then can we isolate the specific effect of smoking?

III. CAUSES, DETERMINANTS, AND DIFFERENTIAL RISK

In an overview of research conducted by demographers in the area of mortality, Eileen Crimmins (1993, p. 589) wrote the following:

> Mortality research remains largely descriptive, as was the fertility research of 30 years ago. Because no "theory" of mortality currently exists, this area is likely to undergo considerable theoretical development in the coming decade. Currently, the emphasis is placed on describing differences according to race, ethnicity, and social class. New independent variables, such as social support, are suggested regularly as possible explanatory variables for mortality differences, but the mechanisms through which these variables might work often remain unspecified.

This comment adequately summarizes the situation of field research in this area, even if some recent attempts to test the effect of maternal education (Sandiford *et al*., 1995) or the evaluation of the respective effects of contextual determinants (collective services such as water or electricity supply) and individual determinants (socioeconomic and behavioral characteristics) are interesting in explaining, for example, rural-urban differentials in child mortality in Brazil (Sastry, 1997).

Research on the determinants or causes of infant mortality is confronted to the relative scarcity of data on the phenomenon and the highly complex interrelations of its factors at several levels of observation and analysis demonstrated by demographers and by epidemiologists.

Developed countries generally have civil registration or medical birth registry systems that ensure complete coverage (almost exhaustive coverage for medical registers) of births and deaths. Organized for administrative or medical purposes, these information systems enable research to at least associate the sociodemographic characteristics of a child's parents (e.g., occupation, nationality, education) and certain biologic variables (e.g., mother's age, parity, last birth interval, twin status, sex of the child) with the risk of infant mortality or the risk of low birth weight or preterm delivery (Table 46–2).

In many cases, the possibility of calculating differential infant mortality risks or risks of low-birth-weight or preterm births depends on being able to match death records having less information with the corresponding birth records (Gourbin, 1996; Gourbin and Masuy-Stroobant, 1995). Medical birth registers are generally limited to recording early neonatal deaths, and if they contain other medical data (e.g., maternal pathologies, parity, gestation period, child's

TABLE 46–2 Social Inequality, Fetal Mortality, and Infant Mortality per 1000 Births in Europe

Social characteristics	England and Wales (1995)		Austria (1995)		Belgium (1994)		Hungary (1995)		Portugal (1995)		Romania (1995)	
	FM	IM	FM	IM	FM	IM	FM	IM	FM	IM	FM	IM
Total population	5.5	6.1	4.4	5.4	4.3	7.5	3.5	10.7	6.9	7.5	6.2	21.2
Marital status of mother												
Married	5.2	5.4	4.1	5.4	4.1	7.2	3.2	9.4	6.0		6.0	
Not married	6.2	7.6	5.0	5.4	5.6	9.0	4.5	15.7	10.9		7.0	
Relative risk × 100	119	141	122	100	137	125	141	167	182		117	
Father's profession or social class (for children born in wedlock)												
Highest level	4.2	4.2			2.5	5.6						
Lowest level	8.5	7.7			6.4	10.7						
Relative risk × 100	202	183			256	191						
Mother's level of education												
Highest level			3.2	4.5			2.1	7.3		2.9	4.9	9.2
Lowest level			5.0	8.2			5.7	27.5		9.8	7.8	51.6
Relative risk × 100			156	182			271	377		338	159	561

Data from civil registration, published statistics (England and Wales, Hungary), available records (Austria, Portugal, Romania), and from analysis of private data by the author (Belgium).

birth order, precise circumstances of birth, baby's situation at birth, care of the newborn), social data are often absent. In certain countries and especially in most northern European countries, these data are systematically matched with those of civil registers making it possible to analyze data from both sources. The systematic recording of the individual identification code of the population register on each administrative document makes different types of cross-tabulations possible in most northern European countries. Consequently, it is by matching data from medical registers with census data that measurement of the joint effects of parents' level of education on infant mortality became possible in Norway (Arntzen *et al.*, 1993).

If these databases are available to researchers, they enable them at most to measure very precisely, for total populations, the differential risks associated with the several available variables, to establish the possible interactions between fertility behavior, past pregnancies' outcomes, and social identification variables, or to measure the net effect of these individual variables (after controlling for confounders) on mortality or on the indicators of child health at birth (e.g., weight, gestational age). They are, however, strictly limited to an individual level of analysis and can only rarely test hypotheses regarding the mechanisms (of prevention) explaining how the mother's education could protect or even improve the health of children.

In developing countries, those for which most of the theories and explanatory frameworks were developed (see Chapter 56), most studies on infant and child mortality are based on retrospective data from secondary analyses of surveys (initially the World Fertility Surveys, then the successive waves of Demographic and Health Surveys), whose primary aim was to measure and analyze fertility. Numerous studies were therefore based on the fertility history of women and the results of successive pregnancies and then, depending on available data, went on to develop valid estimates of infant and child mortality before measuring its differentials (Table 46–3). If the statistical rarity of death forced them to conduct mortality analysis on the 5 or even 10 years preceding the survey, they however present the undeniable advantage of providing different levels of analysis: individual, household and community or contextual levels (as in the example of Brazil cited earlier).

To the question about what are the determinants of infant mortality or at least what are the proven determinants, which are significantly related to the risk of infant mortality in several contexts (as far as confounders are taken into account) and for which a plausible mechanism for action could be described (if not tested in the field), the answer remains difficult. An impressive number of risk criteria have been published and are regularly updated, but, for instance, only a vague response is available to the question concerning the pathways through which early prenatal

TABLE 46–3 Demographic Characteristics of Infant Mortality in High-Mortality Countries

Demographic characteristics	Niger DHS 1998	Burkina Faso DHS 1999	Bangladesh DHS 1997	Kenya DHS 1998	Bolivia DHS 1998	Indonesia DHS 1997	Guatemala DHS 1999	Vietnam DHS 1997
Total mortality								
Infant mortality	135.8	108.6	89.6	70.7	73.5	52.2	49.1	34.8
Child mortality	193.0	129.6	42.0	37.1	27.6	19.4	16.3	11.5
Sex of child								
Girl	130.5	100.9	84.3	66.8	69.2	44.9	48.1	26.9
Boy	140.9	116.1	94.9	74.5	77.6	59.1	50.0	42.0
Relative risk × 100	108	115	113	112	112	132	104	156
Birth order								
Birth order 4, 5, or 6	121.6	97.0	82.4	67.3	78.8	49.7	47.8	36.7
First birth	162.3	144.3	108.6	62.7	59.2	47.7	52.4	35.4
Relative risk × 100	133	149	132	93	75	96	110	96
Last birth interval								
4 years or more	79.3	48.1	55.5	44.3	32.8	31.8	32.7	20.7
Less than 2 years	176.2	154.6	136.1	102.2	122.7	103.4	70.3	60.4
Relative risk × 100	222	321	245	213	374	325	215	292
Mother's age								
20–29 years	126.4	105.9	79.3	58.6	67.3	47.4	41.0	33.7
Less than 20 years	177.6	142.2	106.2	97.0	96.8	63.4	71.1	46.0
Relative risk × 100	141	134	134	166	144	134	173	137
Mother's level of education								
Secondary or higher	70.1	78.5	64.8	40.0	41.3	28.0	41.1	29.0
None	140.9	110.2	98.1	82.2	112.5	77.5	55.7	48.8
Relative risk × 100	201	140	151	206	272	277	136	16.8
Place of residence								
Urban	79.9	67.4	73.0	55.4	53.0	35.7	49.0	23.2
Rural	146.7	113.2	91.2	73.8	99.9	58.0	49.1	36.6
Relative risk × 100	184	168	125	133	188	162	100	158

Risks are calculated per 1000 births for the 10 years preceding the survey.
From Demographic and Health Survey, Survey Indicators STATcompiler, www.measuredhs.com, January 2001.

consultations can reduce infant mortality or the incidence of preterm births, especially when maternal education and place of residence considerably reduce net effects. We are faced with the situation described by Kenneth Wolpin (1997, p. 532) "The result of the absence of theorizing is that researchers are unrestricted in the choice of factors that are simultaneously considered," which makes systematic and comparative analysis particularly difficult.

IV. FROM MORTALITY TO HEALTH: A QUESTION OF WEIGHT, HEIGHT, AND GESTATIONAL AGE

Mortality is most often preceded by a worsening of health or even a state of frailty at birth. Pediatricians have for a long time recognized birth weight as an indicator of a child's frailty. If it is less than 2500 g, it is identified as *low birth weight*; if it is less than 1500 g, it is *very low birth weight*. If weight is not actually a valid measure of maturity for a newborn (e.g., a baby can be born at term, or 37 weeks' gestation, and weigh less than 2500 g, describing a case of intrauterine growth retardation), it is at least more accessible and easier to measure, irrespective of place or conditions of birth–you only need a scale.

Taking birth weight into account in the measurement of infant mortality in low-mortality countries ensures a considerable improvement in spatiotemporal comparability of the phenomenon. This is because of the effect on the overall level of infant mortality of the increasing incidence of differences in legal criteria for birth reporting in civil registers and of the growing and varying tendency to declare very low birth weight births, for which survival chances have improved dramatically because of advances in neonatal intensive care (Masuy-Stroobant and Gourbin, 1995).

TABLE 46–4 Early Neonatal Mortality (0 to 6 Days) per 1000 Live Births by Categories of Birth Weight in Some European Countries in 1990

Weight (g)	Austria	Belgium	Czech Republic	Hungary	Poland[a]
500–999	397.1	387.5	609.3	718.8	754.4
1000–1499	95.2	134.9	257.7	236.6	304.5
1500–1999	41.1	33.9	71.6	48.6	101.7
2000–2499	11.4	11.0	16.4	14.9	24.4
2500–2999	2.3	1.7	3.6	2.8	4.9
3000–3499	0.8	1.1	1.6	1.7	2.2
3500–3999	0.5	0.7	0.9	1.5	1.7
4000–4499	0.9	1.3	1.4	1.4	1.9
4500+	2.6	1.9	2.5	4.4	6.2
Total	3.3	3.4	5.8	8.7	12.4
Infant mortality	7.8	8.0	10.8	14.8	19.4

[a] Including "nonviable" live births. For Poland, the weight categories do not correspond exactly to international recommendations. The figures presented here refer to the following categories: 601–1000 g; 1001–1500 g; 1501–2000, and so on.
From Civil Registration statistics. Unpublished data except that for Poland.

A close relationship is systematically observed between birth weight and the risk of mortality (Table 46–4), including in developing countries. McCormick (1985, cited in Kramer 1987, p. 664) stated the following: "In both developed and developing countries, birth weight is probably the single most important factor that affects neonatal mortality, in addition to being a significant determinant of post-neonatal infant mortality and of infant and childhood morbidity." There are many similarities between the determinants of low birth weight and those identified for infant mortality.

The choice of the risk of low birth weight has the advantage of partially solving the problem of the statistical rarity of death faced by researchers (i.e., the incidence of births weighing below 2500 g is at least 10 times the risk of death in low-mortality countries) and, in the case of developed countries, save time spent in matching births and death records. The disadvantage, however, is that it excludes maternal behavior after delivery. An example is given by the mother's smoking habits, infant feeding practices, and sleeping positions in the study of the determinants of cot death. In developing countries, this indicator of intrauterine growth can be generalized by a measure of its nutritional status at different ages, as suggested by Henry Mosley and Lincoln Chen (1984) (see Chapter 56). The choice of birth weight as an indicator of health at birth offers a much wider range of research themes going far beyond those restricted to mortality. One example of this is a case-control study in Sudan comparing individual and community characteristics of families with low-birth-weight children with families whose children weighed 2500 g or more (Taha *et al.*, 1995).

Acknowledgment

This chapter was translated from the French version by O. D. Chimere-Dan.

References

Arntzen Annett, Magnus Per, and Bakketeig Leiv S., 1993. Different effects of maternal and paternal education on early mortality in Norway, *Paediatric and Perinatal Epidemiology*, vol. 7(4), p. 376–386.

Blanc Ann K., 1991. *Demographic and Health Surveys, world conference, August 5–7, 1991, Washington D.C., executive summary*. Columbia, Maryland, IRD/Macro International, 28 p.

Bouchut Eugène, 1878. *Traité pratique des maladies des nouveau-nés, des enfants à la mamelle et de la seconde enfance*, 7th ed. Paris, Librairie Baillière et Fils, 1127 p.

Brzezinski Zbigniew J. and Szamotulska Katarzyna, 1994. The widening gap in low birthweight rates between extreme social groups in Poland during 1985–90, *Paediatric and Perinatal Epidemiology*, vol. 8(4), p. 373–383.

Clements C. John, Strassburg Marc, Cutts Felicity T., and Torel Carole, 1992. The epidemiology of measles, *World Health Statistics Quarterly*, vol. 45(2/3), p. 285–291.

Crimmins Eileen M., 1993. Demography: the past 30 years, the present, and the future, *Demography*, vol. 30(4), p. 579–591.

Gourbin Catherine, 1996. Inégalités sociales en santé et mortalité foeto-infantiles. Les sources nationales disponibles en Europe, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin, and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui. Chaire Quetelet 1994*, p. 67–96. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 560 p.

Gourbin Catherine and Masuy-Stroobant Godelieve, 1995. Registration of vital data: are live births and stillbirths comparable all over Europe? *Bulletin of the World Health Organization*, vol. 73(4), p. 449–460.

Hill Althea and Hill Kenneth, 1988. La mortalité en Afrique: niveaux, tendances, différences et perspectives, *in*: *L'état de la démographie africaine*, p. 67–84. Liège, Union internationale pour l'étude scientifique de la population, 156 p.

Hoffman Susie and Hatch Maureen C., 1996. Stress, social support and pregnancy outcome: a reassessment based on recent research, *Paediatric and Perinatal Epidemiology*, vol. 10(4), p. 380–405.

Kramer Michael S., 1987. Determinants of low birth weight: methodological assessment and meta-analysis, *Bulletin of the World Health Organization*, vol. 65(5), p. 663–737.

Kronenfeld Jennie Jacobs, Reiser Marc, Glik Deborah C., Alatorre Carlos, and Jackson Kirby, 1997. Safety behaviors of mothers of young children: impact of cognitive, stress and background factors, *Health*, vol. 1(2), p. 205–225.

Levy Michel-Louis, 1999. Tous les pays du monde (1999), *Population et Sociétés*, July-August (348), 8 p.

Masuy-Stroobant Godelieve, 1983. *Les déterminants de la mortalité infantile. La Belgique d'hier et d'aujourd'hui*. Louvain-la-Neuve, CIACO, 540 p.

Masuy-Stroobant Godelieve, 1994. La mortalité infantile en Europe et au Canada. Un problème résolu ? *Cahiers Québecois de Démographie*, vol. 23(2), p. 297–340.

Masuy-Stroobant Godelieve, 1997. Infant health and infant mortality in Europe: lessons from the past and challenges for the future, *in*: Carlo A. Corsini and Pier Paolo Viazzo (eds.), *The decline of infant and child mortality. The European experience: 1750–1990*, p. 1–34. The Hague, Martinus Nijhoff, 258 p.

Masuy-Stroobant Godelieve and Gourbin Catherine, 1995. Infant health and mortality indicators. Their accuracy for monitoring the socio-economic development in the Europe of 1994, *European Journal of Population*, vol. 11, p. 63–84.

McCormick Marie C., 1985. The contribution of low birth weight to infant mortality and childhood morbidity, *New England Journal of Medicine*, vol. 312, p. 82–90.

McKeown Thomas, 1976. *The modern rise of population*. London, Edward Arnold, 169 p.

Morris J.N. and Heady J.A., 1955. Social and biological factors in infant mortality. V. Mortality in relation to the father's occupation 1911–1950, *The Lancet*, March 12, p. 554–559.

Mosley Henry W. and Chen Lincoln C., 1984. An analytical framework for the study of child survival in developing countries, *Population and Development Review*, vol. 10, Suppl., p. 25–45.

Rossi-Espagnet A., Goldstein G.B., and Tabibzadeh I., 1991. Urbanisation et santé dans les pays en développement: un défi pour la santé pour tous, *Rapport Trimestriel des Statistiques Sanitaires Mondiales*, vol. 44(4), p. 186–244.

Sandiford Peter, Cassel Jeffrey, Montenegros Maria-Martha, and Sanchez Germana, 1995. The impact of women's literacy on child health and its interaction with access to health services, *Population Studies*, vol. 49(1), p. 5–17.

Sastry Narayan, 1997. What explains rural-urban differentials in child mortality in Brazil? *Social Science and Medicine*, vol. 44(7), p. 989–1002.

Tabutin Dominique and Willems Michel, 1996. La surmortalité des filles jusqu'à 1940. Un bel exemple des inégalités sexuelles dans l'histoire occidentale, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin, and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui. Chaire Quetelet 1994*, p. 129–177. Louvain-la-Neuve, Academia-Bruylant/L'Harmattan, 560 p.

Taha El Tahir, Gray Ronald H., Abdelwahab Mabyoua Mustafa, and Abdelhafeez Adelrahman, 1995. Distribution and determinants of low birthweight in Central Sudan, *Paediatric and Perinatal Epidemiology*, vol. 9(2), p. 185–200.

Topp Monica, Uldall Peter, and Langhoff-Roos Jens, 1997. Trend in cerebral palsy birth prevalence in eastern Denmark: birth-year period 1979–86, *Paediatric and Perinatal Epidemiology*, vol. 11(4), p. 451–460.

Trygg Kerstin, Lund-Larsen Karin, Sandstad Berit, Hoffman Howard J., Jacobsen Geir, and Bakketeig Leiv S., 1995. Do pregnant smokers eat differently from pregnant non-smokers? *Paediatric and Perinatal Epidemiology*, vol. 9(3), p. 307–319.

Vasey Daniel E., 1997. An estimate of neonatal tetanus mortality in Iceland, *European Journal of Population*, vol. 13(1), p. 49–69.

Wolpin Kenneth I., 1997. Determinants and consequences of the mortality and health of infants and children, *in*: Marc R. Rosenzweig and Oded Stark (eds.), *Handbook of population and family economics*, p. 483–557. Amsterdam, Elsevier Science, 302 p.

Zakharov Serguei V., 1996. The Second World War as a turning point of infant mortality decline in Russia, *in*: Godelieve Masuy-Stroobant, Catherine Gourbin, and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui. Chaire Quetelet 1994*, p. 311–333. Louvain-la-Neuve, Academia-Bruylant/L'Harmattan, 560 p.

C H A P T E R

47

Relationship between Morbidity and Mortality by Cause

VIVIANA EGIDI AND LUISA FROVA

Università degli Studi di Roma "La Sapienza," Rome, Italy

Instituto Nationale di Statistica (ISTAT), Rome, Italy

I. KNOWLEDGE GAINED BY MONITORING THE MORBIDITY PROCESS

When mortality is high, particularly due to infectious and acute diseases, the life table and its indicators constituted quite efficient tools for evaluating the population's state of health. Increased life expectancy and the growing impact of chronic or degenerative diseases impose the need for a more in-depth evaluation of the current trends. Attention is shifting from the considerable progress made in quantitative terms over the past few years, particularly with regard to survival in adult and senile age, toward an analysis that investigates the quality of the years that have been added to the mean life duration. The possibility of monitoring the entire morbidity process from a theoretical and statistical point of view rather than only its terminal event (death) substantially increases knowledge and the possibilities to evaluate differentials and trends as well as allowing the formulation of more appropriate public health policies.

This chapter presents an overview of the available data sources for the evaluation of the health status of a population and the concepts and methods that take into account the relationships between mortality by cause and the incidence, prevalence, and lethality of diseases. We consider the methods of evaluation of survival by health status based on morbidity-mortality life tables and their corresponding healthy life expectancies. The latter refer to different concepts and definitions of health and can be computed according to different methods, taking into account the basic assumptions and the availability of data on incidence or prevalence of diseases.

II. MONITORING MORBIDITY

1. Hypotheses about Longevity and the Quality of Survival

Numerous hypotheses have been put forward concerning the consequences of increased longevity on the quality of survival in adult and senile age. Certain theories maintain that an improvement of therapeutic capability, leading to the lengthening of a patient's life, has not been accompanied by an equally efficient action in the field of disease prevention. The decline in mortality therefore would be the result of a decrease in the probability of death (or lethality) from chronic diseases rather than a reduction in their incidence and would therefore hide a worsening of survival in qualitative terms. This is the *theory of pandemic morbidity* (Gruenberg, 1977, 1980; Kramer, 1980). Partially correcting this hypothesis, Olshansky (1991) maintains that the fall in mortality has merely changed the balance of morbidity processes from fatal to nonfatal ones, observing that the diseases that mostly affect the quality of survival are not necessarily those associated with the highest risk of death.

On the contrary, many experts (Day, 1991; Robine and Colvez, 1991; Fries, 1980, 1983, 1987, 1989) believe that, thanks to a more suitable diet, less tiring lifestyles, greater attention to physical exercise, more widespread awareness about prevention and well-known risk factors (e.g., smoking, alcohol, sedentary habits, high-fat food intake), the state of health remains at good quality levels up until an advanced age and that the morbidity process tends to be more and more concentrated in later life This is the *theory of the contraction of morbidity*.

In between these two extreme viewpoints, a theory called *dynamic equilibrium* has been advanced (Manton, 1982). According to this theory, the increase in life expectancy can be partly explained by a slowing in the rate of progression of chronic diseases. For patients, this produces a longer mean duration of the state of disease, with a subsequent increase in the prevalence of chronic diseases that would remain at less serious levels for longer periods. According to another hypothesis, the dynamic equilibrium could be produced by means of a reduction in the mean duration of disease accompanied by an increase in the mean level of seriousness of the disease itself (Olshansky, 1987).

2. Disease as a Process

According to the state of health, a survival analysis requires the availability of both conceptual and methodologic tools capable of monitoring the entire process from the onset of the disease until the death of an individual or his or her eventual recovery. Numerous efforts have been made to complement traditional indicators of mortality with measures capable of indicating the population's state of health. There are several difficulties involved, for the most part linked to the very nature of the phenomenon, for which even the definition poses a problem. This chapter does not address the question of the wider debate about the different objective and subjective approaches to study the state of health. The goal of describing the relationship between morbidity and cause-specific mortality leads to adoption of an objective approach, according to which the attribution of a state of ill health depends on the presence of a specific disease, which is identified by reference to the International Classification of Disease (ICD), used for the classification of causes of death.

Despite the simplification introduced whereby ill health is identified with a state of disease, the conceptual and methodologic problems that remain are considerable. Disease is not an event that can be identified as taking place at a given time and age of an individual with the same degree of certainty normally encountered when identifying a precise event. The passage from a condition of health to one of disease is rarely evidenced by one precise event. Generally, it can be represented as the beginning of a process–not necessarily completely evident–that goes through a number of phases, developing over the course of time and with reference to the age of the individual concerned, leading to different outcomes (e.g., recovery, chronic state, death). The course of the disease, its duration and outcome, depend on the type of disease and on endogenous factors, such as sex of the individual and age at which the disease was contracted, and exogenous factors, such as the health care facilities available to the individual and his or her ability to use them effectively. During the process, the individual may go through different phases of severity of the disease, both in terms of the risk of death connected with it and in terms of limitations of autonomy and impact on the quality of life. At an advanced age, it is difficult to attribute the state of disease to one specific morbidity process. More often than not, several processes accumulate and interact in the same individual, creating even more complex situations.

III. MORBIDITY PROCESSES AND CAUSES OF DEATH

1. Disease, Death, Lethality, and Recovery: From the Individual to the Population

The concepts and models necessary to evaluate the impact on a population of a particular morbidity process are substantially different from those used in clinical or epidemiologic microanalysis. The objectives of the two types of study are also very different: the former tends to highlight the consequences rather than the causes of particular trends and characteristics of the state of health, often placing the accent on aspects such as the evaluation of the impact of morbidity and its dynamics with respect to health resources available and associated costs. The latter objectives are aimed at describing and interpreting the causal mechanisms underlying morbid states to provide indications regarding the possibilities to intervene in the processes themselves. For instance, an important element in microanalysis is the exact determination of the *moment of onset* of a disease for an individual, an occurrence that often precedes the first manifestations of the process and its diagnosis by a long period (Fig. 47–1).

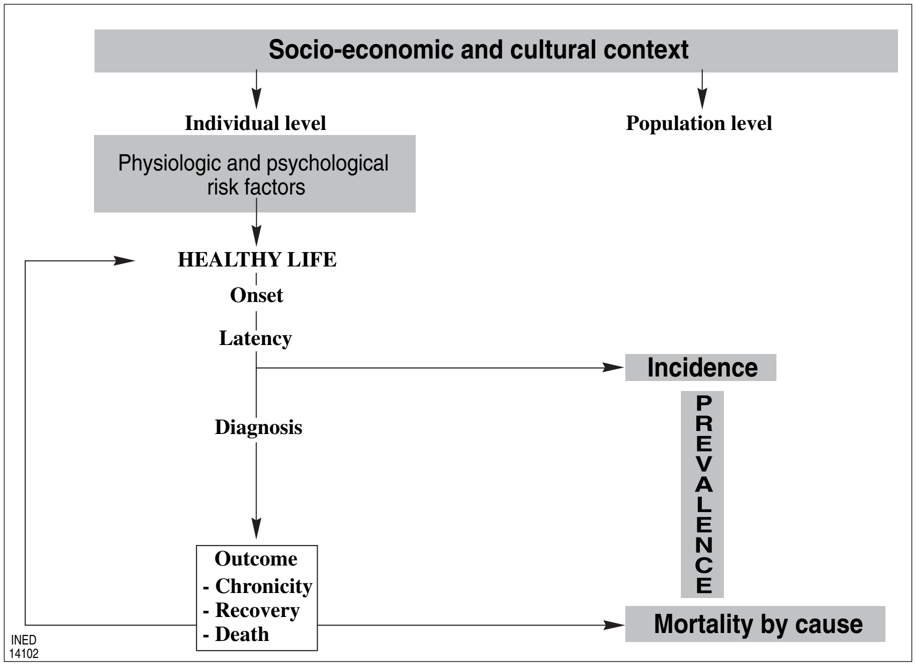


FIGURE 47–1 The morbidity process: context, individual events, and statistical measures.

This element cannot be overlooked, but it is less critical in macroanalysis, especially concerning the economic and social impact of the disease, as the period during which an individual is affected by a disease but is not conscious of it and is not harmed can be ignored because it does not exert any particular demands on society. It becomes important to monitor the latent period with greater accuracy by macroanalysis when the need is to evaluate dynamics of morbidity over time or differentiate its effects between population groups. The earlier age of contraction of a disease and an increase of its duration can be explained by an improvement in diagnostic capabilities and by an increase in health education, which today allows for the identification of a disease much sooner than in the past and in a differential manner in the population. *Recovery*, the transition from a state of disease to that of health, poses considerable problems when determining the exact point in time because it is difficult to establish when a patient ceases to be sick and returns to a healthy state. In this case, however, the macroanalysis approach allows for simplifications that can be used for estimating purposes. For instance, the survivors of a cohort of patients, all of whom contracted the disease at the same age and therefore, at a certain date, exhibit the same duration of the disease, can be considered "statistically recovered" when their relative survival curve becomes parallel to the axis of duration, expressing the nullification of the excess death risk of patients when compared with the general population.

a. Models of Disease-Death Processes

In an attempt to govern the complexity of disease-death processes in the population, models are designed to monitor the numerous elements coming into play. A set of *states*, each referring to a particular state of health, and *transitions* between states, regulated by specific transition probabilities, are assumed to represent the aging process of a cohort exposed to the risk of becoming ill and dying. The two models described subsequently illustrate the logic used, and reference is made to them to illustrate the procedure for constructing morbidity-mortality tables. Before examining the assumptions and characteristics underlying each model, it is useful to briefly discuss the symbols used:

$\alpha(x)$ is the overall mortality rate at age x in the general population.

$\gamma(x)$ is the cause-specific mortality rate at age x in the general population.

$\mu(x)$ is the hazard rate to contract the disease at age x for healthy people (i.e., the incidence rate).

$\lambda(x,\tau)$ is the recovery rate at age x for an individual who contracted the disease at age τ.

$\beta(x,\tau)$ is the mortality rate at age x for an individual who contracted the disease at age τ.

$\delta(x,\tau)$ is the cause-specific mortality rate at age x for an individual who contracted the disease at age τ.

$\xi(x)$ is the mortality rate at age x from causes other than the specific disease in the general population.

When a number of morbidity processes affect the population, rates $\gamma_i(x)$, $\mu_i(x)$, $\lambda_i(x,\tau)$, $\beta_i(x,\tau)$, and $\delta_i(x,\tau)$ are specified with respect to disease i, and $\beta_{i,j}(x,\tau,\rho)$ is the death hazard rate at age x for those having contracted disease i at age τ and disease j at age ρ.

Model A (Fig. 47–2) represents the simplest and most frequently used case, in which the process of elimination of the cohort is described by means of disease-death process and a death risk due to all *other causes* than the specific disease, acting in an instantaneous manner on the healthy and diseased population. A healthy individual at age x is exposed to the risk of falling ill or dying from other causes. Only when an individual is ill is he exposed to the death hazard from the specific cause.

Model B (Fig. 47–3) represents a more complex situation in which the process of elimination of the cohort is broken down by two disease-death processes and an instantaneous risk from other causes. The individual is exposed to the death hazard due to a specific cause only in the case that he is already ill and suffering from the corresponding disease.

The complexity of Figure 47–3 emphasizes the difficult conceptual and operative points of view in generalizing morbidity models to monitor a greater number of diseases. For n diseases, 2^n life states and $n+1$ death states must be contemplated together with all of the corresponding probabilities of transition between them.

Moreover, with regard to the availability of basic information, numerous problems exist. Only for a few illnesses, characterized by high social costs, have research instruments been set up that enable the estimation of incidence, recovery, and lethality rates. Many countries now have registers in which information related to tumors or serious illnesses of the circulatory system is collected. In just a few of these nations, the registers cover the entire national population; most refer to specific subpopulations, often geographically bounded. It is almost impossible to obtain the information required to estimate the probabilities of falling ill from a specific illness in the presence of others, the cause-specific mortality and recovery rates for those that have caught more than one disease. For this reason, it is necessary to introduce a hypothesis, often unrealistic (see Chapter 44), of independence between illnesses, whether it is during the initial stage, the subsequent stages, or the result of the morbidity processes.

IV. RELATIONSHIPS THAT LINK INCIDENCE, PREVALENCE, AND MORTALITY IN A COHORT

The overall incidence, prevalence, and mortality rates for the general population and for patients within the same cohort are very closely linked. In the case of incurable diseases or those having a high fatality rate, for which the probability of the patient returning to a healthy state can be hypothesized as being null

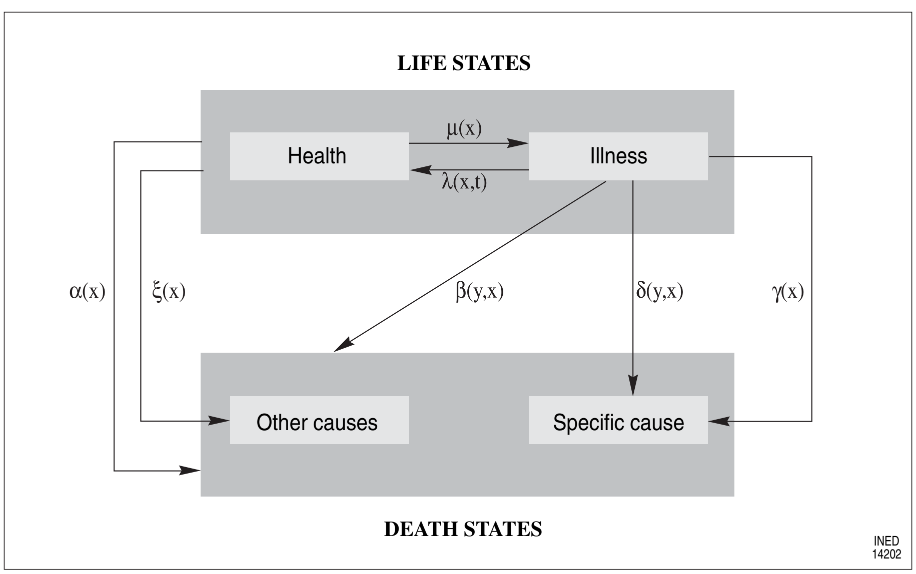


FIGURE 47–2 Model A: one disease-death process and one cause of death.

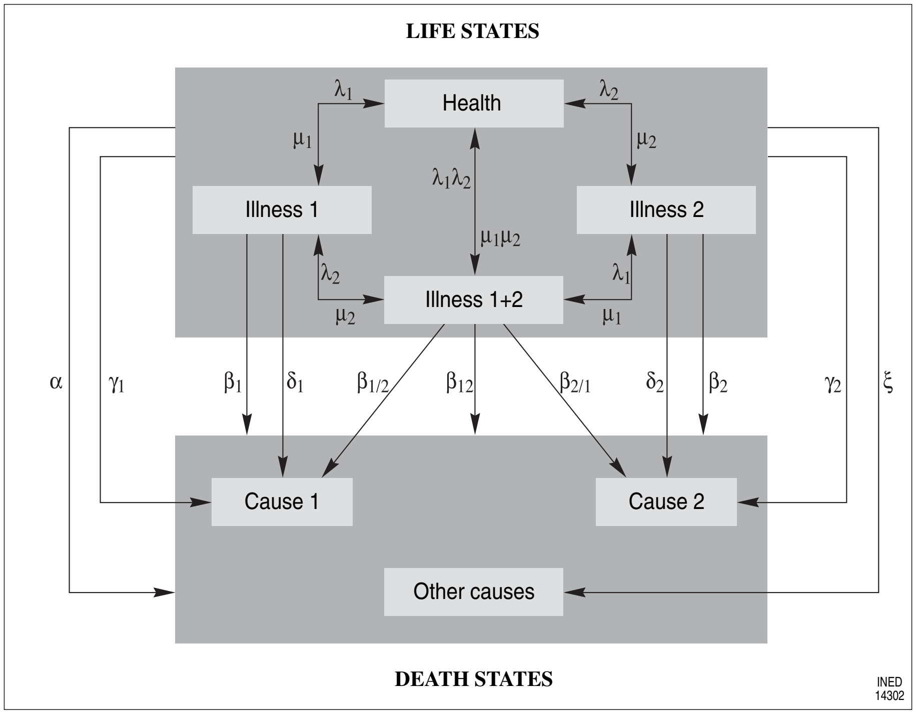


FIGURE 47–3 Model B: two disease-death processes and one cause of death.

or extremely low, knowledge of the relative survival curve for patients with respect to the general population allows the transition from incidence to prevalence to be appraised in a relatively simple manner (Verdecchia *et al*., 1989). With reference to the symbols used in Figure 47–2, the following can be said:

$$\nu(\mathrm{x}) = \int_0^x [1-\nu(\tau)]\mu(\tau)\mathrm{s}(\mathrm{x},\tau)\mathrm{d}\ \tau \qquad (1)$$

In Equation 1, for a selected disease,

$\mu(\tau)$ has been previously defined.
$\nu(x)$ is the prevalence rate at age x (or probability of being sick at age x, conditional on survival until age x).
$s(x,\tau)$ is the probability for an individual, having contracted the disease at age τ, of surviving the additional risk of death between ages τ and x. This is the relative survival function, which is the probability of survival up until age x of an individual who contracted the disease at age τ, $\beta(x,\tau)$, related to the same probability of survival of the general population, $\alpha(x)$.

This information can be expressed as follows:

$$\mathrm{s}(\mathrm{x},\tau) = \exp\left\{-\int_\tau^x [\beta(\mathrm{u},\tau)-\alpha(\mathrm{u})]\,\mathrm{du}\right\}$$

Another important relationship links the instantaneous mortality rate from a specific cause of death in the general population to the incidence, the prevalence, and the death hazards in the general population and in patients:

$$\gamma(\mathrm{x}) = \int_0^x [1-\nu(\tau)]\mu(\tau)\mathrm{s}(\tau,\mathrm{x})\delta(\mathrm{x},\tau)\mathrm{d}\ \tau \qquad (2)$$

The previous equations completely describe the morbidity-mortality process in a cohort, and knowing $\gamma(x)$, $\alpha(x)$, $\delta(x,y)$, and $\beta(x,y)$, they can also be used to estimate the incidence and prevalence of the disease on the basis of complete cohort mortality data. (Verdecchia *et al*., 1989).

The principal assumption of these relationships is the irreversibility of the morbidity processes. However, for practical applications, this assumption is acceptable for many chronic degenerative diseases, reflecting the idea that a patient with a history of a severe disease (e.g., cancer, ischemic heart disease) is expected to make greater use of health resources for the rest of his or her life.

It is evident that these equations are valid within a same birth-cohort. In a population, the past dynamics of the elements in play and the structure of the population, both general and the sick, would prevent the establishment of such precise relationships, except under the hypotheses of population stationarity and risks of illness and death. Despite this, the sense of relationship remains, as is clear from Table 47–1, created from data of incidence, prevalence, and

TABLE 47–1 Incidence and Prevalence Rates Estimated from Patient's and Population Mortality in Italy, 1990

Cancer sites	Males				Females			
	Relative survival[a]	Mortality by cause	Incidence	Prevalence	Relative survival[a]	Mortality by cause	Incidence	Prevalence
Lung	0.08	87.8	105.1	135.4	0.12	13.8	15.6	18.9
Larynx	0.65	8.2	19.3	187.5	0.71	0.4	0.9	10.1
Colon-rectum	0.38	29.0	61.5	237.3	0.40	24.7	53.3	250.0
Stomach	0.19	26.5	35.1	86.7	0.21	16.0	21.7	62.7
Breast					0.72	34.1	112.2	280.9

[a] Crude rates per 100,000 at 5 years since diagnosis.
From Berrino F. *et al*., 1995.

mortality in Italy for the years 1970 and 1990. The wide gap between incidence and prevalence characterizing cancers with the highest relative survival rate is evident. For cancer of the larynx, which allows for survival after 5 years of about 70% for women and 65% for men, the ratio between incidence and prevalence is 1 : 10. For lung cancer, the ratio is very close to 1 (1 : 1.3).

V. SPECIFICATION OF SURVIVAL ACCORDING TO STATE OF HEALTH: THE MORBIDITY-MORTALITY TABLE

Among the synthetic indicators summarizing morbidity that have obtained most success in recent years are life expectancies specified with respect to state of health which integrate information on mortality and morbidity. The *health expectancies*, as these indicators are generally called, can be referred to different concepts and definitions of health, and they can be calculated by various methods. The *perceived life expectancies* are very widely used and constitute a family of measures that use data on perceived health status, the best known of which is the *healthy life expectancy*, namely, the average number of years an individual is expected to live in a state of health defined as good or very good if current patterns of mortality and self-rated morbidity continue to apply. On the basis of the objective approach adopted here, the *disease-free life expectancies* can be estimated; this term refers to the average number of years an individual is expected to live free from one or more specific diseases.

The estimate of these indicators is based on specific morbidity-mortality tables that act as an integrating mechanism of information derived from several sources and studies regarding the morbidity and mortality of the population, including them in a model based on a cohort and life course perspective such as that presented in Figures 47–2 and 47–3. Substantially, the tables are made up of a set of life table survival curves related to the different morbidity processes being examined. When longitudinal data are available, the model can be used as a model of survivorship to describe the elimination process of a cohort of healthy individuals due to disease and death. When longitudinal data are not available—and this is the most frequent case—an assumption of stationary population dynamics can be adopted to construct a period morbidity-mortality table based on the disease and death risks observed in a certain period, which is a different and complementary way of evaluating the impact of a disease on a given population. Like a traditional period life table, the period morbidity-mortality table reflects the cross-sectional experience of individuals who belong to different cohorts. The table makes it possible to obtain measures summarizing morbidity and mortality of a population, to show the net effect of the age-structure, and to evaluate the impact of such morbidity on the quality of survival at that moment.

Comparisons of period morbidity-mortality life tables can show more rapidly and vividly period-specific effects, such as a better accuracy in diagnostic procedures or the development of new therapies; on the other hand, cohort influences (i.e., the cohorts' modification in life styles, smoking habits, or physical activity) are more clearly evident in comparisons of cohort morbidity-mortality tables.

For construction of morbidity-mortality tables, it is theoretically possible to use various methods that can generally be distinguished according to two main principles: prevalence methods and incidence methods. The conventional method, or Sullivan method, belongs to the former, whereas the multistate morbidity-mortality tables are an example of the latter.

1. The Sullivan Method

The Sullivan method (1971) is certainly the simplest and most widespread method for constructing a morbidity-mortality table and, as a result, of indicators of life expectancy by state of health. It can be applied in under the most diverse conditions because it requires only a life table of the population referring to a period in which prevalence rates by age are available (whether objective or subjective in nature). In this regard, it must be emphasized that this type of method is the only one that can be used when health data definitions are based on individual perception and on self-evaluation for which the basic information is represented by prevalence rates of states of health perceived as being "not good," gathered by means of interview surveys conducted directly with individuals themselves. In these cases, concepts such as disease-specific incidence and curve of survival of patients, necessary when using incidence methods, have no meaning at all.

The prevalence method is based on the distribution of the years lived from a life table, which is the $l(x)$ column of the table, into the different states of health in proportion to the rates of prevalence of the disease (or condition), $\nu_i^*(x,x + \Delta x)$, surveyed at a specific moment regarding a specific population, and belonging to a specific age bracket. The years spent in a state of illness (or with a condition) between the age x and $x + \Delta x$ can be calculated using the following equation:

$$L_i(x, x+\Delta x) = \int_x^{x+\Delta x} L_i(u)du = \nu_i^*(x, x+\Delta x) \times \int_x^{x+\Delta x} L(u)du$$

The years spent in a healthy state can be determined as follows:

$$L_h(x,x + \Delta x) = L(x,x + \Delta x) - L_i(x,x + \Delta x)$$

From the $L_i(x,x + \Delta x)$ term, it is easy to calculate synthetic measures such as *healthy life expectancy* or *disease-free life expectancy*. The parsimonious nature of the method and the extreme simplicity of the calculations necessary for the estimates are discounted by a rather strict system of assumptions (i.e., stationarity of risks of mortality and morbidity of the population) and by the inevitable approximation brought about by the fact of operating with prevalence rates (i.e., stock rates) that may, especially for processes with a shorter duration, greatly underestimate the real diffusion of the disease in the population and, as a consequence, the person-years spent in unhealthy conditions (Alho, 1992; Freeman and Hutchison, 1980; Keiding, 1991; Preston, 1987). Moreover, in the case of long-term pathologies, the hypothesis of stationarity is unacceptable because it relates to long periods due to high rates of survival. The cross-sectional prevalence data used by this method imposes serious limits due to the influence of the current states of morbidity and of those of previous years (Brouard and Robine, 1992).

2. Multistate Life Tables for Chronic Diseases

With incidence methods, the morbidity-mortality table is calculated following the logical evolution of the morbidity process from its onset to its final outcome, favorable or unfavorable, through competition with other morbidity processes and other risks of elimination to which the population is exposed. These methods make it possible to evaluate the consequences of the dynamics that in the past few years have affected the different aspects of the morbidity process: disease onset, its duration, and its lethality. The same logic on which these methods are based assumes the adoption of the definitions of objective health based on the absence of certain diseases or specific disabilities or handicaps for which it makes sense to assume a morbidity process that, arising or diagnosed at a specific moment, develops over the course of time toward the three possible outcomes (i.e., death, recovery, chronic state) (see Fig. 47–1).

One of the essential limits of application of these methods is imposed by the availability of adequate data. They assume the availability of very detailed information concerning the probability of transition of individuals between the various states of health, deducible only by means of continuous longitudinal surveys (e.g., follow-up studies, registers), and they are therefore difficult to have available except for certain important diseases and in countries that are very advanced in the field of collecting health information.

Let us suppose that the transition rates between all states comprising the model are available. The multistate multiple-decrement life table proposed by A. H. Pollard (1980) extends the concepts of the aggregate life table to cause-specific mortality and foresees that the healthy individual can leave this state for reasons of disease or death. This model, similar to that illustrated in Figure 47–1, consists of two double-decrement tables: Individuals remain in the disease-free group or depart from the disease-free group by contracting a disease or by dying from causes other than the disease under study. Those who contract disease cannot recover and remain ill until they die of the disease or die of other causes. Only the group of patients is exposed to the risk of death from the specific disease, whereas both of the groups are exposed to the same risk of death from all other causes of death (the process of disease-death and the death hazard from other causes are considered to be independent of each other).

This method of constructing the morbidity-mortality table undoubtedly constitutes a considerable step forward compared with the previous method based on prevalence rates, even though it does not take into account the possibility of recovery from the disease and return to a healthy state. However, this possibility is considered in the multistate increment-decrement model that provides for the monitoring of all possible states of transition when entering and leaving the different states (Duchêne 2002; Pollard J. H. 1990; Rogers 1989; Haberman 1983, 1984; Golini and Egidi, 1982; Chiang 1968). In these models, modifications of the state of health are represented by means of the transitions between the various states regulated by specific probabilities. The probability each individual has to transit from one state to another in a determined period is assumed to be independent from the transitions made previously and is conditioned by the last state in which the individual was found. These processes are principally based on Markov chain models to calculate transition matrices describing all transitions between states that may occur throughout an individual's life span.

To consider a greater number of interfering variables, methods for estimating probabilities of transitions are becoming more and more sophisticated. For example, to take into account the duration of the disease (i.e., the period of stay in a given state), semi-markovian models have been developed.

The incidence methods make it possible to construct a morbidity-mortality table and to estimate all of the analytic and synthetic variables and indices associated with it. To explain the relations existing within the morbidity-mortality tables, reference is made to Model A and Model B and to their respective symbols, assuming for simplicity the irreversibility of morbidity processes ($\lambda_i = 0$, for all i). It is assumed that mortality rates from other causes are independent of the morbidity process and that both the healthy and the ill are exposed to the same risk $\xi(x)$ of death from other causes. In Model A, this relationship is calculated as follows:

$$\xi(x) = \alpha(x) - \gamma(x)$$

In Model B,

$$\xi(x) = \alpha(x) - \gamma_1(x) - \gamma_2(x).$$

a. Model A

The set of survivors, $l(x)$, in the morbidity-mortality table at each age is obtained as the sum of the healthy survivors, $l_h(x)$, and diseased survivors, $l_i(x)$:

$$l(x) = l_h(x) + l_i(x)$$

In this case,

$l_h(x) = l_h(0) \times \exp\left\{-\int_0^x [\mu(u) + \xi(u)]du\right\}$ are the healthy survivors at age x

$l_i(x,\tau) = l_h(\tau) \times \mu(\tau) \times \exp\left\{-\int_\tau^x \beta(u,\tau)du\right\}$ are the patients who, having contracted the disease at age τ, are still alive at age x

$l_i(x) = \int_0^x l_i(x,\tau)d\tau$ are the diseased survivors at age x, whatever the duration of the disease.

The morbidity-mortality table therefore provides the structure of the population by state of health (healthy or diseased) for each age. On the basis of the survivors specified according to the state of health, the specific life expectancies can be calculated. This also makes it possible to assess the healthy survivors to obtain two other interesting indicators: the life expectancy and the disease-free life expectancy for healthy individuals at age x. The first indicator is estimated using the following equation:

$$e_h^T(x) = \frac{\int_x^\omega l_h^T(u)du}{l_h(x)} \quad \text{where}$$

$$l_h^T(u) = l_h(x) \times \exp\left\{-\int_x^u \alpha(s)d(s)\right\}$$

The second is obtained as follows:

$$e_h^h(x) = \frac{\int_x^\omega l_h(u)du}{l_h(x)} = e_h(x) \times \frac{l(x)}{l_h(x)}$$

In the previous equation, $e_h(x)$ is the disease-free life expectancy at age x.

Table 47–2 shows, as an example, total and "disease-free" life expectancy for two important diseases, cancer and major coronary events. Morbidity-mortality tables, based on Model A, are elaborated for each disease. Because Italy does not have incidence data, incidence was estimated by means of Equations 1 and 2 applied to overall and cause-specific cohort mortality data and patients' relative survival observed in existing local registers.

The increase in life expectancy seen in Italy between 1970 and 1990 of 5.3 years for men and 6.2 years for women should be scaled down considering the cancer-free life expectancy of 4.3 and 4.9 years, respectively.

TABLE 47–2 Life Expectancy, Cancer-Free Life Expectancy, and Major Coronary Events–Free Life Expectancy in Italy, 1970 and 1990: Model A

Age	Years	Males			Females		
		Total life expectancy	Life expectancy free from		Total life expectancy	Life expectancy free from	
			Cancer	MCE		Cancer	MCE
Birth	1970	68.8	67.4	67.5	74.5	72.3	74.0
	1990	74.1	71.7	73.1	80.7	77.2	80.5
65 years	1970	13.3	12.1	12.2	16.0	14.3	15.5
	1990	15.5	13.4	14.7	19.3	16.7	19.1
80 years	1970	5.8	5.0	5.0	6.6	5.6	6.2
	1990	7.4	5.8	6.8	8.6	7.0	8.5

MCE, major coronary event.

Moreover, the life expectancy advantage of women is considerably reduced: of the 6.6 years that on average a woman lives more than a man, only 5.5 years are cancer free. On the contrary, the dynamics over the last 20 years regarding major coronary events has been favorable, and the increase in disease-free life expectancy is greater than the overall value (i.e., 5.6 years for men and 6.5 years for women). At 80 years of age, as much as 92% of the additional 5.8 years of life expectancy of men is free of major coronary events (99% for women), whereas only 78% (81% for women) is free of cancer.

b. Model B

The set of survivors in the table is obtained from the sum of the survivors generated by four distinct processes:

1. Healthy survivors, $l_h(x)$
2. Patients with disease 1, $l_{i1}(x)$
3. Patients with disease 2, $l_{i2}(x)$
4. Patients with both diseases, $l_{i1,2}(x)$

These terms are added as follows:

$$l(x) = l_h(x) + l_{i1}(x) + l_{i2}(x) + l_{i1,2}(x)$$

From the previous equation,

$$l_h(x) = l_h(0) \times \exp\left\{-\int_0^x [\mu_1(u) + \mu_2(u) + \xi(u)]du\right\}$$

$$l_{i1}(x,\tau) = l_h(\tau) \times \mu_1(\tau) \times \exp\left\{-\int_\tau^x [\beta_1(u,\tau) + \mu_2(u)]du\right\}$$

$$l_{i2}(x,\rho) = l_h(\rho) \times \mu_1(\rho) \times \exp\left\{-\int_\rho^x [\beta_2(u,\rho) + \mu_1(u)]du\right\}$$

$$l_{i1,2}(x,\tau,\rho) = l_{i1}(\rho,\tau) \times \mu_2(\rho) \times \exp\left\{-\int_\rho^x [\beta_{1,2}(u,\tau,\rho)]du\right\}, \quad \text{if} \quad \tau < \rho$$

$$l_{i1,2}(x,\tau,\rho) = l_{i2}(\tau,\rho) \times \mu_1(\tau) \times \exp\left\{-\int_\tau^x [\beta_{1,2}(u,\tau,\rho)]du\right\}, \quad \text{if} \quad \rho < \tau$$

In the previous equations, $\tau(x)$ and $\rho(x)$ are, respectively, the age at onset of disease 1 and of disease 2. Details of the formulas are not given, but they are nothing more than extensions of those developed in Model A.

A worthwhile comment can be made with regard to the example shown in Table 47–3 regarding life expectancy for the overall and healthy-only Italian population for 1970 and 1990 calculated using this model. When considering the two morbidity processes (i.e., cancer and major coronary events) acting together, disease-free survival is reduced to 70.8 years for men and to 77.0 years for women, a lower number of years than estimated separately for the two processes (see Table 47–2). By reaching 65 years of age in a healthy state, a man's life expectancy increases from 15.5 years to 17.3 years (from 19.3 years to 21.1 years for a woman). For a woman 80 years old, who can look forward to a further 8.6 years, being in a healthy state leads to a life expectancy of more than 10 years, and of these, as many as 8.4 are free of cancer and major coronary events.

3. Other Indicators Derived from Morbidity-Mortality Tables

Morbidity-mortality tables constitute an extremely important tool for the evaluation of the impact of incidence and prevalence of specific diseases or states of health perceived as "not good" on the quality of survival. Besides the calculation of the traditional summary index representing life expectancy by state

TABLE 47–3 Life Expectancy and Cancer and Major Coronary Events–Free Life Expectancy Calculated for Total and Healthy Individuals in Italy, 1970 and 1990: Model B

Age	Male life expectancy				Female life expectancy			
	l(x)		$l_h(x)$		l(x)		$l_h(x)$	
	Total e	Disease-free e_h	Total e_h^T	Disease-free e_h^h	Total e	Disease-free $_h$	Total e_h^T	Disease-free e_h^h
Birth								
1970	68.8	66.2	68.8	66.2	74.5	71.8	74.5	71.8
1990	74.1	70.8	74.1	70.8	80.7	77.0	80.7	77.0
65 years								
1970	13.3	11.1	14.7	12.3	16.0	13.8	17.3	15.0
1990	15.5	12.5	17.3	13.4	19.3	16.4	21.1	18
80 years								
1970	5.8	4.3	7.5	5.6	6.6	5.3	8.0	6.4
1990	7.4	5.3	9.9	7.2	8.6	6.9	10.4	8.4

of health, other indicators such as the mean duration of disease, mean age of disease onset, and the probability of being "healthy" at a given age, can be calculated. Variables such as the structure of patients by age and duration of disease and net of the age structure of the population can be considered as useful information, which is especially necessary to better guide actions of health policies and to evaluate their impact on health conditions of the population.

A morbidity measure of particular interest when considering specific diseases is the probability that an individual will reach a specific age x without contracting the specific disease:

$$p_h(x) = \frac{l_h(x)}{l(0)} = \exp\left\{-\int_0^x \mu(u) + \xi(u)du\right\}$$

Another is the disease-free life expectancy at birth:

$$e_h(0) = \int_0^\omega \frac{l_h(x)}{l(0)}dx$$

Another measure is the probability of survival after a given period of duration of the disease, x, for an individual who has been diagnosed at age τ (i.e., the proportion of cohort members who survive all risks of mortality from diagnosis to duration x):

$$_x p_\tau^i(\tau + x) = \frac{l_i(\tau + x, \tau)}{l_i(\tau)} = \exp\left\{-\int_\tau^{\tau+x} \beta(u,\tau)du\right\}$$

The case fatality-rate is the proportion of the cohort of individuals who have been diagnosed at age τ who succumb to the disease as an underlying cause of death:

$$CF_\tau = \frac{\int_\tau^w l_i(u,\tau)\delta(u,\tau)du}{l_i(\tau,\tau)}$$

These relations are verified within a cohort but may be generalized for a population through a suitable system of hypotheses and the availability of information collected on two successive occasions that permits estimation of growth rates in the population with or without a disease by duration since diagnosis or by age (Preston, 1987).

When measures of prevalence according to age are available (certainly the most common case of availability of cross-sectional surveys), it is possible to estimate, for example, the probability that an individual will reach a specific age x without contracting a specific disease:

$$p_h^*(x) = \frac{H(x)}{B} \times \exp\left\{\int_0^x r^H(j)dj\right\}$$

In the equation, $H(x)$ is the number of individuals who are free from a specific disease at age x; B is the number of births, and $r^H(j)$ is the growth rate of the number of persons of age j without the disease. In this case,

$$e_h^*(0) = \int_0^w \frac{H(x)}{B} \times \exp\left\{\int_0^x r^H(j)dj\right\}dx$$

This equation provides the disease-free life expectancy at birth.

In a stationary population, duration-specific rates of growth in the number of cases must be zero, and the period probability of not developing the disease and the disease-free life can be derived directly from observed data on disease-free persons and on the annual number of births. When the population is not stationary, a simple growth correction is required. The growth adjustment, $r^H(j)$, can be estimated from two disease prevalence surveys.

CONCLUSIONS

Even the brief review of relationships between morbidity and mortality developed in this chapter demonstrates that there is a vast interpretative and analytic potential offered to demographic study by examining the morbidity process, besides the obvious benefits it affords to epidemiologic and clinical studies. It is necessary to delve deeper into this complex field, but continued study can provide a wealth of information, if one wishes to abandon the merely descriptive viewpoint of an analysis of mortality, of its evolution, of its differentiation between population groups, and the possibility of its further future reduction. Every phase of the process that precedes the event of death is influenced by endogenous and exogenous factors that condition its evolution, modifying the duration and outcome and influencing the consequences the disease produces on the autonomy of the individual and the possibility to conduct an acceptable life.

A fundamental problem for the development of these studies is in many countries still represented by the lack of availability of statistical data capable of accounting for the diverse and numerous aspects of the morbidity process, a lack that can be justified by the complexity of the instruments that must be used to correctly collect this type of information, from both theoretical and conceptual viewpoints. The diffidence shown for a long time by medical science with respect to the possibilities offered by analyses conducted on very large population groups and the similar reluctance shown by statisticians and demographers with regard to analyses of a predominantly qualitative nature have represented considerable obstacles to cooperation, without which it is difficult to make the qualitative jump that can lead to an improvement in understanding the complex individual and social mechanisms that influence the quality of human survival and the possibility of further extending the life span.

Much progress has been made in gathering adequate statistical information and in creating methodologic tools that are better adapted to representing the complexity of the phenomena. The road ahead is long in conceptual and methodologic terms. It involves the exploration of new dimensions in the state of health, the study of the relationships with lifestyles and changes in lifestyles, and the identification of interdependencies among the different diseases and conditions that may favorably influence the outcome.

References

ALHO Juha M., 1992. On prevalence, incidence, and duration in general stable populations. *Biometrics*, vol. 48, p. 587–592.

BERRINO Franco, SANT Milena, VERDECCHIA Arduino, CAPOCACCIA Riccardo, HAKULINEN Timo, and ESTÉVE Jacques, 1995. *Cancer survival in Europe*, IARC Scientific Publication No. 132. Lyon, International Agency for Research on Cancer.

BROUARD Nicolas and ROBINE Jean-Marie, 1992. A method for calculation of health expectancy applied to longitudinal surveys of the elderly, *in*: Jean Marie Robine, Madeleine Blanchet, and Edward J. Dowd (eds.), *Health expectancy*. London, OPCS, HMSO, p. 87–94.

CHIANG Ching Long, 1968. *Introduction to stochastic processes in biostatistics*. New York, John Wiley and Sons, 313 p.

DAY Lincoln H., 1991 Upper-age longevity in low mortality countries: a dissenting view, *in*: W. Lutz (ed.), *Future demographic trends in Europe and in North America*. New York, Academic Press, p. 117–144.

DUCHÊNE Josianne, 2002. The measurement of mortality by cause and of morbidity, *in*: Guillaume Wunsch, Michel Mouchart, and Josianne Duchêne (eds.), *The life table. Modelling survival and death*, p. 207–244. Dordrecht, Kluwer.

FREEMAN Jonathan and HUTCHISON George B., 1980. Prevalence, incidence and duration. *American Journal of Epidemiology*, 112(5), p. 707–723.

FRIES James F., 1980. Aging, natural death, and the compression of the morbidity. *New England Journal of Medicine*, vol. 303, p. 130–135.

FRIES James F., 1983. The compression of morbidity. *Milbank Quarterly*, 61(3), p. 397–419.

FRIES James F., 1987. An introduction to the compression of morbidity. *Gerentologica Perspecta*, vol. 1, p. 5–8.

FRIES James F., 1989. The compression of morbidity: near or far? *Milbank Quarterly*, 67(2), p. 208–232.

GOLINI Antonio and EGIDI Viviana, 1982. Effetti delle variazioni di morbosità sulla mortalità e sull'ammontare e la struttura di una popolazione. *Statistica*, vol. 42(4), p. 479–512.

GRUENBERG Ernest M., 1977. The failure of success. *Milbank Memorial Foundation Quarterly/Health and Society*, vol. 55, p. 3–24.

GRUENBERG Ernest M., 1980. Epidemiology of senile dementia, *Advances in Neurology*, vol. 19, p. 437–457.

HABERMAN Steven, 1983. Decrement tables and the measurement of morbidity: I., *Journal of the Institute of Actuaries*, vol. 110, p. 361–381.

HABERMAN Steven, 1984. Decrement tables and the measurement of morbidity: II, *Journal of the Institute of Actuaries*, vol. 111, p. 73–86.

HOEM Jan M. and JENSEN Ulla Funck, 1981. Multistate life table methodology: a probabilistic critique. Working paper no. 8. Copenhagen, Laboratory of Actuarial Mathematics, University of Copenhagen, 98 p.

KEIDING Niels, 1991. Age-specific incidence and prevalence: a statistical perspective, *Journal of the Royal Statistical Society A*, vol. 154, Part 3, p. 371–412.

KRAMER Michael, 1980. The rising pandemic of mental disorders and associated chronic diseases and disabilities. *Psychiatria Scandinavica* 62, Suppl. 285, p. 282–297.

Manton Kenneth G., 1982. Chancing concepts of morbidity and mortality in the elderly population, *Milbank Memorial Foundation Quarterly/Health and Society*, vol. 60(2), p. 183–244.

Olshansky S. Jay, 1987. The compression of mortality and morbidity: comments on debate, *Gerontologica Perspecta* vol. 1, p. 19–23.

Olshansky S. Jay, Rudberg Mark A., Carnes Bruce A., Cassel Christine K., and Brody Jacob, 1991. Trading off longer for worsening health. The expansion of morbidity hypothesis, *Journal of Aging and Health* vol. 3(2), p. 194–216.

Pollard Alfred H., 1980. The interaction between morbidity and mortality, *Journal of the Institute of Actuaries*, vol. 107(3), p. 233–302, summary of discussion, p. 303–313.

Pollard John H., 1990. Morbidity and mortality: differentials and their trends, *European Journal of Population*, vol. 6(3), p. 205–225.

Preston Samuel H., 1987. Relations among standard epidemiologic measures in a population, *American Journal of Epidemiology*, vol. 126(2), p. 336–345.

Robine Jean-Marie and Colvez Alain, 1991. Quelle espérance pour quelle vie? *Futuribles*, vol. 155, p. 72–76.

Rogers Richard G., Rogers Andrei, and Belanger Alain 1989. Active life among the elderly in the United States: multistate life-table estimates and population projections, *Milbank Quarterly*, vol. 67(3–4), p. 370–411.

Sullivan Daniel F., 1971. A single index of mortality and morbidity, *HSMHA Health Reports* 86, p. 347–354.

Verdecchia Arduino, Capocaccia Riccardo, Egidi Viviana, and Golini Antonio, 1989. A method for the estimation of chronic disease morbidity and trends from mortality data, *Statistics in Medicine*, vol. 8, p. 201–216.

P A R T

II

ENDOGENOUS AND EXOGENOUS MORTALITY

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

The process resulting in death has a wide range of beginnings. It can be posited that the human organism has an inherently finite life span and that each individual is biologically predestined to die once his or her time has come. However, the time of death may be precipitated or delayed by external events or by human behavior itself. Life and death are the result of a subtle balance between endogenous and exogenous factors, which the next five chapters will endeavor to take stock of.

Jean Bourgeois-Pichat was among the earliest demographers in the early 1950s to differentiate endogenous and exogenous causes of death. The former were hereditary characteristics and so were not readily curable, whereas the latter were environmental and could be acted on. The concepts and approaches have evolved greatly since then, and the changes are reviewed by Jacques Vallin and Giovanni Berlinguer in Chapter 48.

Endogenous mortality is grounded on genetic makeup. In Chapter 49, Lamberto Soliani and Enzo Lucchetti outline the main genetic factors that predispose to good or ill health, reviewing the genetic diseases that may lead to below-average life spans and discussing what can be done about them.

Exogenous causes of mortality can be broadly categorized as those attributable to natural environmental threats, possibly as a result of human action, and those resulting from individuals' behavior in response to such threats or to changes in their own bodies. Francis Sartor reviews environmental factors in Chapter 50, and Jacques Vallin, Graziella Caselli, and Pierre Surault assess behaviors, lifestyles, and sociocultural factors in Chapter 51.

Modern medicine's efforts to deliver on public demands for care and prevention are examined by Paolo Vineis in Chapter 52.

CHAPTER

48

From Endogenous Mortality to the Maximum Human Life Span

JACQUES VALLIN AND GIOVANNI BERLINGUER

Institut national d'études démographiques (INED), Paris, France

Università degli Studi di Roma "La Sapienza," Rome, Italy

The medical approach to causes of death, a subject that has been extensively discussed in previous chapters and particularly so by France Meslé in Chapter 42, has always maintained a close relationship with diseases that physicians endeavor to cure but that can also lead to death. Consequently, the method of classifying causes of death has always been closely linked to the method of classifying diseases because of the need, as Charles Nam (1990) wrote, "to examine morbidity in conjunction to mortality and to become cognizant of morbid conditions contributing to the mortal state." It has been this way since antiquity. However, since the beginning, the idea of classifying causes of death has gone hand in hand with research into the fundamental principles common to diseases and causes of death. Often, this led to emphasis on dichotomous criteria such as "internal or external," "natural or violent," "essential or accidental," and "predisposed or dissociated." The supposed explanatory capacity of such fundamental distinctions, reflecting the ideas of the time, stimulated etiologic research, with one theory being replaced by another, but they have, by imposing presuppositions on a whole era, also occasionally led to some epistemologic roadblocks, restraining the progress of ideas. In this way, the miasmatic interpretation of numerous fever-type lethal syndromes confused the understanding of their microbial origin and even the concept of infectious disease until the latter half of the 19th century. In its turn, the germ theory, which after the work of Louis Pasteur was finally maintained at the beginning of the 20th century and which advanced the view that behind each disease was the *presence* of a microbe, delayed the onset of research on the explanation of disease due to the *absence* of certain agents (e.g., vitamin deficiency) or to other causes. The progress of etiopathogenic knowledge and its preventive or therapeutic properties has periodically challenged the successive inclinations to enclose causes of death within a simple principle of fundamental etiologic classification (Vallin and Meslé, 1988).

By echoing the historical evolution of demographic thinking in this area in this chapter, we are concentrating on the question of whether certain causes are inescapable because they are inherent to the nature of human beings, whereas others are avoidable, resulting from the action of external agents or from a deleterious controllable process. The main interest of such a distinction is to steer us toward reflection on the maximum human life span, a consideration that, far from reaching definite answers, is leading to a very open debate: Are we now, with a life expectancy of 80 years or more at birth, near to the end of the *health transition* (see Chapter 57) as biologists such as James Fries or demographers such as Jay Olshansky are claiming? Or on the contrary, can we expect to encounter new battle fronts against death that will in the future open the way to substantial progress of a different kind to what has preceded?

In an era when life expectancy at birth did not reach 30 years, Condorcet, in his splendid *Esquisse d'un Tableau Historique des Progrès de L'esprit Humain*, wrote the following:

> One feels that the progress made in medicine, having become more efficient thanks to the advance of reason and social order, must in time eliminate transmittable and contagious diseases, as well as general diseases due to climate, food and working conditions. This should logically be the case for all the other illnesses the distant causes of which we will one day know. Would it be absurd to suppose that this perfecting of the human race should be expected to go on indefinitely, that there must come a time when death would only be the result of either extraordinary accidents or of the progressively slower destruction of the vital forces and that finally the average length of time between birth and this destruction has no specifiable duration? There is no doubt that man will not become immortal; but is there no reason to believe that the interval between the moment when his life begins and the time when, naturally and without the intervention of an illness or an accident, he usually finds it difficult to go on living, continually increase? (1795, translation from French to English by Paul Belle).

Although absolutely prophetic, this vision nevertheless contained one fundamental contradiction: while rejecting in advance the idea that humans could one day attain immortality, this philosopher of the French Revolution imagined no less than that human life could "increase continuously." Current theories continue to confront this contradiction. Condorcet's text also contains the beginnings of the two fundamental hypotheses that separate immortality from the existence of an absolute limit to a human life. According to him, with progress, death could in effect only be the result of "exceptional accidents" or the slow destruction of "vital forces." If death only occurred because of exceptional accidents, it would open the way to immortality to a small number of individuals. If death only occurred in cases of accident and therefore was independent of age, it would strike only a constant fraction of each generation at random each year. It would suffice therefore that this fraction be low enough and the generation be large enough to enable at least a few individuals to live for an almost infinite period. If death occurred as a result of a slow degeneration of the vital forces, at the end of this process, all individuals would die. By slowing down the process, life could be prolonged, but without stopping this process completely, humans will never be immortal. Acknowledging the existence of an intrinsic process of degeneration is to acknowledge death as inevitable, resulting directly from the biologic aging of the organism, and the existence of an imperative maximum human life span, an intrinsic limit to the species.

The concept of *endogenous mortality*, as opposed to that of *exogenous mortality*, which results from external factors, brings together causes of death that are intrinsic to the organism but do not coincide entirely with the concept of biologic aging for two reasons. Even if the endogenous factors of mortality are conceived implicitly as those that command the degeneration of the vital forces (Rose, 1991; Kirkwood, 1990), random endogenous factors, independent of age and which provoke death while by-passing the aging process, can be imagined. Above all, an aging process that is not entirely endogenous but, on the contrary, results in an interaction between the human organism and its environment can also be imagined (Finch and Kirkwood, 2000).

From a demographic point of view, there are two real types of limits on the life span within a species: *maximum longevity* or the maximum age that even the most fortunate individual could not exceed, and the average age at which all individuals in the best possible health conditions die, which Jean Bourgeois-Pichat called *the maximum biologic life expectancy limit*. If this second limit depends on the first, it is not entirely reduced to maximum longevity because it is just as dependent on the distribution of individual life span this side of maximum longevity. In the assumption of biologic aging, a very simple model takes into account both these aspects of the maximum human life span: Gompertz's law. It is from this model that Jean Bourgeois-Pichat and Paul Vincent tried to measure the maximum human life span. However, today's debate is about the extent to which this law applies to the human species.

I. GOMPERTZ'S LAW: MEASUREMENT FOR BIOLOGIC AGING

1. The Force of Mortality and the Aging of the Human Organism

As early as 1825, Benjamin Gompertz put forward a formula for defining the probabilities of death according to "*the progressive deterioration of the human being*" with age (see Chapter 11). He estimated that the force of mortality, μ_x, followed an exponential function of age:

$$\mu_x = Bc^x$$

In the equation, B and c are constants. He saw in this mathematical model a description of "two generally co-existing causes; the one, chance, without previous disposition to death or deterioration; the other, a deterioration, or an increased inability to withstand destruction" (Gompertz, 1825, p. 517).

This formula does not require any preconceived idea on the maximum life span. If the values B and c are fixed, μ_x tends toward infinity, and it is easy to imagine that the chances for survival are minute beyond a certain age. Strictly speaking, when the force of mortality tends toward infinity, the probability of dying, ${}_1q_x$. tends toward 1, without ever completely reaching it. This therefore makes the survival of at least a few rare individuals beyond any limits possible in theory. Nevertheless, the exponential constraint is such that for this probability to have the slightest chance of occurring, it would have to be in the presence of a population of an immeasurable size, which means that for a given value of the B and c parameters, the model fixes an age limit on the survival of a generation. However, for this age to vary, new values for this parameter are all that are needed. In this sense, the model does not impose a precise limit. This adjustment is based on the idea that, in a given context, mortality is essentially an exponential function of age that translates into the idea that the human organism is doomed to age, which leads unavoidably to death.

2. The Gap between the Model and Reality: External Causes

Gompertz's model gives a good description of reality only beyond a certain age. This is why other authors have tried to refine Gompertz's model to bring it closer to real data. William Makeham (1860, 1867) believed that some causes of death were independent of age and proposed to extrapolate the force of mortality at old ages on the basis of a modified version of Gompertz's law:

$$\mu_x = A + Bc^x$$

In the equation, the parameter A represents the risk of dying for all causes that are independent of age (Thatcher *et al*., 1998).

The hypothesis of causes of death being independent from age is questionable, and numerous authors have tried to suggest other refinements to better account for the complexity of the existing relationship between risk of death and age (Duchêne, 1980). Adjustment of the mortality curve using Gompertz's formula always comes up against the fact that causes of death do not always vary in the same way with age and that certain causes do not follow the exponential schema at all.

Can it not be imagined that mortality is a combination of at least two components: a fundamental component, intrinsic to the process of degeneration (i.e., biologic aging), which the exponential model describes perfectly, and one or several complementary components of external causes that might or might not be linked to age but in a nonexponential way? Jean Bourgeois-Pichat volunteered to divide the endogenous and exogenous causes to attempt to answer this question.

II. JEAN BOURGEOIS-PICHAT: SEARCH FOR THE "HARD ROCK"

If statistics on deaths from sufficiently reliable causes are available, it should be possible to distinguish those related to degeneration of the organism and those related to external surroundings.[1] In measuring mortality due to the first causes, it should be possible to simultaneously obtain a measurement of the biologic limit of human life. Acknowledging the fact that the approach by declared causes, often subject to caution, was most frequently insufficient, especially with regard to infant mortality, Jean Bourgeois-Pichat also extended Gompertz's idea of a mathematical link between mortality and age to infants to specify the border between endogenous and exogenous factors.

1. From Endogenous Infant Mortality to the Maximum Biologic Life Table

a. Endogenous and Exogenous Infant Mortality

Bourgeois-Pichat's work initially revolved around the first year of life, establishing what he called *the biometric analysis of infant mortality* (Bourgeois-Pichat, 1951a, 1951b).

Some infant deaths clearly result from the physical constitution of the newborn (i.e., congenital deficiencies) or from conditions of childbirth, reputed to be unavoidable at that time (e.g., obstetric trauma); both are cases of endogenous mortality. Others result from external aggression or inadequate care (e.g., infectious diseases, malnutrition, violent deaths); these are cases of exogenous mortality. Using statistics from medical causes of death, it is possible to separate the two groups. Most frequently (and even more so before the 1950s), the results are not satisfactory because there is a great margin of uncertainty (undeclared or ill-defined causes) and because endogenous causes tended to be overestimated for a long time, especially during the period when the idea of congenital deficiency was very fashionable. This is why Jean Bourgeois-Pichat proposed to define the proportion of endogenous deaths in infant mortality using a mathematical adjustment of deaths by age. His method is based on a double hypothesis:

1. The (cumulated) distribution by age of exogenous deaths between 1 and 12 months is a simple mathematical function of age.
2. All endogenous deaths occur before the age of 1 month.

[1] This paragraph draws largely on a previous contribution to an INED publication (Vallin, 1993).

Several trials in countries with relatively reliable mortality statistics enabled the investigator to establish that the number of cumulated exogenous deaths (E_n) between birth and age n (in days) responded well to the following type of function:

$$E_n = b \cdot \log^3(n + 1)$$

This corroborates the first hypothesis.

He also established that the same function fits beyond the age of 1 month, except for one constant, the total D_n of deaths from all causes:

$$D_n = a + b \cdot \log^3(n + 1)$$

This verified the second hypothesis. Thereafter, the parameter a is no less than the number of deaths attributable to endogenous causes.

Transformed into a graph with an adequate age scale, $\log^3(n + 1)$, the cumulated number of deaths by age between 1 and 12 months easily lends itself to a line adjustment whose intersection with the axis of deaths gives the number of endogenous deaths (Fig. 48–1).

From the start, Jean Bourgeois-Pichat was aware that endogenous mortality defined in this way was not immutable. He ascertained that it was not the same everywhere and that it had a tendency to decline, although significantly slower than exogenous mortality. It was therefore not completely impervious to health improvement; it was merely much more rebellious. Whereas it was conceivable that exogenous mortality could be eradicated one day, it seemed to him impossible not come up against an endogenous life span limit that would be difficult to overcome. It is this notion of a limit that Jean Bourgeois-Pichat (1952) endeavored to specify in his work on infant mortality while widening his field of research to mortality as a whole.

b. The Maximum Biologic Life Table

Is it possible to establish a life table that in some way constitutes the insuperable limit, at least according to current medical knowledge? For infant mortality, this seemed to have been the case. In Norway, the country that was in the lead, the process of decline of endogenous mortality seemed to be nearing its end. To obtain an estimate of this near-insuperable limit, it was enough to extrapolate this tendency. In this way, Jean Bourgois-Pichat retained the limit of infant mortality rate at 13 per 1000 for boys and 9 per 1000 for girls. These limits have been easily surpassed in several countries since then, but at the time, in a country such as France, where the infant mortality rate was 50 per 1000, they seemed far in the future.

What about the other ages? Still drawing on data from Norway, a country where mortality was particularly low and the causes of death relatively well recorded, and by separating deaths that appeared to be essentially exogenous (i.e., infectious diseases, respiratory diseases, and accidents) from those which seemed essentially to occur with the normal wear of the organism (all other causes), Jean Bourgeois-Pichat ascertained that between 30 and 90 years, the quinquennial mortality rate of this second group adhered absolutely to Gompertz's law (i.e., a line on a semi-logarithmic graph). Knowing that it concerned a country

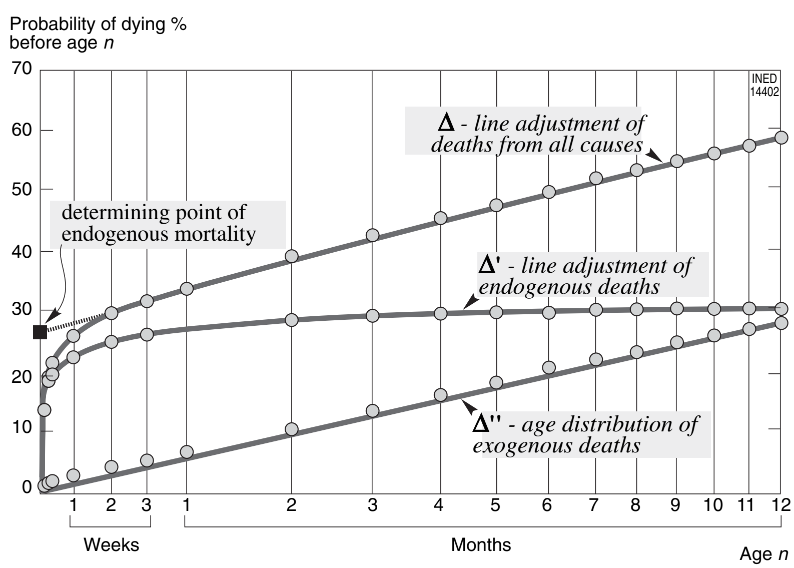


FIGURE 48–1 Separation of endogenous and exogenous infant mortality, according to Jean Bourgeois-Pichat (1951).

with the lowest rates, he assimilated these rates to the endogenous mortality limit at these ages, presuming that beyond the age of 30, noninfectious or nonaccidental mortality in that country reflected the natural consequences of the deterioration of the organism. Between 1 and 29 years, separating the endogenous from the exogenous declared causes of death seemed too risky to him, even in a country such as Norway, whereas it was very difficult to measure beyond 90 years, taking into account the risks of small numbers and problems with data quality. However, taking into account the extremely accurate observations between the ages of 30 and 90, it was enough to retropolate the line adjustment obtained for endogenous mortality downward for these ages to obtain the rate between 1 and 30 years and extrapolate the same upward, beyond 90 years, to complete the life table. At very old ages, for which all mortality is endogenous, he verified that his adjustment coincided with the results obtained by Paul Vincent (1951) on the mortality of the very old.

In terms of life expectancy, the limit fixed in this way was 76.3 for men and 78.2 for women. One-half century later, these figures may make us smile because this theoretical maximum for women has already been exceeded in several countries, with female life expectancies superior to 80 years and the figure for men now on the way to exceeding his estimate. Nevertheless, at the time, this left a wide margin for future progress, because in France, for example, according to the most recent life table (1946 to 1949), French women could hope to gain 10.8 years and French men to gain 14.4 years, according to this "law."

To get across the significance of his research, Jean Bourgeois-Pichat (1952) concluded his *Essay on the Maximum Biological Human Life Span* with a geologic comparison. One can, he wrote "compare mortality to a piece of ground composed of two unequally friable rocks. The process of erosion takes place to begin with, on the more tender, leaving bare the profile of the hard rock."[2]

2. The Hard Rock Is Not Unassailable

The hard rock is not unassailable. Neither are the methods.

a. Biometric Analysis of Infant Mortality Confronted with the Facts

Paradoxically, the contradictions exposed in Bourgeois-Pichat's law by the facts are without doubt, the most interesting contributions of this "law," which is now obsolete. It is these exceptions that have given meaning to a rule, which has been inept at predicting the limit of infant mortality decline.

Late Exogenous Excess Mortality

As early as 1957, Jean-Noël Biraben and Louis Henry showed that in some Mediterranean populations, the graphical adjustments proposed by Jean Bourgeois-Pichat could not be obtained with one single line, but with two. The points corresponding to the first months aligned themselves with a lower slope than the points corresponding to the last months. They attributed this contradiction to the Bourgeois-Pichat law to a particularity in the health situation of the Mediterranean countries. Because milk is more difficult to preserve in hot temperatures than in cool ones, infant excess mortality from diarrhea after weaning occurred during the summer. In other words, exogenous mortality increased abruptly after weaning, thereby breaking Bourgeois-Pichat's line at around the third month (Biraben and Henry 1957).

The same phenomenon was observed in Senegal by Pierre Cantrelle a short time later, thanks to the creation of one of the first *demographic surveillance systems*[3] in the districts of Niakhar and Paos-Koto (Cantrelle, 1969). This time, because weaning occurred later, the break in the alignment occurred after the sixth month. Excess exogenous mortality was prolonged to beyond the first birthday, and the risk of death was still very high during the second year. This phenomenon, confirmed by surveys of seasonal mortality (Cantrelle and Leridon, 1969), is now well known and characterizes the profile of mortality in several developing countries where infant and child mortality is, at equal levels of life expectancy, very much higher than in the European-type schema. The exception to Bourgeois-Pichat's law is particularly interesting because it provided an improved interpretation of observations that up to then tended to be held in doubt because of the poor quality of the data.

Better still is the fact that in this type of situation the respective proportions of *normal exogenous mortality* and *late excess exogenous mortality* essentially linked to nutritional problems can be measured through the break in alignment shown by comparing the two lines of adjustment. This characteristic was revealed to be particularly useful in the analysis of infant mortality by sex in some countries. In Algeria in 1969 to 1970, the infant mortality rate was 142 per 1000 boys and 141

[2] "Comparer la mortalité à un terrain composé de deux roches inégalement friables. L'action de l'érosion s'exerce d'abord sur la plus tendre, mettant à nu le profil de la roche dure."

[3] This is one of the terms used for permanent observation systems in which births, deaths, migrations etc within a population are continuously recorded after an initial survey (see Chapters 121 and 124).

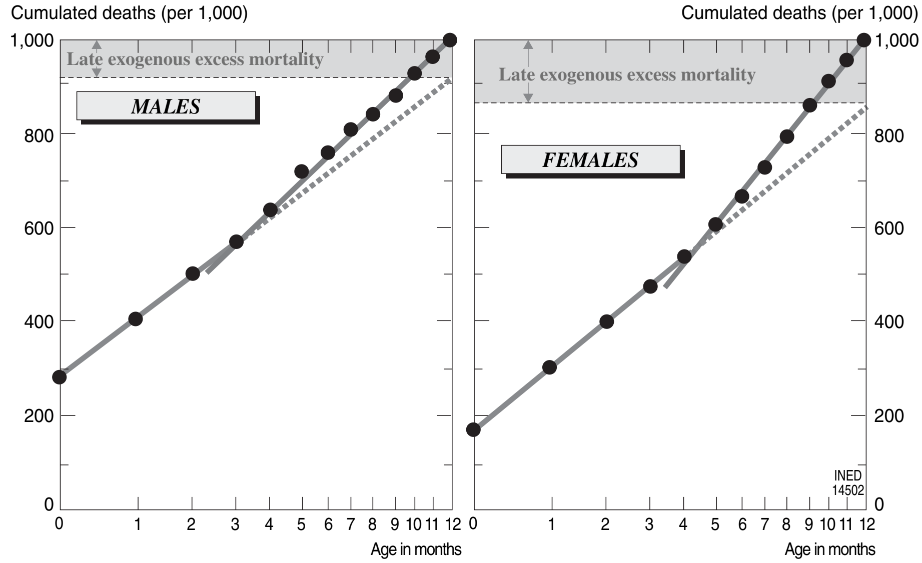


FIGURE 48–2 Biometrical analysis of child mortality: Algeria, 1969 to 1970. (From Vallin, 1975.)

per 1000 girls. Meanwhile, thanks to the Bourgeois-Pichat method, it was possible to establish that endogenous mortality was 40 and 24 per 1000, respectively. The excess male mortality was 67% while normal exogenous mortality was 89 per 1000 and 97 per 1000, and the surplus exogenous mortality associated with weaning was 13 and 20 per 1000, and the excess female mortality was 10% to 50% (Fig. 48–2). In this way, it became apparent that excess mortality in small girls, observed beyond 1 year, took effect from the first year and that it was entirely exogenous and affected particularly by late excess exogenous mortality linked to weaning and therefore to family behavior (Vallin, 1975).

Early Exogenous Excess Mortality

Other types of exceptions to the Bourgeois-Pichat law have been revealed. For example, for ancient Europe, Louis Henry and Yves Blayo observed a different break in the alignment, the adjustment line of the first points having a stronger slope than the line adjusting the last points, here again, the exception is interesting as it allowed the researchers to put forward a hypothesis of early exogenous excess mortality associated with certain respiratory diseases, particularly active in winter (Henry and Blayo, 1967). This same phenomenon was observed in France in certain administrative regions (particularly in Normandy) between the two world wars (Prioux and Vallin, 1976).

These two types of exceptions, whether they concern early or late exogenous excess mortality, only affect the principle of alignment of exogenous deaths, Jean Bourgeois-Pichat's first hypothesis. The second hypothesis (i.e., no endogenous deaths beyond 1 month) remains intact, and the principle of determining endogenous mortality by a (wisely chosen) adjustment line persists.

Postponement of Endogenous Deaths

The same does not apply to the exception of recently characterized populations with very low infant mortality. A break in the alignment can be seen (Fig. 48–3), but the higher slope of the first line cannot be interpreted as a reflection of early exogenous excess mortality. It is due to the fact that endogenous mortality is no longer specific to the first month. In a significant number of cases, newborns who previously would have died very early survive some extra weeks or months, and the second hypothesis of Jean Bourgeois-Pichat is no longer verified. The law no longer applies at all.

b. Causes of Death, Endogenous Mortality, and Avoidable Mortality

Beyond the biometric analysis of infant mortality, measuring the maximum biologic life span according to Jean Bourgeois-Pichat depends entirely on the manner in which the causes of death indexed by number of deaths are distributed between causes considered endogenous and those considered exogenous. This is the most perishable aspect of Jean Bourgeois-Pichat's calculations. Retaining all diseases of the respiratory systems, for example, as exogenous, whereas

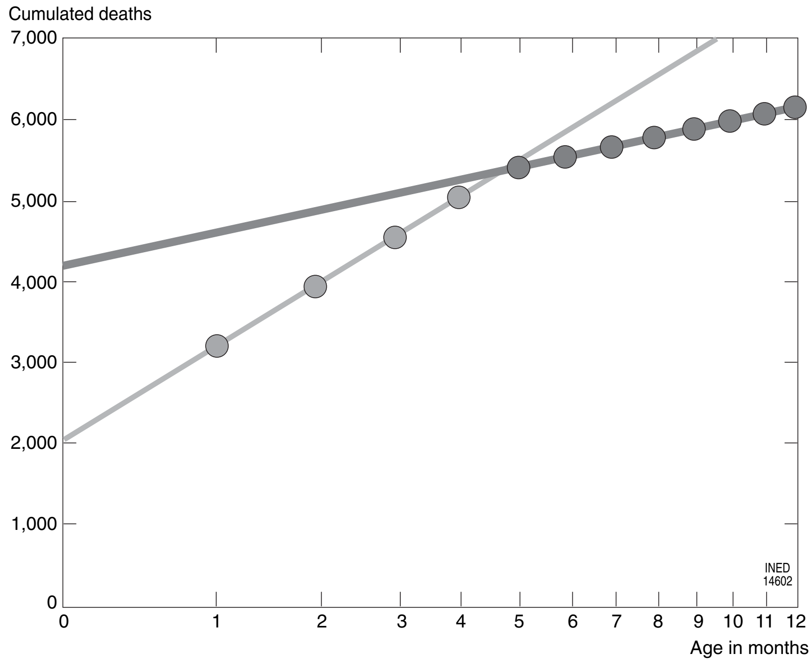


FIGURE 48–3 Biometrical analysis of child mortality: France, 1988. (Source: Vallin, 1993.)

several of them are chronic diseases not necessarily originating from a process of external aggression, is obviously not ideal. Inversely, retaining all cancers and cardiovascular diseases as endogenous, whereas some of these diseases are largely caused by smoking or alcoholism, without adding the henceforth-acknowledged influence of dietary habits or lifestyle, becomes problematic. What about all the other nonaccidental causes, all classified as endogenous? It is at the very least a very rough approximation. The author was aware of this, but he thought that one way or the other, the opposite effects would cancel out and be sufficient to save the approach. In reality, there is no guarantee of compensation effects and imbalance can obviously vary.

There is worse than this. Beyond this approximation, unfortunate but perfectible, there is a slide from the notion of *endogenous* deaths toward the notion of *avoidable* deaths. In trying to define a maximum biological life span, Jean Bourgeois-Pichat relied on the idea that all exogenous deaths, caused by external aggressions are avoidable and that all endogenous biologic deaths are unavoidable. However, it is not certain that we will be capable one day of avoiding all exogenous deaths (because the difficulty in eliminating the significant remainder of infectious diseases or coping with new diseases such as acquired immunodeficiency syndrome shows), and it is becoming more apparent that truly endogenous causes, such as genetic diseases, can be overcome.

c. Hard Rock and Limits: Where Does the Inevitable Come In?

We should acknowledge Jean Bourgeois-Pichat for having always stated that he was reasoning "according to current medical knowledge." The *hard rock* was not unassailable. It merely needed to be processed differently to the *soft rock*, and it required new medical discoveries. This is why Jean Bourgeois-Pichat published an update of his results in 1978, to measure the evolution of the hard rock under the effects of medical progress. To this end, he repeated the same calculations, with the same hypotheses, using data from Norway and more recent statistics. The result was at the very least surprising: new limits of life expectancy were establishing themselves, 80.5 years for women and 73.8 years for men, instead of 78.2 and 76.3 according to the 1952 calculation. Even more surprising was the researcher's comment: "During the last 25 years, while the soft rock has been disappearing, the level of the hard rock has risen for men and crumbled for women."[4] The idea that the increase of 2.3 years in the average maximum life expectancy for women was a result of the erosion of the hard rock under the influence of new medical technology could possibly be accepted, but how could medical discoveries and their spread explain the reduction observed in men?

[4] Durant les 25 dernières années, alors que la roche tendre disparaissait, le niveau de la roche dure s'est élevé pour les hommes et s'est effondré pour les femmes.

Has erosion ever been seen to increase the size of mountains?

In reality, Bourgeois-Pichat's maximum biologic life span included certain completely avoidable causes of death "according to current medical knowledge," but these were on the increase during the period under consideration. The author himself mentioned the increase in male mortality from cancers and cardiovascular diseases. This obviously did not result from an evolution in medical knowledge, but from increase in smoking and increase in the consumption of animal fat and some other social phenomena that did not fulfill a "biologic necessity." This is the major limitation of Jean Bourgeois-Pichat's work. However, is it not more a limitation in form than in essence? Bernard Benjamin (1982), for example, in taking up Jean Bourgeois-Pichat's idea and trying to improve the translation in terms of causes of death, came up with a life expectancy limit of 81.3 for men and 87.1 for women in 1982. These figures are themselves very close to the life expectancy obtained in France in 1978 by looking only at the deaths by "degeneration" resulting from an etiologic reclassification of causes of death: 80.9 for men and 85.7 for women (Vallin and Meslé 1988). Criticism of the quantitative results should not make us lose sight of the essential: the idea of the existence of a hard rock, an unassailable limit according to current medical knowledge probably is true, as is the hope that technologic progress can reduce this hard rock. However, if the hard rock is assailable, how far can we go?

III. IS THERE A LIMIT?

Since the 18th century, it seems that medical progress has barely enabled the average length of life (i.e., life expectancy) to approach the theoretically possible maximum (i.e., longevity). A greater proportion of each generation survives to increasingly older ages, with most deaths occurring in an increasingly older, but also limited, age range, without significantly changing the maximum age at death. This is called the *rectangularization* of the survival curve, a phenomenon that is supposed to set an unassailable limit on life expectancy at birth. However, two questions are far from answered. Is human longevity really intangible? If it evolves, will it be with or without calling Gompertz's law into question?

1. The Limits of Rectangularization

Figure 48–4 illustrates the phenomenon of the rectangularization of the survival curve from French female life tables of the 18th century (Blayo, 1975) and of the present day (Vallin and Meslé, 2001b) with a projection up to 2100 (Vallin and Meslé, 2001b). Starting as a concave, the curve becomes more and more

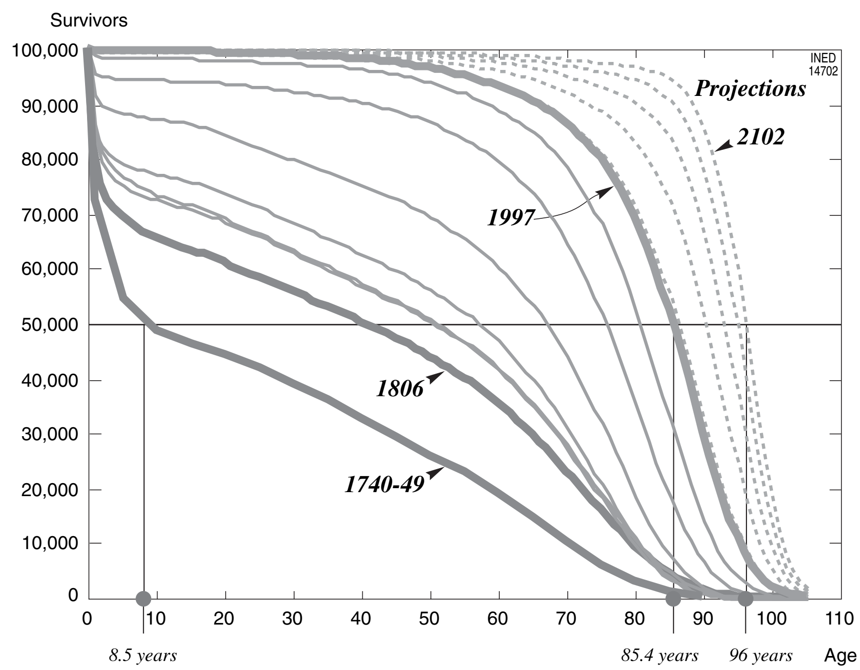


FIGURE 48–4 Rectangularization of the survival curve: plot for females in France since the 18th century and projections up to 2100.

convex with time. An amazing fact is that the median age at death, the age at which 50% of the initial number of an imaginary birth cohort under the conditions of the life table of the period (see Chapters 4 and 20) would survive, increased from only 8 years in 1740–1749 to more than 85 years in 1997. In France at the end of the 20th century, very few women died before the age of 60; on the contrary, most died between the ages of 75 and 95 years. If the recent trend of mortality decline continues in the same pace until the end of the century, the median age at death will reach 96 years in 2100.

Various researchers have proposed different indicators to measure more precisely this phenomenon of rectangularization of the survival curve and numerous reviews of these indicators have been published. Table 48–1 lists the available indicators drawn up by John Wilmoth and Shiro Horiuchi (1999) and completed by Jean-Marie Robine (2000, 2001) to include more recent research by Vaïnö Kannisto (2000a, 2000b).

John Wilmoth and Shiro Horiuchi, after having inventoried and tested ten indicators, recommend using the interquartile range (*IQR*) for its pertinence and ease of interpretation (Wilmoth and Horiuchi, 1999). Vaïnö Kannisto recommended using a new indicator that he named *C*50: the smallest age range necessary to regroup 50% of deaths. Figure 48–5 compares the evolution of these two indicators for French females from the middle of the 18th century until today and prolongs the comparison on the basis of projections up to 2100 (Meslé and Vallin, 2002).[5]

TABLE 48–1 Principal Indicators Proposed for Measuring the Dispersal of Life Spans and the Rectangularity of the Survival Curve

Indicators Inventoried by Wilmoth and Horiuchi (1999)
Proportion occupied by the surface of a fixed rectangle under the survival curve
Proportion occupied by the surface of a moving rectangle under the survival curve
Maximum decline in numbers of the survival curve between two ages
Maximum acceleration in the decline of the survival curve from one age to another
Maximum deceleration in the decline of the survival curve from one age to another
"Prolate" indicator (Eakin and Witten, 1995)
Interquartile range
Standard deviation
Gini's coefficient
Keyfitz's H (Keyfitz and Golini, 1975)
Other Indicators Inventoried by Jean-Marie Robine (2001)
Coefficient of variation (Nusselder and Mackenbach, 1996)
Numerator of Keyfitz's H (Nusselder and Mackenbach, 1996)
Age at various centiles of the distribution of deaths (Manton and Stallard, 1996)
Standard deviation of ages at death below the mode (Kannisto, 2000a, 2000b)
Standard deviation of ages at death of the last quartile (Kannisto, 2000a, 2000b)
Smallest age range assembling a given proportion of deaths (Kannisto, 2000a, 2000b)

The *IQR* and *C*50 have both declined rapidly during the first half of the 20th century, reflecting the scale of the phenomenon of rectangularization of the survival curve observed during this period. However, Figure 48–5 shows the superiority of Kannisto's *C*50 over the *IQR*, insofar as the latter overestimates the speed of rectangularization by measuring the concentration of deaths over too wide an age range while the rectangularization is still relatively modest. This results from the fact that the *IQR* does not, contrary to *C*50, necessarily identify the smallest age range that assembles 50% of deaths. On the other hand, neither *C*50 nor *IQR* are capable of measuring the phenomenon of rectangularization of the survival curve already perceptible in the 18th and 19th centuries. Not only are these two indicators almost stable from 1820 to 1880, except for annual fluctuations, but they also greatly increase from 1740 to 1820, and this latter phenomenon is even more apparent with Kannisto's *C*50 than with the *IQR*.

This disadvantage reflects the important role played by the decline in infant and child mortality in the initial improvements in life expectancy. In the case of *IQR*, while infant and child mortality remained high, the second quartile begins at a very early age, and the *IQR* is relatively short. From 1740 to 1820, the *IQR* rose as infant mortality declined. Then followed a period in which the *IQR* related to a relatively flat portion of the survival curve, with a slope that had hardly varied since the 19th century, which explains the relatively high level of the plateau observed between 1820 and 1900. In the case of *C*50, the influence of child mortality is stronger still because research on even the smallest age range involves the inclusion of infant mortality itself as soon as it goes beyond a certain threshold, and the higher the level of infant mortality, the smaller the *C*50 which includes it, as we can see in the tables for the 18th century.

This gravitation of both *C*50 and *IQR* toward infant mortality attracts attention toward a part of the mor-

[5] The calculation is based on completed annual life tables for the period 1806 to 1997 and for the 1997 to 2100 projection (Vallin and Meslé, 2001b) and summarized decennial tables between 1740 to 1749 and 1800 to 1809 (Blayo, 1975).

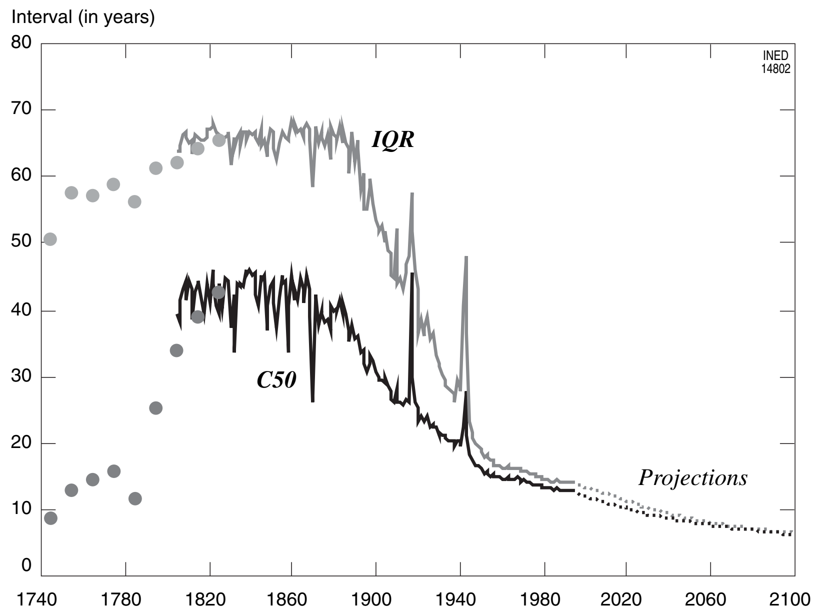


FIGURE 48–5 Evolution of the interquartile range and of Kannisto's *C*50 for females since the 18th century in France and projections up to 2102. (From Meslé and Vallin, 2002.)

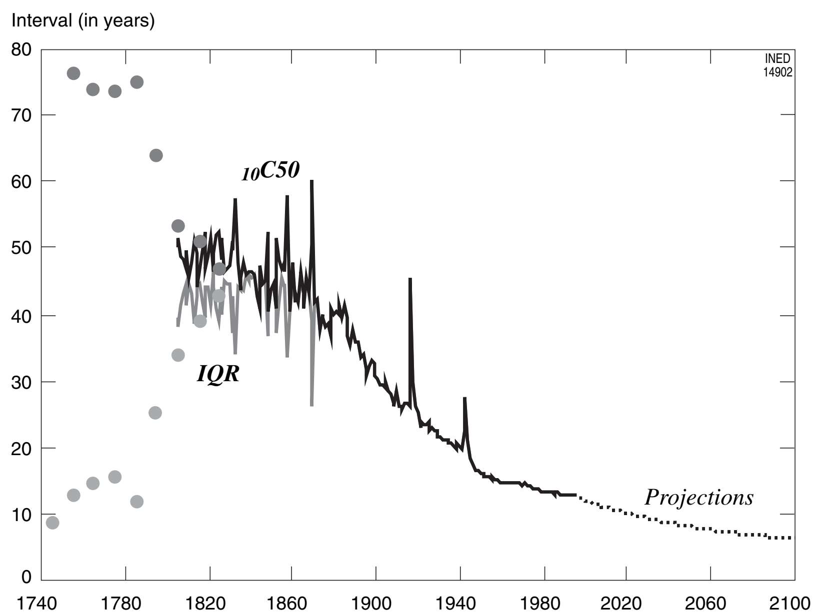


FIGURE 48–6 Evolution since the 18th century and projections up to 2102 of $_{10}C50$ compared with Kannisto's *C*50 for females in France. (From Meslé and Vallin, 2002.)

tality curve that has nothing to do with the phenomena that they are supposed to measure; the evolution of mortality at ages where the risk of death increases with age is what is of interest in the rectangularization of the survival curve. When we need to measure the age range that assembles 50% of deaths, this inconvenience can be overcome by limiting research on the smallest age range that assembles a given proportion of life table deaths to ages of more than 10 years, where the lowest point of the mortality curve is generally situated, thereby creating a new indicator called $_{10}C50$ (Meslé and Vallin, 2002).

Figure 48–6 compares the evolution of $_{10}C50$ with that of Kannisto's *C*50 and illustrates the advantage of $_{10}C50$ over the preceding one (and *a fortiori* over the *IQR*). It also calls to mind that the advantage in relation to *C*50 is valid only when mortality before 10 years is high enough to impose a lower boundary inferior at this age. In all other cases, $_{10}C50$ is, by construction, equal to Kannisto's *C*50.

However, $_{10}C50$ is itself not totally independent of the level of childhood mortality. In some cases in which more than 50% of deaths occur before the age of 10 years, it simply does not exist. Nevertheless, this is hardly an inconvenience because during the whole historical period that corresponds to health transition in France, it appears only once, in the 1740 to 1749 life table, a period of severe food shortages. It is therefore possible to say that in studying the highest concentration of 50% of deaths, this indicator monitors the process of rectangularization entirely.

The difference between $_{10}C50$ and Kannisto's *C*50 is interesting in the sensitivity to infant mortality. In Figure 48–6, it is established that during the 19th century, the fluctuations of both indicators are almost symmetrically opposed. Because *C*50 has the age of 0 as the lower boundary, each point of infant mortality is translated by a narrowing of the age range because most of the deaths are concentrated in the first ages. On the contrary, these same peaks of infant mortality produce a widening of $_{10}C50$ because the increased proportion of childhood deaths reduce those of deaths after 10 years by the same amount, and all other things being equal, a larger age range is required to embrace 50% of deaths. This phenomenon is particularly visible in 1834 with the cholera epidemic and in 1871 with the *Paris Commune*. This difference argues once more in favor of $_{10}C50$ to the extent that when *C*50 escapes the gravitation toward childhood mortality, it behaves exactly like $_{10}C50$. The same is true of the infant mortality crisis of 1918 (i.e., Spanish influenza) or that of 1944 to 1945 (i.e., bombings and food crisis).

Since the 18th century, in 3 centuries of health transition, the progression in the phenomenon of rectangularization has not ceased (except for historical accidents, which have sometimes severely disrupted the trend), a phenomenon that goes hand in hand with constant elevation of the modal age at death: Death always occurs later in an increasingly declining age range. This is clearly seen in the French data, but an international comparison shows that the French case is by no means exceptional (Fig. 48–7).

There is no indication of an imminent reopening of the age range at death. On the contrary, the mortality projections obtained for France by extrapolating the recent trends of decline in the probabilities of dying by age over a century indicate the possibility of a new phase of rapid rectangularization in which $_{10}C50$ could drop for French women, from a little less than 12 years in 1997 to slightly more than 6 in 2012 (see Fig. 48–6). However, it is only a projection based on the hypothesis of a constant life expectancy of 105 years. Other scenarios can be imagined in which mortality declines more and more rapidly at very old ages, opening the way for a new expansion of ages at death and therefore to a derectangularization of the survival curve. However, in the case of a projection (Brutel, 2001), assuming that beyond 100 years mortality becomes

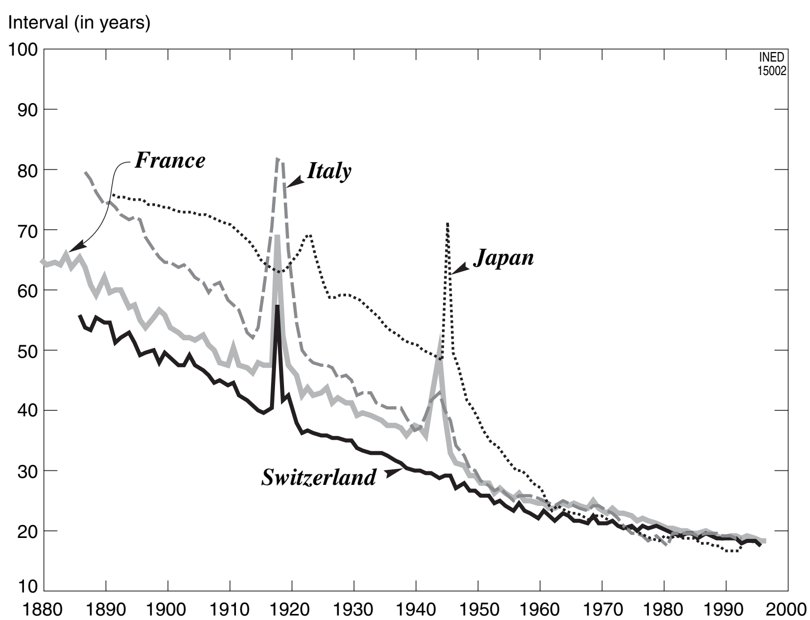


FIGURE 48–7 Trend of $_{10}C50$ for females since the end of the 19th century in four developed countries. (From Meslé and Vallin, 2002.)

independent of age and lowers with time, ${}_{10}C50$ continues to drop for French women during the whole of the 21st century (Meslé and Vallin, 2002). Even with a strong hypothesis of a re-expansion of ages at death after 100 years, the phenomenon of rectangularization could still continue for a long time, and we are no doubt a long way from seeing it appear through indicators such as ${}_{10}C50$. To be capable of perceiving a possible future expansion of ages at death, indicators based exclusively on very old ages (100 years and older) would have to be used; it is precisely at these ages that it quickly becomes impossible to measure mortality for want of data of sufficient quality.

Secular trends give the impression that the progress of life expectancy is constrained by the impossibility of exceeding a certain level imposed by the maximum human life span: The nearer this maximum life span is approached, the more the potential for future progress is exhausted. At the extreme, supposing that nonmeasurable human longevity was effectively characterized by an intangible life expectancy of 1.5 years at age 105. Supposing we had succeeded in eliminating all mortality before this age, life expectancy could reach 106.5 years, but it is not only about an absolute, totally unassailable limit; it is also about a completely illusory objective, because the total eradication of all mortality between 0 and 105 years is doubtless beyond reach.

Some biologists and demographers believe that life expectancy should come up against a nearly unassailable limit before 105 years, or even 100 years. Irreducible mortality by senescence begins much earlier and progresses exponentially with age. The average age at death, or life expectancy at birth, therefore cannot equal the maximum life span. Biologically, it is obliged to remain significantly less. At the beginning of the 1980s, James Fries (1980) and Jay Olshansky and Brian Ault (1986) estimated that this limit should be about 85 years. It is also approximately at this level of life expectancy that United Nations experts have put a ceiling on all future trends for their world population projections, country by country (see Chapter 77). This point of view appears today to be excessively pessimistic, because current life expectancy, for women at least, is already approaching this limit in several countries, nearing 84 years in France and 85 years in Japan, and this progress does not seem to be running out of steam. In a later article, Jay Olshansky appreciably changed his attitude and is now trying to estimate the time it would take for life expectancy to reach 85 or even 100 years in various countries. Japanese women probably will reach the limit of 85 years in about 2005, French women in 2014, French men in 2052, and Japanese men in 2060, but American women will reach it in 2152 and American men in 2239 (Olshansky *et al.*, 2001). A life expectancy of 100 years could be attained by French women in 2106 and even by American men in 2577. It is therefore only a matter of time! If the remaining perspectives for the simple rectangularization of the survival curve remain open, what is the case for the fundamental constraint of a fixed longevity?

2. Fixed or Evolutionary Longevity?

The measurement of human longevity is, alas, a field dominated by much uncertainty. It is also the fundamental crux of the debate on the future of mortality: Does an endogenous biologic limit to the human life span exist?

a. First Attempts to Estimate the Maximum Human Life Span

Convinced of the existence of a limit for the human life span, Paul Vincent tried to estimate it at the beginning of the 1950s. It was also on this issue that he developed the method called *the extinct generations* (see Chapter 11) to measure the mortality of the old (Vincent, 1950). Having through this method calculated as carefully as possible the risks of death at very old ages in four European countries (i.e., France, Sweden, Netherlands, and Switzerland) and based on Gompertz's principle (although without referring to it at all), he adjusted the age-specific probabilities of death observed between 98 and 104 years and prolonged the exponential upward, presuming that the age corresponding to the point where the line extrapolated from the logarithms of these probabilities reached unity was a precise estimate of the extreme limit of life. No one can survive at a probability of dying of 1. Using cumulated data from these four European countries, the author estimated the current limit of the human life span to be a little less than 110 years (Vincent, 1951), not really leaving any chance for Jeanne Calment to become famous.

However, 20 years later, Françoise Depoid (1973) repeated the same operation on the same countries with later data and found a limit of 117.3 for men and 119.3 for women. Did this mean that human longevity had increased?

At the root of these studies, there is a calculation problem. By extrapolating the line from the logarithms of the probabilities of dying, Paul Vincent neglected the fact in Gompertz's law that an exponential increase with age applies to the force of mortality and not to the probabilities of death. When the force of mortality tends toward infinity, the probability of dying tends toward 1 (see Chapter 11). There is therefore not an age (if not infinity) where the probability of dying can precisely reach unity or any justification for a linear extrapolation of the adjustment line of the logarithms of the probabilities of death (Le Bras, 1976; De Finetti, 1953). It is true that the practical range of this difference between instantaneous force of mortality and annual probability of dying is minimal when we look at ages with low probabilities of death, but at old ages, the values of the probability of dying are high enough for it to grow in size. The slope of the line adjustment of logarithms of death probabilities therefore is likely to decrease if more calculations are based on empirical mortality data at the oldest ages. This is one of the possible reasons for the difference of 9 years between the age limit measured by Paul Vincent and that estimated by Françoise Depoid 20 years later. This does not remove from the fact that measuring a maximum age in this way is illusory and if a limit does exist, this method is incapable of indicating its possible trend.

This does not however invalidate the existence of a maximum human life span nor answer the question of a possible postponement of this limit in the last decades, as several recent observations like the emergence since the 1980s of a non-negligible number of "supercentenarians", individuals who have reached the age of 110 years, might suggest (Thatcher, 2001; Meslé *et al*., 2000).

b. Biologic Aging and Longevity

Assuming there is a biologic limit to human longevity, can it evolve with time? If longevity constitutes a biologic characteristic of the species, its evolution is based on the principles of Darwin: the probabilities of mutations at the time of reproduction and a selection of favorable mutations (or the elimination of unfavorable mutations) by mortality (Kirkwood and Rose, 1991; Rose, 1991; Kirkwood, 1990). According to James Carey (2000), human longevity increased significantly during the evolution of *Homo* and since the appearance of *Homo sapiens*. Although it did not exceed the age of menopause by more than 7 to 11 years for *Homo habilis*, it did by 15 to 18 years for *Homo erectus*, with the selection process retaining the favorable aspects of survival well beyond the age of menopause (notably for the rearing of children and grandchildren). In this way, longevity reached 60 to 63 years for *Homo erectus* before reaching more than 70 years among the first *Homo sapiens*, which leads to the conclusion that it has increased again in our species, because it is now situated around the 122 years of Jeanne Calment.

We know that the mechanisms of evolution at the origin of this lengthening of the maximum human life span can take effect only in the long term, with the succession of generations. Not only is the reproductive cycle of mankind long, but for 2 centuries, by greatly reducing fertility and mortality, people have considerably reduced the field of action of their own evolution. How can we imagine at this precise moment that the process liable to increase human longevity has accelerated to the point where change is visible in a few decades of statistical observation, as the results of Paul Vincent and Françoise Depoid would suggest? It is extremely doubtful.

There may be another explanation. Human longevity may not be a biologic characteristic directly determined by genetic heritage; it may result mainly from a balance, albeit ephemeral, between the potential of this heritage and environmental constraints. Is it not possible to draw on recent biologic breakthroughs in the field of aging to shed light on demographers' questions on longevity? This question will be asked more specifically in Chapter 49 about the genetic factors of mortality, but as Graziella Caselli, James Vaupel and Anatoli Yashin (2000) wrote:

> Ageing and longevity are very typical questions on today's agenda. The two processes are closely intertwined. Biologists and geneticists who study the mechanisms implied and try to establish at what stage one is old as well the real possibilities of further extending longevity, have come up with about 300 or so theories on ageing. According to Medvedev (1990) these can be grouped under four main theories: programmed genetic ageing; age associated changes; primary damage; evolutionary theories (Marigliano, 1995). Some of these theories focus on "*how we age*", i.e., on the mechanisms of free-radical damage, telomere shortening, errors in cell division, hardening of arteries, etc. that characterize aging. Other theories focus on *why we age* (i.e., on the fundamental evolutionary forces that underlie the fact that humans and other species are not immortal) (Finch and Kirkwood, 2000; Rose, 1991; Kirkwood, 1990; Williams 1957). Demographers are interested in *when we die*. That is, demographers want to know why some people die at age 60, others at 80, and a few at 100 and why the number of people dying at older ages is rapidly increasing. The theories of how we age have not yet shed much light on this kind of question. The current theories of why we age are so abstract and general that these theories shed virtually no light at all on the age-patterns of mortality and changes in these patterns over time.

For the time being, we should be satisfied with believing that human longevity is possibly gifted with a certain elasticity according to the evolution of the environment. What global change, such as climatic modifications, have produced indirectly in the long term in the species through selection, humans' own actions are in the process of doing, at least on an individual level, by modifying his relationship with his environment. In other words, could not progress in health have prolonged not only the average life span (i.e., life expectancy), but also its maximum span (i.e., longevity), without invoking an evolution of the genetic heritage of humanity, at this stage?

c. *Indicators of Recent Trends*

It is possible to assert that important changes are being produced at the level of mortality at old age.[6] Cases of extreme longevity have always been fascinating. According to the Bible, Adam, the first man, lived 930 years, and others after him lived even longer, such as Enoch, who died at 965 years, and Methuselah, the champion of champions, who died at 969 years! Logically, the ages at procreation were also much higher than today. According to Genesis, Adam was 130 when he "knew" Eve to give her Seth, her third child, and he knew her many more times after that. All this is no more than legend. The Bible is full of epic images, and the age of the patriarchs from before the flood was meant to strike the imagination, rather than to recount reality. For the successors after the flood, the estimates are much more reasonable: 275 years for Abraham, 180 for Isaac, 147 for Jacob, and 110 for Joseph.

In the more recent era, the centenarians of the Caucasus and those from several Andean valleys have been in the news, without mentioning the American Georgians. If they appeared credible for some time, we know today that they did not represent reality. They are three examples of a classic phenomenon: In many societies, the very old inspire deference and respect, and when the civil status of individuals is not well recorded, the ages declared for the very old are most of the time overestimated. For a legend to be born, all that is required is an argument to render the facts plausible (e.g., the effects of living on a high mountain or eating yogurt, the hypothesis of a genetic heritage specific to an isolated people). The longevity of the Georgians of the Caucasus was very flattering to Stalin, who did his utmost to maintain the myth (Bennett and Garson, 1986). It has also been shown that the first information from the Indians at Vilcabamba largely overestimated the ages (Mazess and Forman, 1979). As for the American Georgians of 120, 130, and even 140 years, this is simply a reflection of imprecision in census age declaration particularly in the black population among whom, at the end of the 19th century, it was rare for births to be recorded.

It is extremely improbable that until recent decades, anyone could have lived much longer than 100 years. It was even rare for this age to be reached. On the contrary, centenarians today are counted by thousands, and we are almost certain that at least one woman, Jeanne Calment,[7] lived for 122 years (Robine and Allard, 1999). Does this mean that human longevity is increasing?

It could initially be believed on reading about the secular evolution of the maximum age of death observed in a country such as Sweden that has for a long time had an excellent system for registering births (Fig. 48–8).

The oldest age at death observed in a year necessarily fluctuates. It is an extreme performance, an event, which cannot be established each year at the same level. Nevertheless, despite the very strong fluctuations observed in Figure 48–8, as much for men as for women, the trend is very clearly rising: in the middle of the 19th century, the maximum age at death oscillated between 100 and 105 years for women and between 97 and 102 for men, whereas over the past 20 years, it has been approximately 107 to 112 years for women and 103 to 109 years for men. This phenomenon has recently accelerated; the slope of the regression line obtained for the 1970 to 1997 is significantly higher than that obtained for the 1851 to 1969 period (Wilmoth *et al.*, 2000).

This does not necessarily mean that human longevity is increasing. There is a simple statistical game linked to the evolution of the number of elderly people, itself partly associated with medical progress, at the root of this phenomenon. The greater the number of people living to 90 years, the greater, with equal chances of survival, the probability of at least one person reaching 95, 100, or 105 years. For a very long time, human populations increased only very slowly as a result of the small surplus of a very high fertility over an equally high mortality. The population pyramid was very pointed at the peak; those older than 60 years represented barely 5% of the total population. The very old were rare, and the probability of a centenarian emerging was extremely low.

From the end of the 18th century, Europe's population went through a period of exceptional population growth, which in itself would have been sufficient for the emergence of an increasing number of very old people. This increase went hand in hand with aging of the population, which widened access to very old ages. Initially, the simple reduction of infant mortality that occurred 60 years earlier significantly increased the number of the generations entering old age. Better still, in the 1960s, survival beyond 60 years greatly

[6] This section uses extensive passages from an article that appeared in *Population et Sociétés* (Vallin and Meslé, 2001a).

[7] Doyenne of humanity, probably of all time, Jeanne Calment, born February 21, 1875, and died on August 4, 1997, having spent all her life in Arles. No one has lived to her age since. On the men's side, the validated record is that of Thomas Kristian Mortensen, born August 16, 1882, and died April 25, 1998, at the age of 115 years.

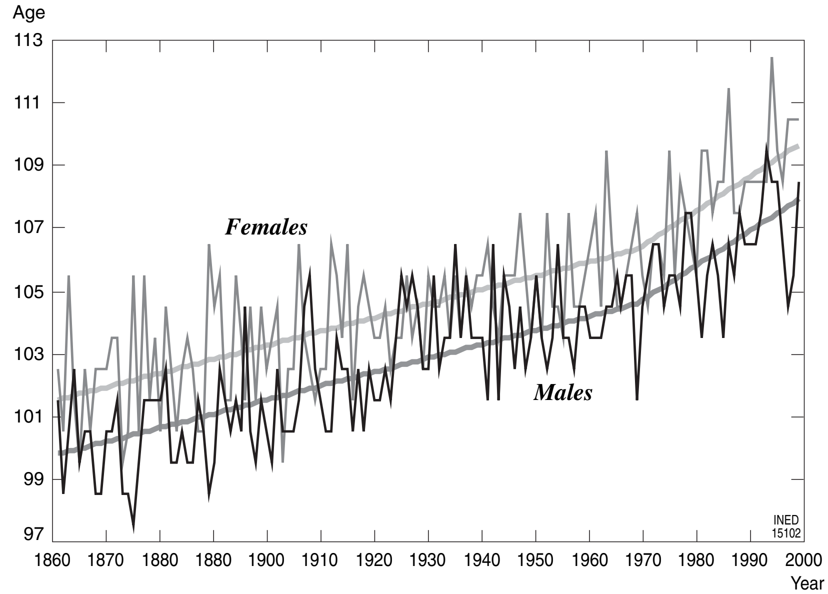


FIGURE 48–8 Evolution according to sex of the oldest age at death observed in a year since 1851 in Sweden. (From Wilmoth *et al.*, 2000.)

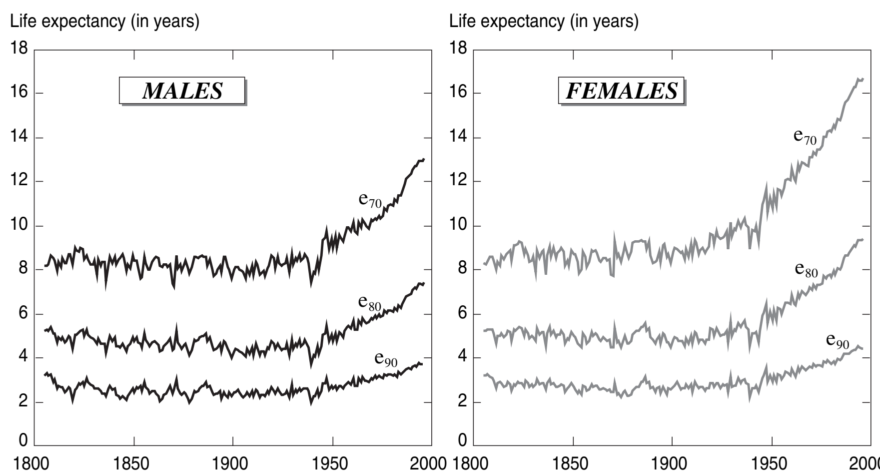


FIGURE 48–9 Evolution according to sex of life expectancies at 70, 80, and 90 years since 1806 in France. (From Vallin and Meslé 2001a.)

increased, thereby increasing the numbers at 70, 80, and 90 years by as much again and increasing the possibility of survival of a person to an age never previously reached. Since 1969 in Sweden, the mortality decline after 70 years has been responsible for 95% of the increase in the maximum age at death (Wilmoth *et al.*, 2000).

The more mortality decline settles at older ages, the more the question begs to be asked whether, beyond the effect of numbers, there is not also a real increase in the maximum life span due to a new phase in medical progress.

The increase in life expectancy at 70, 80, and 90 years has been spectacular since the 1950s, whereas it was previously nil, especially for men (Fig. 48–9). In France, the life expectancy of men at age 70, which had consistently oscillated between 7 and 9 years from the beginning of the 19th century up to the Second World War, exceeded 13 years in 1997. For women, progress started a little earlier, between the two wars, when the level was more or less a constant of 8 to 10 years, to nearly 17 years today. The same change in the maximum life expectancy can also be observed at 80 and 90 years, and if at these ages recent developments

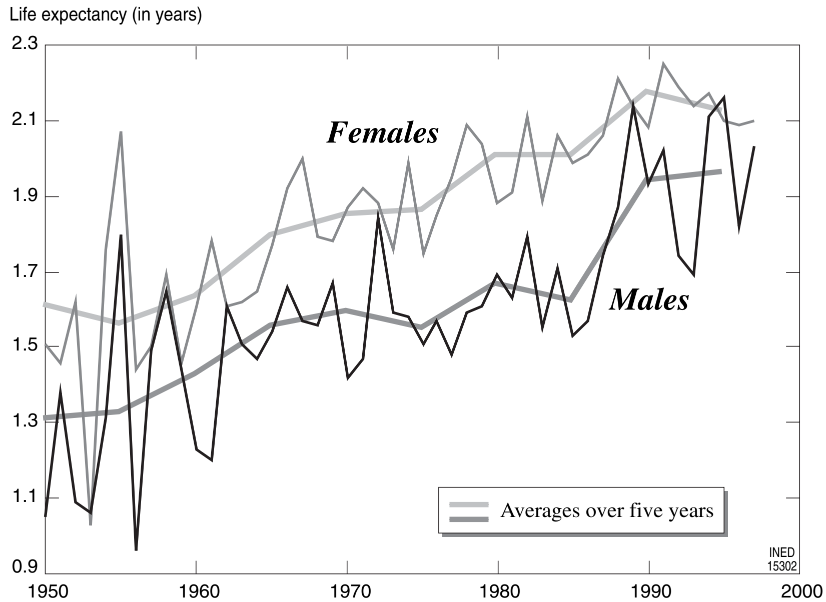


FIGURE 48–10 Evolution of life expectancy at 100 years since 1950 in France. (From Vallin and Meslé, 2001a.)

seem slower than at 70 years, it is only because of the arithmetic scale of the graph. The relative increase is just as important: an increase of more than 60% for men and nearly double for women.

Even at a 100 years, life expectancy has been noticeably increasing for several decades (Fig. 48–10). At this age, the annual figures fluctuate much more because they are based on reduced numbers, and only recent years are shown here to accompany the annual results of moving averages over 5 years. Nevertheless, the trend is clearly on the increase. At 1.1 year at the beginning of the 1950s, increased life expectancy for men at 100 years is 2 years. At the same time, women's life expectancy has increased from approximately 1.4 to 2.1 years.

This sudden increase in survival at very old age is an enigma. Does this mean that medical progress will henceforth affect the maximum human life span (i.e., longevity) and not only the average life span (i.e., life expectancy at birth)?

IV. DOES GOMPERTZ'S LAW APPLY TO THE OLDEST AGES?

The double assumption of an extreme intangible maximum human life span and an exponential growth in mortality by senescence well before this limit, is frequently being strongly contested. One of the first to do this was the biologist, Roy Walford who, in *The Maximum Life Span* (1984), suggested that initially life expectancy would approach the current maximum life span, but that soon, due to a technologic jump provoked by biologic progress, this limit could be pushed back and that it was not absurd to imagine that life expectancy could reach 150 years at the end of the 21st century. Other researchers believe that certain living beings (e.g., queen bees, some species of flies, sequoias) are escaping from the communal rule of the exponential increase of risk of death with age and would be nearly immortal were they not subjected to the risk of violent death (Klarsfeld and Revah, 1999). Better still, the massive experiments carried out on the fruit fly during the last 15 years have shown that beyond a certain age, mortality declines with age (Carey 1997; Vaupel and Carey, 1993), which incites a demographer such as James Vaupel to ask the following question: "Shall we soon witness a situation where certain individuals of the human race, as in the case of these flies, live long enough to reach a stage where mortality declines with age?" (Vaupel *et al*., 1998). A lucky few could then dream of immortality.

Even if it seems premature to judge the concrete reality of lengthening human longevity, let us say that this could occur in two quite different ways, only the second of which calls Gompertz's law into question.

1. Is the Curve Shifting to the Right?

Gompertz's law, if it can be interpreted as a model of characteristic biologic aging of humans, does not prejudice the fixity of longevity or place a limit to mortality decline. The fixity of longevity, which constrains the rectangularization of the survival curve, assumes that because all the mortality curves should converge toward the same force of mortality at the extreme age, mortality decline will be accompanied by an increase in the slope of the curve. For the limit to be constantly pushed back, it therefore suffices that the mortality decline does not modify the slope of the mortality curve, with the latter shifting entirely toward the right. Figure 48–11 shows the evolution of the French female mortality curve since the beginning of the 19th century. As mortality previously declined, the slope of the mortality curve increased. Between 1810 to 1814 and 1950 to 1954, mortality very clearly declined much more for those younger than 70 years compared with those older than 70 years, and the slope of the probabilities of death increased considerably. This evolution seems to corroborate the existence of an intangible maximum life span. On the contrary, since 1950, mortality has been declining rapidly for older ages. The rate of decline remains a little less than for younger ages, and the slope of the mortality curve continues to increase, although less so than previously, whereas anyone would logically have expected this phenomenon to entrench itself.

This assessment is not enough to prove a real lengthening of longevity, but the trend firmly supports such a hypothesis. However, we would have to be able to show that it not only occurs between 70 and 100 years as is seen in Figure 48–11, but also beyond 100 years. Here, we come up against the rarity of reliable data on mortality after 100 years. We are only now beginning to have access to some acceptable measurements, and there is a desperate lack of experience to judge the evolution in the shape of the curve beyond that age.

2. Is Progress with Aging Slowing Down?

Certain researchers believe that mortality curve is shifting toward the right and that Gompertz's law no longer applies beyond a certain age. Shiro Horiuchi and John Wilmoth (1998) showed that the increase of mortality with age tends to slow down and even stop

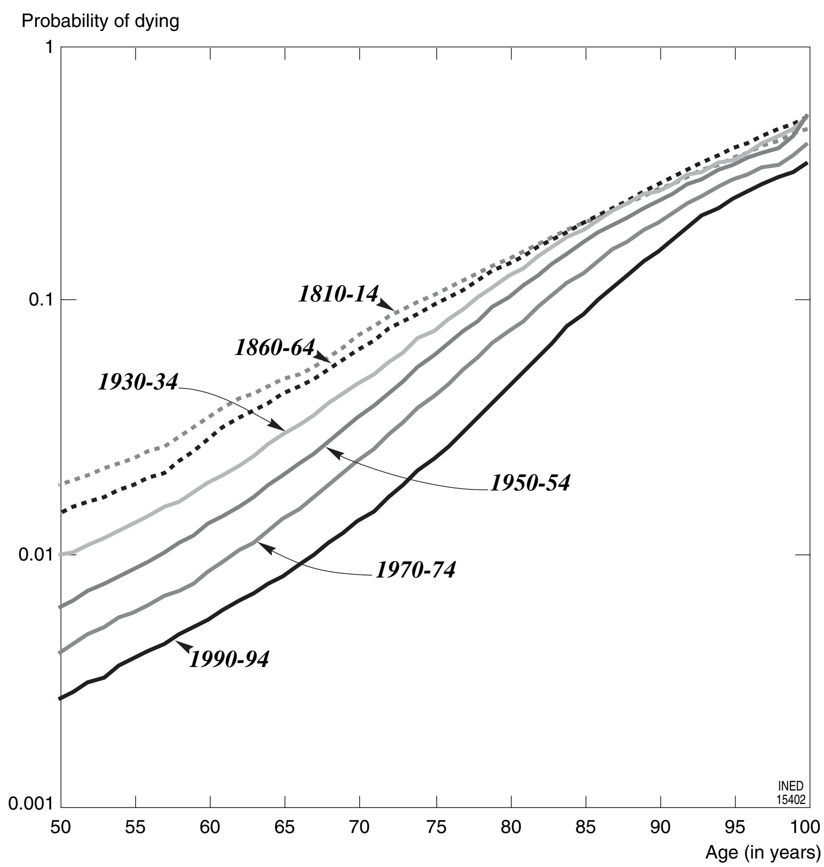


FIGURE 48–11 Modification of the mortality curve between 50 and 100 years since the beginning of the 19th century. (From Vallin and Meslé, 2001.)

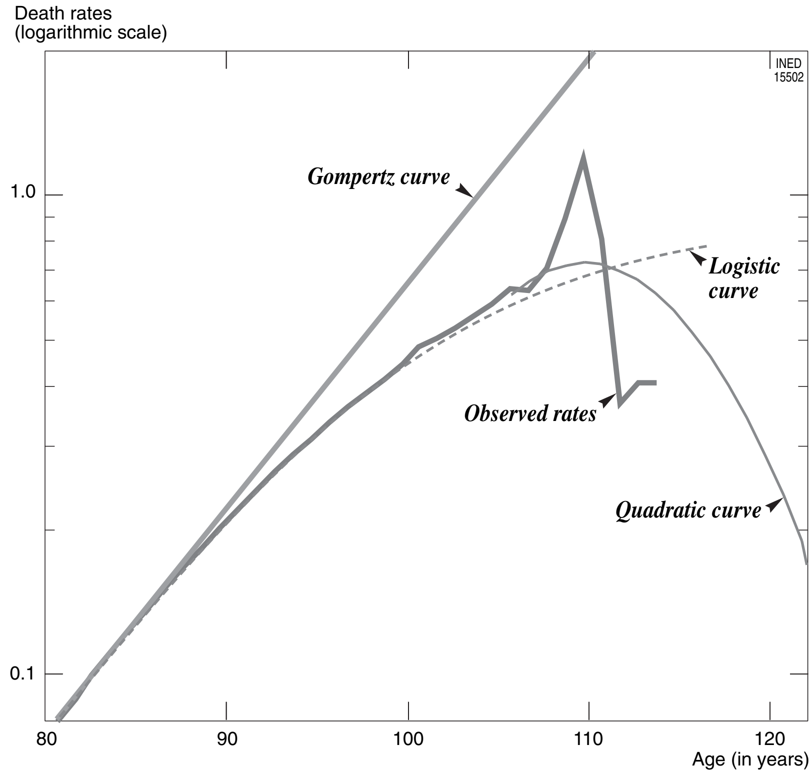


FIGURE 48–12 How do we adjust the risks of death observed at very old ages? Possible alternatives to Gompertz's law, according to James Vaupel (2001). The *bold solid line* indicates rates observed in cumulated data from Japan and the 14 European countries.

at very old ages. Is this a real phenomenon characterizing the force of mortality, or is it a result of bias, increasing with the risk of mortality at these ages, in the estimate of this instantaneous force based on discontinuous measurements? If this phenomenon is real, is the rapid reduction of heterogeneity in the survivors exposed to a very high mortality enough to explain this deviation from Gompertz's law, with each homogenous subgroup remaining subject to this law? Jean-Marie Robine (2001) has proposed two other hypotheses that could explain the interruption in the increase of mortality with age: the general improvement in the environment and in living conditions that has occurred during the last half century in industrialized countries and the fact that if elderly people are very fragile, they are also more frequently being placed in a protective environment, like crystal jewels carefully kept away from any risk of damage. Nevertheless, the available data are too tenuous to give definite answers to these questions. The door is wide open to the imagination of extreme scenarios. Based particularly on the example provided in the experiments performed on flies, James Vaupel (2001) has put forward the hypothesis that the mortality curve might be expected to turn around at very old ages for humans, too (Fig. 48–12).

The current data are too fragile to enable a choice between these extreme models of adjustment. This is why the development of research on centenarians and supercentenarians, those rare individuals who have life spans that exceed 110 years, has begun (Robine and Vaupel, 2001; Thatcher, 2001; (Meslé *et al*., 2000).

CONCLUSIONS

With Jean Bourgeois-Pichat's attempts, we have seen how difficult and even inaccessible is the definition of endogenous mortality linked essentially to human biology. Most causes demonstrate endogenous and exogenous aspects, and it is rarely possible to make a distinction between them to define the biologic life span limit of humans. Can we at least rely on a genetically determined fixity of longevity to estimate, in accordance with the principle of rectangularization of the mortality curve, what could be the unassailable rectangle or another way of defining the maximum human life span? It is possible that even this framework, about whose dimensions there is a great incertitude, could explode under the effect of the relatively new phenomenon of mortality decline among the very old. It is not certain that Gompertz's law can hold the road indefinitely at very old ages.

Perhaps the biologic material of which living organisms are made does not necessarily define, as it is believed, the vital process whose progressive deterioration leads ineluctably to death. Endogenous factors cannot, by their very nature, be dissociated from the environment with which they interact. Moreover, is this not what we learn from the laws of evolution? When a mutation changes an element of genetic heritage, its own environment determines whether it is retained for transmission to its heirs.

What do biologists have to say today about death? Contrary to what we have believed for so long, biologists think that death is not a biologic necessity but that evolution simply did not choose immortality because it was not essential to the survival of the species (Klarsfeld and Revah, 1999).

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Benjamin Bernard, 1982. The span of life, *Journal of the Institute of the Actuaries*, vol. 109, p. 319–340.

Bennett Neil and Garson Lea Keil, 1986. Extraordinary longevity in the Soviet Union: fact or artefact, *The Gerontologist*, vol. 26(4), p. 358–361.

Biraben Jean-Noël and Henry Louis, 1957. La mortalité des jeunes enfants dans les pays méditerranéens, *Population*, vol. 11(4), p. 615–645.

Blayo Yves, 1975. La mortalité en France de 1750 à 1829, *Population*, special issue Démographie historique, vol. 30, p. 123–142.

Bourgeois-Pichat Jean, 1951a. La mesure de la mortalité infantile: I. Principes et méthodes. *Population*, vol. 6(2), p. 233–248.

Bourgeois-Pichat Jean, 1951b. La mesure de la mortalité infantile: II. Les causes de décès. *Population*, vol. 6(3), p. 459–480.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Bourgeois-Pichat Jean, 1978. Future outlook for mortality decline in the world, *Population Bulletin of the United Nations*, vol. 11, p. 12–41.

Bourgeois-Pichat Jean, 1981. Les projections de population à long terme: perspectives de la mortalité et de la fécondité, *in*: *Projections démographiques: problèmes et solutions*, p. 101–143. New York, United Nations, 316 p. (Doc. TCD/SEM-81/3.)

Bourgeois-Pichat Jean, 1988. Du XX^{e} au XXI^{e} siècle: l'Europe et sa population après l'an 2000, *Population*, vol. 43(1), p. 9–44.

Brutel Chantal, 2001. Projections de population à l'horizon 2050: un vieillissement inéluctable, *INSEE première*, vol. 762, p. 1–4.

Cantrelle Pierre, 1969. *Étude démographique dans la région du Siné Saloum (Sénégal)*. Paris, ORSTOM, 121 p.

Cantrelle Pierre and Leridon Henri, 1969. Mortalité au sevrage, mortalité saisonnière. Un exemple: Pakao (Senegal), *Population*, vol. 24(2), p. 339–343.

Carey James R., 1997. What demographers can learn from fruit fly actuarial models and biology, *Demography*, vol. 34(1), p. 17–30.

Carey James R. and Judge Debra S., 2001. Principles of biodemography with special reference to human longevity, *Population, An English Selection*, vol. 13(1), p. 9–40.

Carey James R. and Judge Debra S., 2001. Life span extension in humans is self-reinforceing: a general theory of longevity, *Population and Development Review*, vol. 27(3), p. 411–436.

Caselli Graziella, Vaupel James W., and Yashin Anatoli, 2000. *Longevity, heterogeneity and selection*, *in*: *Proceedings of XL riunione scientifica della società italiana di statistica*, p. 49–72, SIS 2000, Florence, 437 p.

Condorcet Marie Jean Antoine Nicolas Cariat Marquis de, 1795 (an III). *Esquisse d'un tableau historique des progrès de l'esprit humain*. Paris, Agasse, VIII +389 p. (Text revised in 1933 and introduced by O. H. Prieur, Paris, Boivin et Cie. XXIV, 239 p.)

de Finetti Bruno, 1953. A proposito di una ricerca sull'andamento della mortalià nell età senili, *Rivista italiana di economie, demografia e statistica*, vol. 7(2–4), p. 21–35.

Depoid Françoise, 1973. La mortalité des grands vieillards, *Population*, vol. 29(4–5), p. 755–792.

Duchêne Josianne, 1980. *Un essai de modélisation de la répartition des décès selon l'âge et la cause dans les pays industrialisés*. Louvain-la-Neuve, Cabay, 82 p. (Département de démographie de l'Université Catholique de Louvain.)

Duvillard (de Duran) Emmanuel E., 1806. *Analyse et tableaux de l'influence de la petite vérole sur la mortalité à chaque âge, et de celle qu'un préservatif tel que la vaccine peut avoir sur la population et la longévité*. Paris, Imprimerie nationale, 210 p.

Eakin Tim and Witten Matthew, 1995. How square is the survival curve of a given species? *Experimental Gerontology*, vol. 30(1), p. 33–64.

Finch Caleb and Kirkwood Thomas B. L., 2000. *Chance, development and ageing*. Oxford, UK, Oxford University Press, 278 p.

Fries James F., 1980. Aging, natural death, and the compression of morbidity, *New England Journal of Medicine*, vol. 303, p. 130–135.

Gompertz Benjamin, 1825. On the nature of the function expressive of the law of the human mortality and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Statistical Society*, vol. 27(2), p. 513–583.

Henry Louis, 1984. *Démographie, analyse et modèles*. Paris, INED, 341 p.

Henry Louis and Blayo Yves, 1967. Données démographiques sur la Bretagne et l'Anjou de 1740 à 1829. *Annales de démographie historique*, p. 91–172.

Horiuchi Shiro and Wilmoth John R., 1998. Deceleration in the age pattern of mortality at older ages, *Demography*, vol. 35(4), p. 391–412.

Kannisto Vaïnö, 2000a. Measuring the compression of mortality, *Demographic research*, vol. 3(6), 23 p. (available online: www.demographic-research.org).

Kannisto Vaïnö, 2000b. Mode and dispersion of the length of life, *Population. An English Selection*, vol. 13(1), p. 159–172.

Keyfitz Nathan and Golini Antonio, 1975. Mortality comparisons: the male-female ratio, *Genus*, vol. 31(1–4), p. 1–34.

Kirkwood Thomas B. L., 1990. The disposable soma theory of ageing, *in*: David E. Harrison (ed.), *Genetic effects on aging, II*, p. 9–19. Caldwell, NJ, Telford Press, 573 p.

Kirkwood Thomas B. L. and Rose Michael R., 1991. Evolution of senescence: late survival sacrificed for reproduction, *Philosophical Transaction of the Royal Society (London) [Biology]*, vol. 332, p. 15–24.

Klarsfeld André and Revah Frédéric, 1999. *Biologie de la mort*. Paris, Odile Jacob, 290 p.

Le Bras Hervé, 1976. Lois de mortalité et âge limite, *Population*, vol. 31(3), p. 654–692.

Makeham William M., 1860. On the law of mortality and the construction of annuity tables, *The Assurance Magazine and Journal of the Institute of Actuaries*, vol. 8, p. 301–310.

Makeham William M., 1867. On the law of mortality, *Journal of the Institute of the Actuaries*, vol. 13, p. 325–367.

Manton Kenneth G. and Stallard Eric, 1996. Longevity in the United States: age and sex-specific evidence on life span limits from mortality patterns 1960–1990, *Journal of Gerontology: Biological Science*, vol. 51A(55), p. B362–B375.

Manton Kenneth G. and Yashin Anatoli I., 2000. *Mechanisms of aging and mortality: the search for new paradigms*. Odense monographs on population aging, no. 7. Odense, Odense University Press, 174 p.

Marigliano Vicenzo, 1995. *Invecchiamento e longevità*. Rome, CESI, 356 p.

Mazess Richard B. and Forman Sylvia H., 1979. Longevity and age exaggeration in Vilcabamba, Ecuador, *Journal of Gerontology*, vol. 34, p. 94–98.

McGue Matt, Vaupel James W., Holm Niels V., and Harvald Bent, 1993. Longevity is moderately heritable in a sample of Danish twins born 1870–1880, *Journal of Gerontology and Biological Sciences*, vol. 48(6), p. B237–B244.

Medvedev Zhores A., 1990. An attempt at a rational classification of theories of ageing, *Biology Review*, vol. 65(3), p. 375–398.

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIX^{e} siècle. *Population*, vol. 44(6), p. 1121–1158.

Meslé France and Vallin Jacques, 2002. Comment améliorer la précision des tables de mortalité aux grands âges? Le cas de la France, *Population F*, vol 57(4–5), p. 603–632.

Meslé France and Vallin Jacques, 2002. *Montée de l'espérance de vie et concentration des âges au décès*. Paris, INED, 20 p. (Dossiers et recherches no.108.)

Meslé France and Vallin Jacques, 2003. Increase in life expectancy and concentration of ages at death, *in*: Jean-Marie Robine *et al*. (eds.), *Determining health expectancies*. Chichester, John Wiley & Sons, 444 p.

Meslé France, Vallin Jacques and Robine Jean-Marie, 2000. Vivre plus de 110 ans en France, *Gérontologie et Société*, vol. 94, p. 101–120.

Nam Charles B., 1990. Mortality differentials from a multiple-cause-of death perspective, *in*: Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analysis of mortality*, p. 328–342. International studies in demography. Oxford, Clarendon Press, 416 p.

Nusselder Wilma J. and Mackenbach Johan P., 1996. Rectangularisation of survival curve in the Netherlands, 1950–1992, *Gerontologist*, vol. 36(6), p. 773–781.

Olshansky S. Jay and Ault A. Brian, 1986. The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases, *The Milbank Quarterly*, vol. 64(3), p. 355–391.

Olshansky S. Jay, Carnes Bruce A., and Desesquelles Aline, 2001. Prospects for human longevity, *Science*, vol. 291, p. 1491–1492.

Pressat Roland, 1961. *L'analyse démographique*. Paris, Presses Universitaires de France, 402 p.

Pressat Roland, 1974. Les tables de mortalité en l'absence de certaines causes de décès, *Canadian Studies in Population*, vol. 1, p. 61–72.

Prioux France and Vallin Jacques, 1976. évolution des disparités de la mortalité infantile entre les départements français, *in*: *Les disparités démographiques régionales*, p. 391–399. Paris, CNRS, 614 p. (CNRS 935 National Seminars, Fifth National Demography Seminar.)

Robine Jean-Marie, 2000. *L'évolution de la distribution des durées de vie individuelle*. Montpellier, Équipe démographie et santé, INSERM, 10 p. (Paper presented at the AIDELF Seminar, Byblos, October 10–13, 2000.)

Robine Jean-Marie, 2001. Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: the case of France, *Population, An English Selection*, vol. 13(1), p. 173–194.

Robine Jean-Marie and Allard Michel, 1999. Jeanne Calment: validation of the duration of her life, *in*: Bernard Jeune and James W. Vaupel (eds.), *Validation of exceptional longevity*, p. 145–162. Odense, Odense University Press, 249 p.

Robine Jean-Marie and Vaupel James W., 2001. Supercentenarians: slower ageing individuals or senile elderly? *Experimental Gerontology*, vol. 36, p. 915–930.

Rose Michael R., 1991. *The evolutionary biology of ageing*. Oxford, Oxford University Press, 221 p.

Thatcher Roger, 2001. The demography of centenarians in England and Wales, *Population, An English Selection*, vol. 13(1), p. 139–156.

Thatcher Roger, Kannisto Väinö, and Vaupel James W., 1998. *The force of mortality at ages 80 to 120*. Odense monographs on population aging, vol. 5. Odense, Odense University Press, 104 p.

Vallin Jacques, 1975. La mortalité en Algérie, *Population*, vol. 30(6), p. 1023–1046.

Vallin Jacques, 1993. De la mortalité endogène à l'allongement de la vie: hommage à Jean Bourgeois-Pichat, *in:* Jacques Vallin (ed.), *L'avenir de l'espérance de vie*, p. 23–36. Paris, INED-PUF, 106 p. (Congrès et colloques no. 12).

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978*. Paris, INED, PUF, 608 p. + 7 volumes annexes (travaux et documents, cahier no. 115).

Vallin Jacques and Meslé France, 2001a. Vivre au-delà de 100 ans, *Population et Sociétés*, vol. 365, p. 1–4.

Vallin Jacques and Meslé France, 2001b. *Tables de mortalité françaises pour les XIX^{e} et XX^{e} siècles et projections pour le XXI^{e} siècle*. Paris, INED, 102 p. + CD-rom (données statistiques, no. 4, 2001).

Vaupel James W., 2001. Demographic insights into longevity, *Population, An English Selection*, vol. 13(1), p. 245–260.

Vaupel James W. and Carey James R., 1993. Compositional interpretation of medfly mortality, *Science*, vol. 260, p. 1666–1667.

Vaupel James W., Carey James R., Christensen Kaar, Johnson Thomas E., Yashin Anatoli I., Holm Niels V., Iachine Ivan A., Kannistö Vaïno, Khanzaeli Aziz A., Liedo Pablo, Longo Valter D., Zeng Yi, Manton Kenneth J., and Curtsinger James W., 1998. Biodemographic trajectories of longevity, *Science*, vol. 280, p. 855–860.

Vincent Paul, 1950. Probabilités de mortalité aux âges élevés; tables de mortalité de vieillards, *Bulletin de l'Institut international de statistique*, vol. 32(2), p. 377–388.

Vincent Paul, 1951. La mortalité des vieillards, *Population*, vol. 6(2), p. 181–204.

Walford Roy, 1984. *La vie la plus longue*. Paris, Laffont, 250 p. (Translated from English, *Maximum life span*. New York, WW Norton, 1983.)

Williams George C., 1957. Pleiotropy, natural selection and the evolution of senescence, *Evolution*, vol. 11, p. 398–411.

Wilmoth John R. and Horiuchi Shiro, 1999. Rectangularization revisited: variability of age at death within human populations, *Demography*, vol. 36(4), p.475–495.

Wilmoth John R., Deegan Leo J., Lundström Hans, and Horiuchi Shiro, 2000. Increase of maximum life-span in Sweden, 1861–1999, *Science*, vol. 289(29), p. 2366–2368.

Yashin Anatoli I. and Iashin Ivan A., 1995. Genetic analysis of durations: correlated frailty model applied to survival of Danish twins, *Genetic Epidemiology*, vol. 12, p. 529–538.

Yashin Anatoli I. and Iashin Ivan A., 1997. How frailty models can be used for evaluating longevity limits: taking advantage of interdisciplinary approach, *Demography*, vol. 34(1), p. 31–48.

CHAPTER

49

Genetic Factors in Mortality

LAMBERTO SOLIANI AND ENZO LUCCHETTI

Università degli Studi di Parma, Parma, Italy

I. METHODS AND PROBLEMS IN HUMAN GENETICS

Does having long-lived parents increase the probability of living longer? Although the idea is widely accepted, there is neither proof of the connection nor a clear consensus on the mechanisms of this relationship. What is the fundamental factor determining long life? Is it the absence of hereditary pathologies, whose presence reduces the length of life, or is longevity inherited as a characteristic in and of itself (Vandenbroucke *et al*., 1984; Glasser, 1981; Hammond *et al*., 1971)?

Comparisons between ethnic groups using epidemiologic and demographic data produce similar conclusions. In Florence, for example, deaths from lung cancer are less common, and life expectancy is longer among immigrants from the south of Italy, especially those born in Sicily (Buiatti *et al*., 1985); in Australia, Italians have lower rates of coronary diseases than other national groups (Armstrong *et al*., 1983). In England, numerous differences in cause-specific mortality have been identified among people originating from different Commonwealth countries; analogous observations have been made in Israel between Jews of different provenance. Interesting correlations, often independent of socioeconomic and environmental conditions, have been uncovered in the United States by comparing the territorial concentration of national origin groups to the geographic distribution of principal causes of death (Mason *et al*., 1981; Morin *et al*., 1984). It is easy to assume the existence of hereditary factors that influence or determine the average life span, but in all these examples, it is difficult to separate biologic components from cultural components.

By 1918, before the development of modern genetics, Ronald Fisher (1890–1962) had interpreted observed correlations of certain individual characteristics within families as an indication of their hereditary character. Classic models of the study of hereditary pathologies are based on analysis of a family group (i.e., presence of the disease among ancestors, descendants, or collateral relatives) or of a population (i.e., distributions between groups of individuals as a function of the family ties that unite them). These methods have led to the identification of a number of hereditary diseases and the discovery of their modes of transmission.

Developments in biology over past decades have opened the way for a more precise exploration of the symptoms and modes of transmission of pathologies and for identification of the genes responsible for the diseases and the biochemical mechanisms at work: the precise location of genes, the mechanisms of expression, the genetic anomalies affecting the mutated alleles, and the process by which diseases occur. Cytogenetics has provided detailed descriptions of the morphology and structure of chromosomes, the physical seat of genes. Molecular biology has identified the biochemical underpinnings of numerous hereditary diseases. Understanding the structure of DNA, the process of genetic decoding, and the sequencing of DNA has made it possible in many cases to trace the alteration of genetic information behind the appearance of the disease. Genetic engineering introduces the possibility of direct intervention to replace or repair genes that carry anomalies.

Medical science has been enriched by diagnostic techniques that take information directly from the fetus (e.g., amniocentesis, culture of amniotic cells from the fetus) in addition to ultrasound methods that allow observation of the fetus during pregnancy. Genetic consultation has moved beyond primitive probabilistic diagnosis, based on analysis of the family tree and family memory, to the certainty of direct observations of the fetus and its genetic makeup. The fine tuning of methods that are inexpensive and only moderately invasive has made it possible in many countries to begin systematic programs of complete neonatal examinations for the diseases that are the most common and that have the most serious pathologic and social consequences.

II. GENETIC DISEASES

Diseases with genetic origins can be divided into three principal groups: pathologies caused by anomalies in the number or structure of chromosomes; pathologies stemming from mutation (i.e., variation in the informative contents) of a gene; and pathologies or predispositions to disease due to multifactorial inheritance. From the demographic point of view, the most important genetic diseases are those with a high frequency of occurrence, those that cause death at young ages, and those that are geographically diversified. We therefore limit this discussion of the genetic aspects of mortality to factors that have a significant influence on life expectancy and that can explain a non-negligible proportion of mortality differences between families and between ethnic groups.

1. Chromosomal Disorders

The karyotype of human somatic cells is made up of 46 chromosomes; it is produced by the union of two haploid cells, one from the egg and one from the sperm, with 23 chromosomes each. The resulting diploid cell contains 23 pairs of chromosomes; the elements of each pair (i.e., homologous chromosomes, one from the mother and one from the father) have the same structure and carry information on the same characteristics. All individuals therefore carry two pieces of information for each trait. The pair of chromosomes labeled sexual (the 22 other pairs are called autosomal) is an exception. In men, the pair contains a larger chromosome called the X chromosome and a smaller chromosome called the Y chromosome, and in women, the pair is made up of two X chromosomes. Among men, the X chromosome necessarily comes from the mother and the Y chromosome from the father; among women, one X chromosome comes from the father and the other from the mother.

Meiosis, the process of formation of the gametes, guarantees that each germinal cell (*gamete*) carries 23 chromosomes, exactly one element of each homologous pair; the union of two gametes (i.e., fertilization, formation of the *zygote*) reestablishes the number of 46 chromosomes. Chromosomal disorders result from anomalies in the number or structure of chromosomes. Incidents during the course of meiosis can produce gametes with an abnormal number of chromosomes. The zygote may, for example, contain an extra chromosome (i.e., *trisomy*, or three copies of one chromosome) or one chromosome too few (i.e., *monosomy*, or only one copy of a chromosome). More significant variations can involve several chromosomes, but in general, these defects do not reach the point of manifestation because they cause alterations that are incompatible with further zygotic development.

The chromosome's structure can also be altered by insertions (i.e., repetition of a section of DNA), deletions (i.e., missing sections of DNA), inversions (i.e., section rotated 180 degrees), and translocations (i.e., section transferred to another chromosome). Organisms can adjust for some structural modifications because the information contained on one section of DNA is present and even active if it was transposed onto another chromosome, allowing normal phenotypes to be expressed. Gametes produced by a healthy carrier can nevertheless contain abnormalities, because a chromosome may be missing a piece or possess extra copies, with effects analogous to those produced by monosomies and trisomies.

Alterations in the number of chromosomes usually result from accidents during the formation of gametes; they do not generally have a genetic component and are rarely repeated for the same progenitor. They appear more frequently as the progenitor's age increases. However, chromosomal anomalies of a structural nature can be repeated during the production of germ cells and can produce the same pathologies among several children of the same parent, even if the anomalies are compensated for and have no effects on the healthy carrier.

Chromosomal anomalies, even when they are compatible with the zygote's development, inevitably lead to very serious and complex illnesses. These anomalies are common at the moment of conception, but most malformations are eliminated during the earliest phases of development of the fertilized egg.

By combining data from numerous analyses, it has been estimated that about 15% of confirmed pregnancies end in spontaneous abortion, and that in almost one-half of these cases, the fetus carries a chromosomal anomaly. For the earliest miscarriages, occurring before 4 weeks' gestation, the frequency of chromosomal anomalies is between 90% and 100%; it falls to 60% for terminations in the fifth through eighth week and to 32% for terminations in weeks 9 through 12. Trisomies (52%) constitute the most common type of anomaly. Spontaneous abortions at 16 to 18 weeks' gestation show karyotypes with trisomy of chromosome 21 (i.e., Down syndrome) or with trisomy 18, which most often results in normal karyotypes. Autosomal monosomies are rarely observed in miscarriages. Their effects appear to be so serious that they make it impossible for the embryo to develop, and the embryo is eliminated at a very early stage, probably before implantation.

In 7% of couples who are sterile or who have had several miscarriages, one of the partners is a carrier of a chromosomal rearrangement that has been compensated for; the individual is healthy but produces abnormal germinal cells. This frequency is 20 times higher than that observed among couples who have not experienced spontaneous abortions.

Among infants born alive, the frequency of chromosomal anomalies is on the order of 1%. According to epidemiologic analyses, which normally exclude deaths in the first 24 hours, 1 birth in 160 to 200 presents an abnormal karyotype. Clinical manifestations are observed in 35% of cases; the most common are autosomal trisomies, certain variations in the number of sexual chromosomes (notably Klinefelter's syndrome and Turner's syndrome), and noncompensated transpositions.

The best known syndrome, Down syndrome, is caused by trisomy of chromosome 21. It appears once in every 450 to 750 births. It leads to a complex pathology, causing increased infant mortality or premature aging, and persons with Down syndrome require lifelong medical treatment. As in almost all syndromes of chromosomal aberration, there is variation in the expression of the various pathologies associated with the anomaly. Cardiac effects are common, and the risk of acute leukemia is 20 times higher than in individuals with normal karyotypes.

Chromosomal anomalies appear more frequently as the mother's age increases. It has been estimated that in the absence of voluntary abortion, clinically significant chromosomal anomalies would appear in about 0.5% of births to mothers age 35 years, 1.5% of births to mothers age 40 years, and 5% of births to mothers age 45 years.

Among live births, chromosomal anomalies are relatively uncommon. Their frequency is declining in developed countries with the spread of regular examinations and therapeutic abortion.

2. Single-Gene Disorders

Single-gene disorders are caused by mutations of genes or modifications in the nucleotide sequence of DNA. Insertions and deletions or substitutions of one nucleotide for another that modify the sequence can change the information contained in the gene; the related function (i.e., synthesis of a structural or functional enzyme or protein) may then be carried out incorrectly, leading to the appearance of a pathology. These point-specific mutations can be sex-linked or autosomal, depending on whether the gene in question is situated on a sex chromosome. Mutations are further distinguished by whether they are dominant (i.e., manifestation of disease when only one of the two alleles is mutated) or recessive (i.e., disease appears only when both alleles are mutated).

The catalog of mendelian inheritance among humans established by Victor A. McKusick in 1988 included 1443 dominant autosomal characteristics, 626 recessive traits, 139 traits linked to the X chromosome, and 2000 traits that has not been completely identified or validated. Great progress has been made during past decade in mapping the human genome, notably by the Human Genome Project, which was dedicated to complete identification of the human genome. The hereditary diseases most important for mortality include Tay-Sachs disease, cystic fibrosis, sickle cell anemia, hypercholesterolemia, and phenylketonuria; we provide a brief profile of these conditions to illustrate the diversity of their modes of genetic transmission, their symptoms, and their diffusion.

Tay-Sachs disease was first described by the English ophthalmologist Warren Tay in 1881 and the American neurologist Bernard Sachs in 1887. It is common among Ashkenazi Jews, established in northeast Poland, southern Lithuania, and parts of European Russia, but it is rare among Sephardic and Asian Jews. In the United States and in Canada, Tay-Sachs disease has an incidence of 1 case in 5000 births among Jews, as opposed to 1 case in 550,000 among non-Jews. Caused by a recessive autosomal trait, the disease is characterized by the accumulation of a particular type of lipid, gangliosides, in the neurons and the brain. Children with the disease have normal reactions at birth, but they deteriorate rapidly starting in the first year of life. By age 18 months, children can become blind, gradually lose hearing, have serious difficulties in digesting food, develop macrocephaly, or develop general paralysis. Most victims lose mobility by their second year. Death, often by respiratory infection, follows between the ages of 3 and 4.

Cystic fibrosis is one of the most common childhood diseases among whites, but it is extremely rare among other racial groups. It is caused by a recessive mutation of a gene on chromosome 5. This disease, like Tay-Sachs disease, is expressed only in homozygotes. It affects 1 newborn in 2500 in the general population; the frequency among healthy heterozygotes is 1 in 25. The disease leads to respiratory complications; in the absence of treatment, 95% of people with the disease die within 5 years. With treatment, life expectancy rises to 12 to 15 years for women and 15 to 20 years for men.

Among blacks, the most common autosomal recessive disease is *sickle cell anemia*, caused by the mutation of a gene on chromosome 11. This gene directs the production of hemoglobin, which becomes defective with the genetic mutation and is no longer able to transport oxygen. Red blood cells tend to cluster in the narrowest capillaries, preventing the transport of blood to tissues and organs. Homozygotes usually die between ages 15 and 30 years and rarely live past age 45. In West Africa, one child in 100 has the disease, and 1 person in 6 is a carrier. In the whole of the African continent, it is estimated that about 100 000 people die annually from sickle cell anemia.

A Mediterranean form, thalassemia major or Cooley's anemia, is common in malarial zones of that region. Heterozygotes (i.e., thalassemia minor) have overall advantages relative to healthy individuals, because they have a lower susceptibility to malaria. In the homozygote state, the allele produces serious anemia, which until recently was fatal in early childhood. The allele is still common in several formerly malarial areas, including the Po delta and several regions of southern Italy, Sardinia, and Greece, although the elimination of malaria there has removed the conditions favoring heterozygotes and suppressed the mechanism of conservation of the allele. In Asia Minor, the disease affects 1% of Armenians and Syrians, among whom one person in six is a heterozygote.

Familial hypercholesterolemia is an example of a disease linked to a dominant autosomal allele. Affecting an average of 1 in 500 individuals, it is responsible for 5% of all heart attacks before age 60. It is particularly common among Lebanese, South Africans, and French Canadians, isolated populations exhibiting evident founder effects. In adulthood, heterozygotes have a predisposition to heart attacks that is 25 times greater than that of normal individuals; among carriers, 50% of men and 35% of women die before age 60 years.

Phenylketonuria is a hereditary metabolic disease with strong biochemical and genetic heterogeneity, caused by an unknown number of DNA mutations. In Western industrialized countries, the incidence of the disease ranges from 1 case in 6000 to 1 in 20,000 newborns. For some time, a simple, inexpensive diagnostic test has been part of the examinations normally carried out at birth; the test permits a straightforward and effective intervention, starting in the first week of life, involving the adoption of a specific dietary regimen that can neutralize the risk of irreversible brain damage.

3. Multifactorial Diseases

Most phenotypic characteristics (e.g., height, bodily proportions, shape of the face and physiognomic characteristics, skin color, blood pressure), like much pathology, do not have a simple genetic origin but rather are determined by the interaction of several genes with each other and with the environment. Multifactorial heredity is usually expressed as a predisposition toward a disease whose phenotypic manifestation also depends on environmental context or behavior (e.g., diet, environment, lifestyle). Analyses of multifactorial heredity aim both to separate genetic factors from environmental determinants (to evaluate their relative weight) and to identify the genes involved and their modes of interaction. For diseases, analysis also seeks to identify people or groups at risk who may benefit from an intervention.

Coronary atherosclerosis, for example, recurs in families; its frequency in case families is two to six times higher than in control families, suggesting the presence of a hereditary component. Family affinity is stronger when the onset of the disease is early (i.e., the strength of the correlation increases as age at onset falls); the hereditary component is most marked in the most serious or earliest forms. The hypothesis is confirmed by the presence of a strong genetic component in specific pathologies (e.g., hyperlipidemia, hypertension, diabetes) associated with coronary diseases. The fact that family affinity is more marked when the victim is a woman suggests that sex plays a role, but this may just as well come from behavior as from biologic sex (see Chapter 53). The complex genetic-environmental interaction is confirmed by the fact that, even in the family of origin of spouses of a case family, disease frequency is stronger than in control families. In the absence of shared genetic makeup, family affinity may arise from a simple commonality of habits, behaviors, lifestyles, and diet (factors that also influence marital choice) and, from there, shared environmental conditions likely to provoke the disease.

Comparison of twins historically has constituted an elegant method of evaluating the genetic component of measurable characteristics. Monozygotic twins come from the duplication of the same zygote and share the same genetic makeup. Any differences between monozygotic twins are therefore due to environment and not to genes. Dizygotic twins are equivalent to ordinary brothers and sisters and, like them, share only 50% of their genetic makeup; the only difference between dizygotic twins and other siblings is that twins are born at the same time and undergo more strictly similar environmental conditions.

Family analyses and twin studies have uncovered a large number of multifactorial genetic diseases. Among the most common are schizophrenia, several forms of diabetes, hypertension, and cardiovascular diseases. Age is an important cofactor; its role is clearer for cardiovascular and neoplastic disease and for endocrine and metabolic disorders. With age and accumulated exposure to environmental factors—ranging from tobacco consumption to radiation and pollution—the immune system loses capacity. This results in rapid growth of the risk of disease.

The association between disease and the various modes of phenotypic expression of the same characteristic (i.e., polymorphic characteristics such as blood type or the human leukocyte antigens [HLA] system) is another promising instrument for classifying multifactorial pathologies. Groups of infectious diseases (e.g., plague; cholera; chronic infections like tuberculosis and syphilis; intestinal infections, particularly among children; infantile and juvenile tropical diseases), still important today in certain countries, strike less frequently or are less lethal for those who belong to certain phenotypic groups. For the most part, these are statistical observations whose biochemical mechanisms are not known.

The HLA system, an ensemble of several highly polyallelic genes situated on chromosome 6, regulates individuals' capacity for immune defense. Several hundred cases of strong statistical association between diseases and specific alleles of the HLA system are known, including ankylosing spondylitis, psoriasis vulgaris, rheumatoid arthritis, juvenile diabetes mellitus, and multiple sclerosis. Even allergies, autoimmune diseases, and immunodeficiencies (including severe combined immunodeficiency, which is inherited as a recessive monosomy in 50% of cases and leads to death by bacterial, viral, or mycotic infection as early as the first months of life) are understood as the consequences of antigen-antibody reactions that depend on the functioning of the HLA system. These diseases should not be confused with acquired immune deficiency syndrome (AIDS), which causes its complicated activation of vital cells in the immune system under the effects of a retrovirus, not from direct causes or from genetic predispositions.

Several generalizations can be made based on current knowledge of human tumors. Mendelian inheritance is rare, although it is suggested in the case of certain types of observed frequencies, such as for neurofibromatosis, multiple polyposis, and various neoplastic endocrine syndromes. The probability of a neoplastic transformation may be correlated with turnover of the somatic cells involved; we know from the biology of carcinogenesis that neoplasms increase with age. Chromosomal anomalies are common in cancerous cells but rarely specific; in certain cases, the risk of neoplasm is associated with a higher frequency of chromosomal ruptures, especially if these involve the systems of control of cellular division. When tumors appear, somatic cells are more often involved than germinal cells. Simple or polygenic heredity is a cause in 20% of tumor cases. For the most part, there is a general interaction of environmental agents, and the precise relations of cause and effect are unknown.

4. Mitochondrial DNA and Degenerative Diseases

Extranuclear genetic information (DNA) is contained in the mitochondria, cytoplasmic organelles that control the process of oxidation and energy metabolism, furnishing close to 90% of the energy necessary for cells and for the organism as a whole. Mitochondria contain their own DNA (mtDNA), and the structure, organization, and genome sequence of human mitochondrial mtDNA, as well as certain associated structural and regulatory functions, are known. The tissues and organs most often affected by a shortage of cellular energy are the central nervous system and, in decreasing order of the seriousness of symptoms, cardiac and skeletal musculature, kidneys, and endocrine tissues.

Mitochondria are inherited exclusively from the mother in the ovum, which contains an abundance of them. Mitochondria in the sperm are rare, degrade rapidly, and usually do not penetrate the oocyte during fertilization. The genetic information contained in the mitochondria of every cell of every individual comes from the mother. Because of the transmission method, there can be notable variation in the number and type of mitochondria among children of the same woman; as a result, even among siblings, the pathologies associated with the mitochondrial genome can have very different symptoms and levels of seriousness.

The list of diseases that can be provoked by mutations of mtDNA or that are at least are related to mitochondrial functioning continues to grow (Wallace *et al*., 1997). Among the diseases that are the most widespread and the most likely to be associated with mutations in mtDNA are Alzheimer's disease (i.e., progressive loss of cognitive capacity), chronic progressive external ophthalmoplegia (i.e., paralysis of the ocular muscles and mitochondrial myopathy), Leber's hereditary optic neuropathy (i.e., permanent or temporary blindness, provoked by the deterioration of the optic nerve), and Leigh's syndrome (i.e., progressive loss of motor and verbal capacity, with degeneration of basal ganglia, potentially lethal in childhood).

III. THEORIES ON AGING AND THE LIFE SPAN

Aging and death can be seen as biologically programmed phenomena, just like development after conception and growth during adolescence and up until maturity. Hormones, enzymes, and antibodies change during the individual's life, both in response to the environment and in reaction to genetic programs. Changes interact and accumulate, causing cellular anomalies, cell death, tissue dysfunction, and degeneration of the organism, resulting in senescence and death (Hayflich, 1965, 1980). There is no shortage of examples of biologically programmed death being activated at a given stage of development of the human organism. Around the seventh week of fetal life, for example, mullerian-inhibiting substance stimulates the degeneration of the primitive female reproductive system in boys; similarly, certain specific functions, such as the functions of genes that code for the synthesis of various types of hemoglobin, disappear or are activated at different ages.

The idea that aging and the individual's natural death depend on characteristics inscribed in cells dates from the end of the 19th century. The zoologist August Weismann wrote, as early as 1891, that death takes place to the extent to which worn tissue can no longer renew itself, and the capacity for growth through cellular division is limited and not eternal. In the early 20th century, Alexis Carrel's experiments on somatic cells in chickens, which seemed to subdivide infinitely, gave credit to the hypothesis of *potential immortality* of cells cultivated in vitro. In the 1960s, Leonard Hayflick showed that normal diploid fibroblasts (i.e., embryonic mesenchyme cells that later become the fundamental amorphous substance) in mammals are capable of only a finite number of mitoses (i.e., divisions); this number is estimated at 50 ± 10 for the human embryo. Other observations seem to confirm the hypothesis that the duration of cellular life is biologically programmed. When donors are old, their cells cease to divide; the species with the shortest life spans enjoy a lower number of divisions, although the relationship is not proportional; and the cells of individuals affected with forms of premature aging have reduced capacity for division (Schneider and Mitsui, 1976; Martin *et al*., 1970; Epstein *et al*., 1966).

The idea of being able to predict a person's life span has fascinated researchers. Recall Raymond Pearl's assertion (1928, p. 151): "The length of life depends inversely on the rate of living," in which *rate of living* is used to mean the rhythm of energy consumption. Others, including Charles Rose and Benjamin Bell (1971), have found a positive correlation between the length of life and the rate of activity.

Two principal groups of hypotheses address the molecular and genetic mechanisms likely to determine aging and death. The regulation of the dynamics of mitosis (i.e., duplication of somatic cells) depends on the mitotic clock, and cells lose their capacity for division because of the accumulation of errors, meaning that mitosis is no longer possible after a certain level is surpassed. Somatic mutations likely contribute to the process of natural aging, in the same way that they contribute to carcinogenesis, which also depends on age. The mechanisms and interactions with other cellular processes are still unknown; even the causes to which the errors should be attributed are not clear (Flodin, 1984).

In recent years, it has been proposed that "the hour of death" can be estimated from the length of telomeres; the ends of chromosomes, the telomeres, may regulate the capacity for cellular duplication. Although it is not known whether the phenomenon is a cause or an effect of aging, the telomere does shorten over the course of an individual's life. Various research projects have tried to quantify the observed relationship between the length of the telomere (easily measurable in a cellular culture) and the length of life remaining for an individual. The passage from cells to organism raises scientific objections, because not all cells in the organism have the same capacity for replication. However, by cloning an individual from cells whose telomeres have been elongated, some researchers believe that an individual may inherit greater longevity. It may be possible to positively affect the length of life by replacing sick organs with healthy new organs derived by culturing cells with the capacity for longer life.

In an article published in *Science* with the very promising title of "Extension of Life-Span by Introduction of Telomerase into Normal Human Cells," Andrea G. Bodnar and colleagues (1998) extended the telomeric theory of aging to the point of completion by demonstrating that an enzyme, telomerase, can prevent degradation of the telomere. In young cells, telomerase is abundant, and telomeres renew themselves rapidly; in aged cells, in which telomerase is scarcer or absent, the telomeres become progressively shorter, and the cell stops dividing and dies. Scientists have been able to clone telomerase extracted from the sperm and the egg and introduce it into human cells from different tissues; this procedure has resulted in an increased number of mitotic divisions.[1] Stéphanie M. Cascio (2001) wrote:

> Normal somatic cells have a finite life span due in part to their inability to maintain telomere length and chromosome stability. Immortalization strategies based on recent advances in telomere biology and aging research have led to the creation of genetically stable, nontumorigenic immortalized cell lines. Reversible immortalization, using the Cre-lox recombination and excision system, has been developed for the expansion of primary cells for cell based clinical therapies. Immortalized human hepatocyte cell lines with differentiated liver functions would find broad applications in biomedical research, especially for pharmacology and toxicology, artificial liver support, and hepatocyte transplantation.

In presenting these results, some members of the press have regarded telomerase as a veritable fountain of youth—a label that is scientifically misleading but strongly appealing to the sensitive cord of expectations aroused by discoveries in the biotechnology industry. Another promising aspect of this discovery relates to cancerous cells, whose high capacity for division may be related to a high concentration of telomerase. Useful therapeutic developments may be generated by reducing the quantity of telomerase in tumor cells, causing them to age and die after a few divisions.

Another theory of aging, which does not necessarily contradict the other theories, is based on mitochondrial functioning. The process of energy production in cells can generate free radicals, which may increase the frequency of mutations of mitochondrial DNA. The accumulation of mutations may reduce the capacity for energy production and accelerate aging. It has been shown that the production of energy declines with age as somatic mutations linked to mitochondrial DNA increase, especially in the brain, skeletal and cardiac muscle, skin, and liver (Wallace, 1992). Proponents of this hypothesis have demonstrated that laboratory animals submitted to a restrictive diet producing a lower amount of free radicals accumulate fewer damages in their mtDNA, staying healthier and living longer than controls allowed to eat freely.

Free radicals attack all cellular structures: the nucleus, the mitochondria, and especially the membrane. The cellular membrane protects the cell, regulates the exchange of substances between environments inside and outside the cell, and activates numerous, vital biochemical reactions. Free radicals, which are produced during normal metabolic processes, can multiply under the effects of environmental factors (e.g., pollution), lifestyle factors (e.g., tobacco use, high-calorie diets that are rich in toxic residue), and disease (e.g., inflammations, medications). They reach elevated levels in cancerous cells. Their action can be counteracted by antioxidants, substances usually found in a diet rich in polyunsaturated fats that may also block the growth of cancerous cells. Pharmaceutical companies propose a number of multivitamins and minerals that are expected to reduce free radicals and their aging effects. This hypothesis demonstrates the importance of environment, both external and internal to individuals and their cells, in the biologic determination of the life span.

IV. SEX

In societies in which the effect of environmental factors on mortality has been reduced, we have seen a long period during which women's life expectancy has grown more rapidly than men's (see Chapter 53). Explanations for this difference have been sought in the factors differentiating the behavior and characteristics of the two sexes (Vallin, 1993 and 1999). At the biologic level, there are four types of sexual differentiation among humans: chromosomes (XX or XY), gonads (ovary or testicles), phenotype (internal and external sexual traits), and psyche.

Women and men differ in their chromosomal makeup. For women, the pair of sex chromosomes contains two identical elements, X chromosomes; in men, a single X chromosome is paired with a smaller chromosome, called Y. The primary function of genetic factors determining sex is to produce the gonads, which spur the development of other sexual organs and secondary sexual characteristics. Unlike other species, in humans it has been shown that the presence of a Y chromosome, not the absence of an X, determines the development of the testicles; it is not the lack of a "feminine" chromosome that determines the male but the presence of a specific "masculine" chromosome.

[1] The authors specify (p. 352) that biotechnologic applications of this phenomenon may promote the extension of human life: "Cloned normal diploid cells could replace established tumor cell lines in studies of biochemical and physiological aspects of growth and differentiation; long-lived normal human cells could be used for the production of normal or engineered biotechnology products; and expanded populations of normal or genetically engineered rejuvenated cells could be used for autologous or allogenic cell and gene therapy. The ability to extend cellular life-span, while maintaining the diploid status, growth characteristics, and gene expression pattern typical of young normal cells, has important implications for biological research, the pharmaceutical industry, and medicine."

More than 100 genes are attributed with certainty to the human X chromosome. These genes control various characteristics and functions: red blood cells, vision, hearing, the nervous system, the muscular system, teeth, skin, and the metabolism of glucose. Pathologies caused by recessive alleles located on the X chromosome have different probabilities of expressing themselves according to sex. Because men are hemizygous, the allele carried on the unique X chromosome is necessarily expressed; among women, the effects of a harmful or lethal allele can be counteracted by the presence of a normal allele on the second X chromosome. A heterozygote woman carrying a harmful or lethal recessive allele will not develop the associated disease (i.e., healthy carrier), but one-half of her male children will be affected by the disease. Her daughters will all be healthy, but one-half will be carriers of the damaged allele.

The best-known example of recessive sex-linked heredity is that of hemophilia A, a disease affecting the male descendants of Queen Victoria and carried by her daughters, healthy carriers, into many European royal families. A simple numeric example suffices to give the order of magnitude of the difference between male and female risks of developing the disease. A damaged allele with a frequency of 1 in 1000 will be expressed once in every thousand male births and only once in a million female births. Even if the disease is lethal and no affected individual can reproduce, the population will carry about 2 per 1000 female healthy carriers, and one-half of their children will receive the lethal gene (and be automatically sick if they are male or healthy carriers if girls).

As early as the 1920s, despite the still imprecise definition of pathologies at that time, it was observed that there are more men than women among individuals showing mental retardation. Since the late 1970s, it has been known that a form of mental retardation is linked to the X chromosome. This is the most common form of hereditary mental retardation after Down syndrome. However, the biochemical mechanism is not known.

Men's greater risk of developing hereditary diseases or dysfunctions linked to the X chromosome has often been invoked to explain lower male life expectancy. Errors in the duplication of DNA accumulate in somatic cells over time. When these errors involve genes on the X chromosome, they are always expressed in men, because of their hemizygous condition, but they can remain silent among women thanks to their second X chromosome. Women therefore benefit from a greater resistance to anomalies that occur over the course of life. The hypothesis is not disproved by the theory of random inactivation of one of the female's two X chromosomes (i.e., the female, despite possessing two X chromosomes, would have the same number of active genes as the male as the result of a compensation mechanism inactivating one of the two X chromosomes).

The X chromosome shows less polymorphism than other autosomes. It seems to be better preserved over the course of evolution, possibly by the greater intensity of selection to which it is subjected among males, in whom abnormal mutations are always expressed.

Based on numerous clinical and experimental observations, it is clear that sexual hormones have a strong influence on the metabolism of lipids and lipoproteins and on atherosclerotic diseases; this influence produces health advantages in women of reproductive age, especially in more developed countries. This would explain the lower mortality of women, as well as the smaller social and geographic differences in mortality observed among women. From the psychological or behavioral point of view, women often show greater concern for their health.

V. BIOMETRIC MODELS

In the early 20th century, Ronald Fisher (1918) proposed a method for estimating coefficients of heritability, based on the variance and correlation coefficients between parents and children, between brothers and sisters, or between spouses. Relative to historical methods based on biometrics, modern genetics has the advantage of identifying the biochemical causes of hereditary diseases, providing clear indications for diagnosis and treatment. Knowing the substance whose absence (or excess) leads to the disease is an extremely important piece of operational information; administering (or reducing) the substance makes it possible to eliminate the symptoms of the disease. Being able to identify the genetic defects behind the shortage (or excess) of the incriminating substance also leaves open the possibility of eradicating the cause itself. Nevertheless, neither the genealogic or mendelian approach nor biometric methods of inquiry among families and populations have lost their value; they are still the most appropriate methods for evaluating the relative weight of genetic and environmental components in the case of multifactorial characteristics or pathologies, especially for preliminary analyses.

In the early 1950s, Jean Bourgeois-Pichat (1952) proposed a graphical method for dividing infant mortality into portions attributable to "endogenous" deaths (i.e., deaths from hereditary defects, congenital malformations, obstetric trauma, or similar causes) and "exogenous" deaths (i.e., linked to external factors) (see Chapter 48). This method remains one of the most elegant demographic models for evaluating the biologic and environmental components of mortality. In this method, a logarithmic function of age in days is plotted along the x-axis of a system of cartesian axes, and the cumulative number of deaths occurring by each age is plotted along the y-axis. This produces a curve that has a logarithmic shape for the first months of life and then becomes rectilinear up until age 1 year. This curve can be interpreted as the result of the rapid drop-off in endogenous deaths after the first month of life combined with the linear distribution of exogenous deaths (as a function of transformed age) over the course of the following months. The trend line of the graph at age 1 month can then be extended back to the y-axis to estimate the number and the relative incidence of endogenous deaths (see Figure 48–1 in Chapter 48). The evolution of these two components of infant mortality illustrates their different sensitivities to medical progress. Exogenous mortality diminished rapidly with the development of modern medicine and the improvement in sanitation and nutrition conditions, whereas endogenous mortality was more resistant and seemed to be heading asymptotically toward a limit below which it is very difficult to drop. The decline in infant mortality and the resulting increase in life expectancy come largely from declines in exogenous mortality; there has been a concomitant increase in the relative proportion of endogenous deaths, on which the influence of medical progress has been much lower.

Demographic analysis of mortality rates by age can be used to measure differences between populations and to evaluate the effects of genotype. The standard population method of calculating standardized or normalized rates eliminates the effects of differing age structures (see Chapters 15 and 42). By applying this method separately by sex to different ethnic groups with similar socioeconomic characteristics, it is possible to evaluate the total effect of ethnic difference, a largely hereditary factor (Kitagawa and Hauser, 1973; Kitagawa, 1964).

The existence of a hereditary component in the determination of life span implies that populations with similar socioeconomic and environmental conditions would maintain relatively constant mortality differences at older ages and that these differences would be correlated with the distance between their genetic makeups. Italy, thanks to its geographic configuration, the historical separation between regions, and its migratory flows, presents optimal conditions for a territorial analysis. As early as the 1940s, Livio Livi (1941) held that territorial differences in mortality at older ages were the result of biologic differences between populations showing significantly different patterns of longevity. Based on the same idea, Corrado Gini predicted with great foresight a longer life expectancy for southern regions once infant mortality was reduced to the level then prevalent in the northern regions. The strong regional correlation between genetic distances and the slope of the mortality curve for older ages in Italy furnishes direct demonstration of this thesis (Soliani, 1984).

These correlations are necessary but not sufficient to demonstrate that ethnic or territorial differences in mortality are solely attributable to heredity (Caselli and Egidi, 1980, 1988). Within the nuclear family and across the family tree, values and behaviors that surely influence both cause of death and survival, such as nutrition and lifestyle, are transmitted along with genes.

VI. PERSPECTIVES

Despite scientific progress in formal and molecular genetics and in medicine, it must be recognized, as Albert Jacquard (1980) observed, that the commonly accepted role of biology in the determination of human longevity has never been conclusively demonstrated.

Knowledge about many biologic and genetic aspects of specific diseases has increased. It is now possible to counteract the effects of many causes that reduce longevity by means of medication or by acting on environmental factors. With the recent transfer of research from universities to industry (and more generally, from the public sector to the private sector), capital investment, and the organization of groups, there has been a significant increase in the implementation of practical and marketable solutions. Medicine, biotechnology, and biodemographic research have opened highly promising paths. One approach using gene therapies seeks to eliminate the negative effects of hereditary defects and acquired illnesses and to identify and to provide genes that are positively advantageous for survival. A second approach, building on the reconstitution of personal genealogies in regions with good health data, allows the identification at birth of people at risk of diseases that are at least associated with family predispositions (i.e., prevalence determined by culture or environment) if not fully hereditary (i.e., prevalence determined by genetics).

In medicine and biotechnology, a historical change is presaged by the announcement of the possibility of direct action to cure or treat hereditary or acquired diseases using genetic information carried by the individual. The conceptual phase is almost complete, although the development of applications in humans is still in the early stages (Brenner, 1995). In principal, it is possible to replace a defective gene with its normal version; in practice, the desired insertion is not yet feasible in humans. Certain genes are useful only if their expression is regulated such that the specified protein is produced in an exact quantity, at the right moment, and for the necessary period. Happily, in many cases, such precise regulation is unnecessary; in these cases, the solution may be more straightforward.

Genetic research is primarily directed at pathologies caused by gene mutations in the germline, but its discoveries also apply to many acquired diseases that have a genetic component (i.e., predisposition) or result from mutations in information at the level of somatic lines. Research can then attempt to correct or eliminate the harmful effects of these diseases. For instance, more than one-half of the clinical experiments in genetic therapy deal with cancer, which is only partially hereditary. Many of these approaches have not yet gone beyond preliminary clinical tests, but the medical research world considers their future developments to be very promising (Rosenfield and Curiel, 1996).

Biotechnology applications to correct genetic information are still limited to intervention in somatic cells. These techniques cannot be applied to germ cells until there is more complete information on their consequences for future generations. It is already possible to produce twins by cloning, although this is not done in humans for bioethical reasons. In theory, gene therapy affecting the germline could correct pathologic alterations in all of the children of a couple and in all their descendants.

It is important not to confound the cell and the organism. For example, it would be possible to prevent the aging of certain cells by intervening in the key genes coding telomerase, thereby treating certain diseases and lengthening the average life span, but we would not have discovered the key to aging and to longevity. The human organism is a unique and extremely complex whole. Genetics should make it possible to understand the phenomena of aging better and to delay them selectively or to accelerate them (e.g. to hasten the death of cancerous cells). These gains would result in increased life expectancy, comparable to that which is obtained by successively replacing the defective parts of an automobile, but we would still not be able to extend the biologic limits of the life span.

Bioengineering holds out the prospect of discovering and using genes to guarantee a better use of actual vital potential and to increase longevity. Protection against the formation of atherosclerotic plaque, the obstruction that blocks blood flow, is carried out by a genetically controlled element regulating blood levels of a coagulation agent called factor VII. The action of this protective element can go so far as to reduce the risk of cardiovascular disease by a factor of 15. The lower incidence of cardiovascular diseases in Italy relative to the countries of Northern Europe may result from the higher allelic frequency of synthesis of this protective element (20% in Italy). Epidemiologic studies have already shown that a reduction in blood levels of factor VII obtained by the administration of low doses of an anticoagulant similarly diminishes the risk of cardiovascular incident. We now understand the basic genetic structures, and we can foresee, at least in theory, the recourse to genetic engineering to insert protective alleles in the genetic makeup of individuals.

In April 2000, the theory of telomeres was unexpectedly relaunched by an article in *Science* on the results of a new cloning method applied to six calves. Among the first animals cloned, such as the famous sheep Dolly, starting at a young age, cells in the clones were the same age as the donor cells (Shiels *et al*., 1999). Their telomeres were short, and they benefited from fewer replications than normal for an individual of the same age.[2] This posed a problem of a metaphysical nature: "Are her cells older than she is?" (Vogel, 2000). With the cloning of six calves from the nucleus of an adult cow, the results were contrary to all expectations. These calves had telomeres that were not only longer than the donor's, but also longer than those of ordinary bovines of the same age and even longer, in some cases, than those of calves at birth (Lanza *et al*., 2000). The cause is unknown (possibilities include random variations, different species, different types of cells, and different methods of nuclear transfer), but the researchers involved spoke of cellular rejuvenation: In these clones, cells divided an average of 93 times, whereas in a normal fetus, they divide an average of 61 times. In the journal's editorial, Gretchen Vogel (2000) cites Robert Lanza's remark underlining the importance of the discovery: "If this increased lifespan extends to the whole animal, there is a real possibility that cloned animals might live as much as 50% longer than their normal counterparts, up to 180 to 200 years in the case of humans."[3]

[2] The authors specify: "Recent veterinary examination of the nuclear-transfer animals has confirmed that they are healthy and typical for sheep of their breeds, despite having a shorter mean TRF (terminal restriction fragment) length."

[3] The authors add (p. 665): "Here we investigate whether the production of live offspring is possible by nuclear transfer of late-passage somatic cells and whether the epigenic changes seen in the donor cells, such as telomere shortening and loss of replicative lifespan, are reflected in the resultant organism."

Moving from animals to humans, the possibility of regenerating telomeres and "rejuvenating" cells is important not for cloning new individuals, which is ethically forbidden, but for replacing diseased or aged cells and organs with young and healthy cells. These results make it possible to imagine creating a clone that in 6 or 7 days would become a cluster of multifunctional cells at a relatively early stage of life, able to develop into any organ or tissue—from the brain to the heart—after transplantation into the failing organ of a sick person. Three methods of forming such omnipotent cells are known, starting from the union of a woman's ovum and a human cell, the union of an animal ovum and a human cell, or human cells that have not yet specialized.

The first technique, known as human-human, consists of emptying an ovum of its nucleus and replacing it with the nucleus of an adult cell. Even if the objective is only to form a pre-embryo (in medicine, the word *embryo* is used only after 14 days of gestation) to use cells that have been alive several days for exclusively therapeutic goals, many fear that the next step will be to create a human clone; more radically, this ensemble of still-undifferentiated cells is already seen as an individual. This situation invokes universal law and the laws of numerous countries that forbid the deliberate creation and destruction of a human being for the benefit of another. With the second method, called human-animal, for which it is an animal ovum whose nucleus is withdrawn and replaced with the nucleus of an adult human cell, the resulting cells adapt too slowly in the diseased human organism. The third path takes progenitor cells called *stem cells* from adult tissues (in this case, from the blood) and transforms them into mature cells that can be reprogrammed according to the patient's needs. This method is more complex than the first two, but it seems to many to be the only method that can reasonably be applied.

The laws in many countries have been overtaken by the rapid evolution of techniques for which we now lack clear and coherent norms. Even the sentence "The creation of human embryos for research purposes is prohibited" (article 18, paragraph 2) in the European Bioethics Convention of 1996 can be interpreted in different ways. For some, it forbids all use of pre-embryos, even for medical purposes, whereas for others, it forbids only the formation of embryos or individuals through cloning. Future scientific and medical results depend on the outcome of this political debate and the more precise guidelines on human cloning that will be provided by ethics committees.

Developments in computing have opened new horizons for personal biodemographic research, allowing applications that have until now been inconceivable. Reconstitutions of individual histories can be extended to large populations and, most importantly, enriched by automatic linkages with different cause-of-death databases from hospital, clinical, and pharmaceutical data. Demographic history, genealogic history, and medical history can be brought together to identify people at risk. In the United States and in Canada (Bouchard *et al*., 1991), linking millions of data on the populations of entire regions has made it possible to identify people at risk and have them tested for genetic predispositions to cardiac diseases, depression, osteoporosis, obesity, asthma, and different types of cancer. This is an example of mass testing that may become current practice in the near future. Thanks to now-classic statistical techniques, it should soon be possible to measure the heredity of biologic and genetic factors of other characteristics that we do not yet know with precision, such as the length of life.

Acknowledgment

This chapter was translated from the French version by Sarah R. Hayford.

References

Armstrong Bruce, Margetts Berens, Margetts B.M., Masarei J.R.L., and Hopkins S.A., 1983. Coronary risk factors in Italian migrants to Australia, *American Journal of Epidemiology*, vol. 118(5), p. 651–658.

Bodnar Andrea G., Ouellette Michel, Frolkis Maria, Holt Shawn E., Chiu Choy-Pik, Morin Greb B., Harley Calvin B., Shay Jerry W., Lichtsteiner Serge, and Wright Woodring E., 1998. Extension of life-span by introduction of telomerase into normal human cells, *Science*, vol. 279, p. 349–352 (www.sciencemag.org).

Bouchard Gérard, de Bbraekeller Marc (ed.), 1991. *Histoire d'un génome: population et génétique dans l'est du Québec*. Québec, Presses de l'Université du Québec, 607 p.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalitá biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Brenner Malcolm K., 1995. Human somatic gene therapy: progress and problems, *Journal of Internal Medicine*, vol. 237(3), p. 229–239.

Buiatti Eva, Geddes Marco, Kriebel David, Santucci Marco, and Biggeri Annibale, 1985. A case control study of lung cancer in Florence, Italy. II. Effect of migration from the south, *Journal of Epidemiology and Community Health*, vol. 39(2), p. 251–255.

Cascio Stephanie, 2001. Novel strategies for immortalization of human hepatocytes, *Artificial Organs*, vol. 25(7), p. 529–538.

Caselli Graziella and Egidi Vivviana, 1980. *Le differenze territoriali di mortalità in Italia. Tavole di mortalità provinciali 1971–1972*. Rome, Istituto di Demografia, 168 p.

Caselli Graziella and Egidi Viviana, 1988. Les variations géographiques de la mortalité, *in*: Jacques Vallin and France Meslé (eds.), *Les causes de décès en France de 1925 à 1978*, p. 507–607. Paris, INED, PUF, 607 p. (travaux et documents, cahier 115).

Epstein Charles J., Martin George M., Schultz Amelia L., and Motulsky Arno G., 1966. Werner's symdrome: a review of its symptomatology, natural history, pathological features, genetics and relationship to the natural aging process, *Medicine*, vol. 45(3), p. 177–221.

Fisher Ronald A., 1918. The correlation between relatives on the supposition of Mendelian inheritance, *Transaction of the Royal Society* (Edinburgh), vol. 52, p. 399–433.

Flodin Nestor W., 1984. The senescence of postmitotic mammalian cells: a cell clock hypothesis, *Mechanism of Ageing and Development*, vol. 27, p. 15–27.

Glasser Michael, 1981. Is longevity inherited? *Journal of Chronic Diseases*, vol. 34, p. 439–444.

Hammond E. Cuyler, Garfinkel Laurence, and Seidman Herbert, 1971. Longevity of parents and grandparents in relation to coronary heart disease and associated variables, *Circulation*, vol. 43(1), p. 31–44.

Hartl Daniel L., 1983. *Genetica Umana*. Bologna, Zanichelli, 487 p. (translation of. *Human Genetics*. Harper & Row, Inc. 1983).

Hayflick Leonard, 1965. The limited in vitro lifetime of diploid cell strains, *Experimental Cell Research* vol. 37, p. 614–636.

Hayflick Leonard, 1980. Biological aspects of aging, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 223–259. Liège, UIESP, Ordina éditions, 484 p. (Proceedings of the conference on biological and social aspects of mortality and the length of life, Fiuggi Terme, Italy, May 1980.)

Jacquard Albert, 1980. Heritability of human longevity, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of lfe*, p. 303–314. Liège, UIESP, Ordina éditions, 484 p. (Proceedings of the conference on biological and social aspects of mortality and the length of life, Fiuggi Terme, Italy, May 1980.)

Kitagawa Evelyn M., 1964. Standardized comparisons in population research, *Demography*, vol. 1(1), p. 296–315.

Kitagawa Evelyn M. and Hauser Philip M., 1973. *Differential mortality in the United States: a study in socioeconomic epidemiology*. Cambridge, MA, Harvard University Press, 205 p.

Lanza Robert P., Cibelli Jose B., Blackwell Catherine, Cristofalo Vincent J., Francis Mary K., Baerlocher Gabriela M., Mak Jennifer, Schertzer Michael, Chavez Elizabeth A., Sawyer Nancy, Lansdorp Peter M., and West Michael D., 2000. Extension of cell life-span and telomere length in animals cloned from senescent somatic cells, *Science* vol. 288, p. 665–668 (www.sciencemag.org).

Livi Livio, 1941. *Trattato di demografia. I fattori bio-demografici nell'ordinamento sociale*. Padova, CEDAM, Casa Editrice dott. Antonio Milani, 268 p.

Mange Arthur P. and Johansen-Mange Elaine, 1990. *Genetics: human aspects*, 2nd ed. Sunderland, MA, Sinauer Associates, 591 p.

Martin George M., Sprague Curtis A., and Epstein Charles J., 1970. Replicative lifespan of cultivated human cells: effect of donor's age, tissue and genotype, *Laboratory Investigation: A Journal of Experimental Pathology*, vol. 23(1), p. 86–92.

Mason Thomas J., Fraumeni Joseph J.R., Hoover Robert, and Blot William J., 1981. *An atlas of mortality from selected diseases*. Washington DC, National Institutes of Health, 309 p. (DHHS publication no. 81-2397.)

McKusick Victor A., 1988. *Mendelian inheritance in man: catalogs of autosomal dominant, autosomal recessive and x-linked phenotypes*, 8th ed. Baltimore, Johns Hopkins University Press, 1626 p.

Morin Michele, Pickle Linda, and Mason Thomas, 1984. Geographic patterns of ethnic groups in the United States. *American Journal of Public Health*, vol. 74(2), p. 133–139.

Pearl Raymond, 1928. *The rate of living: being an account of some experimental studies on the biology of life duration*. London, University of London Press, 185 p. (Reprinted in the New York, Knopf, 8th ed.)

Rose Charles L., Bell Benjamin, 1971. *Predicting longevity: methodology and critique*. Lexington (Massachusetts), Heath Lexington books, 265 p.

Rosenfield Maryland E. and Curiel David T., 1996. Gene therapy strategies for novel cancer therapeutics, *Current Opinion in Oncology*, vol. 8(1), p. 72–77.

Schneider Edward L. and Mitsui Yohio, 1976. The relationship between in vitro cellular ageing and in vivo human age, *Proceedings of the National Academy of Science of the USA*, vol. 73, p. 3584–3588.

Shiels Paul G., Kind Alexander J., Campbell Keith H., Waddington David, Wilmut Ian, Colman Alan, and Schnieke Angelika E., 1999. Analysis of telomere lengths in cloned sheep, *Nature*, vol. 399, p. 316–317 (www.nature.com).

Soliani Lamberto, 1984. L'ipotesi biologica sulle diffrenze territoriali di mortalità in Italia, *in*: *Atti della XXXII riunione scientifica della Società Italiana di Statistica*, Sorrento, 11–13 Aprile, vol. 2, p. 271–283. Rome, Società Italiana di Statistica, 528 p.

Vallin Jacques, 1993. Social change and mortality decline: women's advantage regained or achieved? *in*: Nora Federici, Carol Manson, and Solvi Sogner (eds.), *Women's position and demographic change in the course of development*, p. 190–212. International studies in population. Oxford, Clarendon Press, 368 p.

Vallin Jacques, 1999. *Mortalité, sexe et genre*. Liège, UIESP, 42 p. (Le genre dans les études de population.)

Vandenbroucke Jan, Maatros Ans W., van der Heide-Wessel Cornelia, and van der Heide Roelof, 1984. Parental survival and independent predictor of longevity in middle-aged persons, *American Journal of Epidemiology*, vol. 119(5), p. 742–750.

Vogel Friedrich and Motulsky Arno G., 1986. *Human genetics—problems and approaches*. Berlin, Springer-Verlag, 807 p.

Vogel Gretchen, 2000. In contrast to Dolly, cloning resets telomere clock in cattle, *Science*, vol. 288, p. 586–587 (www.sciencemag.org).

Wallace Douglas C., 1992. Mitochondrial genetics: a paradigm for aging and degenerative diseases? *Science*, vol. 256, p. 628–632.

Wallace Douglas C., Brown Michael D., and Lott Marie T., 1997. Mitochondrial genetics, *in*: David L. Rimoin, Michael J. Connor, and Reed E. Pyeritz (eds.), *Emery and Rimoin's principles and practice of medical genetics*. New York, Churchill-Livingstone, 2037 p.

Weismann August, 1891. *Essays upon heredity and kindred biological problems*, vol. 1, 2nd ed. Oxford, UK, Clarendon Press, 455 p.

CHAPTER

50

The Environmental Factors of Mortality

FRANCIS SARTOR

Institut Scientifique de la Sante Publique, Brussels, Belgium

The International Council for the French Language proposed the following definition for the term *environment*: "the group of physical, chemical, and biologic elements and the social elements liable, at a given time, to have a direct or indirect, immediate or long-term effect on living beings and on human activities" (Aguesse, 1971). In the broadest sense, the environment combines three elements: a socioeconomic component, a psychosocial component, and an ecologic component. The first governs human activities aimed at the production, distribution, and consumption of goods and wealth. The second influences attitudes and individual behavior in various areas such as diet and addiction (e.g., tobacco, alcohol, drugs). The ecologic component refers, among other things, to the physicochemical qualities of the environment surrounding people's homes. From all the evidence, exogenous deaths are caused by the environmental characteristics of a population.

Neither the influence of the three environmental components on mortality nor *a fortiori* their interactions are studied in this chapter. The discussion is limited to an inventory of the physical, chemical, and biologic elements that are causes of avoidable deaths. The influences of the social component and lifestyles on exogenous mortality are studied in other chapters, notably Chapter 51. Every year, earthquakes, volcanic eruptions, hurricanes, avalanches, and landslides cause many premature deaths on a global scale. The impact on exogenous mortality of these natural disasters associated with structural elements of the environment is not considered in this chapter.

The influence of the environment's harmful physicochemical elements on mortality can be quantified by estimating the absolute number of premature deaths or the reduction in life expectancy attributable to these risk factors, or both. An impact study of this kind would also define the contribution of the elements in question to exogenous mortality. For that purpose, we need reliable data concerning the scale of the risks and the size of the populations exposed to the risk factors under consideration. This fundamental information is lacking for most of the risk factors discussed in this chapter, and projecting any kind of estimate using our current information is unrealistic.

The qualitative study of causes of exogenous deaths associated with the physicochemical factors of the environment also poses methodologic problems. In this chapter, we establish that mortality increases particularly in cases of severe environmental pollution. Excess mortality generally occurs in the most vulnerable population groups: young children, the elderly, and fragile people (e.g., invalids). In a population exposed on a single occasion or over a prolonged period to harmful environmental factors, we observe a sequence of biologic effects of different types, ordered according to the increasing gravity of the effects. At the beginning of the process, we find signs of biochemical changes measurable at cellular levels among a large number of people. The significance of these early effects is sometimes difficult to establish. The reversible symptoms, such as coughs, that are observable in the population at a lower frequency follow in this hierarchy. The established cases of illnesses that do or do not necessitate hospitalization, define the next terms of the sequence, and they concern an even smaller number of persons. Death constitutes the final effect of the process. The frequency of each of these effects in the population is therefore assumed to vary in inverse relation to their seriousness. In some conditions of exposure, only morbidity may be affected without any immediate effects on mortality. However, even a temporary increase in morbidity can be the source of a decrease in life expectancy.

Establishing the link between an environmental factor and such a diluted and delayed excess mortality proves to be a very difficult task within the framework of classic demographic methods. Moreover, the causes of death do not clearly reflect the initial causes of illness, especially when the period of latency is very long, as is the case for cancer. To analyze the causes of such illnesses, it is necessary to have individual data, whereas demographers generally use aggregate data. The link between air pollution and lung cancer, for instance, can be studied by examining the evolution of the mortality rates of lung cancer in several towns according to risk indicators such as the degree of urbanization and industrialization of those towns. Some sources of systematic error inherent in such an approach cannot be avoided, even if the populations studied are comparable from the point of view of smoking frequencies. The influence of individual risk factors, such as duration of residence, exposure to carcinogenic elements in the work environment, mobility, and characteristics of the housing, are not easy to assess using aggregate data. Analytic epidemiologic studies using individual data collected during case-control or cohort surveys can better assimilate confounding elements. From a methodologic point of view, the difference between analytic epidemiologic studies and demographic studies is that the former rely on population samples, whereas the latter usually study entire populations. Theoretically, it would be possible to carry out demographic studies using individual data as long as it is possible to match information contained in different data sets, such as national registers, cancer registers, and computerized social security records. Nevertheless, this approach remains problematic in many countries because of issues such as data confidentiality and the protection of privacy. In many public health problems related to physicochemical environmental elements, the possibility of a reduction in life expectancy with an increase in morbidity cannot be excluded. It is for this reason that the influence of environmental pollution on mortality is in many cases examined by means of morbidity.

The structure of the chapter is based on the essential ecologic components: the physical, chemical, and biologic elements. This may be questionable, particularly because these elements usually do not act independently. Moreover, the distinction between physical elements and chemical elements is not always clear; radon which will be considered as a physical factor and a chemical factor, is an example of this lack of distinction, which can be confusing. As a product affiliated to uranium and radium, radon can be classified as a physical agent with radioactive properties, but when it is found in the air in a dwelling, it can be considered a chemical pollutant of a gaseous state emitting radioactive particles. Other subdivisions could have been used to write this chapter.

The chapter title suggests that we could first study the influence of the natural environment on human health, followed by the effects of human-induced environmental changes on morbidity and mortality. This method is not necessarily the most interesting, and even though it helps to fill certain gaps inherent in the approach used in this chapter, it can cause other problems. In the first place, we stumble over the difficult problem of defining the meaning of natural environment. Then there is the inevitable overlap due to the interaction of elements. In what context should the problem of excess mortality be treated when it is due to heat waves if the temperature is considered as a factor of the natural environment and air pollution from ozone caused by traffic is responsible for the high temperatures and exacerbate the effect of the same temperature on health? In health problems linked to the environment, there will always be a conjunction or an interaction between various elements of the ecologic component to which socioeconomic and psychosocial influences must be added. The main characteristic of health problems linked to the environment is the contribution of complex phenomena of different types. A global vision of these phenomena is therefore necessary to evaluate the scale of these problems and to measure the impact of each of the various risk factors. The principal objective of studying environment and health relationships is to identify these risk factors, not to make a systematic classification of the nature of these elements.

I. PHYSICAL FACTORS

1. Climatic Conditions

The extent of daily deaths from all causes varies according to outside temperature according to a V- or U-shaped curve (Kunst *et al.*, 1993; Kilbourne, 1990).

The *optimal temperature*, also called the *comfortable temperature*, at which the risk of death is minimal depends on geographic location. For daily deaths from all causes recorded in St. Louis, Missouri (Kilbourne, 1990), the optimal temperature is about 25°C, whereas it is 17°C according to mortality data collected in the Netherlands (Kunst *et al*., 1993). This difference suggests the existence of a phenomenon called *acclimatization*. The same phenomena, the U-shaped curve and geographic variations of optimal temperatures, have also been observed to affect cardiovascular and cerebrovascular diseases (Pan *et al*., 1995; Rogot and Padgett, 1976). Cardiovascular and cerebrovascular excess mortality in extreme temperatures can be explained by the phenomenon of the body's adaptation to cold and heat. Cold and heat trigger different mechanisms that increase blood pressure, heart rate, and the blood's ability to clot. The number of deaths caused by respiratory diseases increases during the winter months, and this constitutes a second factor that contributes to excess mortality from all causes at low temperatures. In the latter case, it is the reduction of the body's resistance to microbiologic elements and the increase in contagiousness due to a greater propensity for confinement, which increases the risk of death.

Death by hyperthermia (i.e., death due to increased body temperature after severe dehydration) is the second cause of death during heat waves, after deaths caused by cardiovascular and cerebrovascular diseases. Death by hyperthermia generally affects isolated people and invalids who cannot feed or take care of themselves (although air conditioning is a factor that protects from hyperthermia). Death by hyperthermia can also be the result of negligence (e.g., children abandoned in a car). The share of excess mortality attributable to hyperthermia nevertheless varies considerably according to each heat wave and does not seem to exceed 50% of total excess mortality (Kilbourne, 1990). Daily mortality appears to be influenced by other meteorologic factors, such as wind direction, humidity, rainfall, and snow (Kunst *et al*., 1993; Rogot and Padgett, 1976). Contrary to common belief, an increase in humidity during heat waves may be a protective factor in temperate regions (Kunst *et al*., 1993). Studies of the effect of global warming on mortality suggest a possible decrease in cardiovascular deaths during the winter among the elderly living in areas with moderate winter weather.

Deaths from skin cancer in the United States and Canada, an indirect method of measuring exposure to solar radiation, vary inversely with latitude; an increase of 10% in the annual intensity of irradiation by ultraviolet rays could therefore be linked to an excess mortality from skin cancers of 4% at 50° and 9% at 30° latitude (IARC, 1992; Elwood *et al*., 1974). Mortality data recorded in Australia and New Zealand for European immigrants, numerous studies relating to geographic variation of mortality, and incidences of these illnesses, as well as experimental data from animals conducted at the International Agency for Research on Cancer (IARC), have concluded that solar radiation is carcinogenic for humans (IARC, 1992). Between 65% and 80% of melanoma cases can be attributed to overexposure to sunlight (Rivers, 1996; Koh, 1995). The incidences of skin cancer and certain other ocular diseases such as cataracts may increase during the decades to come because of more intense irradiation of the earth; part of the ultraviolet rays are absorbed by the stratospheric ozone layer, whose depth is diminishing (Armstrong, 1994).

The degree of humidity and the temperature are the principal factors governing the proliferation of microscopic fungi (i.e., mold) likely to produce natural toxins, called *mycotoxins*, on certain foodstuff. One of these, aflatoxin B_1 can contaminate corn, wheat, barley, oats, rye, rice, sorghum, groundnuts (i.e., peanuts), and cotton plant seeds. Several studies, carried out mainly in Africa, have showed an increase of primary cancer of the liver in humans in accordance with the level of food contaminated by aflatoxin B_1 (WHO, 1980).

Desertification is the main cause of famines, which contribute greatly to mortality. The Sahel is a zone 300 to 500 km wide that borders on the south of the Sahara and forms the transition between the desert and the tropical forest on the west coast of Africa. In this semiarid zone, the soil needs to be humid during 2 months of the year for seeds to germinate and for roots to form. Otherwise, the entire harvest is delayed until a more favorable period. In the Sahel, rainfall has been decreasing for more than a century, with periods of extreme drought occurring, notably between 1915 and 1918 and between 1968 and 1972. The decrease in rainfall partly explains the encroachment of the desert by 100 km observed between 1900 and 1975 in the Sudan (Holdgate *et al*., 1982). In recent decades, overgrazing and the increase of agricultural land through deforestation have favored soil erosion and increased desertification. Certain regions in Australia, Brazil, Iran, Pakistan, Afghanistan, the Middle East, Northern Africa, and Libya are also affected by this phenomenon. In 1982, it was estimated that the food supply of 80 million people living in the regions recently affected by desert encroachment was insufficient as a result of the decline in agricultural production (Holdgate *et al*., 1982).

2. Radioactivity

Uranium and radium are two radioactive elements that naturally occur in soil, rocks, and water. Through their disintegration, the inert gas radon is formed, which is itself radioactive and has a life span of 4 days. Radon emits α particles, which are responsible for its carcinogenic action. In the 16th century, Agricola, in his book *De Re Metallica*, stated that exposure to radon was one of the causes of lung cancer among miners in central Europe. Between 1970 and 1980 in many countries, air pollution from radon inside dwellings was the main source of exposure to ionizing radiation; the radon was generated by radium in the soil and the rocks of the land on which the dwellings were built, or it came from the radium in some of the building materials, such as plaster. The extrapolations obtained from the risks observed in miners exposed to radon suggest that between 11% and 13% of all deaths from lung cancer each year of people living in single-family dwellings in the United States can be attributed to radon exposure (Lubin, 1994). A case-control study carried out in Sweden has confirmed that this radioactive element is an important risk factor for lung cancer (Pershagen *et al*., 1994). The results of this survey were not, however, corroborated by an independent study carried out at the same time in Canada (Látourneau *et al*., 1994). A definitive response to this important public health problem is expected in the coming years as at least 19 case-control studies on the effects of radon have been undertaken in recent years (Lubin, 1994). A preliminary meta-analysis carried out on eight case-control studies suggest that the risk of lung cancer increases by 14% for levels of radioactivity usually found in a significant number of dwellings. About 6%, 28%, and up to 40%, respectively, of dwellings in the United States, Stockholm, and in some regions of Finland are affected by this problem (Lubin and Boice, 1997).

Pollution of the environment by radioactive substances around nuclear power plants and factories can also be a risk factor for ill health. In England, geographic analysis of mortality statistics revealed abnormally high incidences of leukemia in children living in a village near the Sellafield nuclear power plant, which from 1952 had been producing plutonium and processing radioactive waste from nuclear power plants all over the country (Gardner and Winter, 1984). A possible relationship between radioactive pollution of the environment and leukemia was then confirmed by the results of a study of two cohorts of children (Gardner *et al*., 1987a, 1987b). A case-control survey later suggested that the appearance of leukemia in these children resulted from their fathers' exposure to ionizing radiation in the workplace before conception (Gardner *et al*., 1990). The increase in cases of leukemia in the areas surrounding nuclear plants continues to be problematic, because several studies carried out in France and the United States have not been able to reproduce the association observed in England (Gardner, 1992; Hill and Laplanche, 1990; Viel and Richardson, 1990).

Serious nuclear accidents can have disastrous consequences for the environment, as the example of Chernobyl shows. One of the plant's reactors exploded on April 25, 1986, and it caused a fire that lasted for 10 days. This accident caused serious soil pollution in the area around the plant by the fallout of nonvolatile radioactive elements such as strontium and plutonium. The air, greatly polluted by the radioactive isotopes of iodine and cesium and aided by the wind, contaminated the soil of most European countries. In the Netherlands, for example, the activity of cesium 137 in milk was 10 times higher in 1986 and 1987 than it was between 1980 and 1985 (WHO, 1995). Iodine 131 was only detected temporarily in the air because of its short life span. The immediate and long-term consequences of this catastrophe on morbidity and mortality are not precisely known. The size of the population still irradiated by weak doses varies according to estimates between 8 and 17.5 million people living mainly in the Ukraine, Byelorussia, and Russia (Ginzburg and Reis, 1991). A million people probably received a higher dose of radiation: the 135,000 people living within a 30-kilometer radius of the plant and evacuated the day after the catastrophe; the 100,000 people evacuated in the year after the fire; the 10,000 people still living in radioactive zones; and the 400,000 people and an unknown number of military personnel who participated in the clearing operations without any protection against ionizing radiation.

The initial explosion caused the death of 2 workers, whereas the deaths due to irradiation and burns of 30 firemen and workers participating in the clearing operations were recorded in the 3 months after the fire. In 1990, the number of deaths directly attributed to irradiation was put at 250 (WHO, 1995; Ginzburg and Reis, 1991). According to experts from the U.S. Department of Energy, the radioactive contamination of the planet caused by this accident is likely to result in 28,000 new cases of cancer (Ginzburg and Reis, 1991). A sharp increase in the cases of thyroid cancer in Ukrainian and Byelorussian children was observed in the 5 years after the passing of the radioactive cloud over these two regions and has been attributed to the presence of radioactive iodine in the cloud (Weinberg *et al*., 1995; WHO, 1995). This increase may nevertheless be artificial as a result of the improvement in the recording of epidemiologic data after the accident.

This argument appears to be corroborated by the negative results obtained during the clinical and radiologic examination of the thyroids of the people living in villages at distances of 30 to 300 km from Chernobyl that were carried out in 1990 (Mettler *et al*., 1992). Significant variations in the incidences of congenital malformations and leukemia in children have not been recorded in most European countries where radioactive fallout was observed. The immediate public health problem seemed to be the psychological consequences of the evacuation of populations; the number of voluntary abortions also increased in certain European countries after the disaster. According to the World Health Organization (WHO, 1995), it will probably not be possible to evaluate the long-term impact of this accident on morbidity and mortality.

II. CHEMICAL FACTORS

1. Soil Quality

Most elements of the stratosphere interact in the soil; water, air, climatic and geologic elements, and living organisms participate in the elaboration of the soil. Vegetable species that develop in soil transform solar energy into living matter using chlorophyllous photosynthesis. The physicochemical composition of these vegetable species, the first link in the earth's food chain, is influenced by the nature of soils. Water infiltration in soils also affects their physicochemical qualities.

In certain cases, soil type can be the primary cause of a disease. Humans and animals need certain food components to stay in good health. These components are indispensable to the functioning of the organism and are referred to as *essential*. Among these essential components are several amino acids, some fatty acids, vitamins, and 14 minerals, also called *trace elements*. The minerals include iron, iodine, fluoride, copper, manganese, zinc, cobalt, molybdenum, selenium, chromium, nickel, tin vanadium, and silicon. To meet the organism's needs, the daily dose of each of these elements should be well defined, because an insufficient dose is a cause of deficiency diseases and an overdose can transform these essential elements into toxic substances.

The variations of geochemical conditions and an unequal distribution of these trace elements all over the world are responsible for large differences in the doses of essential elements found in the human diet. The health consequences of an insufficient supply of these elements, which is directly linked to soil type, are well known for at least three trace elements: selenium, fluoride, and copper. In the Kesham region of China, an endemic cardiomyopathy that particularly affects children has existed since the beginning of the 20th century. It was discovered that this disease, which has a lethality rate of about 50%, is caused by a deficiency of selenium in the diet, which is a result of the extreme penury of this metal in the soil. Excess selenium also has deleterious effects. Toxic effects have been recorded in workers exposed to aerosols containing selenium in its elementary state and in people having eaten foods rich in selenium (WHO, 1987b; Högberg and Alexander, 1986). Fluoride is well known for its capacity to prevent carries and tooth decay. In some tropical and subtropical regions, where the water and the soil are rich in fluoride, there are zones where skeletal fluorosis, caused by excessive doses of fluoride, is endemic (WHO, 1985). This bone disease is also a serious problem in certain areas of several developing countries, where it causes serious disability in a substantial proportion of the population. Iodine is a component of thyroid hormones, and the iodine content of food varies according to the composition of the soil from which it is derived. Severe iodine deficiency causes simple goiter, cretinism, a decrease in fertility, and an increase in perinatal and infant mortality (Delange, 1994). In Europe, it is estimated that between 50 and 100 million people living mainly in mountainous regions are affected by this deficiency. Iodism indicates poisoning due to a prolonged absorption of iodized medicines; environmental factors do not seem to play a decisive role in the cause of this disease.

Soils are also contaminated by various chemical substances that are not essential components for living organisms. Some pollutants can last for several years, even several centuries, and will slowly be accumulated in the earth's food chain. This bioaccumulation can become a health threat by increasing the risk of diseases such as cancer. The substances in question are pesticides, organic components such as PCBs and dioxins, some inorganic components such as nitrates, and some heavy metals such as cadmium. For most of the population, these substances are absorbed through ingestion of food and water. Nevertheless, inhaling these substances can be a supplementary source of risk for certain categories of people; the air in some rural zones can contain measurable quantities of pesticides as a result of spraying and as a consequence of the evaporation of the pesticides deposited in water, vegetation, and the soil. Only one association of a causal nature has been established: between exposure to arsenic-based pesticides and pulmonary cancers (WHO, 1995). PCBs belong to a complex group of aromatic chlorinated hydrocarbons, which have a number of uses as thermal transfer fluids, organic dilutants, plasticizers, lubricating inks, flame retarders for fireproofing, paint additives, liquid plugs, immersion oils, dielectric adhesives, and fluids for condensers and transformers. PCBs are concentrated in the fat of marine and land animals because of their high solubility in lipids. The possibility of an increased risk of liver cancer associated to exposure to PCBs has been evoked in workers exposed by their occupation and in populations accidentally exposed to PCBs. In contrast to PCBs, dioxins have never been synthesized for any use. They are by-products that appear during the synthesis of other chemical substances (i.e., PCBs, chlorinated phenol and their by-products, and especially herbicides). Dioxins do not exist naturally but are present everywhere in the environment. A small increase in the risk of some cancers has been recorded among workers exposed to dioxins. Nevertheless, it is estimated that there is hardly a measurable risk incurred by the general population as they are exposed to quantities of dioxin of between two and three times lower than doses absorbed by workers exposed by their occupation.

Nitrates and nitrites are groupings of atoms that are part of the nitrogen cycle, and they exist naturally. Their accumulation in the environment, particularly in the soil, is the result of massive uses of nitrogenous fertilizer in agriculture, the increase in quantities of nitrogenous waste originating from the production of cattle and poultry, and from the treatment of urban waste. Vegetables are usually the main source of nitrates. The concern associated to this type of soil pollution comes from the fact that nitrates and nitrites can form nitrosamines in the presence of amines. The formation of these derivatives can take place in the digestive tract or in the environment. Most nitrosamines have proved to be carcinogenic in animals. Moreover, studies have shown that the geographic distribution of deaths from gastric cancer is linked to the presence of nitrates. Because they are based on the analysis of aggregate data without excluding the influence of confounding factors, the relation of cause and effect remains controversial.

The toxic effects caused by the absorption of large doses of cadmium through the food chain are an example of the possible repercussions of soil pollution by heavy metals on human health. In Japan at the end of the Second World War, certain bone diseases were more common among people living in several villages in the Jinzu river basin (Toyama region). This disease is characterized by intense pain in the bones, from which was derived the name Itai-Itai (i.e., ouch-ouch in Japanese). The pain was the result of a weakening of the bones, which was likely to cause multiple, spontaneous fractures in the most serious cases (Kjellström, 1986). The kidney function of most of those affected with the disease was also damaged. This disease affected mostly multiparous menopausal women. According to an epidemiologic survey carried out in 1967, the prevalence of the Itai-Itai disease was approximately 7% in a total population of 6700 residents older than 30 years who lived in the Jinzu basin zone. Approximately 100 deaths were attributed to this disease. Organs of the autopsied patients contained very high quantities of cadmium, lead, and zinc. The cadmium content of rice cultivated in the region was also very high. The cause of Itai-Itai disease was the major pollution of soil and water from waste produced in a zinc and lead mine upstream from the villages where the cases of this disease had been recorded.

2. Water Quality

The expansion of industrial activities and certain modern agricultural practices, such as heavy use of fertilizers and pesticides, have in the last few decades modified the quality of water as much in rivers, natural lakes, and dams (i.e., surface water) as in superficial or deep aquifer layers (i.e., underground water). The level and complexity of this universal problem was exposed in a WHO publication that recommended a guideline value (WHO, 1996).[1]

Noncarcinogenic pollutants present in small quantities in drinking water accumulate slowly in the targeted organs of the body. For example, lead accumulates in the bones, and cadmium accumulates in the kidneys and the liver. In this way, very small doses can cause minor biochemical modifications, which can be measured with the help of specific biologic tracers. Enzymatic inactivity has been observed for small environmental exposures to lead, but the physiologic-pathologic significance of these effects is still not clearly established. When there is an increase in the dose absorbed, preliminary symptoms (sometimes unspecific) are observed, such as neuropsychological disorders or mild anemia in the case of exposure to lead in the surroundings, followed by chronic poisoning for which the prognosis is not necessarily fatal. It is therefore improbable that trace amounts of noncarcinogenic pollutants account for a substantial difference in the risk of deaths from all causes or from specific pathologies in the general population, provided the intensity of the exposure does not reach

[1] A guideline value represents the concentration of a constituent that does not result in any significant risk to health of the consumer over a lifetime of consumption (WHO, 1996).

exceptional levels and that the exposed population is not large.

The quality of diagnosis and medical data recorded in the medical certificate can also be an important factor in the observation of this type of problem, as the following example shows. Using biologic tracers, an overload of lead was found in 40% of the inhabitants of a well-defined region in the east of Belgium, where the drinking water was severely contaminated by lead piping (Sartor and Rondia, 1980 and 1981; Sartor *et al*., 1981). This situation had existed since the beginning of the 20th century, and the volume of the overload suggested that numerous cases of lead poisoning, also known as *saturnism* in the medical literature, must have occurred in the region. An excess mortality from renal diseases was not excluded. A survey that would enable the registration of cases of saturnism could not be carried out because of a confidentiality issue with the medical records and the polymorphism of the symptoms of poisoning which could affect the exactitude of the medical diagnosis and possibly the cause of death. The number of deaths from all causes was not higher than in other bordering regions after several confounding factors were taken into account (Sartor, 1979). Infant mortality in particular did not seem to be affected, even though this is the group most vulnerable to the effects of lead. Only male cerebrovascular mortality appeared to increase (Sartor and Rondia, 1983). It was in 1987, after an autopsy, that the first fatal case of lead poisoning of a young child living in this region was reported; the child had initially been hospitalized for acute encephalopathy attributed to another cause (Heinrich *et al*., 1987).

It is sometimes possible to observe an increase in the cases of a disease specifically caused by a pollutant without the mortality related to that disease being affected. In a region in the center of Slovakia that had major pollution of water, air, and soil by arsenic, the incidence of skin cancer, with the exception of melanoma, was twice as high as in other Slovak regions between 1975 and 1984 (Plesko, 1989). Because of the low lethality of this type of cancer, there was no geographic variation in the mortality rates for skin cancer. In effect, in situations such as this, the number of deaths attributed to a particular disease is independent of the number of cases of the disease. The site of the tumor provoked by the ingestion of arsenic is relatively specific to this pollutant, but this is not the case for other carcinogenic substances in the environment. The potential confounding factors must first be taken into consideration, followed by the dose to which the population has been exposed and the relatively long period of latency that separates the beginning of the process of carcinogenicity and the tumorous propagation. From all evidence, death certificates, which provide the classic data used to study mortality, do not provide information that is precise enough on this subject. It is therefore difficult to univocally establish possible links between the level of environmental pollution and certain long term effects such as cancer based on mortality studies of the general population.

Moreover, the ecologic nature of mortality analyses (i.e., use of aggregate data) can be a source of premature causal interpretations. In the 1970s, the results of about 50 ecologic surveys carried out in various countries suggested that the extent of cardiovascular mortality was inversely correlated to the level of mineralization in drinking water (Neri *et al*., 1974). Two major hypotheses have been proposed to explain this association: the protective role of calcium, magnesium, or lithium, which are more abundant in water with high mineral content, and the toxicity of elements present in water pipes, such as lead and cadmium, which are often more abundant in water low in minerals. Because surveys based on individual observations have proved to be more convincing than ecologic surveys in showing a cause-effect link, a longitudinal study and two case-control studies were undertaken. The results of these three surveys, based on individual data, invalidated the link observed in the ecologic studies (Comstock, 1988).

In Japan, a major pollution of water by mercury salts proved to be an indisputable cause of avoidable death. From 1953 to 1956, numerous cases of poisoning occurred in a fishing community on the bay of Minamata. The cause was attributed to the consumption of fish contaminated by methyl mercury formed by liquid waste from a factory producing vinyl chloride. In 1985, 2758 cases of the Minamata disease were recorded, 656 (25%) of which were fatal. Deaths from all causes, from liver cancer, and from liver and kidney diseases in the exposed population (50,000 inhabitants) for both sexes were found to be higher than those recorded in the two neighboring sample regions (Tamashiro *et al*., 1985 ; Tamashiro *et al*., 1987).

Pollution of underground water by nitrates as a result of massive use of fertilizer is another example of an environmental factor liable to provoke avoidable morbidity and possibly avoidable mortality. The quantity of nitrates, particularly in artesian well water, can easily exceed the guideline value recommended by the WHO (WHO, 1996). Ingestion of large quantities of nitrates, especially by infants, can be the source of a sometimes-lethal asphyxia called methemoglobinemia, which is caused by a denaturation of hemoglobin. In the United States, 278 cases of methemoglobinemia attributable to the pollution of drinking water, of which 39 were fatal, were recorded between 1945 and 1950 (Walton, 1951). Because reporting of this disease was not obligatory and the declarants were volunteers, the incidence of this disease was obviously underestimated. In Hungary at the beginning of the 1970s, 200 to 300 cases of methemoglobinemia were reported annually in zones with low-quality drinking water (Rudnai, 1992). At the end of the 1980s, only about 30 cases were recorded as a result of the preventive measures that had been put in place. Nomadic populations seem to be those most at risk for this condition; between 1987 and 1992, the Slovakian authorities recorded 129 nonfatal cases of methemoglobinemia, most of which had occurred in Gipsy camps using unregulated wells. As with most health problems linked to environmental factors, the effect of water pollution by nitrates on morbidity and mortality is difficult to quantify as a result of the varying reliability of the collected data (e.g., methods of recording and declaring cases of illness or deaths, rate of coverage of the exposed population, accuracy of diagnosis).

3. Air Quality

Certain meteorologic conditions can constitute an environmental factor capable of causing a disastrous incidence of air pollution. Thick fog associated with an increase in temperature with altitude (i.e., an inversion of the thermic gradient), lack of wind, and temperatures slightly below 0°C prevent the dispersal of atmospheric pollutants and promote their accumulation in the surrounding air. Atmospheric conditions such as these were behind three serious occurrences of air pollution during which an increase in mortality was observed. The first incident, reported by Batta and colleagues (1933), occurred in December 1930, in the industrial zone of the Meuse Valley in Belgium, where 27 different types of factories (e.g., coking plants, steelworks, zinc industry, production of sulfuric and phosphoric acid, glassworks, crystal manufacturers) were located over a distance of 30 km. The quantity of sulfur dioxide and sulfuric acid in the surrounding air, estimated using simple dispersion models, seemed to have been between several multiples of 10 and 1000 mg/m^3 of air. The crude death rate observed in this region during this period of pollution was at least 10 times higher than the expected rate, and pulmonary necrosis was discovered during autopsies carried out on persons in good health who died suddenly during this episode.

Historically, this was also the first study that suggested that air pollution is an environmental factor liable to increase the risk of death. In 1948, a similar incident occurred in Donora, a small industrial center near Pittsburgh, Pennsylvania, during which the mortality rate increased 10-fold (Lipfert, 1994). The third episode occurred in London in December 1952, and it was the most documented case for the volume of exposure and for public health indicators (Lipfert, 1994). Instruments for the continuous measurement of the quantity of dioxide and sulfur in the breathable particles in the surrounding air were available at the time. The average daily quantities of these two pollutants, commonly found in London during the winter, amounted to 0.5 mg/m^3. During the first half of December 1952, they increased 10-fold to between 4 and 5 mg/m^3. In the London agglomeration, where the average number of deaths per fortnight was a little over 4000 (100,000 per year), more than 8000 deaths were recorded during the first half of December 1952, which corresponds to a standardized mortality ratio of 200% for a population of 10 million, whose usual mortality rate was 1%. This phenomenon affected all age groups, although it was more intense among children and elderly people (Ministry of Health, 1954). Respiratory and cardiovascular diseases stood out as the main causes of excess mortality. An increase in the daily rate of hospitalizations and the requests for refunding medical services indicated a significant rise in morbidity during this period. Because burning coal rich in sulfur in a hearth was a fairly widespread method of domestic heating at the time, the degree of pollution of the indoor air was probably comparable to that of the outdoor air. This supplementary argument supported the likelihood of a cause-effect relationship between air pollution and excess mortality.

Various legislative measures have progressively been put in place in most industrialized countries to avoid the harmful effects of air pollution. Their aim is to reduce the sulfur content in fossil fuel used in the production of electricity and for domestic heating, improve the conditions of dispersal of pollutants (increasing the height of chimneys), purify gaseous emissions, inspect individual boilers, and possibly cease the production of electricity when critical levels of air pollution have been exceeded. Producing electricity from natural gas and nuclear plants and instituting measures to save energy have significantly reduced pollution by sulfur dioxide and breathable particles in most countries in Western Europe and in North America. The quantity of these two pollutants in the air of these countries is, on average, 100 times lower than the quantities recorded in London in 1952. Several surveys based on ecologic analyses of prolonged time series have nevertheless suggested that an increase in the degree of air pollution by sulfur dioxide and fine particles (i.e., particles that can be physically assimilated to spherical particles of a diameter of

10 μm and that are capable of reaching the pulmonary cells), at the low levels currently measured, could provoke a significant increase in deaths from all causes, particularly from respiratory and cardiovascular diseases (Katsouyanni *et al*., 1997; Schwartz, 1991, 1993; Dockery *et al*., 1992; Schwartz and Dockery, 1992a and 1992b; Pope *et al*., 1992). An increase in air pollution by fine particles of 0.1 mg/m^3 could therefore be the cause of a daily excess mortality of between 4% and 16%. Meanwhile, the causal nature of this combination remains controversial (Sartor, 1998; Gamble, 1998; Gamble and Lewis, 1996; Moolgavkar and Luebeck, 1996).

From the 1970's onward, public health problems linked to air pollution by motor vehicle exhaust fumes became a source of concern. These gases contain significant quantities of toxic agents such as lead, nitrogen dioxide, carbon monoxide, and some volatile organic compounds such as benzene and some hydrocarbon by-products. It is estimated that one-half of the nitrogen oxide emissions come from motor vehicle exhaust fumes (Stanners and Bourdeau, 1995). During the summer, as a result of strong sunshine, nitrogen oxide and the volatile organic compounds become precursors in the formation of tropospheric ozone. The increase in air pollution from the latter also can be a cause of excess mortality (Sartor *et al*., 1995 and 1997; Touloumi *et al*., 1997; Moolgavkar *et al*., 1995; Kinney and Ozkanyak, 1991).

The fact that the connection between the presence of certain toxic substances and mortality is generally difficult to establish has already been emphasized in this chapter. This is particularly true for carcinogenic effects caused by the inhalation of substances such as benzene, aromatic polycyclical hydrocarbons from motor vehicle traffic, and coal combustion. It has been known for a long time that mortality from bronchial cancer is higher in urban or industrial areas than in rural areas, and an often repeated hypothesis is that this difference is attributable to the presence of larger quantities of carcinogenic substances in the air of cities. At the end of the 1970's, by extrapolating from mortality data collected from workers exposed to such substances, it was estimated that the combined effect of all these carcinogenic pollutants present in the air in cities could only be translated into the occurrence of five cases of lung cancer per 100,000 inhabitants per year (Doll, 1978). Several case-control and longitudinal epidemiologic surveys have corroborated this estimate by showing that the carcinogenic substances in the air can, independently of other risk factors and in a worst-case scenario, double the risk of lung cancer in the general population (Cohen and Pope, 1995; Barbone *et al*., 1995; Engholm *et al*., 1996).

Serious air pollution can also be caused by a major chemical accident that generates a massive emission of a toxic substance into the environment as a result of a serious technical breakdown. The accidents in Seveso, Bhopal, and Chernobyl are the three most famous examples of modern industrial disasters that caused serious pollution of the environment, particularly of the air. The dioxin liberated in the region of Seveso on June 10, 1976, did not have an immediate effect on mortality and caused an epidemic of chlorinated acne in children younger than 15 years, which was observed a few days after the accident (Caramaschi *et al*., 1981). A longitudinal study of the exposed population continues to look for any carcinogenic effects due to this exposure to dioxin (Bertazzi *et al*., 1989, 1992, 1993; Pesatori *et al*., 1993). The latest data analysis shows an excess mortality from diseases of the circulatory and respiratory systems, probably due to a combination of exposure to dioxin and the psychosocial consequences of the accident. Deaths from diabetes are also higher, but a connection to the dioxin exposure remains to be established. The increase in deaths from gastrointestinal cancers and cancers of the lymphatic and hematopoietic tissues are indisputably attributable to the exposure to dioxin (Bertazzi *et al*., 1998).

During the night of December 2, 1984, the air in the region of Bhopal, India (900,000 inhabitants), was severely polluted by the volatilization of 27 tons of methyl isocyanate, which was stocked in the tank of a chemical factory. Two hundred thousand people were poisoned by this gas, which is a serious irritant for the respiratory, ocular, and cutaneous mucous membranes. The official assessment of this catastrophe, an underestimate according to some Indian experts, is at least 2500 immediate deaths (Dhara, 1992). According to the local authorities, the mortality rate for the exposed population remained higher than that for a neighboring population that was not exposed until 1990; the number of deaths attributable to this air pollution in the 5 years after the disaster rose to 1100 supplementary deaths, mainly caused by respiratory complications. At the end of the 1980's, the prevalence of respiratory, ocular, and gastrointestinal symptoms in the exposed population was also much higher than in a neighboring sample population. The fire in the Chernobyl nuclear plant, mentioned earlier in this chapter, further illustrates the cross-border nature of air pollution. Air can play the role of a vector in transporting harmful substances to regions far from where the emission occurred.

The quantity of pollutants measured in the outdoor air, in the specimens taken from outside dwellings, has been used as an indicator of exposure in most ecologic surveys showing the effects of air pollution on morbidity or mortality. However, a large proportion of the population of industrial countries spends most of their lives inside work premises and dwellings, where the air quality is not necessarily the same as that of the outdoor air. The degree of air pollution from outside pollutants found inside work premises depends on the rate of infiltration of the outdoor air, the size of particles, and the chemical reactivity of the pollutants with the materials for constructing and furnishing the walls in the premises. The rate of infiltration can vary considerably according to the type and the use of premises. The finest particles penetrate the interior of premises more easily and can remain suspended in the air for a long time (Lipfert, 1994). Sulfur dioxide and ozone are gases capable of great chemical reactivity, and they therefore disappear rapidly, whereas carbon monoxide can remain for a long period because of its chemical stability.

It is not easy to estimate inside dwellings the degree of exposure to pollutants present in the outdoor air. Indoor air can be more polluted by certain substances than outdoor air. This is the case for carbon monoxide, nitrogen oxide produced by gas cookers, tobacco smoke, and household dust (i.e., coal burning residue). Other substances that pollute indoor air are not normally present in the outdoor air. These are mainly organic solvents found in building materials, paint, aerosols, and cleaning products. Sources of biologic pollution, such as dust mites, hair from domestic animals, and allergenic plants, are found inside dwellings. Thermal insulation and the use of air conditioning are also factors in the problem of air pollution inside buildings. Even though the agents likely to pollute indoor air and the levels of pollution that they can cause are known, the statistics of morbidity and mortality are only available for certain pollutants, such as radon (discussed earlier) and carbon monoxide. Between 1978 and 1988, the number of deaths attributed to carbon monoxide in the United States (total of 11,500 deaths) represented one-fourth of deaths due to accidental poisoning; for comparison, heroin overdose caused the deaths of approximately 6000 people during the same period (Cobb and Etzel, 1991). Carbon monoxide caused by vehicle exhaust fumes was responsible for 57% of these deaths, and the toxin produced by faults in domestic heating appliances were responsible for 20% of these deaths; accidental poisoning was not recorded as a cause in 20% of the death certificates studied.

The various varieties of asbestos are derivatives of fibrous silicate. Asbestos was abundantly used for its insulating properties from the beginning of the 20th century by the building industry. These fibers are carcinogenic agents reputed to be the only known cause of mesothelioma of the pleura, a rare cancer that affects one or two people per million in the general population per year; this form of malignant tumor cannot be attributed to smoking, even though it can exacerbate the carcinogenic effects of asbestos. The mortality rate from mesothelioma of the pleura is considered to be the best indicator of past exposure to asbestos. Numerous experts expect to see an increase in mortality from this cancer over the next 20 years, particularly among construction workers (Price, 1997; Peto *et al*., 1995). Excess mortality from mesothelioma of the pleura has also been observed in Turkish villages where dwellings were built using silicate materials similar to asbestos (Baris *et al*., 1992).

The role of exposure to asbestos in the development of mesothelioma of the pleura was discovered during epidemiologic surveys on industrial toxicology. These surveys look at occupational intoxication, mainly caused by air pollution of the workplace by the chemicals used or produced industrially. The number of substances that may cause occupational poisoning is impressive. More than 1500 chemical compounds are listed in the index of Robert Lauwerys' reference book (1990). Despite the existence of occupational hazards, mortality from all causes in the population of workers (age standardized) is generally inferior to that of the general population (Table 50–1).

The healthy worker effect comes from a selection, based partly on medical criteria, of subjects in good health; it seems to be greater for nontumorous diseases than for cancers. The observation of cohorts of workers more particularly exposed to certain substances has nevertheless indicated that mortality attributable to certain types of cancer has revealed itself as being significantly higher than in the general population (see Table 50–1). In this way, surveys on mortality have helped to reveal the carcinogenic effects on humans of substances such as benzene, nickel, arsenic, vinyl chloride, and tar.

III. INFECTIOUS DISEASES LINKED TO THE ENVIRONMENT

In this section, we discuss only the pathogenic germs that can be transmitted to humans as a result of environmental conditions and that can therefore be eliminated with appropriate measures. On a global level, the number of deaths caused by these infectious diseases is much greater than for those that can be attributed to the physical and chemical factors discussed earlier in this chapter.

The principal health risk linked to the pollution of drinking water comes from the contamination of the

TABLE 50–1 Deaths from All Causes and Specific Mortality in Some Cohorts of Workers Exposed to Specific Toxic Substances

Industrial activity	Standardized mortality ratio (%)[a]			
	All causes	From all cancers	From specific cancers	Site
Rubber	83 (5079)	93 (986)	129 (55)	Bone marrow
Vinyl chloride	75 (352)	102 (79)	412 (7)	Liver
			280 (7)	Brain
Steelworks	82 (4716)	92 (1008)	283 (40)[c]	Lungs
Nickel (Scotland)	90 (67)[b]	320 (38)	1600 (16)	Lungs
			1100 (11)	Nose
Nickel (Canada)	80 (245)	130 (54)	220 (19)	Lungs
			3680 (7)	Nose
Arsenic	111 (324)	147 (69)	300 (32)	Lungs

[a] The numbers of deaths observed are given in parentheses.
[b] Deaths from all other causes except cancer.
[c] Workers in coking plants.
From Monson, 1980; NAS, 1975, 1977.

water supply by bacteria, viruses, protozoa, or parasites. The presence in drinking water of germs pathogenic to humans is always caused by fecal contamination, except in the case of the guinea worm, which is responsible for dracunculiasis. The eradication of these diseases depends on sanitation, the sanitary evacuation of excreta and the purification of domestic used water. The infectious diseases transmitted in water include diarrheal diseases, cholera, viral hepatitis, and salmonellosis. However, water is not the only channel of transmission; these diseases can also be carried by contaminated food. Some diseases can be transmitted by the inhalation of water (e.g., legionnaire's disease) or by skin contact (e.g., bilharziasis). On a global level, mortality from water-transmitted diseases remains a major public health problem. In 1985, noncholeraic diarrheal diseases in children younger than 5 years caused 3.5 million deaths in developing countries, which was about 10% of the total 39 million deaths in that year (Lopez, 1990).

Deaths from cholera remain fewer than deaths from diarrheal diseases. In 1991, more than one-half million cases of cholera, of which 16,000 were fatal, were reported by 52 countries (Siméant, 1992). The lethality rate of cholera decreased on a global level during the seventh pandemic, decreasing from 50% in 1960 to 3.3% in 1991. Several factors have contributed to this development: improvement in notification of cases (i.e., more exhaustive recording of the benign cases which previously escaped notification), the reduced virulence of the El Tor biotype, improved knowledge about the disease, and rapid implementation of effective means of combating it (Siméant, 1992). It is nevertheless difficult to specify the portion of deaths from diarrheal diseases and from cholera directly attributable to the bad bacteriologic quality of water because of the lack of available epidemiologic data. In industrial countries, bacteriologic water pollution remains a public health problem, although the effect on mortality cannot be compared with that in developing countries. In the United States for example, 291 epidemic outbreaks of water-transmitted diseases were recorded between 1981 and 1990, causing four deaths (Craun, 1992).

Pathogenic germs can also be transmitted to humans through other organisms living in water or air, which serve as intermediary hosts or media. The two most frequently occurring diseases in tropical and subtropical zones are transmitted in this way: bilharziasis and malaria. Bilharziasis is a parasitic disease caused by the presence of a worm, the *Schistosoma*, in certain veins of the host organism. It is transmitted through skin contact with water containing larvae that developed in intermediary hosts, which in this case are certain types of aquatic snails. This parasitic disease is endemic in the rural regions of 74 countries, where 250 million people are infected. Given the diversity of the organs infested with the parasite, it is difficult to evaluate the level of mortality attributable to this parasitosis. After a latent period of 5 to 15 years, those suffering from bilharziasis develop serious diseases of the bladder, brain, liver, spleen, intestines, and lungs as a result of the parasitic activity of *Schistosoma*. Moreover, presence of the parasite in the bladder causes a particular type of cancer of this organ, which is the main cause of death in young adult Egyptians.

Malaria, an Italian word that means "bad air," is an infection with the *Plasmodium*, a parasitically living, unicellular species transmitted to humans through a mosquito bite. The regions where endemic malaria still exists or where no antimalarial campaign has ever taken place contain up to 500 million inhabitants (9% of the world's population) and are mainly situated in Africa. According to estimates by the WHO (1992), more than 80% of the clinical cases and 90% of the carriers of the parasite are in tropical Africa. However, only very incomplete and erratic information on deaths attributable to malaria is available. The WHO's African Region declared 800,000 children's deaths per year in 1991, but child mortality had declined in certain zones as a result of widespread use of antimalarial drugs and of progress in education and social development.

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Aguesse Pierre, 1971. *Clefs pour l'écologie*. Paris, Seghers, 238 p.

Armstrong Bruce K., 1994. Stratospheric ozone and health, *International Journal of Epidemiology*, vol. 23(5), p. 873–885.

Barbone Fabio, Bovenzi Massimo, Cavalieri Furio, and Stanta Giorgio, 1995. Air pollution and lung cancer in Trieste, Italy, *American Journal of Epidemiology*, vol. 141, p. 1161–1169.

Baris Yusif I., Simonato Lorenzo, Saracci Rodolfo, and Winkelman Regina, 1992. The epidemic of respiratory cancer associated with erionite fibres in the Cappadocian region of Turkey, *in*: Paul Elliott, Jack Cuzik, Dallas English, and Richard Stern (eds.), *Geographical and environmental epidemiology: methods for small-area studies*, p. 310–322. Oxford, UK, Oxford University Press, 382 p.

Batta George, Firket Jean, and Leclerc Edmond, 1933. *Les problèmes de pollution de l'atmosphère*. Liège, Georges Thone, 462 p.

Bertazzi Pier Alberto, Brenucci Ilaria, Brambilla Gabriella, Consonni Dario, and Pesatori Angela, 1998. The Seveso studies on early long-term effects of dioxin exposure: a review, *Environmental Health Perspectives*, vol. 106, Suppl 2, p. 625–633.

Bertazzi Pier Alberto, Pesatori Angela, Consonni Dario, Tironi Adriana, Landi Maria T., and Zochetti Carlo, 1993. Cancer incidence in a population accidentally exposed to 2,3,7,8-tetrachlorodibenzo-*para*-dioxin, *Epidemiology*, vol 4, p. 398–406.

Bertazzi Pier Alberto, Zocchetti Carlo, Pesatori Angela C., Guercilena Stefano, Consonni Dario, Tironi Adriana, and Landi Maria T., 1992. Mortality of a young population after accidental exposure to 2,3,7,8-tetrachlorodibenzodioxin, *International Journal of Epidemiology*, vol. 21(1), p. 118–123.

Bertazzi Pier Alberto, Zocchetti Carlo, Pesatori Angela, Guercilena Stefano, Sanarico Maurizio, and Radice Laura, 1989. Ten-year mortality study of the population involved in the Seveso incident in 1976, *American Journal of Epidemiology*, vol. 129(6), p. 1187–1200.

Bochichio Francesco, McLaughlin James P., and Piermatei Silvana, 1995. *European Collaborative action: indoor air quality and its impact on man, report no. 15. Radon in indoor air*. Luxembourg, European Commission, Directorate-General XIII, 50 p.

Caramaschi F., del Corno G., Favaretti C., Giambelluca S.E., Montesarchio E., and Fara G.M., 1981. Chloracne following environmental contamination by TCDD in Seveso, Italy, *International Journal of Epidemiology*, vol. 10, p. 135–143.

Cobb Nathaniel and Etzel Ruth, 1991. Unintentional carbon monoxide-related deaths in the United States, 1979 through 1988, *Journal of the American Medical Association*, vol. 266, p. 659–663.

Cohen Aaron and Pope Arden III, 1995. Lung cancer and air pollution, *Environmental Health Perspectives*, vol. 103, Suppl. 8, p. 219–224.

Comstock George W., 1988. Soft water/hard arteries: an interpretation of ecologic findings, *in*: Leon Gordis (ed.), *Epidemiology and health risk assessment*, p. 248–255. New York, Oxford University Press, 294 p.

Craun Gunther F., 1992. Waterborne disease outbreaks in the United States of America: causes and prevention, *World Health Statistics Quarterly*, vol. 45, p. 192–199.

Delange François, 1994. The disorders induced by iodine deficiency, *Thyroid*, vol. 4, p. 107–128.

Dhara Ramana, 1992. Health effects of the Bophal gas leak: a review, *Epidemiologia e Prevenzione*, vol. 52, p. 22–31.

Dockery Douglas W., Schwartz Joel, and Spengler John D., 1992. Air pollution and daily mortality: associations with particulates and acid aerosols,*Environmental Research*, vol. 59, p. 362–373.

Doll Richard, 1978. Atmospheric pollution and lung cancer, *Environmental Health Perspectives*, vol. 22, p. 23–31.

Elwood Mark J., Lee John A.H., Walter Stephen D., Mo T., and Green Axel E.S., 1974. Relationship of melanoma and other skin cancer mortality to latitude and ultraviolet radiation in the United States and Canada, *International Journal of Epidemiology*, vol. 3, p. 325–332.

Engholm Gerda, Palmgren Finn, and Lynge Elsebeth, 1996. Lung cancer, smoking, and environment: a cohort study of the Danish population, *British Medical Journal*, vol. 312, p. 1259–1263.

Gamble John, 1998. $PM_{2.5}$ and mortality in long-term prospective cohort studies: cause-effect or statistical associations? *Environmental Health Perspectives*, vol. 106, p. 535–549.

Gamble John and Lewis Jeffrey, 1996. Health and respirable particulate (PM_{10}) air pollution: a causal or statistical association? *Environmental Health Perspectives*, vol. 104, p. 838–850.

Gardner Martin J., 1992. Childhood leukaemia around the Sellafield nuclear plant, *in*: Paul Elliott, Jack Cuzik, Dallas English, and Richard Stern (eds.), *Geographical and environmental epidemiology: methods for small-area studies*, p. 291–309. Oxford, Oxford University Press, 382 p.

Gardner Martin J., Hall Andrew J., Downes Susan, and Terrell John D., 1987a. Follow up study of children born elsewhere but attending schools in Seascale, West Cumbria (schools cohort), *British Medical Journal*, vol. 295, p. 819–822.

Gardner Martin J., Hall Andrew J., Downes Susan, and Terrell John D., 1987b. Follow up study of children born to mothers resident in Seascale, West Cumbria (birth cohort), *British Medical Journal*, vol. 295, p. 822–827.

Gardner Martin J., Snee Michael P., Hall Andrew J., Powell Caroline A., Downes Susan, and Terrell John D., 1990. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria, *British Medical Journal*, vol. 300, p. 423–429.

Gardner Martin J. and Winter P.D., 1984. Mortality in Cumberland during 1959–78 with reference to cancer in young people around Windscale,*Lancet*, vol. 1, p. 216–217.

Ginzburg Harold M. and Reis Eric, 1991. Consequences of the nuclear power plant accident at Chernobyl, *Public Health Reports*, vol. 106, p. 32–40.

Heinrichs C.L., Geubelle F., Lambrechts L., Guidi O., and Fissette M., 1987. A propos d'un cas mortel d'intoxication au plomb chez l'enfant, *Revue Médicale de Liège*, vol. 42(119), p. 769–773.

Hill Catherine and Laplanche Agnès, 1990. Overall mortality and cancer mortality around French nuclear sites, *Nature*, vol. 347, p. 755–757.

Högberg Johan and Alexander Jan, 1986. Selenium, *in*: Lars Fridberg, Gunnar Nordberg, Velimir Vouk (eds.), *Handbook on toxicology of metals: specific metals*, vol. 2, p. 482–520. Amsterdam, Elsevier, 704 p.

Holdgate Martin, Kassas Mohammed, and White Gilbert (ed.), 1982. *The World Environment 1972–1982. A report by the United Nations environment programme*. Dublin, Brunswick Press Ltd, UNEP, 637 p.

IARC, 1992. *Solar and ultraviolet rariation*. IARC monographs on the evaluation of carcinogenic risks for humans, vol. 55. Lyon, International Agency for Research on Cancer, 316 p.

Katsouyanni Klea, Touloumi Giotta, Spix Claudia, Schwartz Joel, Balducci Franck, Medina Silvya, Rossi G., Wojtyniak Bogdan, Sunyer Jordi, Bacharova Ljuba, Schouten Jan P., Pönkä Antti, and Anderson H. Ross, 1997. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project, *British Medical Journal*, vol. 314, p. 1658–1663.

Kilbourne Edwin M., 1990. Atmospheric change and temperature-related health effects, *in*: James C. White, William Wagner, and Carole N Beal (eds.), *Global atmospheric change and public health*, p. 89–100. New York, Elsevier, 300 p.

Kinney Patrick L. and özkanyak Halûk, 1991. Associations of daily mortality and air pollution in Los Angeles county, *Environmental Research*, vol. 54, p. 99–120.

Kjellström Tord, 1986. Itai-Itai disease, *in*: Lars Fridberg, Carl-Gunnar Elinder, Tord Kjellström, Gunnar Nordberg (eds.), *Cadmium and health: a toxicological and epidemiological appraisal: effects and response*, vol. 2, p. 257–290. Boca Raton, FL, CRC Press, 307 p.

Koh Howard K., 1995. Preventive strategies and research for ultraviolet-associated cancer, *Environmental Health Perpectives*, vol. 103, Suppl. 8, p. 255–257.

Kunst Anton E., Looman Caspar W.N., and Mackenbach Johan P., 1993. Outdoor air temperature and mortality in the Netherlands: a time series analysis, *American Journal of Epidemiology*, vol. 137, p. 331–347.

Lauwerys Robert, 1990. *Toxicologie industrielle et intoxications professionnelles*. Paris, Masson, 693 p.

Látourneau Esther G., Krewski David, Choi N. Won, Goddard Michael J., McGregor Richard G., Zielinski J.M., and Du J., 1994. Case-control study of residential radon and lung cancer in Winnipeg, Manitoba, Canada, *American Journal of Epidemiology*, vol. 140(4), p. 310–322.

Lipfert Frederic W., 1994. *Air pollution and community health: a critical review and data sourcebook*. New York, Van Nostrand Reinhold, 556 p.

Lopez Alan D., 1990. Causes of death: an assessment of global patterns of mortality around 1985, *World Health Statistics Quarterly*, vol. 43, p. 91–104.

Lubin Jay H., 1994. Lung cancer and exposure to residential radon, *American Journal of Epidemiology*, vol.140(4), p. 323–332.

Lubin Jay and Boice John, 1997. Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies, *Journal of the National Cancer Institute*, vol. 89, p. 49–57.

Mettler Fred A., Williamson Michael R., Royal Henry D., Hurley James R., Khafagi Frederick, Sheppard Michael C., Beral Valerie, Reeves Gillian, Saenger Eugene L., Yokoyama Naokata, Parshin Vladimir, Griaznova Elena A., Taranenko Margarita, Chesin Vadim, and Cheban Anatoly, 1992. Thyroids nodules in the population living around Chernobyl, *Journal of the American Medical Association*, vol. 268, p. 616–619.

Ministry of Health, 1954. *Mortality and morbidity during the London Fog of December 1952*. London, HSMO, 61 p. (Reports on public health and medical subjects, no. 95.)

Monson Richard R., 1980. *Occupational epidemiology*. Boca Raton, FL, CRC Press, 219 p.

Moolgavkar Suresh H. and Luebeck E. Georg, 1996. A critical review of the evidence on particulate air pollution and mortality, *Epidemiology*, vol. 7, p. 420–428.

Moolgavkar Suresh H., Luebeck E. Georg, Hall Thomas A., and Anderson Elizabeth E., 1995. Air pollution and daily mortality in Philadelphia, *Epidemiology*, vol. 6, p. 476–484.

NAS, Committee on Medical and Biologic Effects of Environmental Pollutants, 1975. *Nickel*. Washington, DC, Printing and Publishing Office of the National Academy of Sciences, 277 p.

NAS, Committee on Medical and Biologic Effects of Environmental Pollutants, 1977. *Arsenic*. Washington, DC, Printing and Publishing Office of the National Academy of Sciences, 332 p.

Neri Luciano C., Hewitt David, and Schreiber George, 1974. Can epidemiology elucidate the water story? *American Journal of Epidemiology*, vol. 99, p. 75–88.

Pan Wen-Harn, Li Lung-An, and Tsai Ming-Jan, 1995. Temperature extremes and mortality from coronary heart diseases and cerebral infarction in elderly Chinese, *Lancet*, vol. 345, p. 353–355.

Pershagen Göran, Åkerblom Gustav, Axelson Olav, Clavensjö Bertil, Damber Lena, Desai Gunilla, Enflo Anita, Lagarbe Frádáric, Mellander Hans, Svartengren Magnus, and Swedjemark Gun Astri, 1994. Residential radon exposure and lung cancer in Sweden, *The New England Journal of Medicine*, vol. 330, p. 159–164.

Pesatori Angela C., Consonni Dario, Tironi Adriana, Zochetti Carlo, Fini Angela, and Bertazzi Piero Alberto, 1993. Cancer in a young population in a dioxin-contaminated area, *International Journal of Epidemiology*, vol. 22(6), p. 1010–1013.

Peto Julian, Hodgson John T., Matthews Fiona, and Iones Jacqueline, 1995. Continuing increase in mesothelioma mortality in Britain, *Lancet*, vol. 345, p. 535–539.

Plesko Ivan, 1989. *Atlas of cancer occurrence in Slovakia*. Bratislava, Slovenska Akadámia Vied, 286 p. (in Slovakian, with a summary in English).

Pope C. Arden III, Schwartz Joel, and Ransome Michael R., 1992. Daily mortality and PM10 pollution in Utah Valley, *Archives of Environmental Health*, vol. 47(3), p. 211–217.

Price Bertam, 1997. Analysis of current trends in the United States mesothelioma incidence, *American Journal of Epidemiology*, vol 145(3), p. 211–218.

Rivers Jason K., 1996. Melanoma, *Lancet*, vol. 347, p. 803–806.

Rogot Eugene and Padgett Stephen, 1976. Associations of coronary and stroke mortality with temperature and snowfall in selected areas of the Unites States, 1962–1966, *American Journal of Epidemiology*, vol. 103, p. 565–575.

Rudnai Peter, 1992. Surveillance for environmental disease in Hungary, *Mortality, Morbidity Weekly Report*, vol. 41, Suppl., p. 167–170.

Sartor Francis, 1979. Contribution de l'analyse démographique dans les enquêtes de toxicologie épidémiologique. étude de la mortalitá "toutes causes" dans une rágion à eau douce, *Archives Belges de Mádecine Sociale, Hygiène, Mádecine du Travail et Médecine Légale*, vol. 37, p. 537–561.

Sartor Francis, 1998. Impact de la pollution de l'air sur la mortalité: limites de l'analyse des séries chronologiques, *in*: AIDELF (ed.), *Morbidité, mortalité: problèmes de mesure, facteurs d'évolution, essai de prospective*, p. 281–290. Paris, Presses Universitaires de France, 737 p. (Actes du Colloque international de Sinaia, 3–6 Septembre 1996.)

Sartor Francis, Demuth Claude, Snacken René, and Walckiers Denise, 1997. Mortality in the elderly and ambient ozone concentration during the hot summer, 1994, in Belgium, *Environmental Research*, vol. 72, p. 109–117.

Sartor Francis and Rondia Désiré, 1980. Blood lead levels and age: a study in two male urban population not occupationally exposed, *Archives of Environmental Health*, vol. 35, p. 110–116.

Sartor Francis and Rondia Désiré, 1981. Setting legislative norms for environmental lead exposure: results of an epidemiological survey in the East of Belgium, *Toxicology Letters*, vol. 7, p. 251–257.

Sartor Francis and Rondia Désiré, 1983. Hardness of municipal waters and cardiovascular mortality in four small Belgian towns, *in*: Mohamed Abdulla and Baboo Nair (eds.), *International symposium: health effects and interactions of essential and toxic elements*, Lund, June 13–18, p. 140. Bromley, Chartwell-Bratt, 227 p.

Sartor Francis, Snacken René, Demuth Claude, and Walckiers Denise, 1995. Temperature, ambient ozone levels, and mortality during the hot summer 1994 in Belgium, *Environmental Research*, vol. 70, p. 105–113.

Sartor Francis, van Beneden Pierre, and Rondia Désiré, 1981. Water lead concentrations and measurements of health risk in a soft drinking water area, *in*: Proceedings of the third international conference on heavy metals, Amsterdam 1981, p. 39–43. Edinburgh, CEP Consultants.

Schwartz Joel, 1991. Particulate air pollution and daily mortality in Detroit, *Environmental Research*, vol. 56, p. 204–213.

Schwartz Joel, 1993. Air pollution and daily mortality in Birmingham, AL, *American Journal of Epidemiology*, vol. 137, p. 1136–1147.

Schwartz Joel and Dockery Douglas W., 1992a. Increased mortality in Philadelphia associated with daily air pollution concentrations, *American Review of Respiratory Diseases*, vol 145, p. 600–604.

Schwartz Joel and Dockery Douglas W., 1992b. Particulate air pollution and daily mortality in Steubenville, Ohio, *American Journal of Epidemiology*, vol. 135, p. 12–19.

Simeant Silvère, 1992. Choléra, 1991: Vieil ennemi, nouveau visage, *Rapport trimestriel de statistiques sanitaires mondiales*, vol. 45, p. 208–219.

Stanners David, Bourdeau Philippe (eds.), 1995. *Europe's environment. The Dobris assessment*. Luxembourg, Office for Official Publications of the European Communities, 676 p.

Tamashiro Hidehiko, Arakaki Mikio, Akagi Hirokatsu, Futatsuka Makoto, and Roht Lewis, 1985. Mortality and survival for Minamata disease, *International Journal of Epidemiology*, vol. 14, p. 582–588.

Tamashiro Hidehiko, Fukutomi K., and Lee E.S., 1987. Methylmercury exposure and mortality in Japan: a life table analysis, *Archives of Environmental Health*, vol. 42, p. 100–107.

Touloumi Giotta, Katsouyanni Klea, Zmirou Denis, Schwartz Joel, Spix Claudia, Ponce de Leon Antonio, Tobias Aurelio, Quennel Philippe, Rabzcenko D., Bacharova Ljuba, Bisanti L, Vonk Judith M., and Ponka Antti, 1997. Short-term effects of ambient oxidants exposure on mortality: a combine analysis within the APHEA project, *American Journal of Epidemiology*, vol. 146(2), p. 177–185.

Viel Jean-François and Richardson Silvia T., 1990. Childhood leukaemia around the La Hague nuclear waste reprocessing plant, *British Medical Journal*, vol. 300, p. 580–581.

Walton Graham, 1951. Survey of the literature relating to infant methemoglobinemia due to nitrate-contaminated water, *American Journal of Public Health*, vol. 41, p. 986–996.

Weinberg Armin D., Kripalani Sunil, McCarthy Philip L., and Schull W. Jack, 1995. Caring for survivors of the Chernobyl disaster. What the clinician should know, *Journal of the American Medical Association*, vol. 274, p. 408–412.

WHO, 1980. *Mycotoxines. Critères d'hygiène de l'environnement 11*. Geneva, World Health Organization, 142 p.

WHO, 1985. *Fluor et fluorures. Critères d'hygiène de l'environnement 36*. Geneva, World Health Organization, 150 p.

WHO, 1987a. *Air quality guidelines for Europe*. Copenhagen, World Health Organization, Regional Office for Europe, 426 p.

WHO, 1987b. *Selenium. Environmental health criteria 58*. Geneva, World Health Organization, 306 p.

WHO, 1996. *Guidelines for drinking-water quality, vol. 2 Health criteria and other supporting information*. Geneva, World Health Organization, 973 p.

WHO, Division of Control of Tropical Diseases, 1992. World malaria situation 1990, *World Health Statistics Quarterly*, vol. 45, p. 257–266.

WHO, European Centre for Environment and Health, 1995. *Concern for Europe's tomorrow. Health and the environment in the World Health Organization European Region*. Stuttgart, Wissenschaftliche Verslagsgesellschaft, 573 p.

C H A P T E R

51

Behavior, Lifestyles, and Sociocultural Factors of Mortality

JACQUES VALLIN, GRAZIELLA CASELLI, AND PIERRE SURAULT

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Groupe d'études sur démographiques, economiques et socials, Université de Poitiers, Poitiers, France

Although health and mortality depend partly on the natural or humanly adapted environment in which individuals live, people can also encourage or curtail the incidence of diseases and death by their behavior and living habits (Stanley and Joske, 1980). However, this behavior is largely determined by attitudes generated by cultural inheritance, living conditions, and policies directed at local and national levels, especially in the realm of health. In the following sections, we examine the different elements in individual behavior that have a bearing on health and then concentrate on the influence of surroundings, living conditions, and collective action on the expression of attitudes that stimulate this behavior.

I. BEHAVIOR THAT INFLUENCES HEALTH AND MORTALITY

The preservation and deterioration of current and future health are affected by every activity of daily life, such as washing, smoking cigarettes, eating pasta and fruit, crossing the road, working in a chemical factory, heating the house, having unprotected sexual relations, practicing a sport, having social relations with other people. The list of individual behaviorals having an effect on health and mortality is long, but we shall try to identify the most significant aspects by distinguishing within the individual activity the care devoted to the body, sexual behavior, organization of the living and professional environment, and relations with other people.

1. Bodily Care and Leisure Pursuits

A certain number of human activities are directed expressly to the care of the body. The same clearly applies to using health services, whether in an attempt to cure a disease or to prevent its development. Care of the body necessitates eating, drinking, resting, and exercising, actions that, although designed to keep the body in good working order, may also cause the state of health to deteriorate.

a. Resort to Medical Services

The behavior most immediately allied to health and survival is the resort to care in the case of sickness. The individual decision to seek treatment may often determine a cure and the preservation of future chances of survival. Although recovery from benign or even serious diseases may occur without any treatment, resort to care increases the chance of a faster, more complete cure with fewer complications. In addition to immediate survival, recourse to care also preserves the capacity for later survival.

Resort to care is not enough. The treatment must be effective. Generally, the result will not be the same if the sick person resorts to traditional forms of healing or to modern scientific medicine, although there is a margin of uncertainty. Sometimes, there has been a tendency to age-old solutions that had merit and instead to regard modern medicine, which remains an art that is subject to error, as sacred. In some cases, sick people given up for lost by the medical faculty have been cured by turning to alternative or complementary forms of medicine, such as homoeopathy, acupuncture, and phytotherapy. It is also true that certain elements of the pharmacopeia are far from having proved themselves, and studies have shown that the placebo effect is sometimes as important as the medication itself (Shapiro and Shapiro, 1997). Nevertheless, the healing of sick people depends on their attitudes about the resources that health services place at their disposal, whether in visiting a physician or in following a prescribed treatment. This does not mean that sick persons should display blind confidence in their physicians, but rather that their ability to discuss their symptoms with them to establish a diagnosis obtain a prescription play an important part in the success of the procedure.

What is valid for the treatment of a disease applies equally to its prevention. Although prevention within a population is largely the concern of those in charge of public health, it also depends on individual behavior. Vaccination campaigns can be organized and the most useful of them can even be made obligatory, but the vaccine coverage of a population depends very much on individual behavior. However, vaccination is one of the preventative measures mostly easily imposed from above. In the area of cardiovascular diseases, cancers, and vehicular accidents, individual intervention is even more conclusive, because preventative techniques require the vigilance of the patient (e.g., prevention of breast cancer) and because successful prevention necessitates a change in behavior (e.g., food regimen, wearing a car seat belt).

The effects of resorting to medicine, however, are not always beneficial. Many types of harmful effects may injure health and cause death. Analysts must take account of the iatrogenic effects of care given, such as secondary effects of medication or treatment or a surgical or anesthetic accident. Some cases of deleterious effects, such as that of thalidomide, are well known (Lenz and Knapp, 1962), but others doubtlessly have had more discreet consequences (e.g., harmful effects of chemotherapy, unwise use of cortisone) (Meyler and Herxheimer, 1972). There is sometimes even a risk involved in crossing the threshold of a hospital, where it is difficult, especially in developing countries, to ensure perfect asepsis; some infectious diseases, such as certain forms of hepatitis, are rarely contracted other than in a hospital. The transmission of malaria and of acquired immunodeficiency syndrome (AIDS) through the transfusion of contaminated blood also occurs in such medical centers. Individual behavior clearly is at work in the abuse of drugs, including those sometimes designated as comfort drugs (Meyler and Herxheimer, 1972). The most flagrant case is that of "psychopharmacy."

b. Individual Hygiene

Situated on the frontier between infectious diseases and bodily comfort are all the practices of individual hygiene. Following the discoveries of Louis Pasteur, there is no need to stress the merits of soap and detergent in keeping the body and clothes protected from pathogenic agents. Progress in hygiene has contributed to the decline in infantile and maternal mortality and in diseases such as typhus and tuberculosis.

The same action that in most cases is favorable to health may be harmful in others. Certain types of soap may have been factors in propagating cholera (Jean-Noël Biraben, personal communication, 2001). More common, however, are the effects of excessive zeal in matters of hygiene, consisting of overuse of certain practices (especially cosmetics) or using without adequate protection certain chemical products that are too aggressive, such as the case of skin diseases caused by domestic detergents (Leung and Greaves, 2000; Little, 1977).

c. Food Habits

Although a person must eat to survive, eating is also a pleasure that can easily lead to excesses. Health depends on the quantity of food consumed, the quality of the food, and the balance of the diet.

Quantity

For thousands of years, humans have faced the threat of famine. Avoiding famine was little concerned with individual behavior but was much more affected by systems of agricultural production and storage, the distribution of food resources, and disturbances in obtaining food because of climatic or political events.[1] This problem remains to some extent that of certain disadvantaged populations because they inhabit regions or countries that are very poor and economically fragile or because they are at risk for famine because of political struggles or military aggression.

In more developed countries, quantity may be a problem at the margins of society, especially for those

[1] In the Ukraine, for example, life expectancy at birth fell to 7 years for men and 10 years for women during the great famine of 1933, provoked by Stalin's "deculakisation" campaign (Vallin *et al.*, 2002).

who have no access to adequate resources. This situation does not usually depend directly on individual behavior.

Two types of individual behavior may seriously damage health through the quantity of food consumed: anorexia and bulimia. Taken to extremes, these types of behavior are diseases in themselves. Without going as far as the extremes of pathologic cases, insufficiency (especially among certain old people) and increasingly excessive food consumption lead to consequences that are harmful to health. Obesity is recognized as a major risk factor for cardiovascular diseases, diabetes, and other diseases (Bender *et al.*, 1998; Solomon and Manson, 1997; Truswell, 1982), and it is a scourge of modern, developed nations.

Quality

Although gradual reduction in the risks of food shortage or famine has greatly contributed to the extension of life expectancy, going beyond a certain threshold of food sufficiency, without there necessarily being an excessive quantity, has sometimes led to the emergence of new health problems. A traditional diet consists of a small number of basic elements, but access to increasingly abundant supplies has been conducive in some cases to excessive consumption of certain ingredients that favor certain pathologies. In the 1950s, there was a rise in mortality from cerebrovascular diseases coincident with a rise in the living standard in countries such as Portugal, Scotland, and Japan, where salt plays an important part in traditional food. In Hungary, access to more abundant food supplies led rapidly to excessive consumption of fats and a consequent rise in cardiovascular mortality (Meslé and Vallin, 1992; Shigematsu and Yananawa, 1985). Another type of adverse effect arising from reduction in scarcity may perhaps be illustrated by the case of Northern Italy in the 19th century. The introduction of maize on a large scale enabled people to eat more abundantly by making polenta the national dish, but this was to the detriment of other elements in the traditional diet, and this excessively monotonous regimen gave rise to a veritable epidemic of pellagra (Livi-Bacci, 1986). Subsequently, knowledge about the connection between diet and health has greatly advanced food behavior and plays an increasing role in the level of life expectancy.

The part played by vitamins in a healthy diet was established from the late 19th century onward. Scurvy, a disease rife among seamen living exclusively on preserved foods, was remedied by the consumption of lemons rich in vitamin C (Davidson *et al.*, 1979; Lind, 1753). It was a lack of tryptophan or niacin (i.e., nicotinamide, vitamin B_5, or vitamin PP) that caused the outbreaks of pellagra in Northern Italy and in South Carolina (Goldberger *et al.*, 1920). In the 20th century, the part played by vitamins came to the forefront in the prevention of certain cancers and in protection from the effects of free radicals (Holmes *et al.*, 1992; Cutler, 1984). Equally positive is the case of foodstuffs containing vegetable fiber in the prevention of certain cancers, especially of the colon (Modan and Lubin, 1980). The public is increasingly attentive to the presences of vitamins, fiber, amino acids, and other components of food, and individual life expectancy may be influenced by such behavior.

The consumption of animal fats (i.e., saturated fatty acids) encourages the development of cardiovascular diseases (Stamler, 1973; Keys, 1970) and some cancers. This relationship is, with rare exceptions, apparent in almost all developed countries (Artaud-Wild *et al.*, 1993). In the United States and in central and northern Europe, where the basic diet is very rich in animal fats, cardiovascular mortality is traditionally higher than in the Mediterranean countries, where there is a greater focus on the consumption of vegetable oil and fish (Pica and Uemura, 1982). In France, the consumption of olive oil seems to have a protective effect against cancer of the colon (Meslé, 1983). Consumer habits concerning sugar also have a strong influence on health and mortality, especially through obesity (Truswell, 1982) and diabetes (Sokolov, 1997; Wunderlich, 1982; Power *et al.*, 1971).

Methods of preparation and eating are also important. Cooking over charcoal favors certain cancers (McBean and Speckmann, 1982); frying in oil causes digestive conditions, especially stomach cancers (Ji *et al.*, 1998); and the ingestion of maté (South America) or of boiling hot tea (China, Japan) cause cancers of the esophagus (Kinjo *et al.*, 1998; Day, 1982).

Balanced Diet

Beyond the harmful character of certain nutrients or the positive effects of others, the general balance of the diet has to be taken into account. One poison may neutralize another, and variety of culinary ingredients may introduce positive synergies. This explains the success of the Mediterranean diet (Kushi *et al.*, 1995; Keys *et al.*, 1986) and elucidates the French paradox (Criqui and Ringel, 1994; Renaud and de Lorgeril,1992; Hegsted and Ausman, 1988; Tunsdall-Pedoe, 1988; Richard, 1987; Saint-Leger *et al.*, 1979), even though certain investigators consider that this paradox results from a later increase than elsewhere in the consumption of saturated fatty acids (Law and Wald, 1999) or even believe that it does not exist (Ducimetière and Richard, 1992). How else can the fact be explained that cardiovascular mortality, which is already exceptionally low at a national level in France, is even lower in the southeast, the region known for cassoulet (i.e., traditional dish based on fatty pork and duck), conserve of duck, foie gras, and good wine?

Food Safety

In addition to quantity, quality, and general balance, healthy eating requires vigilance in matters of hygiene (i.e., food safety). Although food safety is of current interest, especially because of the occurrence of bovine spongiform encephalopathy (not to mention *Listeria* or the possible effects of Genetically Modified Organisms or GMOs), it is far from new. The fear of food poisoning (e.g., poisonous mushrooms, ergot, botulism, saturnism) has occupied an important place in the preparation of food. Although the invention of the feeding bottle enabled certain unweaned infants to be saved, its general use was at first the cause of much infant mortality. It was only with the practice of sterilization that these deaths were avoided. Some important progress is still needed in developing countries as breast-feeding declines.

d. Use of Alcohol, Tobacco, Stimulants, Narcotics, and Tranquilizers

Within the realm of food hygiene comes the consumption of alcohol and that of various substances having antidepressive, stimulant, or tranquilizing effects. Although these substances at first may satisfy physical or psychological needs, they quickly can become the object of excessive consumption and harmful to health.

Alcoholism

The consumption of alcohol, especially of wine, which forms part of the diet in many societies, is not *a priori* a danger to health. A number of studies have shown that regular but moderate consumption of wine may have a protective effect against certain cardiovascular conditions, particularly myocardial infarction (Renaud and de Lorgeril, 1992; Poikolainen, 1982; Saint-Leger *et al.*, 1979). Abuse of alcohol can cause severe harm (Sournia, 1986). Certain countries, such as France (Nizard and Munoz-Perez, 1993; Alderman, 1956, 1964) or Poland (Meslé, 1991; Okolski, 1985a, 1985b) are particularly affected; the most extreme case is that of Russia (Shkolnikov and Nemtsov, 1997; Meslé *et al.*, 1994).

The effects of alcohol vary according to the type of consumption. In a country such as France, there is the practice of regularly consuming excessive quantities of drinks with a low alcohol content, which progressively degrades certain organs, especially the liver, generating an excess mortality due to cirrhosis of the liver. In a country such as Poland or Russia, alcohol is mostly consumed in draughts of strong alcohol absorbed in high quantities and mostly causes death by acute alcoholism and alcoholic psychosis. In all cases, alcoholism has serious repercussions in the area of cardiovascular diseases (Koskenvuo *et al.*, 1986; Leclercq *et al.*, 1979) and certain cancers (e.g., esophagus, upper respiratory tract, digestive tract) (Tuyns, 1982).

Indirectly, alcohol kills through violent deaths. Vehicular accidents are frequently the result of loss of vigilance and diminished reflexes caused by the consumption of alcohol (Got *et al.*, 1984; Norman, 1962). Alcoholism equally encourages suicide and homicide (Ritson, 1977). Alcohol ingestion during pregnancy also compromises development of the fetus (Lee and Leichter, 1982).

Tabacism

Unlike alcohol, the consumption of tobacco is harmful to health from the very first cigarette (Peto, 1986). Tobacco may sometimes contribute to a sense of well-being, but it is always at the price of health risks (Guignon, 1994; Lopez, 1993; Peto *et al.*, 1992; Hill, 1988; Masironi and Rothwell, 1988; Benjamin, 1982). According to the U.S. Surgeon-General (1979), in the early 18th century, J. J. Holland and, then in 1795, S. T. Soemerset drew attention to the relationship between tobacco use and cancer of the lip. At the beginning of the 1950s, following the clinical studies of Frans Mueller (1939), Ernst Wynder Evarts Graham (1950), and Richard Doll and Bradford Hill (1950, 1952, 1954), the relationship between tabacism and lung cancer was firmly established. The effects of tobacco on cancer of the lungs, together with hypercholesterolemia and myocardial infarction, was an example chosen by Daniel Schwartz (1989) to illustrate the epidemiologic approach in causality research.

Although it is known that the consumption of tobacco affects life expectancy through many causes of death, bronchial and lung cancers are most commonly linked to tobacco use and are the dominant tobacco-related causes of death in many populations (Doll and Peto, 1981). In Europe and the United States, tobacco-related deaths became widespread during World War I, when tobacco was systematically distributed to the soldiers in trenches, until the period between 1960 and 1970, when anti-tobacco policies were at last developed (Vallin and Meslé, 2001; Caselli *et al.*, 1999). In addition to bronchial and lung cancer, the use of tobacco products greatly increases the risk of developing cancers of the upper respiratory and digestive tracts, bladder, prostate, and breast (Rosenberg, 1987; The Surgeon General, 1982); certain cardiovascular (especially ischemic heart diseases) or cerebrovascular diseases (Kaprio, 1984; The Surgeon General, 1983; Garfinkel, 1979); and chronic diseases of the respiratory tract (Wilhelmsson *et al*., 1981; Weill and Diem, 1978).

Tobacco is a drug in the strong sense of the term, in that it creates dependence (Ravenholt, 1984). To start smoking is to accept the risk of creating an irrepressible need, an addiction that leads to smoking more and more and prevents the consumer from ceasing to smoke through severance crises. Most frequently, the smoker wishing to abandon tabacism must resort to medical and psychological assistance.

Tabacism also harms nonsmokers forced to inhale the smoke from their neighbors in confined areas (Helsing and Comstock, 1987). Smoke from tobacco may be particularly dangerous for the young children of smoking parents (Poswillo and Alberman, 1992) and for the fetus within the womb of smoking mothers (Oakley, 1989; Sachs, 1987).

Alcohol-Tobacco Synergy

Already harmful individually, alcoholism and tabacism are even more harmful when combined. The synergy produced increases the risks of cancers of the upper respiratory and digestive tracts (Austin, 1982; Mahboubi and Sayed, 1982; Tuyns, 1982; Wynder and Bross, 1961). The synergistic effects of alcohol and tobacco explain the high death rate from carcinomas observed for French men (Nizard and Munoz-Perez, 1993).

Stimulants

To a lesser degree, the abuse of fortifying and stimulating drinks such as coffee may engender health problems (Debry, 1993, 1994; Garattini, 1993; Saint-Arroman, 1845, 1846). These substances are particularly inadvisable for persons suffering from cardiovascular and other illnesses.

Narcotics

Narcotics include heroin, cocaine, and synthetic drugs such as crack and ecstasy (Solanto *et al*., 2001; Harvey and Kosofsky, 1998; Inciardi and Harrison, 1998; Karch 1996; Malizia *et al*., 1993; Ellinwood and Kilbey, 1977). These so-called hard drugs are dangerous for physical and mental health; they can cause death, especially in the case of an overdose. Even more than tobacco, they create dependence on the part of the consumer that can be severed only by detoxification and rehabilitation (Carton *et al*., 1994; Cohen, 1990).

The so-called soft drugs (e.g., hashish, marijuana, quat) are more controversial. These drugs do not immediately appear to be dangerous to health, and their likelihood of encouraging a shift to harder drugs has been challenged (Nahas, 1976; Berke and Hernton, 1974). Even if they are not as dangerous as alcohol or tobacco, soft drugs have serious indirect effects, such as respiratory tract irritation and an increased risk of vehicular accidents (Roques, 1999).

Drug taking by intravenous injection is indirectly a cause of death because it is one of the major avenues for the introduction of human immunodeficiency virus (HIV), the AIDS virus (De Rossi *et al*., 1988), and hepatitis viruses (Bortolotti *et al*., 1986).

Tranquilizers

Tranquilizers are substances that may help in overcoming difficult stages in life and thereby favor survival. However, they still are drugs, and making use of them without strict medical control can cause serious health disorders (Bish *et al*., 1996; Hughes *et al*., 1994).

e. Leisure and Physical Activity

Rest for the body and mind is essential for the preservation of health, but leisure time can be misspent. Increasingly productive machines have freed people from the burden of work and have enabled long periods of rest and leisure. In relying too much on energy-saving devices, people can deprive their bodies of the exercise necessary for good functioning and create opportunities for new pathologies.

Has there ever been a finer discovery than that which enables the control of electricity? Did not Lenin define communism as "socialism plus electricity"? However, in almost eliminating the difference between day and night, this marvelous invention has opened the door wide to the disturbance of the activity-rest cycle imposed by the earth's rotation. Management of sleep has become the object of individual or collective living behavior, and the alterations can have consequences on health.

Another invention, an extension of the first, is the loudspeaker, which adolescents have adopted to enjoy the excitement arising from the excess of decibels in hard rock music, a practice associated with a high risk for deafness (Tubiana and Amadio, 2000).

Behavior concerning exposure to the sun is of considerable importance in preventing certain conditions. Solar radiation facilitates the fixation by the organism of indispensable substances such as iron and calcium. Exposure to sunlight promotes the synthesis of vitamin D, essential for calcium fixation. Lack of exposure to the sun is one of the well-known causes of rachitis (Paunier, 1984), and it may be encouraged by too much protection from solar radiation. However, too much exposure provokes skin cancers (Dubertret *et al.*, 1995; Blum, 1959; Dubreuilh, 1907). The incidence of all types of skin cancers has greatly increased among northern populations due to their infatuation with sun tanning (Scotto and Fraumeni, 1982), and the incidence varies with latitude (Lee, 1982).

Even though the Olympiads were a Greek invention many centuries before our time, sport can be considered as the modern antidote for the reduction of physical activity in work. In this sense, the practice of sport is largely beneficial to health (Surault, 1989, 1996; Paffenbarger *et al.*, 1978). It is an important factor in the prevention of cardiovascular diseases (Paffenbarger *et al.*, 1986; Kannel, 1967; Morris *et al.*, 1953a, 1953b) and osteoarticular diseases, especially osteoporosis (Daniels, 2000). However, sporting activities may have serious consequences, even leading to death. The risk is even greater when the sport is practiced intermittently and without adequate training; skiing accidents are typical. Professional athletes, with or without doping (Yesalis, 2000; Clark, 1990), incur accidents at work and sports-related diseases, to the extent of generating a new medical specialty: sports medicine (Magnin and Cornu, 1997).

2. Sexual Behavior

Sexual activity, primarily necessary for the reproduction of the species but also useful for individual development, encompasses a panoply of risks that largely depend on behavior. The most prominent risk is that of sexually transmitted diseases, such as syphilis, gonococcemia, blennorrhagia, yaws, and AIDS. In a strictly monogamous society, these diseases would not qualify for mention, at least those for which other ways of transmission are insignificant.

Diffusion of these diseases largely depends on sexual behavior. In the case of AIDS, the reason for the very rapid transmission of the virus among homosexual populations was the frequency of multiple partnerships and the high renewal rate of partners (Konings and Caraël, 1997; Caraël, 1995; Anderson *et al.*, 1990; Anderson, 1988). If in sub-Saharan Africa the epidemic quickly spread among the heterosexual populations, it was because they widely practiced multiple partnerships and frequently changed partners (Brouard, 1987, 1994). Behavior regarding the choice of partners also affects the spread of disease. Nicolas Brouard (1994) has shown that the dynamism of the African epidemic is largely a result of the diversity of age differences between partners; a strictly constant difference would in the end entail the disappearance of the disease.

As with the spread of disease, prevention depends greatly on sexual behavior, whether of contaminated persons avoiding transmission of the disease or of healthy persons avoiding contraction of disease. In both cases, two types of elementary precautions may be very effective: abstinence (or at least observation of strictly reciprocal monogamy) and systematic protection of the relationship (i.e., condom or pessary). Both cases entail a general change in behavior, which is often difficult to attain.

Sexually transmissible diseases are not the only area of sexual life for which behavior may have an effect on health and mortality. Certain practices in traditional societies, allied or not to initiation rites, may be disastrous for health. These extend from more or less serious sexual mutilation imposed on women, ranging from the excision of the clitoris to the ablation of the major and minor labia, and to the even more barbarous practice of infibulation, which is extensively practiced in the countries within the Horn of Africa. Besides the suffering and deprivation of pleasure that they cause, these practices increase women's risks for infection, gynecologic diseases, and obstetric difficulties (Locoh, 1998; WHO, 1997, 1998).

Modern contraception under medical supervision has a positive effect on health (Rosenberg, 1991). Consultations concerning adoption or the follow-up of contraceptive methods have enabled progress to be made in many countries in the prevention of uterine and breast cancer, for which early detection is decisive (Wingo, 1993; La Vecchia, 1990; Hatton *et al.*, 1985a, 1985b). Abortion, depending on whether it is practiced officially in a hospital environment or clandestinely in more or less questionable venues, involves very different risks for the health and survival of the mother (Abou-Zhar and Royston, 1991; Measham *et al.*, 1981). In the words of Axel Mundigo (1996): "The legal status of the abortion determines to a large extent whether it takes place under safe or unsafe conditions."

3. Life Setting and Professional Activity

Other areas in which lifestyles may influence health include habitat, transport, work, and migration.

a. Habitat

It is unnecessary to dwell at length on the importance of housing conditions. From the prehistoric cave to the apartment block of modern times, the average conditions of comfort and quality of life settings have greatly evolved, but there remain an immense variety of situations. Access to a certain type of accommodation is not solely a matter of individual behavior (we return to this topic later), but it may have a pronounced effect on health.

Certain elements of comfort are of great importance. In urban housing in northern countries, the lack of ventilation, insufficient exposure to sunlight, and overcrowding formerly played major roles in the development of tuberculosis (McDougall, 1949; Greenwood, 1926), and improved standards have considerably contributed to a decline in unhealthy living conditions. In Sweden, the introduction of double-pane windows, leading to air-tight enclosure, probably played a part in propagating tuberculosis (Greenwood, 1924). The widespread adoption of central heating systems and the use of polyvalent vaccines have probably contributed to the decrease in great influenza epidemics. Nevertheless, the precariousness and unhealthiness of some housing remains a major factor in public health (Burridge and Ormandy, 1993), especially in developing countries. On the margins of society, the extreme situation of those with no fixed abode involves serious health problems (Barow *et al*., 1999).

b. Transport

Whatever means of transport is employed, traveling involves risk, but it varies from one form to another and depends how the chosen means is used. Generally, travel by rail is less dangerous than by air or sea, which are themselves less dangerous than traveling on roads (Vallin and Chesnais, 1975). On the road, the pedestrian, cyclist, car driver, motorcyclist, and public transport user have different risks. All of them, especially those holding handlebars or a steering wheel, run risks more or less consistent with their behavior. It is in the attempt to minimize these risks that highway rules and regulations have been developed and safety measures increased. The most marked differences in behavior essentially concern respect for these rules. Obeying speed limits (Vallin and Chesnais, 1975) and wearing a seat belt appear to be most important (Chesnais and Vallin, 1977; Nelson, 1974). It is quite significant that in all the European countries, in a surprisingly simultaneous way, the ascending curve of road deaths reversed around 1974, the year of the first oil crisis (Vallin and Meslé, 2001). The sudden rise in fuel prices aroused awareness in most people of the value of speed limits that could spare human lives and their pockets. Moreover, alcohol use by drivers considerably increases the risk of vehicular accidents (Got *et al*., 1984; Norman, 1962).

c. Work

Working conditions are an important factor in public health (Costa and Segnan, 1988; Volkoff, 1987; Cassou *et al*., 1985; Chesnais, 1985a, 1985b; Duclos, 1984; Dejours, 1980; Vallin, 1979a). This situation became a public health issue with the advent of industrialization (Haines, 1991; Dublin, 1917), and in this area, where industry (Goldsmith, 1982) or agriculture (Laurell, 1982) are concerned, there is an immense diversity of situations. Occupations that correlate with serious risks include building, chemical manufacturing, and mining, not to mention the police or the army in time of war. For the same type of employment, there is an immense variety of situations because of diversity in the state of economic development, the social surroundings, and respect for the standards in force (by the employer and by the employee himself). Industrial injuries (Héran-Le Roy, 1999; Davis and Friedman, 1989; Molinié and Volkoff, 1984; Leplat and Cuny, 1974) and officially recognized professional diseases (Couturier, 2000; Martin, 1975) are often only the visible part of the iceberg. Apart from the appearance of specific diseases, working conditions contribute in various ways to the progressive deterioration in the state of health. The rhythm of work, cleanliness of the working area, difficulty of the tasks, and stress may cause premature weakening of the body (Surault, 1995a; Vachon, 1987; Wolf and Finestone, 1986). However, safety at work and appropriate sanitary provisions are factors in conserving health, and part of a balanced life is achieved by good social integration through work (Baldge and Krieger, 2000). Unemployment has a negative effect on health (Valkonen and Martikainen, 1995; Gélot, 1992; Smith, 1987; Platt, 1984; Bungener *et al*., 1982).

d. Migration

Migration has an effect on health and survival. It is quite often motivated by the desire to improve living conditions and, in the case of refugees and asylum seekers, even to escape from life-threatening situations. The act of migrating itself often involves an important part of the risk, especially in the case of forced (refugees) or clandestine migration. The risks undertaken are sometimes enormous, whether they concern long journeys plagued with difficulties (e.g., drama of *boat people*), the horrible destiny of the 58 Chinese found dead in a container in Dover in June 2000 (e.g., see *Le Monde*, 23 June), or the amazing adventure of the young Senegalese who traveled for 5 hours in the undercarriage compartment of an Airbus[2] between Dakar and Paris in January 1999, in a temperature of –50° C.

[2] Unfortunately, he did not profit from his incredible exploit and was expelled without further consideration (*Le Monde*, March 16, 1999).

At the place of arrival, migrants usually encounter health conditions that are better than those from which they came. Admittedly, the extent of this phenomenon is difficult to measure, because although the death rate of migrants is lower than at the place of departure, it partly results from the fact that migration attracts individuals in good health and capable of "making the journey" (Trovato, 1998; Courbage, 1996). In the case of South-to-North migration, health conditions are better in the country of arrival than that of departure. This does not necessarily mean that they are the same for the migrant as for the native of the country of refuge (Rice, 1999; Trovato, 1991). Most frequently, the migrant accepts less favorable living conditions and accepts work that is rejected by the native inhabitants (Piore, 1979) and that is riskier (Travato, 1991).

South-to-North migration is not the only model. Internal migrations have sometimes produced (in northern countries) or still produce (in southern countries) various effects on health. At a certain stage of economic and social development, urbanization attracted people in search of employment to cities, where they were crowded in unhealthy conditions worse than they had known in their country districts. In 19th century Europe, this resulted in marked urban excess mortality (Williams, 1989; Rödel, 1988; Robert, 1988; Davis, 1973). Today, this situation is partly compensated in southern countries by the existence within cities of minimum health provisions, which are not found in outlying country districts (Vallin, 1989, 1992), and mortality is often much higher in the rural environment (Mozumder *et al.*, 1997; Gichuhi, 1993).

Migration is sometimes a decisive factor in transmitting diseases (McNeil, 1980). It has played a big part in the spread of AIDS in Africa (Lalou and Piché, 1994).

4. Relationships

Personal relationships influence health and survival. We provide three examples in the following sections.

a. Raising Children

It is well known that the behavior of mothers and more generally of parents toward children from birth to the beginning of adult life is decisive for the physical and mental health of their offspring. From the stage of pregnancy onward, the maternal and family environment may be favorable or otherwise for the unborn child (e.g., medical follow-up of the pregnancy, type of activity of the mother, maternity leave, consumption of alcohol, tobacco, or drugs) (Inciardi *et al.*, 1997; Klerman and Johnson, 1991). The type of feeding (i.e., breast-feeding or other forms) (Palloni and Tienda, 1986), weaning conditions (Beverly, 1982; Cantrelle and Leridon, 1980), quality and quantity of nourishment, resort to medical care, hygiene, and accident prevention depend very much on the family environment and especially on the mother. Some studies show, for example, that within the same family, different treatment may be given to children according to their sex, especially in societies in which the preference for boys is very strong, resulting in excess mortality for girls. This has been especially demonstrated in Bangladesh (Chen *et al.*, 1981), in Algeria (Vallin, 1979b), and in China (Calot and Caselli, 1988, 1989). This phenomenon was formerly evident in Europe (Pinnelli and Mancini, 1992, 1999; Tabutin, 1978).

Children also may be treated differently according to other characteristics. In sub-Saharan Africa, twins are often treated differently from single children, whether for worse or better, according to whether they are culturally venerated or cursed (Pison, 1989, 1992).

It was recently discovered that the sudden death of infants, for a long time unexplained, depended to a large extent on how the babies were put to sleep. It sufficed to lay them on their back or side instead of on their stomachs for this phenomenon to decline decisively (Becroft *et al.*, 2001; Flick *et al.*, 2001; Gessner *et al.*, 2001; Sullivan and Barlow, 2001). It is also known that exposure to nicotine, another factor depending on the attention of the parents, favors the sudden death of the infant (Milerad *et al.*, 1998).

The failure of families to provide for the rearing of children obviously endangers the life of the newborn. In the extreme case, abandonment (*exposure* as they formerly called it) is a flagrant cause of excess infant mortality. This occurrence was formerly well known in Europe (Rollet, 1997; Buccianti, 1993; Corsini, 1991) and still is today in certain developing countries (Seboui, 1990; Mehta, 1982). It has recently been dramatically aggravated by the spread of AIDS, which has increased the cases of abandonment and made the lives of abandoned children even more precarious (Anonymous, 1999). Even when the collective responsibility for abandoned children is suitably assumed, the children continue to have higher than average risks for disease and death. In certain developing countries, the solution has been adopted of entrusting the children to another family for a longer or shorter period. This system may, according to the context, give the child better or worse chances of survival (Bledsoe and Brandon, 1989). The practice of putting an infant out to be nursed may also have positive and negative effects. In 19th century Europe during the course of industrialization, families of workers were frequently led to do this under terrible conditions, which in France were the main cause of the upsurge in infant mortality between 1850 and 1880 (Rollet, 1990).

Beyond infant health and mortality, maternal and family behaviors also have an influence on survival to adulthood, especially through hygiene and by means of health education. All sorts of events experienced in childhood may have repercussions in terms of survival into adulthood (Marmot, 1995; Caselli, 1995; Menahem, 1994; Wilmoth *et al*., 1989, 1990; Horiuchi, 1983).

b. Violence

Violent behavior has an influence on health and mortality. Domestic violence is often little publicized, but it is present everywhere, in all societies and social groups: ill-treated children (Kashani and Allan, 1998; Ariffin *et al*., 1997; Burgess, 1978) and violence toward women (Taylor, 1997; Nowrojee and Manby, 1995), including sexual violence (which is not limited to the mutilation rites described earlier) and physical violence.

Even in the absence of war or major political unrest, homicide is a cause of death that occupies a highly variable place from one country to another and from one epoch to another (Chesnais, 1976). In certain countries, death rates from homicide are very high. This is especially the case in the United States (Lester, 1993; Iskrant and Joliet, 1968) and even more so in Russia (Meslé *et al*., 1994).

Collective violence is also a factor. Victims sometimes number in the millions or tens of millions (Lancaster, 1990; Chesnais, 1976; chap. 33, p.314–340). Evaluation of the excess mortality of war in the Ukraine between 1941 and 1945 produced an estimate that life expectancy fell to 14 years for men and 21 years for women (Vallin *et al*., 2002)!

c. Importance of Relational Networks

Although the family circle is of prime importance for the health and survival of children, the family and the operation of all kinds of relational networks have parts to play at all ages and especially for older people. Loneliness and abandonment are well-known factors in the decline of health and hastening of death (Sabin, 1993), whereas the quality of intergenerational cooperation helps to maintain good health and prolongs life (Avlund *et al*., 1998; Steinbach, 1992; Rogers, 1996; House *et al*., 1988; Pous *et al*., 1988).

In the extreme case, the resort to suicide, an eminently individual act, is rarely unrelated to the family circle. Integration within a sound network of family and social relationships is a determining factor in the prevention of suicide (Besnard, 1997; Surault, 1995b; Lester, 1994; Baudelot and Establet, 1984; Durkheim, 1897).

5. Synergies

This long list of behavioral factors affecting health and mortality is inadequate for summarizing the situational diversity. Although, in theory, each of these factors can act negatively or positively in isolation while all the others remain neutral, most frequently, many factors combine in the same or in opposite directions. When several factors operate together in the same direction, a synergy may develop in which the result exceeds the simple addition of the expected results of each factor acting in isolation. Mention has been made of alcohol-tobacco synergy responsible for the tumors of the upper respiratory and digestive tracts. The numerous examples of this type include speed and alcohol while driving and tobacco use and silicosis among miners. The accumulation of risk behavior is often the cause of great differences in mortality observed between social or family groups (Das Gupta, 1990).

II. SOCIOCULTURAL FACTORS AND ATTITUDES REGARDING THE BODY AND HEALTH

Discussion of behavioral factors in the preceding sections must not allow the fact to be overlooked that health and survival are determined by more than these behaviors. Health and survival depend on processes that are largely determined by the context. Individual behavior is derived from general attitudes regarding the body and health, which are themselves allied to culture. These attitudes are expressed according to the resources available to the individuals, and they are largely conditioned by the place people occupy in society. Sanitary and social policies affect behavior, whether directly through awareness campaigns or indirectly through improvements in the services and their accessibility.

1. Attitudes Regarding the Body: A Question of Culture

Attitudes regarding the body, health, disease, death, and medicine influence lifestyles and behavior in very diverse areas, particularly that of prevention and resort to the health care system. These attitudes are allied to representations that are primarily social.

Although the body has an anatomic and physiologic reality, it is also, as David Le Breton wrote in 1985, a social construction "both in its involvement in the collective scene as in the theories that explain its functioning or the relations it maintains with the person that it embodies. Biomedical knowledge, and official knowledge of the body in our Western societies, is one representation of the body among others, and effective in the practices that it supports" (Le Breton, 1992).[3] There exist as many representations as there are societies and cultures, with as many different actions based on these different representations (Loux, 1979).

The functions of the body, its uses, its image, and the attention paid to it vary according to societies (Bouchard, 1982) and according to regional and ethnic cultures or social class within the larger societies (Surault, 1983). Values and standards lead to the development of "a sort of code of correct behavior toward one's body" (Boltanski, 1971)[4] and "the bodily experience of every person is penetrated through and through by other people and by society, as the source, organ, and support of all culture" (Bernard, 1973).[5] Ethnologists have frequently stressed the extreme variety in the conception of the body from one society to another.

Lucien Sfez (1995) compared conceptions of the body in the United States, Japan, and France, showing how the differences explained the diversity of behavior and attitudes with regard to health. In the United States, for example, the search for "perfect health" justifies development of a new attitude concerning, among other things, food (without cholesterol), sexual practices with a single partner (danger of AIDS), and struggle against tabacism (safety for oneself and for others), as well a priority given to biotechnologies and extensive programs such as the worldwide Genome project for mapping and sequencing human genes (in favor especially of predictive medicine) and the Biosphere II project, a reduced model of the terrestrial biosphere, in which four men and four women sequester themselves for 2 years (1991–1993).[6]

Representations and uses of the body determine the representations of disease and health, as well as the attitudes concerning them. Disease and health, beyond their biologic aspects, are also social facts: their recognition is allied to the cultural values and standards of a given society or social group; depending on the case, a biologic state is perhaps determined but not regarded as a disease (Foucault, 1972).

In industrialized countries, the increasing attention paid to the body explains the demand for perfect health and even for immortality, and it justifies the means of achieving these goals, such as increased health expenditures and public health policies. The more health improves in terms of longevity, disability, incapacity, or handicap, the greater is the extension of the field of disease and the power of the medical disciplines, such that organic or psychological manifestations are increasingly being classified as morbid symptoms (Lachaud, 1975).[7] In France, for example, the average number of diseases declared per person (i.e., prevalent morbidity) doubled in 20 years from 1.6 in 1970 to 3.3 in 1991 (Sermet, 1994), whereas the rate of persons declaring themselves to be in perfect health (i.e., feeling no discomfort in their daily life) is decreasing (19% in 1980 to 15.8% in 1993) (Bocognano *et al.*, 1994).

Health and longevity are increasingly becoming objectives in themselves and no longer just the means for a better quality of life. The more uncommon disease becomes, the more distant the foreseeable horizon of death, the greater the fear of disease and death,[8] and the greater the requirements for medicine and medical technology. This accounts for the growing medicalization of social, family, and psychological problems and of all that arises from discomfort or merely fatigue, increasing the dependency on medications (Reynaud, 1984) and explaining the success of "happiness pills."

This information points to an astonishing paradox. Although the effectiveness of medical interventions is undeniable, it is also possible to imagine that medicalization and the development of preventive practices pushed to extremes can become a constraint, having negative effects on health through an increase in stress and anxiety, which can help to generate serious conditions, including cancers (Amiel-Lebigre

[3] "Tant dans ses mises en jeu sur la scène collective que dans les théories qui en expliquent le fonctionnement ou les relations qu'il entretient avec l'homme qu'il incarne. Le savoir bio-médical, savoir officiel du corps dans nos sociétés occidentales, est une représentation du corps parmi d'autres, efficace dans les pratiques qu'il soutient."

[4] "Une sorte de code des bonnes manières d'être avec son corps."

[5] "L'expérience corporelle de chacun est pénétrée de part en part par autrui et la société, comme source, organe et support de toute culture."

[6] Lucien Sfez defines *perfect health* or *full health* as an already unified bio-eco-religion, a 21st century utopia, armed with all its symbolic power. In France, *prevention biology* encounters active resistance (Agora and CREMMS, 1994; Malet, 1993).

[7] A quarter of a century ago, Ivan Illich (1975) was speaking of "the elimination of the status of health by unlimited multiplication of the roles of sickness" ("l'élimination du statut de santé par la multiplication illimitée des rôles de malade").

[8] In France, according to the annual CREDOC investigations, the fear of serious illness is constantly increasing. It has become the primary worry (before drugs, insecurity, and unemployment), representing the nonacceptance of what it might bring: death.

and Gognalons-Nicolet, 1993; Dutot and Lambrichs, 1988).

A number of studies in industrialized countries have revealed significant differences according to social environment and education level in the representations of and attitudes toward the body, pain, disease, health, and medicine (Adam and Herzlich, 1994; Aiach *et al*., 1992; d'Houtaud and Field, 1989; Surault, 1984; Aiach and Cebe, 1983; Loux, 1983; Aballea, 1982; Letourmy ad Pierret, 1982; Aiach *et al*., 1980; Zborowski, 1972; Herzlich, 1970; Zola, 1966). In less favored circles and for a given pathology, there is a reduced capacity for identifying with the morbid condition, a less intense reflexive relationship with the body, a greater resistance to pain, and a reduced degree of medicalization, leading to a reduced perception and identification of the disease than among better-educated groups and, consequently, a lower rate of declared morbidity. Similarly, a reduced level of resort to care correlates with an equally declared morbidity, less confidence in medical techniques, a broader definition of health,[9] and fear of incapacity (allied to a real or imaginary experience of the disease and to the level of social protection and resources). The same social differences may be observed in preventive practices (Aiach *et al*., 1992).

In the Western world, since the 18th century, the improvement in health and decline in mortality have changed attitudes toward disease, health, and death and reciprocally, according to a circle that remains up to the present "virtuous," without our knowing what development preceded the other. Previously, disease and death, including high mortality among young children, were accepted with resignation. Analysis of death and disease remained almost absent from the literature until the 20th century. According to Philippe Ariès (1971), death appeared "neither as an accident that a procedure of attention and technique might avoid or postpone[10] nor as a subject of individual affliction. It was a matter of destiny that had to be accepted, even assisted, and also appropriately surrounded by ritual ceremony."[11] The preoccupation was not to avoid the end, but to be aware of its coming. The mere idea of a struggle against death was considered to be sacrilegious, as if to oppose divine will.

It was more or less at the time when deaths began to decrease essentially as a consequence of economic development and its effects that attitudes changed. The child was increasingly seen as a completely separate being (an idea almost absent from earlier literature), and the possibility and desire to act on its destiny emerged. In France, with the dissemination of the ideas of 18th century philosophers and the loss of influence of the Catholic Church (recorded especially by Alfred Sauvy [1960]), attitudes changed more rapidly, as evidenced by an early decline in fertility—because happiness, according to Jean-Jacques Rousseau (1782), was no longer to be sought exclusively in the hereafter, but also on Earth.

This process cannot be applied today to countries with a high death rate, especially because they constitute a cultural ensemble that is much more heterogeneous than in the former European countries,[12] even though the spread of Western culture and ideas occurred there long ago, at least among the elite. Nevertheless, the decline in mortality, especially among very young children, has increased the desire to have control over one's own life and that of others, as can be seen by the general tendency to control fertility.

Attitudes toward medicine are also conditioned by the sociocultural context, which explains attitudes about the body, disease, and death (Adam and Herzlich, 1994; Augé and Herzlich, 1983). For most of the population, the body is an instrument for work, and as long as it is as able to function as such, the person is not considered to be ill. Consequently, as Henry Mosley (1985a) points out, more attention is paid to prevention than to therapeutic measures: "Every stage of life and frequently almost every daily activity is accompanied by rituals and actions designed to ensure health and welfare. Furthermore, life may be filled with all sorts of taboos, many of them related to diet, to avoid sickness" (p. 130).

Mosley, in attempting to analyze the effectiveness of programs for primary health care, stressed the influence of attitudes toward disease and traditional systems of belief; taking them into account or ignoring them may influence the programs' effectiveness. On the contrary, problems sometimes encountered in the

[9] As long as somatic symptoms, diagnosed or not, did not prevent work, a person did not consider herself or himself to be ill. Essentially, health had a market value; it enabled work and the income arising from it to be realized. To this can also be added a nonmarket value, because it enabled optimization of nonworking time, particularly that devoted to leisure pursuits (Letourmy and Pierret, 1982).

[10] For centuries, the practice of medicine remained restricted to a very limited circle, that of the Court, nobility, Princes of the Church, and the prosperous middle class.

[11] "Ni comme un accident qu'un effort d'attention et de technique peut éviter ou reculer, ni comme un sujet d'affliction individuelle. C'est un arrêt du destin, qu'il faut admettre et même aider, qu'il convient aussi d'entourer d'une pompe rituelle."

[12] François Laplantine (1986) speaks of "the infinite variety of reactions toward disease according the societies and epochs" ("la variation infinie des réactions à la maladie selon les sociétés et les époques"), even though the etiologic and therapeutic models are those, very limited in number, that were not exclusive from one another in the same society or even with the same individual.

application of medical techniques that have proved their worth in Europe or North America may result from the fact that modern medicine "also incorporates many popular beliefs and folk practices of Western culture" that are often very far removed from those of developing countries (Mosley, 1985a).

Dominique Tabutin and Eliwo Akoto (1989, 1992) have shown that ethnic membership has a great influence on the mortality of children in sub-Saharan Africa, which confirms the importance of cultural variables that affect the care given to children. "The ethnic group, in their estimation, may influence the mortality of children through a whole series of variables, such as the attitude toward disease, death and care services, beliefs relative to the etiology and therapeutic art, practices in the feeding of children, the care given them (Hill and Randall, 1984), family structures, age at delivery . . ."[13] (p. 61). A number of studies provide similar results (Fassin, 1995; Gbenyon and Locoh, 1989, 1992; Pison, 1989, 1992; Raynaut, 1992; Soumoy and Thiltges, 1992; Vallin, 1989).

Despite the progress made in the epidemiologic transition (see Chapter 57), attitudes concerning the body, disease, death, health, and medicine remain extraordinarily varied, just as varied as the cultures are that underlie them. From the denial of death, symbolized by the American practice of presenting the deceased sitting in an armchair with every appearance of life, to the almost impatient desire for it (i.e., the *tamed death* of Philippe Aries [1977]), as evidenced by the lack of concern of certain populations regarding AIDS; from *perfect health* as the guide to life and subject of all conversations to survival-motivated health practices; from disease that is destructive to that which is liberating (Herzlich, 1992), there are an infinite number of representations and intermediate attitudes, all of which influence preventive behavior,[14] resort to treatment, and death.

In a country such as France, the frequency and importance of alcohol consumption is explicable by a culture rooted in history from the Gallo-Roman epoch, when wine production was introduced. Made sacred by religion, wine has also become a socializing agent, a means of social relationships, which furthers a status-enhancing image of the "social utility, relaxation, and ritual conviviality"[15] factor (Chapuis, 1989), handed down from generation to generation.

Death from a vehicular accident provides a good example of the influence of behavioral differences allied to specific cultural characteristics. Whichever indicator is used, mortality in Europe is lower in the countries of the North. Jean-Claude Chesnais (1985a, 1985b) considers that "the high degree of safety there corresponds to a cultural trait shared by the people of the North—supportive societies where tolerance is higher, collective discipline stronger, and prevention consequently easier."[16] The most stringent policies in Europe for the repression of alcoholism at the wheel were developed in Norway and Sweden. In most Latin countries (including France), this mortality, although declining, remains higher; their inhabitants are more individualistic, tend to favor their own interests rather than that of the majority, and therefore adapt themselves less easily to preventive measures. Especially among men, the automobile is not used solely as a means of transport; it is also a symbol of power, virility, and social prestige (Baudrillart, 1976).

Many examples from traditional societies can be cited, such as in the Sahel environment, where red urine (one of the symptoms of bilharziasis) is associated with virility (i.e., small boys playing in the backwaters are more exposed to it than girls), or the theory of the evil eye in certain Mediterranean populations, leading to the confinement of babies to the interior of the poorly lit and badly ventilated dwellings to avoid exposing them to another person's view.

Measles merits particular mention. For a long time, the synergy between malnutrition and infection was conceded to be the sole factor explaining the highly lethal effect of measles in southern countries and especially Africa, whereas it remains a benign disease in northern countries. The undernourished child is less resistant to the virus, and the sick child in losing its appetite creates a situation that transforms measles into a mortal disease (Solimano and Vine, 1992; Mosley, 1985a, 1985b). The contention was that children did not die directly from the measles but from its serious complications when children were undernourished. In 1980, Peter Aaby demonstrated that such was not the case. The lethal effect of measles in Africa results from lifestyle. Children who live in the overcrowded dwellings, where all or most of the family members sleep confined to a single room, are exposed

[13] "L'ethnie, estiment-ils, peut agir sur la mortalité des enfants à travers toute une série de variables telles que l'attitude face à la maladie, à la mort et aux services de soins, les croyances relatives à l'étiologie et à l'art thérapeutique, les pratiques d'alimentation des enfants, les soins qui leur sont accordés (Hill et Randall, 1984), les structures familiales, l'âge à l'accouchement...."

[14] During a recent conference, professor M. Tubiana cited the results of an American investigation, according to which the Seventh Day Adventists, whose religion forbids them to smoke and drink alcohol, benefit from a cancer morbidity rate that is less than 50% of that of the other inhabitants in the same districts.

[15] "D'utilité sociale, de détente, de rituel de convivialité."

[16] "La grande sécurité y correspond à un trait culturel des peuples du Nord: sociétés solidaires où la tolérance est plus grande, la discipline collective plus forte et, par conséquent, la prévention plus aisée."

to and die of measles. A child contaminated under these conditions effectively receives a strong viral dose that increases the risk of having a fatal disease, whereas the same child catching measles when playing outside with his or her friends develops a benign form of measles (Aaby, 1989, 1992). Measles is more often fatal in West Africa, where polygamy is more widespread and the co-spouses and their children sometimes sleep in the same room, than in East Africa, where there is less polygamy and cohabitation of the co-spouses is rare. Culture and living conditions can strongly influence lifestyle, which may directly affect mortality (Wunsch, 1981).

2. To Know and To Act: A Question of Means

Although attitudes regarding the body and health vary according to the cultural context, involving different ways of representing and perceiving disease, and consequently a greater or lesser disposition to prevent or treat health problems within the same country or culturally homogeneous ethnic group or region, the behavior that arises from these attitudes can also greatly vary according to the means at the disposal of the individuals or families. Mortality can be differentiated by three types of means: education and access to information media, material wealth (income or fortune), and existing health services and their availability.

a. Education

Knowledge is one of the primary resources for encouraging the adoption of behavior conducive to health. There is a close link between mortality and the level of education (Kunst and Mackenbach, 1994). All the investigations carried out on social mortality, especially in industrial countries, fully demonstrate this principle[17] (see Chapter 54). It may partly depend on selection by education of the healthiest individuals (i.e., mentally deficient or sick children are less able to undertake studies) or on the indirect relationship between education and health through contextual variables (i.e., persons belonging to disadvantaged social categories are subject to unfavorable health conditions and have difficultly accessing higher levels of education). Increasing levels of education have increasingly positive effects on health. Education provides knowledge about what is logically beneficial for health, and it is in many ways a means of accessing the power and resources necessary for carrying out preservative or recuperative health activities.

Although all people can, because of their education, act directly on their own health, the education of a family member or household may also influence the health and survival of other members. This is particularly true for the mother in relation to her children. In France, for example, infant mortality entered a very rapid and enduring downward phase in the late 19th century. It was largely a result of the new approach to infectious diseases (especially those of infants) after the discoveries of Louis Pasteur, but the decline in infant mortality also reflected the spread knowledge through general public education after the laws of Jules Ferry on compulsory schooling were enacted.

In developing countries, the education of mothers has been recognized by demographers as one of the prime factors in lowering infant mortality (Caldwell and McDonald, 1981; Behm, 1979; Caldwell, 1979; Behm and Primate 1978). Similar to a number of investigators, Henry Mosley (1985a, 1985b) stressed the importance of the educational level of women in generating a social synergy that enabled a change in attitudes toward disease, its prevention, and treatment and, more generally, toward the child and its survival. In particular, observations in Kenya and other African countries showed that maternal education and income level (or poverty) are related to infant survival.

Dominique Tabutin and Eliwo Akoto (1989, 1992) observed that in sub-Saharan Africa, there is a positive relationship between the educational level of the mother and the survival of the children. The intensity of the relationship varies considerably among countries, and it is more apparent for mortality between 1 and 5 years than before 1 year. The effect of the father's educational level is less clear, but it complements that of the mother.

b. Wealth and Economic Affluence

Economic affluence is an obvious factor in the access to health resources. With incomes that are just enough to satisfy the most elementary daily needs, the treatment of a disease may appear all the more superfluous in that it is not always assured of success. The expense entailed in prevention constitutes a gamble that may not seem urgent to undertake.

The effect of income in terms of health behavior is often less than that of education. At an individual level, the priority for health expenditures does not change much according to income. The higher the income, the greater the increase and diversification of needs, and this is not always to the advantage of health expenditures, even though these increase with income.

[17] The recent synthesis study carried out by Kunst and colleagues (2000) for the European Union confirms this fact.

At an equal collective level of income, resort to medical care depends greatly on the accessibility of heath resources. According to the health policies applied and the system of assuming responsibility (e.g., social security) in force, revenue differences will have a greater or lesser influence on the resort to preventive care (Bell, 1985a, 1985b).

c. Accessibility of Health Services

It does not suffice to pay attention to one's health, to be informed about the means of proceeding, or even to have sufficient resources to pay for care. Effective care must exist (e.g., for infectious diseases), and the services must be available. However, we have seen the time required by medical research to arrive at a treatment, largely insufficient, for AIDS. The example of AIDS is symbolic of the difficulty of making an existing treatment accessible in poor countries (Mann *et al*., 1992). As the French Minister of Health Bernard Kouchner declared in 2001, "90% of the medicaments are in the north, whereas 90% of the sick are in the south."[18] Even though the conflict between the big pharmaceutical companies and the governments in southern regions, such as India, Brazil, or South Africa (who wish to produce cheaper medications by freeing themselves from the licenses) is being resolved, treatments for diseases such as AIDS remain beyond the economic reach of most sick persons.

The distribution of the health services and the price of the medication or treatment are not the only issues. Concerns exist for the whole chain of prevention and assumption of responsibility, including the identification of risks, detection and diagnosis of cases, treatment, follow-up, and responsibility for the sick person at health, social, psychological, and even economic levels. Although AIDS is one of the most striking examples of the important part played by the availability of services, the same applies for many of infectious endemics in southern countries, especially Africa, whether it concerns tuberculosis, measles, malaria, or other tropical diseases.

What remains true for infectious diseases in a number of developing countries is everywhere evident for diseases against which the struggle has only more recently been effectively engaged: cardiovascular diseases and cancers. One of the most striking illustrations is the introduction of emergency intervention services (SAMU in France), which are particularly decisive in the treatment of myocardial infarct. With equal attitude and behavior, the chances of a sick person being treated in time and of being able to recover depend on the presence and effectiveness of such services. Similarly, the arrival of scanners (e.g., computed tomography, magnetic resonance imaging) created a fairly marked disparity in the quality of diagnoses and therefore in the adoption of appropriate treatments according to the ease or difficulty (geographic or other) of access to this type of apparatus.

[18] The declaration of "90% des médicaments sont au nord alors que 90% des malades sont au sud" was made for the first time in the *Journée Mondiale du Sida* on December 1, 2000 and repeated many times since (e.g., see *Le Figaro* of January 30, 2001 or *Le Monde* of July 7, 2001).

3. Social Conditioning

a. Social Categories

Attitudes toward the body and behavior in health matters are largely conditioned by the social surroundings (Valkonen, 1987). This issue is addressed in detail in Chapter 54. It suffices here to stress that social adherence influences almost all of the aspects of health previously described. Each social group has its own culture that affects the attitudes and behavior of its members, and adherence to a social group affects access to the means of action, like education and economic resources. However, the relationship between social class and mortality considerably exceeds the limits of the behavioral mortality factors alone. Some of the differences in mortality observed between social categories arise from selection according to health, which influences access to other social categories. For example, health problems may compromise access to good training, which compromises access to good professional qualifications and therefore to interesting and well-paid employment. Deficient health makes upward social mobility more difficult, and in this way, membership in the higher social categories itself depends on the state of health. However, social categories are largely founded on professional activity, which itself greatly affects mortality through the conditions of work and other profession-related risk factors. This complexity of the relationship between social membership and mortality is addressed in Chapter 54.

b. Sex

Similar relationships exist for death, sex, and gender. Although the differences in mortality between sexes is to a large extent a question of gender—of "socially constructed sex" that engenders specific masculine and feminine behavior—they also introduce other factors, including biologic factors, which will be treated in Chapter 53. Part of the difference results from cultural distinctions that prevail between men and women, especially in their attitudes about the body and health, but also from the diverse conditions of life imposed on men and women because of the role and status assigned to them by society.

c. Marital Status

The fact of belonging to a group and of having a status, whatever it may be, has consequences on lifestyle that may be important for health. An example is provided by the variations in mortality according to marital status. Almost universally, married people have better chances of survival than people who are celibate, widowed, or separated. Reasons for the difference are complex (Vallin and Nizard, 1977). As for migration and social ascent, marriage selects individuals according to their health; sick or handicapped people find it more difficult to find mates. In certain cases, divorce may act in the same way, especially when the law authorizing it is very restrictive and the procedure is so difficult that a solid physical and mental state of health is necessary for bringing the process to a conclusion. This was the case in France when the law forbade divorce if the spouse was suffering from serious illnesses or mental disease (Vallin and Nizard, 1977). When divorce is freely obtained and if all separations (legal or otherwise) are taken into consideration, deaths are higher among persons who are separated, widowed, or single than among those who are married. Apart from the effect of selection, the difference can be explained at two different levels. Status affects lifestyle, and from this point view, marriage protects health by imposing certain elements of lifestyle and behavior that are favorable to health. At an equal age, tabacism, alcohol abuse, and vehicular accidents are more common among single people or persons living alone than among married people. Similarly, eating behaviors, living conditions, and resort to medical care are better among married people than those living alone. However, the change of status may constitute a shock or trauma that worsens health and decreases mortality among widowed or separated persons (Thierry, 1999; Vallin and Nizard, 1977). According to a study by Valkonen and colleagues (2001), these differences in mortality between types of marital status have tended to increase in European countries.

d. Urbanization

Urbanization may have an influence on health because the city at a certain stage of development may be dangerous (e.g., unhealthiness of accommodation, human concentration encouraging the spread of epidemics, social disintegration encouraging the development of violence) or because, at another stage of development, it is favorable to health activities (e.g., human concentration favoring the advantages of health services close at hand, spread of knowledge about health matters). Urbanization also may be influential because the town or city is a modernizing factor of behavior. In Africa, sanitary progress has come first to the towns. All the differential mortality studies recognize this fact, regarding it as an obvious truism on which it is unnecessary to dwell (Tabutin and Akoto, 1989, 1992; Cantrelle *et al*., 1986; Ewbank *et al*., 1986). As confirmed by the authors of a case study published by the United Nations (Cantrelle *et al*., 1986), of all the demographic, socioeconomic, cultural, and environmental factors, living in an urban environment contributes most to the reduction in mortality. In the town itself, the mortality is not the same for those who have been living there for a long time as for the new arrivals. At Pikine in Senegal, the mortality rate for children born in the agglomeration of Dakar is two times below that for children who arrived after their birth (Antoine and Dough, 1988).

This urban advantage can only partly be explained by the great socioeconomic differences that distinguish the town from the rural world. When argument is founded on equal conditions of income, education, or social or professional category, the degree of advantage often declines but almost never disappears. It is stressed that the city usually encompasses the essential medical resources and that the urban environment facilitates health activities. This is undoubtedly the main source of difference in the 19th century, which was immersed in a medicosanitary context unfavorable to towns.

> But beyond these well known characteristics, one should inquire into the part played by the town, more particularly in Africa, in mixing together populations of various origins, pouring them into a common mould and involving them in more modern ideas, especially through education. Not only is the latter more developed in the town, but it also finds a greater justification there through the economic activities of an urban type less indebted to traditional practices than agriculture. Faced by the very considerable differences in African ethnic cultures, the town offers an alternative that is more open to effective medical and sanitary practices, and doubtlessly constitutes, in Africa more than elsewhere, the veritable crucible for modernization[19] (Vallin, 1989).

[19] "Mais, au-delà de ces caractéristiques bien connues, ne doit-on pas s'interroger sur le rôle que la ville joue plus particulièrement en Afrique en brassant des populations d'origines diverses et en les coulant dans un moule commun empruntant davantage aux idées modernes, notamment par le biais de l'instruction. Non seulement cette dernière est plus développée en ville, mais elle y trouve davantage sa raison d'être dans les activités économiques de type urbain moins redevables des pratiques traditionnelles que l'agriculture. Face aux différences culturelles ethniques, très fortes en Afrique, la ville offre une culture alternative davantage ouverte aux pratiques médicales et sanitaires efficaces et constitue sans doute, en Afrique plus qu'ailleurs, le véritable creuset de la modernisation."

d. Religion

Religion is a social conditioning factor that may affect health and mortality. For example, religious ideas affect the consumption of alcohol and the consumption of tobacco products. Within the same population, membership in different religious groups may lead to differences in mortality (Räsänen *et al*., 1996; Akoto, 1990). Moreover, many wars, massacres, and persecutions have been and still are motivated by religious fanaticism.

4. Health Policies

With equal behavior, mortality depends on the current and past health policies in the broadest sense (Vallin *et al*., 1985a, 1985b), and behavior may vary according to the health policies in effect. Some chapters (Chapters 106 to 109) are devoted to health and social policies and their effects on mortality. We limit this discussion to examples of how policies may influence behavior.

a. Availability of Means

Healthy behaviors depend on the means available for their expression. Health policies can improve the provision of health services, whether through treatment or prevention, by maximizing proximity, quality, and effectiveness of health care (Abbasi, 1999).

This is equally true for the development of collective resources that do not have health as its sole object (McKeown and Lowe, 1966). One example is the provision of water. With equal attitudes and propensities, the most elementary rules of hygiene are followed according to the facilities for access to water. Uses include drinking, food preparation, bathing, washing laundry, and cleaning the premises. The construction of networks for drinking water, extended by the systematic installation of running water in houses, has contributed greatly to the progress of hygiene, even though it continues to vary according to individual behavior. Similarly, since antiquity, the installation of sewage networks and of water treatment has increasingly transferred individual to collective responsibility for the protection of the environment through the treatment of domestic waste and eradication of the causes of unhealthy housing.

In an even wider sense, the economic and social development policies that contribute to a diversification in the supply of goods and services tend to widen the choice of consumers and therefore the possibility of expressing behavior and attitudes (favorable or not) about health. One of the most significant examples is that concerning the diversification of available food products. The choice of products now available in developed countries bears no relation to what was formerly available. Because of this abundance, the diet has developed, for better or for worse, according to individual behavior. Globally, it may be assumed that this diversification has moved in the direction of a more balanced and healthier diet. In parallel with this progress, there is concern, sometimes extreme, about food safety by the public and by public health authorities.

b. Orienting Behavior

Policies may also aim at orienting behavior, whether indirectly by acting on the prices of certain products or directly by developing information and awareness campaigns. Governments may go so far as adopting strict regulations on the use of certain products or of prohibiting them.

Two examples can illustrate the effects of price. In France, in 1957, the Pinay-Rueff plan for *economic and financial stabilization*, which caused a sharp rise in prices and a reduction in salary adjustments, led to a decrease in household consumption, a unique phenomenon in France during the whole second half of the 20th century. Paradoxically, this effect found expression in a sudden increase in life expectancy in 1958, which gained more than one year of life expectancy in a single calendar year, compared with an average of less than 6 months gained between 1947 to 1957 and even less afterward. Analysis of the causes of death showed that this exceptional increase was caused entirely by the decrease in the consumption of table wine after the sharp rise in the price of a liter of red wine (i.e., increased in Paris from 92 centimes to 1.4 francs) (Meslé and Vallin, 1993).

Another striking example is provided by the influence of the price policy carried out by the Organization of the Petroleum Exporting Countries (OPEC) on the behavior of motorists. In most European countries, the ongoing tendency for a rise in driving-related deaths was reversed in 1974, which coincided remarkably with the explosion in oil prices. Higher prices engendered an interest in moderating speed to save money, and the slower speeds reduced the number of fatal accidents (Vallin and Meslé, 2001).

Conversely, the price policy for medical products may affect the resort to preventive treatment. In countries having social security systems, it is more often the collective responsibility for medical expenses that limits the price (or even abolishes it) for the consumer and encourages access to treatment and prevention. Development of these institutions since the Second World War has caused a great expansion of health expenditures, which probably has contributed to the increase in life expectancy. Social mortality studies have nevertheless shown that this policy has not reached its primary target of reducing inequality in respect to death, because social differences in mortality continue to increase. Mortality has fallen for all categories, but it is, *grosso modo*, the most favored categories of people who have profited most. Although reduction in the individual's health costs has enabled better expression of favorable behavior, behavioral differences remain important, and their effects may increase precisely because resources are more available.

Health policies may also act directly on behavior through health education campaigns that explicitly encourage certain changes in behavior. Education may concern long-term work, such as the development of health instruction programs in schools or within the area of adult training. Health education may also act through campaigns that are more directly targeted at specific public health problems, such as vaccination, antismoking, and road safety campaigns. Many studies have shown the important effects of antismoking campaigns on deaths from lung cancer (The Surgeon General, 1989, 1990; Blanc, 1985a, 1985b) or from diseases of the coronary arteries (Puska *et al*., 1983).

Policies may go so far as attempting to restrain behavior through regulation or even prohibition. One of the most important areas for health protection regulation is that concerning the production and use of medications. Each product undergoes a very demanding series of tests and quality controls before being authorized for sale, but the use itself is more or less strictly regulated according to the nature of the product. This is one of the current means of curbing medical overconsumption, which is dangerous. It is also one of the means for combating suicidal behavior, and for the prevention of malicious acts.

Another example of regulation aimed explicitly at protecting health and life is that of road safety, including regulations determining highway codes, drivers' licenses, speed limits, and wearing of helmets or seat belts. Many measures aim at controlling the behavior of those using a dangerous means of transport. Similarly, regulations on the use of tobacco are becoming more and more restrictive, with more and more public places prohibited to smokers.

In certain cases, the regulations go as far as a total ban of a product and of its production, marketing, and use. In most countries, this is the case with certain narcotics, even though the health advantage of such extreme measures in restricting behavior is the object of controversy. Experience shows that governmental policies are never enough for eradicating the needs or curbing consumption, and in raising prices, they encourage illegal production and trafficking. Despite the problems with regulation, it is tempting to consider that deaths from lung cancer would never have reached the same high proportions if tobacco had been treated as a hard drug since the late 19th century.

All measures for alcohol prohibition have failed. The famous attempt at prohibition in the 1920s in the United States, which encouraged fraud and crime more than sobriety, was a fiasco. The immediate effect of the campaign carried out by Mikhael Gorbachev in the last hours of the Soviet Union was to cause an appreciable rise in life expectancy,[20] but prohibition measures were too easily avoided (e.g., fraud, resort to nonalimentary alcohols), and the attempt was quickly abandoned, after which life expectancy fell to even lower levels than before the campaign[21] (Shkolnikov *et al*., 1995, 1996).

CONCLUSIONS

Although health and survival depend primarily on the balance between genetically programmed vital resources and the resources and dangers arising from the environment, they also depend on the way in which men and women control this balance, individually and collectively. Individuals and society in their relationship with the environment have a large measure of freedom, which may lead to great differences between the state of health and the probabilities of survival.

Acknowledgment

This chapter was translated from the French version by Ronald Wilkens.

References

Aaby Peter, 1989. La promiscuité, un facteur déterminant de la mortalité par rougeole en Afrique, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 295–324. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

Aaby Peter, 1992. Overcrowding and intensive exposure: major determinants of variations in measles mortality in Africa, *in*:

[20] In Russia, at 61.7 years in 1984, male life expectancy suddenly increased by more than 3 years during 1985, reaching 64.8 in 1986, whereas since 1965, it has slowly but regularly been decreasing (losing 2.5 years in 20 years).

[21] In 1992, Russian life expectancy had already fallen to its 1985 level (62 years), and with the economic and social crisis, it had fallen to 57.7 years in 1994.

Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 319–347. Oxford, UK, Clarendon Press, 450 p.

Aballea François, 1982. Besoin de santé et classes sociales, *Recherche Sociale*, January-March (81), p. 59–80.

Abbasi Karman, 1999. The World Bank health care strategy, *British Medical Journal*, vol. 318, p. 933–936.

Abou-Zahr Carla and Royston Erica, 1991. *Maternal mortality, a global factbook*. Geneva, World Health Organization/Organisation Mondiale de la Santé, 598 p.

Adam Philippe and Herzlich Claudine, 1994. *Sociologie de la maladie et de la médecine*. Paris, Nathan, 128 p.

Agora and Cremms (eds.), 1994. Idéologies de la prévention, *Agora: ethique, médecine et société*, no. 30, p. 3–102.

Aaiach Pierre, Bon Norbert and Deschamps Jean-Pierre (eds.), 1992. *Comportements et santé. Questions pour la prévention*. Nancy, Presses Universitaires de Nancy, 247 p.

Aiach Pierre and Cebe Dominique, 1983. La perception dans les symptômes: variations dans les déclarations selon différentes procédures, *Cahiers de sociologie et de démographie médicales*, vol. 1, p. 3–27.

Aiach Pierre, Leclerc Annette, and Philippe Alain, 1980. Symptomatologie et recours au médecin: proposition pour une approche de la consommation médicale, *Cahiers de sociologie et de démographie médicales*, vol. 20(4), p. 173–181.

Akoto Eliwo, 1990. Christianisme et inégalités en matière de mortalité des enfants en Afrique noire, *Population*, vol. 45(6), p. 971–992.

Amiel-Lebigre Françoise and Gognalons-Nicolet Maryvonne, 1993. *Entre santé et maladie*. Paris, Presses universitaires de France, 176 p.

Anderson Roy M., 1988. The epidemiology of HIV infection: variable incubation plus infection periods and heterogeneity in sexual activity, *Journal of the Royal Statistical Society A*, vol. 150(1), p. 66–83.

Anderson Roy M., Gupta Sunit, and Ng Weshemi Japeth, 1990. The significance of sexual partner contact networks for the transmission dynamics of HIV, *Journal of Acquired Immune Deficiency Syndromes*, vol. 3(4), p. 417–429.

Anonymous, 1999. AIDS leaving infants without mothers and homes, *AIDS Weekly Plus*, June-July (28), p. 16.

Antoine Philippe and Diouf Pap Dempa, 1988. Urbanisation, scolarisation et mortalité des enfants, *Les annales de l'IFORD*, vol. 1, p. 9–24. (Paper presented at the International Union for the Scientific Study of Population seminar on Mortality and socIety in Africa South of the Sahara, Yaoundé, October 19–23, 1987.)

Aries Philippe, 1971. *Histoire des populations françaises et de leurs attitudes devant la vie depuis le XVIII^e siècle*. Paris, Seuil, 414 p.

Aries Philippe, 1977. *L'homme devant la mort*. Paris, Seuil, 642 p.

Ariffin Rohana, Selvaraju S., Samuel Rachel, and Lourdes C., 1997. *Shame, secrecy and silence: study on rape in Penang*. Pulau Pinang, Malaysia, Women's Crisis Centre, 264 p.

Artaud-Wild Sabine M., Connor Sonja L., Sexton Gary, and Conner William E., 1993. Difference in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland, *Circulation*, vol. 88(5), 2771–2779.

Augé Marc and Herzlich Claudine (eds.), 1983. *Le sens du mal. Anthropologie, histoire, sociologie de la médecine*. Paris, Editions des Archives Contemporaines, 278 p.

Austin Donald F., 1982. Cancer of the larynx, *in*: David Schottenfeld and Joseph F. Fraumeni Jr (eds.), *Cancer epidemiology and prevention*, p. 554–563. Philadelphia, WB Saunders, 1173 p.

Avlund Kirsten, Damsgaard Mogens Trab, and Holstein Bjorn E., 1998. Social relations and mortality. An eleven year follow-up study of 70-year-old men and women in Denmark, *Social Science and Medicine*, vol. 47(5), p. 635–643.

Balge Marci Z. and Krieger Gary R., 2000. *Occupational health & safety*, 3rd ed. Washington, National Safety Council Press, 595 p.

Barrow Susan M., Herman Daniel B., Cordova Pilar, and Struening Elmer L., 1999. Mortality among homeless shelter residents in New York City, *American Journal of Public Health*, vol. 89, p. 529–534.

Baudelot Christian and Establet Roger, 1984. *Durkheim et le suicide*. Paris, Presses universitaires de France, 125 p.

Baudrillart Jean, 1976. *La société de consommation, ses mythes, ses structures*. Paris, Gallimard, 318 p.

Becroft David M., Thompson John M., and Mitchell Edwin A., 2001. Nasal and intrapulmonary haemorrhage in sudden infant death syndrome, *Archives of Disabled Children*, vol. 85(2), p. 116–120.

Behm Hugo, 1979. Socio-economic determinants of infant mortality in Latin America, *in*: World Health Organization (ed.), *Délibération de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité*, p. 139–165. Geneva, New York, WHO, 622 p.

Behm Hugo and Primante Domingo, 1978. Mortalidad en los primeros anos de vida en paises de la America Latina, *Notas de Poblacion*, vol. 6(16), p. 23–44.

Bell David, 1985a. Quelles politiques pour réduire rapidement la mortalité dans les pays les moins avancés? *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 475–478. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Bell David, 1985b. What policies will reduce death rates most rapidly in less developed countries? *in*: Jacques Vallin, Alan Lopez and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 493–506. Liège, Ordina édition, 558 p.

Bender Raff, Trautner Christoph, Spraul Maximilian, and Berger Michael, 1998. Assessment of excess mortality in obesity, *American Journal of Epidemiology*, vol. 147(1), p. 42–47.

Benjamin Bernard, 1982. Smoking and mortality, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 433–447. Liège, Ordina éditions, 474 p.

Berke Joseph and Hernton Calvin, 1974. *The cannabis experience: an interpretative study of the effects of marijuana and hashish*. London, Owen, 288 p.

Bernard Michel, 1973. *Le corps*. Paris, Éditions Universitaires, 141 p.

Besnard Philippe, 1997. Mariage et suicide: la théorie durkheimienne de la régulation conjugale à l'épreuve d'un siècle, *Revue française de sociologie*, vol. 38(4), p. 735–758.

Beverly Winikoff, 1982. Weaning: nutrition, morbidity, and mortality consequences, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 113–150. Liège, Ordina éditions, 474 p.

Bish Alison, Golombok Susan, Hallstrom Cosmo, and Fawcett Sue, 1996. The role of coping strategies in protecting individuals against long-term tranquillizer use, *British Journal of Medicine and Psychology*, vol. 69 (Part 2), p. 101–115.

Blanc Michel, 1985a. Les effets à long terme des programmes d'intervention contre le tabagisme: application à la France, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 238–256. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Blanc Michel, 1985b. The long-term effect of anti-smoking programmes, *in*: Jacques Vallin, Alan Lopez and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 241–260. Liège, Ordina édition, 558 p.

Bledsoe Caroline and Brandon Anastasia, 1989. Le placement des enfants et son influence sur la mortalité, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala-Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 271–294. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

Blum Harold F., 1959. *Carcinogenesis by ultraviolet light*. Princeton, NJ, Princeton University Press, 350 p.

Bocognano Agnès, Grandfils Nathalie, le Fur Philippe, Mizrahi Andrée, and Mizrahi Arié, 1994. *Santé, soins et protection sociale en 1993*. Paris, Centre de recherche, d'etude et de documentation en economie de la sante, 180 p.

Boltanski Luc, 1971. Les usages sociaux du corps, *Annales Economie Sociétés Civilisations*, January-February (1), p. 205–233.

Bortolotti Flavia, Cadrobbi Paolo, Carretta M., Meneghetti Francesco, De Rossi Anita, and Chieco-Bianchi Luigi, 1986. HTLV-III infection in drug abuse with acute viral hepatitis, *in*: *Proceedings of the Second International Conference on AIDS, Paris, 1986*, p. 119. Amsterdam, Elsevier, 144 p.

Bouchard Jacques (ed.), 1982. *Corps et société*. Paris, Éditions Érès, 185 p. (*Actions et recherches sociales*, special issue, March.)

Brouard Nicolas, 1987. Sida: durée d'incubation, taux de croissance, taux de reproduction nette, *Population*, vol. 42(6), p. 797–818.

Brouard Nicolas, 1994. Aspects démographiques et conséquences de l'épidémie de sida en Afrique, *in*: Jacques Vallin (ed.), *Populations africaines et sida*, p. 120–178. Paris, La Découverte, Centre Français sur la Population et le Développement, 224 p.

Buccianti Cinzia, 1993. Senza destino. L'abbandono dei bambini a Siena nei primi decenni dell'Ottocento, *Rivista di Studi Napoleonici*, vol. 30(1), nuova serie, p. 9–83.

Bungener Martine, Horelou-Lafarge Chantal, and Louis Marie-Victoire, 1982. *Chômage et santé*. Paris, Economia, 158 p.

Burgess Ann W., 1978. *Sexual assault of children and adolescents*. Lexington, MA, Heath, 245 p.

Burridge Roger and Ormandy David, 1993. *Unhealthy housing: research, remedies, and reform*. London, E & FN Spon, 443 p.

Caldwell John C., 1979. Education as a factor in mortality decline: an examination of Nigerian data, *Population Studies*, vol. 33(3), p. 395–415.

Caldwell John C. and McDonald Peter, 1981. Influence of maternal education on infant and child mortality: levels and causes, *in*: International Union for the Scientific Study of Population (ed.), *International population conference, Manilla, 1981*, vol. 2, p. 79–96. Liège, International Union for the Scientific Study of Population, 556 p.

Calot Gérard and Caselli Graziella, 1988. *La mortalité en Chine d'après le recensement de 1982*. I. *Analyse selon le sexe et l'âge au niveau national et provincial*. II. *Tables de mortalité par provinces*. Paris, Institut National d'Etudes Demographiques, International Union for the Scientific Study of Population, 72 + 112 p. (Dossiers et recherches no. 16.)

Calot Gérard and Caselli Graziella, 1989. La mortalité en Chine d'après le recensement de 1982. Analyse selon le sexe et l'âge au niveau national et provincial, *Population*, vol. 44(4–5), p. 841–872.

Cantrelle Pierre, Diop Ibrahima L., Garenne Michel, Gueye M., and Sadio A., 1986. Le profil de la mortalité et de ses déterminants au Sénégal, 1960–1980, *in*: United Nations (ed.), *Les causes des changements et des différences de la mortalité dans les pays en développement*: *étude de cas dans cinq pays*, p. 102–136. New York, United Nations, 199 p. (Études démographiques no. 94.)

Cantrelle Pierre and Leridon Henri, 1980. Fécondité, allaitement et mortalité infantile. Différences inter-ethniques dans une même région: Saloum (Sénégal), *Population*, vol. 35(3), p. 623–647.

Caraël Michel, 1995. Sexual behaviour, *in*: John Cleland and Benoît Ferry (eds.), *Sexual behaviour and AIDS in the developing world*, p. 75–123. London, Taylor & Francis, 244 p. (Social aspects of AIDS.)

Carton Solange, Jouvent Roland, and Widlöcher Daniel, 1994. Nicotine dependence and motives for smoking in depression, *Journal of Substance Abuse*, vol. 6, p. 67–76.

Caselli Graziella, 1995. Storie di vita e di supravivenza, *in*: Società Italiana di Statistica et Università degli Studi della Calabria (ed.), *Continuità e discontinuità nei processi demografici. L'Italia nella transizione demografica*, p. 89–105. Catanzaro, Rubbettino, 672 p. (Proceedings of the SIS conference, 1995.)

Caselli Graziella, Meslé France, and Vallin Jacques, 1999. Le triomphe de la médecine, *in*: Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998*, p. 125–181. Paris, Librairie Arthème Fayard, 792 p.

Cassou Bernard, Huez Dominique, Mousel Marie-Laurence, Spitzer Catherine, and Tourachet-Hebrard Annie, 1985. *Les risques du travail*. Paris, La Découverte, 650 p.

Chapuis Robert, 1989. *L'alcool, un mode d'adaptation sociale?* Paris, L'Harmattan, 220 p.

Chen Lincoln C., Huq Embadul, and d'Souza Stan, 1981. Sex bias in the allocation of food and health care in rural Bengladesh, *Population and Development Review*, vol. 7(1), p. 55–70.

Chesnais Jean-Claude, 1976. *Les morts violentes en France*. Paris, Institut national d'etudes demographiques, presses universitaires de France, 346 p. (Travaux et documents, cahier no. 75.)

Chesnais Jean-Claude, 1985a. L'efficacité de la prévention des morts violentes, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*. *Influence des politiques sociales et des politiques de santé sur l'évolution de la mortalité*, p. 257–274. Paris, Institut national d'etudes demographiques, presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Chesnais Jean-Claude, 1985b. The prevention of death from violence: lessons from experience, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 261–280. Liège, Ordina édition, 558 p.

Chesnais Jean-Claude and Vallin Jacques, 1977. évolution récente de la mortalité et de la morbidité dues aux accidents de la route, *Population*, vol. 32(6), p. 1239–1265.

Clark Stewart C., 1990. *Drugs and sports*. Alberta, Canada, Alberta Alcohol and Drug Abuse Commission, 17 p.

Cohen Caroline, 1990. *La dépendance tabagique et les phénomènes comportementaux. Le tabac est-il une drogue?* Lyon, Lyon Méditerranée médical. Médecine du sud-est. (Paper presented at the XVII^{e} Rencontres médicales du sud-est.)

Corsini Carlo A., 1991. Era piovuto dal cielo e la terra l'aveva accolto: il destino del trovatello, *in*: *Enfance abandonnée et société en Europe, XIV^{e}–XX^{e} siècle*, p. 81–119. Rome, école française de Rome, 1236 p. (Collection de l'école francaise de Rome, no. 150.)

Costa Giuseppe and Segnan Nereo, 1988. Mortalità e condizione professionale nello Studio Longitudinale Torinese, *Epidemiologia e Prevenzione*, vol. 10, p. 47–57.

Courbage Youssef and Khlat Myriam, 1996. Mortality and causes of death of Moroccans in France, 1979–91, *Population: An English Selection*, vol. 8, p. 59–94.

Couturier Alain J., 2000. *Occupational and environmental infectious diseases: epidemiology, prevention, and clinical management*. Beverly Farms, MA, OEM Press, 747 p.

Criqui Michael H. and Ringel Brenda, 1994. Does diet or alcohol explain the French paradox? *The Lancet*, vol. 344(8939–8950), p. 1719–1723.

Cutler Richard G., 1984. Antioxidants, aging, and longevity, *in*: William A. Pryor (ed.), *Radicals in biology*, p. 371–428. Orlando, Academic Press, 437 p.

Daniels Dianne, 2000. *Exercises for osteoporosis: over 100 exercises to improve strength, balance, and flexibility*. New York, Hatherleigh Press, 131 p.

Das Gupta Monica, 1990. Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India, *Population Studies*, vol. 44(3), p. 479–505.

Davidson Stanley, Passmore Reginald, Brock J. F., and Truswell A. Stewart, 1979. *Human nutrition and dietetics*, 7th ed. Edinburgh, Churchill-Livingstone, 641 p.

Davis Kingsley, 1973. Cities and mortality, *in*: International Union for the Scientific Study of Population (ed.), *International population conference, Liège, 1973*, vol. 3, p. 259–282. Liège, Ordina éditions, 470 p.

Davis Letitia K. and Friedman Daniel, 1989. *Dying for the job: traumatic occupational deaths in Massachusetts*. Boston, Bureau of Health Statistics, Research and Evaluation, Massachusetts Department of Public Health, 29 p.

Day N. E. and Munoz Nubia, 1982. Oesophagus, *in*: David Schottenfeld and Joseph F Jr Fraumeni (eds.), *Cancer epidemiology and prevention*, p. 596–623. Philadelphia, WB Saunders, 1173 p.

De Rossi Anita, Bortolotti Flavia, Cadrobbi Paolo, and Chieco-Bianchi Luigi, 1988. Trends of HTLV-I and HIV infection in drug addicts, *European Journal of Cancer and Clinical Oncology*, vol. 24(2), p. 279–280.

Debry Gérard, 1993. *Le café et la santé*. Montrouge, John Libbey Eurotext, 560 p.

Debry Gérard, 1994. *Coffee and health*. Montrouge, John Libbey Eurotext, 538 p.

Dejours Christophe, 1980. *Travail: usure mentale. Essai de psychologie du travail*. Paris, Le Centurion, 153 p.

Doll Richard and Hill Bradford A., 1950. Smoking and carcinoma of the lung, *British Medical Journal*, vol. 2, p. 739–747.

Doll Richard and Hill Bradford A., 1952. Study of aetiology of cancer of the lung, *British Medical Journal*, vol. 2, p. 1271–1286.

Doll Richard and Hill Bradford A., 1954. The mortality of doctors in relation with their smoking habits, *British Medical Journal*, vol. 1, p. 1450–1455.

Doll Richard and Peto Richard, 1981. *The causes of cancer*, 2 vol. New York, Oxford University Press, 1197 + 1312 p.

Dubertret Louis, Santus René, and Morlière Patrice, 1995. *Ozone, sun, cancer: molecular and cellular mechanisms, prevention*. Paris, éditions INSERM, 223 p.

Dublin Louis Israel, 1917. *Causes of death by occupation: occupational mortality experience of the Metropolitan Life Insurance Company, Industrial Department, 1911–1913*. Washington, DC, Government Printing Office, 88 p. (Industrial Accidents and Hygiene series, no. 11; Bulletin of the United States Bureau of Labor Statistics, no. 207.)

Dubreuilh William, 1907. épithéliomatose d'origine solaire, *Annales de dermatologie et de syphiligraphie*, vol. 8, p. 387–416.

Ducimetière Pierre and Richard Jacques-Lucien, 1992. Dietary lipids and coronary disease: is there a French paradox? *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 2, p. 195–201.

Duclos Denis, 1984. *La santé et le travail*. Paris, La Découverte, 124 p.

Durkheim Émile, 1897. *Le suicide*, 7th ed. Paris, Presses universitaires de France, 463 p.

Dutot Fabrice and Lambrichs Louise L., 1988. *Les fractures de l'âme. Du bon usage de la maladie*. Paris, Robert Laffont, 365 p.

Ellinwood Everett H. and Kilbey M. Marlyne, 1977. Cocaine and other stimulants, *in*: Duke University Medical Center (ed.), *Conference on contemporary issues in stimulant research (1975)*, p. 545–570. *Advances in behavioral biology*, vol. 21. New York, Plenum Press, 721 p.

Ewbank Douglas, Henin Roushdi, and Kekovole John, 1986. Intégration de la recherche démographique et épidémiologique sur la mortalité au Kenya, *in*: United Nations (ed.), *Les causes des changements et des différences de la mortalité dans les pays en développement: étude de cas dans cinq pays*, p. 39–101. études démographiques no. 94. New York, United Nations, 199 p.

Fassin Didier, 1995. Un monde inégal. La globalisation des disparités sociales devant la santé, *Prévenir*, vol. 28, p. 83–92.

Flick Louise, White Deborah K., Vemulapalli Caroline, Stulac Barbara B., and Kemp James S., 2001. Sleep position and the use of soft bedding during bed sharing among African American infants at increased risk for sudden infant death syndrome, *Journal of Pediatrics*, vol. 138(3), p. 338–343.

Foucault Michel, 1972. Maladie mentale et culture occidentale, *in*: François Steudler (ed.), *Sociologie médicale*, p. 91–104. Paris, Armand Colin, 104 p.

Garattini Silvio, 1993. *Caffeine, coffee, and health*. Monographs of the Mario Negri Institute for Pharmacological Research. New York, Raven Press, 420 p.

Garfinkel Laurence, 1979. Cardiovascular mortality and cigarette smoking, *in*: Lars M. Ramström (ed.), *The smoking epidemic, a matter of worldwide concern*, p. 41–44. Fourth World Conference on Smoking and Health. Stockholm, Almqvist & Wiskell International, 341 p.

Gbenyon Kuakuvi and Locoh Thérèse, 1989. Différences de mortalité entre garçons et filles, *in*: Gilles Pison, Étienne van de Walle, and Mpembélé Sala-Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 221–244. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

Gbenyon Kuakuvi and Locoh Thérèse, 1992. Mortality différences in childhood by sex in sub-Saharan Africa, *in*: Étienne van de Walle, Gilles Pison, and Mpembélé Sala-Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 230–252. International Studies in Demography. Oxford, Clarendon Press, 450 p.

Gélot Didier, 1992. Impact de l'emploi et du chômage sur la santé, *Revue Française des Affaires Sociales*, vol. 1, p. 99–112.

Gessner Bradford D., Ives George C., and Perham-Hester Katherine A., 2001. Association between sudden infant death syndrome and prone sleep position, bed sharing, and sleeping outside an infant crib in Alaska, *Pediatrics*, vol. 108(4), p. 923–927.

Gichuhi Agnès W., 1993. *The impact of individual characteristics, community resources and maternal preventive health behavior on infant mortality in rural Kenya*. Ann Arbor, MI, University Microfilms International, 147 p.

Goldberger Joseph, Wheeler George A., and Sydenstricker Edgar, 1920. A study of the relation of diet to pellagra incidence in seven textile-mill communities of South Carolina in 1916, *Public Health Reports*, vol. 35, p. 647–713.

Goldsmith John R., 1982. Mortality and working conditions in industry, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 173–198. Liège, Ordina éditions, 474 p.

Got Claude, Faverjon Gérard, and Thomas Christian, 1984. Alcool et accidents mortels de la circulation, *Bulletin D'information du HCEIA*, vol. 29(1–2), p. 38–60.

Greenwood Major, 1924. The vital statistics of Sweden and England and Wales: an essay in international comparison, *Journal of the Royal Statistics Society*, vol. 87, p. 493–543.

Greenwood Major, 1926. Oral presentation to the Health Committtee of the League of Nations, cited by John Bowes McDougall, 1949. *Tuberculosis: a global study in social pathology*. Edinburgh, E&S Livingstone, 455 p.

Guignon Nicole, 1994. Les consommations d'alcool, de tabac et de psychotropes en France en 1991–1992, *Solidarité Santé*, vol. 1, p. 171–185.

Haines Michael R., 1991. Conditions of work and the decline of mortality, *in*: Roger Schofield, David Reher, and Alain Bideau (eds.), *The decline of mortality in Europe*, p. 177–195. Oxford, New York, Clarendon Press, 270 p.

Harvey John A. and Kosofsky Barry E., 1998. Cocaine effects on the developing brain. New York, New York Academy of Sciences, 436 p. (*Annals of the New York Academy of Sciences*, vol. 846.)

Hatton Françoise, Flament Robert, and Bouvier-Colle Marie-Hélène, 1985a. La lutte contre les maladies cancéreuses, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 217–238. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Hatton Françoise, Flament Robert, and Bouvier-Colle Marie-Hélène, and Maujol Léone, 1985b. The fight against cancer, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 217–250. Liège, Ordina édition, 558 p.

Hegsted D. Mark and Ausman Lynne M., 1988. Diet alcohol and coronary heart diseases in men, *Journal of Nutrition*, vol. 118, p. 1184–1189.

Helsing Knud J. and Comstock George W., 1987. Heart disease mortality in persons living with smokers, *in*: B. Siefert, H. Edson, M. Fisher, H. Ruden, and J. Wegner (eds.), *Indoor air 87. Proceedings of the fourth international conference on indoor air quality and climate, Berlin, 17–21 August 1987*, 3 vol., p. 29–33. Berlin, Institute for Water, Soil, and Air Hygiene.

Héran-Le Roy Odile, 1999. Les risques professionnels pour la santé, *in*: National Institute of Statistics and Economic Studies (ed.), *Données sociales. La société française. 1999*, p. 264–270. Paris, National Institute of Statistics and Economic Studies, 506 p.

Herzlich Claudine (ed.), 1970. *Médecine, maladie et société*. Paris, La Haye, Mouton, 318 p.

Herzlich Claudine, 1992. *Santé et maladie, analyse d'une représentation sociale*. Paris, Éditions de l'EHESS, 210 p.

Hill Allan and Randall Sara, 1984. Différences géographiques et sociales dans la mortalité infantile et juvénile au Mali, *Population*, vol. 39(6), p. 921–945.

Hill Catherine, 1988. Tabac et santé, *in: Lutter contre le tabagisme*, p. 31–175. Official reports. Paris, La documentation française, 279 p.

Holmes George E., Bernstein Carol, and Bernstein Harris, 1992. Oxidative and other DNA damages as the basis of aging: a review, *Mutation Research*, vol. 275(3–6), p. 305–315.

Horiuchi Shiro, 1983. The long-term impact of war on mortality: old-age mortality of First World War survivors in the Federal Republic of Germany, *United Nations Population Bulletin*, vol. 15, p. 80–92.

House James S. Landis Karl R., and Umberson Debra, 1988. Social relationship and health, *Science*, vol. 241, p. 550–554.

Houtaud (d') Alphonse and Field Mark George, 1989. *La santé. Approche sociologique de ses représentations et de ses fonctions dans la société*. Nancy, Presses Universitaires de Nancy, 256 p.

Hughes T. A., Shone Geoffrey, Lindsay G., and Wiles C. M., 1994. Severe dysphagia associated with major tranquillizer treatment, *Postgraduate Medical Journal*, vol. 70(826), p. 581–583.

Illich Ivan, 1975. *Némésis médicale. L'expropriation de la santé*. Paris, Seuil, 223 p.

Inciardi James A. and Harrison Lana D., 1998. *Heroin in the age of crack-cocaine*. Drugs, Health, and Social Policy series, vol. 6. Thousand Oaks, CA, Sage Publications, 289 p.

Inciardi James A., Surratt Hilary L., and Saum Christine A., 1997. *Cocaine-exposed infants: social, legal, and public health issues*. Drugs, Health, and Social Policy series, vol. 5. Thousand Oaks, CA, Sage Publications, 116 p.

INED, 1995. *Vingt-quatrième rapport sur la situation démographique de la France*. Paris, Institut National d'Etudes Demographiques, 62 p.

Iskrant Albert P. and Joliet Paul V., 1968. *Accidents and homicide*. Cambridge, MA, Harvard University Press, 202 p.

Ji Bu-Tian, Chow Wong-Ho, Yang Gong, McLaughlin Joseph K., Zheng Wei, Shu Xiao-Ou, Jin Fan, Gao Riu-Nie, Gao Yu-Tang, and Fraumeni Joseph F. Jr., 1998. Dietary habits and stomach cancer in Shanghai, China, *International Journal of Cancer*, vol. 76(5), p. 659–664.

Kannel William B., 1967. Habitual level of physical activity and risk of coronary heart disease: the Framingham study, *Canadian Medical Association Journal*, vol. 96, p. 811–812.

Kaprio Jaakko, 1984. *The incidence of coronary heart disease in twin pair discordant for cigarette smoking*. Helsinki, University of Helsinki, Department of Public Health Science, 118 p.

Karch Steven B., 1996. *The pathology of drug abuse*, 2nd ed. Boca Raton, FL, CRC Press, 462 p.

Kashani Javad H. and Allan Wesley D., 1998. *The impact of family violence on children and adolescents*. Developmental clinical psychology and psychiatry, vol. 37. Thousand Oaks, CA, Sage Publications, 111 p.

Keys Ancel, 1970. Coronary heart diseases in seven countries, *Circulation*, vol. 41 and 42, Suppl., p. 1–211.

Keys Angel, Menotti Alessandro, Martti Karvonen J., Aravanis Christ, Blackburn Henry, Buzina Ratko, Hironori Toshima, *et al*., 1986. The diet and the 15-year death rate in the seven country study, *American Journal of Epidemiology*, vol. 124, p. 903–915.

Kinjo Yoshihide, Cui Y., Akiba S., Watanabe Shaw, Yamaguchi Naohito, Sobue T., Mizuno Shoichi, and Beral Valérie, 1998. Mortality risks of oesophageal cancer associated with hot tea, alcohol, tobacco and diet in Japan, *Journal of Epidemiology*, vol. 8(4), p. 235–243.

Klerman Lorraine V. and Johnson Michele A., 1991. *Case studies of substance abusing pregnant women, their infants and children*. Washington, DC, Institute of Medicine, 163 p.

Konings Elke and Ccraël Michel, 1997. Marriage patterns, sexual networks and HIV in the developing world, *in*: International Union for the Scientific Study of Population (ed), *International population conference, Beijing, 1997*, p. 501–433. Liège, International Union for the Scientific Study of Population, 472 p.

Koskenvuo Markku, Kaprio Jaakko, Kesaniemi Antero, and Poikolainen Kari, 1986. Alcohol related diseases associated with ischaemic heart disease: a three-year follow-up of the middleaged male hospital patients, *Alcohol*, vol. 21(3), p. 251–256.

Kunst Anton E., Bos Viviane, and Mackenbach Johan P., 2000. *Monitoring socio-economic inequalities in health in the European Union: guidelines and illustration*. Rotterdam, Erasmus University, 212 p.

Kunst Anton E. and Mackenbach Johan P., 1994. The size of mortality differences associated with educational level in nine industrialized countries, *American Journal of Public Health*, vol. 84, p. 932–937.

Kushi Lawrence H., Lenart Elisabeth B., and Willett Walter C., 1995. Health implications of Mediterranean diets in light of contemporary knowledge. 1. Plants foods and diary products. 2. Meats, wine, fats and oils, *American Journal of Clinical Nutrition*, vol. 61, p. 1507S–1427S.

La Vecchia Carlo, 1990. Oral contraceptives and breast cancer: update of an Italian case-control, *in*: Mann Ronald D. (ed.), *Oral contraceptives and breast cancer. The implications of the present findings for informed consent and informed choice*, p. 245–251. Carnforth, UK, Parthenon Publishing Group, 506 p.

Lachaud Jean-Claude, 1975. *Le pouvoir médical source de maladie*. Toulouse, Privat, 174 p.

Lalou Richard and Piché Victor, 1994. *Migrations et sida en Afrique de l'ouest: un état des connaissances*. Paris, Centre français sur la population et le développement, 52 p. (Les dossiers du Centre Français sur la Population et le Développement, no. 28.)

Lancaster Henri Oliver, 1990. *Expectations of life. A study of the demography, statistics, and history of world mortality*, p. 314–340. New York, Springer-Verlag, 606 p.

Laplantine François, 1986. *Anthropologie de la maladie*. Paris, éditions Payot, 411 p.

Laurel Asa Christina, 1982. Mortality and working conditions in agriculture in developing countries, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 199–220. Liège, Ordina éditions, 474 p.

Law Malcom and Wald Nicolas, 1999. Why heart disease mortality is low in France: the time lag explanation, *British Medical Journal*, vol. 318, p. 1471–1480.

Le Breton David, 1985. *Corps et sociétés: essai de sociologie et d'anthropologie du corps*. Paris, Méridiens-Klincksiek, 230 p.

Le Breton David, 1992. *La sociologie du corps*. Paris, Presses universitaires de France, 128 p.

Leclercq Jean-François, Samuel J. L., Merillon J. P., Eschwege E., Bertrand M., Gourgon R., Grosgogeat Y., Bouvrain Y., and Swynghedauw B., 1979. Enquête épidémiologique multicentrique sur les myocardiopathies primitives: à propos de 380 observations, *Archives des maladies du cœur et des vaisseaux*, vol. 72(11), p. 1302–1310.

Ledermann Sully, 1956. *Alcool, alcoolisme, alcoolisation*, vol. 1. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 314 p. (Travaux et documents, cahier no. 29.)

Ledermann Sully, 1964. *Alcool, alcoolisme, alcoolisation*, vol. 2. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 613 p. (Travaux et documents, cahier no. 41.)

Lee John A. H., 1982. Melanoma and exposure to sunlight, *Epidemiologic Reviews*, vol. 4, p. 110–136.

Lee M. and Leichter J., 1982. Alcohol and the fetus, *in*: Patrice E. F. Jelliffe and Derrik B. Jelliffe (eds.), *Adverse effects of foods*, p. 245–252. New York, Plenum, 614 p.

Lenz Widukind and Knapp Klaus, 1962. Thalidomide embryopathy, *Archives of Environmental Health*, vol. 5, p. 100–105.

Leplat Jacques and Cuny Xavier, 1974. *Les accidents du travail*. Paris, Presses universitaires de France, 125 p.

Lester David, 1993. *Patterns of suicide and homicide in America*. Commack, NY, Nova Science Publishers, 104 p.

Lester David, 1994. The prospective effect of marriage for suicide in men and women, *Italian Journal of Suicidology*, vol. 4(2), p. 83–85.

Letourmy Alain and Pierret Jeanine, 1982. *Inégalités sanitaires et inégalités sociales: approches économique et sociologique*. Paris, CEREBE (Centre de recherche sur le bien-être), 306 p.

Leung Donald Y. M. and Greaves Malcolm W., 2000. *Allergic skin disease: a multidisciplinary approach*. New York, Marcel Dekker, 532 p.

Lind James, 1753 (re-edited in 1772). *A treatise of the scurvy*. Edinburgh, Sands, Murray & Cochran, 476 p.

Little Arthur D., 1977. *Human safety and environmental aspects of major surfacants: a report to the Soap and Detergent Association*. Cambridge, MA, Arthur D. Little, 46 p.

Livi-Bacci Massimo, 1986. Fertility, nutrition, and pellagra: Italy during the vital revolution, *Journal of Interdisciplinary History*, vol. 16(3), p. 431–454.

Locoh Thérèse, 1998. Politiques, opinions et attitudes en matière d'excision en Afrique, *Population*, vol. 53(6), p. 1227–1239.

Lopez Alan D., 1993. On the assessment of smoking-attributable mortality, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference, Montreal 1993*, vol. 1, p. 449–458. Liège, Union Internationale pour l'Etude Scientifique de la Population/International Union for the Scientific Study of Population, 662 p.

Loux Françoise, 1979. *Le corps dans la société traditionnelle*. Paris, Berger-Levrault, 178 p.

Loux Françoise, 1983. *Traditions et soins d'aujourd'hui*. Paris, Inter Editions, 320 p.

Magnin Pierre and Cornu Jean-Yves, 1997. *Médecine du sport: pratiques du sport et accompagnements médicaux*. Paris, Ellipses, 816 p.

Mahboubi E. and Sayed G. M., 1982. Oral cavity and pharynx, *in*: David Schottenfeld and Joseph F. Fraumeni Jr (eds.), *Cancer epidemiology and prevention*, p. 583–595. Philadelphia, WB Saunders, 1173 p.

Malet Émile (ed.), 1993. *Santé publique et libertés individuelles*. Paris, Passages, 236 p.

Malizia Enrico, Borgo Stefania, and Smeriglio Massimo, 1993. *Estasi: aspetti tossicologici, clinici e sociali dell'ecstasy, o metilendiossimetamfetamina*. Rome, CIC edizioni internazionali.

Mann Jonathan, Tarandola David J. M., and Netter Thomas W., 1992. *AIDS in the world. A global report, Harvard*. Cambridge, UK, Cambridge University Press, 1038 p.

Marmot Michael G., 1995. Social differentials in mortality: the Whitehall studies, *in*: Alan Lopez, Graziella Caselli, and Tapani Valkonen (eds.), *Adult mortality in developed countries. From description to explanation*, p. 243–260. International Studies in Demography. Oxford, Clarendon Press, 362 p.

Martin Rolland A., 1975. *Occupational disability: causes, prediction, prevention*. Springfield, IL, Thomas, 206 p.

Masironi Robert and Rothwell Keith, 1988. Tendances et effets du tabagisme dans le monde, *Rapport de Statistiques Sanitaires Mondiales*, vol. 41, p. 228–241.

McBean Lois D. and Speckmann E. W., 1982. Diet, nutrition and cancer, *in*: Patrice E. F. Jelliffe and Derrik B. Jelliffe (eds.), *Adversed effects of foods*, p. 501–528. New York, Plenum, 614 p.

McDougall John Bowes, 1949. *Tuberculosis: a global study in social pathology*. Edinburgh, E&S Livingstone, 455 p.

McKeown Thomas and Lowe Charles R., 1966. *An introduction to social medicine*. Oxford, UK, Blackwell Scientific Publications, 356 p.

McNeil William, 1980. Migration patterns and infection in traditional societies, *in*: Neville F. Stanley and Richard A. Joske (eds.), *Changing disease patterns and human behaviour*, p. 27–36. London, Academic Press, 666 p.

Measham Anthony R., Obaidullah M., Rosenberg Michael J., Rochat Roger W., Khan Atiqur R., and Jabeen Suraiya, 1981. Complications from induced abortion in Bangladesh related to types of practitioner and methods, and impact of mortality, *The Lancet*, vol. 1, p. 199–202.

Menahem Georges, 1994. *Problèmes de l'enfance, statut social et santé des adultes*. Paris, Centre de Recherhce, d'Etude et de Documentation en Economie de la Sante, 221 p.

Mehta Meenakshi N., 1982. Physical abuse of abandoned children in India, *Child Abuse and Neglect*, vol. 6(2), p. 171–175.

Meslé France, 1983. Cancer et alimentation: le cas des cancers de l'intestin et du rectum, *Population*, vol. 38(4–5), p. 733–762.

Meslé France, 1991. La mortalité dans les pays d'Europe de l'Est, *Population*, vol. 46(3), p. 599–650.

Meslé France, Shkolnikov Vladimir, and Vallin Jacques, 1994. Brusque montée des morts violentes en Russie, *Population*, vol. 49(3), p. 780–790.

Meslé France and Vallin Jacques, 1992. *Évolution de la mortalité par cancer et par maladies cardio-vasculaires en Europe depuis 1950*. Paris, Institut National d'Etudes Demographiques, 46 p. (Dossiers et recherches no. 34.)

Meslé France and Vallin Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference. Montreal 1993*, vol. 2, p. 365–382. Liège, International Union for the Scientific Study of Population, 493 p.

Meyler Leopold and Herxheimer Andrew (eds.), 1972. *Side effects of drugs: a survey of unwanted effects of drugs reported in 1968–1971*, vol 7. Amsterdam, Excerpta Medica, 758 p.

Milerad Joseph, Vege Åshild, Opdal Siri H., and Rognum Torleiv O., 1998. Objective measurement of nicotine exposure in victims of sudden infant death syndrome and other unexpected child death, *Journal of Pediatrics*, vol. 133, p. 232–236.

Modan Baruch and Lubin Flora, 1980. Epidemiology of colon cancer: fiber, fats, fallacies and facts, *in*: Gene A. Spiller and Ruth M. Kay (eds.), *Medical aspects of dietary fibre*, p. 119–135. New York, Plenum, 299 p.

Molinié Anne-Marie and Volkoff Serge, 1984. Les accidents du travail, *in*: National Institute of Statistics and Economic Studies (ed.), *Données sociales. La société française. 1984*, p. 160–163. Paris, National Institute of Statistics and Economic Studies, 581 p.

Morris Jeremy N., Heady J. A., Raffle P. A. B., Roberts C. G., and Parks J. W., 1953a. Coronary heart disease and physical activity of work, I. Coronary heart disease in different occupations, *The Lancet*, vol. 2, p. 1053–1057.

Morris Jeremy N., Heady J. A., Raffle P. A. B., Roberts C. G., and Parks J. W., 1953b. Coronary heart disease and physical activity of work, II. Statement and testing of provisional hypothesis, *The Lancet*, vol. 2, p. 1111–1120.

Mosley W. Henry, 1985a. Les soins de santé primaires peuvent-ils réduire la mortalité infantile? Bilan critique de quelques programmes africains et asiatiques, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 101–136. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Mosley W. Henry, 1985b. Will primary health care reduce infant and child mortality? A critique of some current strategies, with special reference to Africa and Asia, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 103–138. Liège, Ordina édition, 558 p.

Mozumder Alam, Khuda el Barkat, Kane Thomas T., and Hill Kenneth, 1997. Determinants of infant and child mortality in rural Bangladesh, *in*: Thomas T. Kane, el Barkat Khuda, and James F. Phillips (eds.), *Reproductive health in rural Bangladesh: policy and programmatic implications*, p. 479–500, monograph no. 7. Dhaka, Bangladesh, International Centre for Diarrhoeal Disease Research, Bangladesh, 508 p.

Mueller Frans H., 1939. Tabakmissbrauch und Lungenkarzinom, *Zeitschrift für Krebsforschung*, vol. 49, p. 57–85.

Mundigo Axel, 1996. *The challenge of induced abortion research: transdisciplinary perspectives*. Liège, International Union for the Scientific Study of Population, 27 p. (Paper presented at the seminar on Socio-cultural and Political Aspects of Abortion in a Changing World, Kovalan, Trivandrum, India, March 25–28, 1996.)

Nahas Gabriel G., 1976. *Haschich, cannabis, et marijuana: le chanvre trompeur*. Paris, Presses universitaires de France, 434 p.

Nelson P. G., 1974. *Pattern of injury survey of automobile accident, Victoria, Australia. June 1971–June 1973*. Melbourne, Royal Australasian College of Surgeons.

Nizard Alfred and Munoz-Perez Francisco, 1993. Alcool, tabac et mortalité en France depuis 1950. Incidence de la consommation d'alcool et de tabac sur la mortalité, *Population*, vol. 47(4), p. 975–1014.

Norman L. G., 1962. *Road traffic accidents: epidemiology, control and prevention*. Public health paper, no. 12. Geneva, World Health Organization/Organisation mondiale de la santé, 110 p.

Nowrojee Binaifer and Manby Bronwen, 1995. *Violence against women in South Africa. The state response to domestic violence and rape*. New York, Human Rights Watch, 132 p.

Oakley Ann, 1989. Smoking in pregnancy: smoke screen or risk factor? Towards a materialist analysis, *Sociology of Health and Illness*, vol. 11, p. 311–335.

Okolski Marek, 1985a. Le cas de la Pologne, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 427–444. Paris, Institut National d'Etudes Demographiques, Presses Universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Okolski Marek, 1985b. The case of Poland, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 445–462. Liège, Ordina édition, 558 p.

OMS/WHO (Organisation mondiale de la santé/ World Health Organization), 1997. *Mutilations sexuelles féminines: déclaration commune OMS/UNICEF/FNUAP*. Geneva, WHO, 20 p.

OMS/WHO (Organisation mondiale de la santé/ World Health Organization), 1998. *Les mutilations sexuelles féminines: aperçu du problème*. Geneva, WHO, 78 p. (Aide mémoire no. 153.)

Paffenbarger Ralph S., Brand Richard J. H., Sholtz Robert I., and Jung Dexter J., 1978. Energy expanditure, cigarette smoking, and blood pressure level as related to death from specific diseases, *American Journal of Epidemiology*, vol. 108(1), p. 12–18.

Paffenbarger Ralph S., Hyde Robert T., Wing Alvin L., and Hsieh Chung-Cheng, 1986. Physical activity, all cause mortality, and longevity of College alumni, *New England Journal of Medicine*, vol. 314, p. 605–613.

Palloni Alberto and Tienda Marta, 1986. The effects of breastfeeding and place of childbearing on mortality at early ages, *Demography*, vol. 23(1), p. 31–52.

Paunier Luc, 1984. Nutrition, vitamines, *in*: Pierre E. Ferrier (ed.), *Précis de pédiatrie*. Paris, Payot, Doin, 1051 p.

Peto Richard, 1986. Influence of dose and duration of smoking on lung cancer rates, *in*: David G. Zaridze and Richard Peto (eds.), *Tobacco: a major international health hazard*, p. 23–33. IARC scientific publications no. 74. Lyon, International Agency for Research on Cancer, 324 p.

Peto Richard, Lopez Alan D., Boreham Jullian, Thun Michael, and Heath Clark, 1992. Mortality from tobacco in developed countries: indirect estimation from national vital statistics, *The Lancet*, vol. 339, p. 1268–1278.

Pinnelli Antonella and Mancini Paola, 1992. Différence de mortalité par sexe de la naissance à la puberté en France: un siècle d'évolution, *Population*, vol. 47(6), p. 1650–1676.

Pinnelli Antonella and Mancini Paola, 1999. Mortality peaks in Italy in the late 19th and early 20th centuries: trends by age and sex, *European Journal of Population*, vol. 14, p. 333–365.

Piore Michael J., 1979. *Birds of passage: migrant labour in industrial societies*. Cambridge, UK, Cambridge University Press, 229 p.

Pisa Zbynet and Uemura Kazuo, 1982. Trends of mortality from ischaemic heart disease and other cardiovascular disease in 27 countries, 1968–1977, *World Health Statistics Quarterly*, vol. 35, p. 11–47.

Pison Gilles, 1989. Les jumeaux: fréquence, statut social et mortalité, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 245–269. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

Pison Gilles, 1992. Twins in sub-Saharan Africa: frequency, social status, and mortality, *in*: Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 253–278. International Studies in Demography. Oxford, Clarendon Press, 450 p.

Platt Stephen, 1984. Unemployment and suicidal behaviour: a review of the literature, *Social Science and Medicine*, vol. 19(2), p. 93–115.

Poikolainen Kari, 1982. Alcohol use and mortality, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 417–433. Liège, Ordina éditions, 474 p.

Poswillo David E. and Alberman Eva D., 1992. *Effects of smoking on the fetus, neonate, and child*. Oxford, UK, Oxford University Press, 230 p.

Pous Jacques, Bocquet Hélène, Grosclaude Pascale, Grand Alain, and Albarède Jean-Louis, 1988. Trajectoires de vieillissement dans une population rurale, *Gérontologie et Société*, vol. 44, p. 25–34.

Power Lawrence, Bakker Doris L., and Cooper Marilyn I., 1971. *Sugar diabetes simplified*, 2nd ed. Ann Arbor, MI, 47 p.

Puska Pekka, Salonen Jukka T., Nissinen Aulikki, Tuomilehto Jaakko, Vartiainen Erkki, Korhonen Heikki, Tnskanen Antti, Ronnqvist Paivi, Koskela Kaj, and Huttunen Jussi, 1983. Change in risk factors for coronary heart disease during 10 years of a community intervention programme (North Karelia project), *British Medical Journal*, vol. 287, p. 1880–1844.

Räsänen Johanna, Kauhanen Jussi, Lakka Timo A., Kaplan George A., and Salonen Jukka T., 1996. Religious affiliation and all-cause mortality: a prospective population study in middle-aged men in eastern Finland, *International Journal of Epidemiology*, vol. 25(6), p. 1244–1249.

Ravenholt Reinert T., 1984. Addiction mortality in the United States, 1980: tobacco, alcohol, and other substances, *Population and Development Review*, vol. 10(4), p. 697–724.

Raynaut Claude, 1992. Prévention sanitaire et reproduction sociale, *in*: Pierre Aiach, Norbert Bon, and Jean-Pierre Deschamps (eds.), *Comportements et santé. Questions pour la prévention*, p. 150–144. Nancy, Presses Universitaires de Nancy, 247 p.

Renaud Serge and de Lorgeril Michel, 1992. Wine, alcohol, platelets, and the french paradox? *The Lancet*, vol. 339, p. 1523–1526.

Reynaud Michel, 1984. *Les toxicomanies*. Paris, Maloine, 509 p.

Rice Pranee L., 1999. *Living in a new community: understanding migrant's health*. Melbourne, Ausmed Publications, 265 p.

Richard Jacques-Lucien, 1987. Les facteurs de risque contemporains: le paradoxe français, *Archives des maladies du coeur*, vol. 80, Suppl., p. 17–21.

Ritson Bruce, 1977. Alcoholism and suicide, *in*: Griffith Edwards and Marcus Grant (eds.), *Alcoholism: new knowledge and new responses*, p. 271–278. London, Croom Helm, 359 p.

Robert Jean Claude, 1988. The city of wealth and death: urban mortality in Montreal 1821–1871, *in*: Wendy Mitchinson and James D. McGinnis (eds.), *Essays in the history of Canadian medicine*, p. 1–38. Toronto, McClelland & Stewart, 218 p.

Rödel Walter G., 1988. Urban mortality and the population's health: Mainz from the seventeenth to the beginning of the twentieth century, *in*: Anders Brändström and Lars Göran Tedebrand (eds.), *Society, health and population during the demographic transition*, p. 193–212. Stockholm, Almqvist & Wiksell, 504 p.

Rogers Richard G., 1996. The effects of family composition, health, and social support linkages on mortality, *Journal of Health and Social Behaviours*, vol. 37(4), p. 326–338.

Rollet Catherine, 1990. *La politique à l'égard de la petite enfance sous la III^e République*. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 594 p. (Travaux et documents, cahier no. 127.)

Rollet Catherine, 1997. Childhood mortality in high-risk groups: some methodological reflections based on French experience, *in*: Carlo A. Corsini and Pier P. Viazzo (eds.), *The decline of infant and child mortality. The European experience: 1750–1990*, p. 213–225. Dordrecht, Martinus Nijhoff, 258 p.

Roques Bernard, 1999. *La dangerosité des drogues: rapport au secrétariat d'Etat à la Santé*. Paris, Odile Jacob and La Documentation française, 316 p.

Rosenberg Michael J., 1987. *Smoking and reproductive health*. Littleton, MA, PSG Publishing, 239 p.

Rosenberg Michael J., 1991. Influence of contraceptives on vaginal ecology and infections, *in*: Benson J. Horowitz and Per-Anders Mardh (eds.), *Vaginitis and vaginosis*, p. 55–62. New York, Wiley-Liss, 274 p.

Rousseau Jean-Jacques, 1782. *Les Confessions de J.-J. Rousseau, suivies des Rêveries du promeneur solitaire*, 2 vol. Geneva, Dufour.

Sabin E. P., 1993. Social relationships and mortality among the elderly, *Journal of Applied Gerontology*, vol. 12(1), p. 45–60.

Sachs Benjamin P., 1987. Sharing the cigarette: the effects of smoking in pregnancy, *in*: Michael J. Rosenberg (ed.), *Smoking and reproductive health*. Littleton, MA, PSG Publishing, 239 p. (Re-edited in *Early Human Development*, 1995, p. 134–143.)

Saint-Arroman Auguste, 1845. *De l'action du café, du thé et du chocolat sur la santé, et de leur influence sur l'intelligence et le moral de l'homme*. Paris, J. Laisné, 64 p.

Saint-Arroman Auguste, 1846. *Coffee, tea, and chocolate: their influence upon the health, the intellect, and the moral nature of man*. Philadelphia, T. Ward, 90 p.

Saint-Leger A. S., Cochrane A. L., and Moore F., 1979. Factors associated with cardiac mortality in developed countries with particular reference to the consumption of wine, *The Lancet*, vol. 1, p. 1017–1020.

Sauvy Alfred, 1960. Essai d'une vue d'ensemble, *in*: Hélène Bergues, Philippe Ariès, étienne Hélin, Louis Henry, Michel Riquet, Alfred Sauvy, and Jean Sutter (eds.), *La prévention des naissances dans la famille. Ses origines dans les temps modernes*, p. 376–391. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 500 p. (Travaux et documents, cahier no. 35.)

Seboui H., Boussoffara R., Kuss J., Amri F., Jaziri N., and Snoussi N., 1990. Abandon-adoption: l'expérience du service de néonatologie du Centre hospitalier universitaire de Sousse, Tunisie, *Pédiatrie*, vol. 45(12), p. 869–973.

Schwartz Daniel, 1989. L'explication en épidémiologie, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales: la recherche des causes en démographie*, p. 127–150. Louvain-la-Neuve, CIACO, 478 p. (Chaire Quetelet, 1987.)

Scotto Joseph and Fraumeni Joseph F., 1982. Skin (other than melanoma), *in*: David Schottenfeld and Fraumeni Joseph F. (ed.), *Cancer epidemiology and prevention*, p. 996–1011. Philadelphia, WB Saunders, 1173 p.

Sermet Catherine, 1994. De quoi souffre-t-on? Description et évolution de la morbidité déclarée 1980–1991, *Solidarité-Santé*, vol. 1, p. 37–56.

Sfez Lucien, 1995. *La santé parfaite. Critique d'une nouvelle utopie*. Paris, Seuil, 398 p.

Shapiro, Arthur K. and Shapiro Elaine, 1997. *The powerful placebo: from ancient priest to modern physician*. Baltimore, MD, Johns Hopkins University Press, 296 p.

Shigematsu Itsuzo and Yanagawa Hiroshi, 1985. The case of Japan, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 399–418. Liège, Ordina édition, 558 p.

Shigematsu Itsuzo and Yanagawa Hiroshi, 1985. Le Japon, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 385–402. Paris, Institut national d'etudes demographiques, presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1995. La crise sanitaire en Russie, *Population*, vol. 50(4–5), p. 907–982.

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1996. Health crisis in Russia, *Population, English Selection*, vol. 8, p. 123–190.

Shkolnikov Vladimir and Nemtsov Alexander, 1997. The anti-alcohol campaign and variations in Russian mortality, *in*: Jose Luis Bbadilla, Christine A. Costello, and Faith Mitchell (eds.), *Premature death in the New Independent States*, p. 239–261. Washington DC, National Academy Press, 504 p.

Smith Richard, 1987. *Unemployment and health: a disaster and a challenge*. Oxford, UK, Oxford University Press, 197 p.

Sokolov E. Ivanovich, 1997. Obesity and diabetes mellitus. Utrecht, VSP, 331 p.

Solanto Mary V., Arnsten Amy Frances T., and Castellanos F. Xavier, 2001. *Stimulant drugs and ADHD: basic and clinical neuroscience*. New York, Oxford University Press, 410 p.

Solimano Giorgio R. and Vine Marty, 1982. Malnutrition, infection and infant mortality, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 83–112. Liège, Ordina éditions, 474 p.

Solomon Caren and Manson Joann E., 1997. Obesity and mortality: a review of the epidemiologic data, *American Journal of Clinical Nutrition*, vol. 66(4), Suppl., p. 1044–1050.

Soumoy Eddy and Thiltges Évelyne, 1992. Réflexions sur les facteurs d'inégalité en matière de santé dans les sociétés en développement, *in*: Association internationale des démographes de langue française (ed.), *Démographie et différences*, p. 150–160. Paris, Presses universitaires de France, 662 p. (Colloques no. 8.)

Sournia Jean-Charles, 1986. *Histoire de l'alcoolisme*. Paris, Flammarion, 322 p.

Stamler Jeremiah, 1973. Epidemiology of coronary heart disease, *The Medical Clinics of North America*, vol. 57, p. 5–46.

Stanley Neville F. and Joske Richard A. (eds.), 1980. *Changing disease patterns and human behaviour*. London, Academic Press, 666 p.

Steinbach Ulrike, 1992. Social networks, institutionalization, and mortality among elderly people in the United States, *Journal of Gerontology*, vol. 47(4), p. S183–S190.

Sullivan Frank M. and Barlow Susan M., 2001. Review of risk factors for sudden infant death syndrome, *Pediatric and perinatal epidemiology*, vol. 15(2), p. 144–200.

Surault Pierre, 1979. *L'inégalité devant la mort*. Paris, Economica, 150 p.

Surault Pierre, 1983. Les déterminants socio-culturels de la morbidité et de la mortalité, *in*: Chaire Quetelet (ed.), *Morbidité et mortalité aux âges adultes dans les pays développés*, p. 193–241. Louvain-la-Neuve, Université Catholique de Louvain, Cabay, 257 p.

Surault Pierre, 1984. Les facteurs de différenciation sociale de la mortalité, *Espace Populations Sociétés*, vol. 3, p. 131–150.

Surault Pierre, 1989. Essai de démographie du sport, *in*: Wladimir Andreff (ed.), *Economie politique du sport*, p. 11–33. Paris, Dalloz, 339 p.

Surault Pierre, 1995a. Inégalités sociales devant la santé et la vie dans les pays du Nord, *in*: Hubert Gérard and Victor Piché (eds.), *La Sociologie des populations*, p. 235–255. Montreal, Presses de l'Université de Montréal and Association des universités partiellement ou entièrement de langue française–université des éseaux d'expression française, 508 p.

Surault Pierre, 1995b. Variations sur les variations du suicide en France, *Population*, vol. 50(4–5), p. 983–1012.

Surault Pierre, 1996. Activités physiques et sportives, santé, bien-être et efficacité économique, *Revue juridique et économique du sport*, vol. 40, p. 113–118.

Tabutin Dominique, 1978. La surmortalité féminine en Europe avant 1950, *Population*, vol. 34(1), p. 121–147.

Tabutin Dominique, 1995. Transitions et théories de mortalité, *in*: Hubert Gérard and Victor Piché (eds.), *La sociologie des populations*, p. 257–288. Montreal, Les Presses de l'Université de Montréal/Association des universités partiellement ou entièrement de langue française–université des réseaux d'expression française, 508 p.

Tabutin Dominique and Akoto Eliwo, 1989. Les inégalités socio-économiques et culturelles devant la mort, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala-Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 35–63. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

Tabutin Dominique and Akoto Eliwo, 1992. Socio-economic and cultural differentials in the mortality of sub-Saharan Africa, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala-Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 32–64. International Studies in Demography. Oxford, UK, Clarendon Press, 450 p.

Taylor Ronald B., 1997. *Preventing violence against women and children*. New York, Milbank Memorial Fund, 55 p.

The Surgeon General, 1979. Smoking and health, *in*: The Surgeon General (ed.), *A report of the Surgeon General*, p. i–xxi. Washington DC, U.S. Department of Health, Education, and Welfare.

The Surgeon General, 1982. *The health consequences of smoking. Cancer*. Department of Health and Human Services publication no. CDC/89-8411. Washington DC, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 304 p.

The Surgeon General, 1983. *The health consequences of smoking. Cardiovascular*. Department of Health and Human Services publication no. CDC/89-8411. Washington DC, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 384 p.

The Surgeon General, 1989. *Reducing the health consequences of smoking: 25 years of progress*. U.S. Department of Health and Human Services, vol. 38(S-2), no. CDC/89-8411. Washington DC, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 32 p.

The Surgeon General, 1990. *The health benefits of smoking cessation*. U.S. Department of Health and Human Services publication, vol. 39(RR-12), no. CDC/90-8416. Washington DC, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 12 p.

Thierry Xavier, 1999. Risque de mortalité et de surmortalité au cours des dix premières années de veuvage, *Population*, vol. 54(2), p. 177–204.

Trovato Franco, 1991. *Violent and accidental mortality among immigrants in Canada*. Population Research Laboratory discussion paper no. 81. Edmonton, Canada, University of Alberta, Department of Sociology, Population Research Laboratory, 52 p.

Trovato Frank, 1998. Nativity, marital status and mortality in Canada, *Canadian Review of Sociology and Anthropology/Revue canadienne de sociologie et d'anthropologie*, vol. 35(1), p. 65–91.

Truswell A. Stuart, 1982. Epidemiology of over-nutrition, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 393–416. Liège, Ordina éditions, 474 p.

Tubiana Raoul and Amadio Peter C., 2000. *Medical problems of the instrumentalist musician*. Londres, Martin Dunitz, 626 p.

Tunstall-Pedoe Hugh, 1988. Autres pays, autres moeurs: theories on why the French have less heart diseases than the British, *British Medical Journal*, vol. 297, p. 1559–1560.

Tuyns Albert J., 1982. Alcohol, *in*: Schottenfeld David and Fraumeni Joseph F. Jr (eds.), *Cancer epidemiology and prevention*, p. 293–303. Philadelphia, WB Saunders, 1173 p.

Vachon Mary L. S., 1987. *Occupational stress in the care of the critically ill, the dying, and the bereaved*. Washington, Hemisphere Publishing, 271 p. (Series on death education, aging, and health care.)

Valkonen Tapani, 1987. Social inequality in the face of death, *in*: European Association for Population Studies (ed.), *European Population Conference, Jyväskylä 1987. Plenary sessions*, vol. 1, p. 201–261. Helsinki, International Union for the Scientific Study of Population, European Association for Population Studies, Finnish Consortium, Statistics Finland, 434 p.

Valkonen Tapani and Martikainen Pekka, 1995. The association between unemployment and mortality: causation or selection?, *in*: Alan Lopez, Graziella Caselli, and Tapani Valkonen (eds.), *Adult mortality in developed countries. From description to explanation*, p. 201–222. International Studies in Demography. Oxford, Clarendon Press, 362 p.

Valkonen Tapani, Martikainen Pekka, and Blomgren Jenni, 2001. *Increasing marital status differences in mortality among elderly persons: an international comparison*. Helsinki, University of Helsinki, 15 p. (Paper presented at the European Association for Population Studies seminar, Trends in Differentials in Morbidity and Mortality: Analysis and Explanation, Pontignano, Italy, April 20–23, 2001.)

Vallin Jacques, 1979a. Facteurs socio-économiques de la mortalité dans les pays développés, *in*: Organisation mondiale de la santé/World Health Organization (ed.), *Délibération de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité. El Colegio de Mexico, Mexico, June 19–25, 1979*, p. 266–302. Geneva, WHO, 622 p.

Vallin Jacques, 1979b. Un fait social: la surmortalité des petites filles en Algérie, *in*: Association maghrébine d'étude de la population (ed.), *Actes du III[e] colloque de démographie maghrébine, Tunis, Avril 24–28 , 1978*, vol. 1, p. 189–215. Tunis, Association Maghrébine pour l'étude de la Population, 2 vol., 625 + 641 p.

Vallin Jacques, 1989. Théorie(s) de la baisse de la mortalité et situation africaine, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala-Diakanda (eds.), *Mortalité et Société en Afrique au sud du Sahara*, p. 399–431. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no.124.)

Vallin Jacques, 1992. Theories of mortality decline in the African situation, *in*: Gilles Pison, Étienne van de Walle, and Mpembele Sala-Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 505–436. Oxford, UK, Clarendon Press, 450 p.

Vallin Jacques and Chesnais Jean-Claude, 1975. Les accidents de la route en France. Mortalité et morbidité depuis 1953, *Population*, vol. 30(3), p. 443–478.

Vallin Jacques and Lopez Alan, 1985a. *La lutte contre la mort. Influence des politiques sociales et des politiques de santé sur l'évolution de la mortalité*. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Vallin Jacques, Lopez Alan, and Behm Hugo (eds.), 1985b. *Health policy, social policy and mortality prospects*, p. 399–418. Liège, Ordina édition, 558 p.

Vallin Jacques and Meslé France, 2001. évolution de la mortalité en Europe depuis 1950: mortalité par âge, sexe et causes de décès, *in*: *Tendances en matière de mortalité et mortalité différentielle*, p. 33–188. Strasbourg, Éditions du Conseil de l'Europe, 344 p. (études démographiques no. 36, also published in English: *Trends in mortality and differential mortality*, p. 31–186. Strasbourg, Council of Europe Publishing, 334 p.)

Vallin Jacques, Meslé France, Adamets Serguei and Pyrozhkov Serhii, 2002. A new estimation of Ukrainian losses during the 30's and the 40's crises, *Population Studies*, vol 56(3), p. 249–264.

Vallin Jacques and Nizard Alfred, 1977. La mortalité selon l'état matrimonial: mariage sélection ou mariage protection? *Population*, vol. 32, special issue, Hommage à Louis Henry, p. 95–125.

Volkoff Serge, 1987. Conditions de travail et santé, *in*: National Institute of Statistics and Economic Studies (ed.), *Données sociales. La société française. 1987*, p. 450–455. Paris, National Institute of Statistics and Economic Studies, 625 p.

Weill Hans and Diem J. E., 1978. Relationship between cigarette smoking and occupational pulmonary disease, *Bulletin européen de physiopathologie respiratoire*, vol. 14, p. 137–143.

Wilhelmsson Claes, Vedin Anders, Ulvenstam Göran, Aberg Anders, Johansson Saga, and Wilhelmsen Lars, 1981. Tobacco smoking and mortality from various causes (with discussion), *in*: Harry Boström and Nils Ljungstedt (eds.), *Medical aspects of mortality statistics*, p. 328–344. Stockholm, Almqvist & Wiksell, 387 p.

Williams Naomi, 1989. *Infant and child mortality in urban areas of nineteenth century England and Wales: a record-linkage study*. Ann Arbor, MI, University Microfilms International, 347 p.

Wilmoth John, Vallin Jacques, and Caselli Graziella, 1989. Quand certaines générations ont une mortalité différente de celle que l'on pourrait attendre, *Population*, vol. 44(2), p. 335–376.

Wilmoth John, Vallin Jacques, and Caselli Graziella, 1990. When does a cohort's mortality differ from what we might expect, *Population, an English selection*, vol. 45(2), p. 93–126.

Wingo Phyllis A., 1993. *Five-year survival of women with breast cancer according to prior use of oral contraceptives*. Ann Arbor, MI, University Microfilms International, 323 p.

Wolf Stewart and Finestone Albert J., 1986. *Occupational stress: health and performance at work*. Littleton, MA, PSG Publishing, 243 p.

Wunderlich Ray C., 1982. Sugar and your health: nutritional problems, diabetes, and low blood sugar. St. Petersburg, FL, Johnny Reads, Good Health Publications, 507 p.

Wunsch Guillaume, 1981. Life-styles and death-styles: differentials and consequences of mortality trends, *Genus*, vol 37(1), p. 41–56.

WYNDER Ernst L. and BROSS Irwin J., 1961. A study of etiological factors in cancer of the esophagus, *Cancer*, vol. 14, p. 389–413.
WYNDER Ernst L. and GRAHAM Evarts A., 1950. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma, *Journal of American Medical Association*, vol. 143, p. 329–336.
YESALIS Charles E. (ed.), 2000. *Anabolic steroids in sport and exercise*, 2nd ed. Champaign, IL, Human Kinetics, 493 p.
ZBOROWSKI Mark, 1972. La diversité des attitudes culturelles à l'égard de la douleur, *in*: François Steudler (ed.), *Sociologie médicale*, p. 121–143. Paris, Armand Colin, 398 p.
ZOLA Irving K., 1966. Culture and symptoms: an analysis of patient's presenting complaints, *American Sociological Review*, vol. 31, p. 615–630.

CHAPTER

52

The Medical Response: Medical Technologies and Public Expectations

PAOLO VINEIS

Dipartimento di Scienze Biomediche e Oncologia Umana, Università di Torino, Turin, Italy

Conservative estimates suggest that medical treatment can be scientifically proved effective in only 20% of cases.[1] This fact was behind the major impetus given in recent years to what is known as *evidence-based medicine* (EBM), although the failure of scientific research and medicine to live up to users' expectations is not new. As far back as 1900, the *San Francisco Chronicle* was leading with articles on biomedical issues that are still headline news today, such as baby gender selection, life extension technologies, and new cancer treatments.

After Mary Shelley's *Frankenstein* (1818), the late 19th and early 20th centuries were a time of intense cross-fertilization between science and literature, and the collective imagination centered on the *ambivalence of effects*, an issue extensively explored by contemporary bioethics, particularly by Hans Jonas (1985) and Daniel Callahan (1999). Marshall Berman (1983) argued that Goethe's *Faust* (1790) was an early demonstration of how the noblest ends may spawn the most dreadful deeds.

Examples of the creation of expectations that far exceed the real capacity to deliver practical responses to patients' needs and health problems abound. Jacques Loeb (1912) of the Rockefeller Institute of New York, for example, was in 1910 claiming the ability to recreate life from organic molecules, just as biotechnology developer Craig Venter (2000) was doing very recently. Press reaction to Jacques Loeb's pronouncements varied from enthusiasm to concern. Scientists and academics stressed the benefits and potentials of its clinical application, and the British Association for the Advancement of Science meeting in 1912 was given over to Jacques Loeb's discoveries under the banner of *life per recipe* (Turney, 1998).

Various examples of "technology assessment" are found in Harry Collins and Trevor Pinch's book, *The Golem at Large* (1998), which seeks to analyze the complex relationships among technologic developments, the public image of them, and their real quality-of-life effects. Examples include the Challenger shuttle disaster, the aftermath of the Chernobyl accident, economic forecasting abilities, acquired immunodeficiency syndrome (AIDS) research, and the role of gay self-help groups. The latter is particularly telling as one of the very first examples of a group of biomedical technology users being set up by singularly qualified individuals really able to influence scientific research strategies. The activists of Act-Up have successfully called into question the traditional model of the randomized clinical trial (i.e., phase III studies), thereby prompting the U.S. Food and Drug Administration to authorize the use of a treatment for which no phase III study evidence was available. John Turney (1998) and Harry Collins and Trevor Pinch (1998) assert that the promises of science, their interpretation by the media, and the role of the public have been

[1] This does not mean that such treatments are only 20% effective. They are very likely much more effective, but the proof is lacking.

indissociable for at least a century. The authors of *The Golem at Large* argue that there is no special insight, as claimed, for example, by the leading scientist Richard Feynman (cited in Collins and Pinch, 1998) with reference to the Challenger space shuttle accident, that would enable any technical problem to be foreseen and solved "by a scientist who is smart enough" (p. 151). Harry Collins and Trevor Pinch showed that there was no such thing as an exclusively technical problem; they demonstrated that every situation contained a component of randomness and subjective assessment. They also looked at risk analysis. The skills normally employed in risk analyses are highly specific and sectorial, and they have regard neither to the subjective aspects nor any assessment of risk *acceptability*. The fact that we find 10,000 deaths widely dispersed in space and spread out over time (e.g., road deaths in a year) more acceptable than 10,000 deaths concentrated in a short period in a specific place (e.g., a war) is surely not a technical problem. For Harry Collins and Trevor Pinch, "to give them unconditional respect is to make science and technology a fetish" (p. 153).

I. MEDICAL TECHNOLOGY'S CONTRIBUTION TO MORTALITY DECLINE

How has therapeutic progress helped improve Western public health? Thomas McKeown (1990) showed that most of the remarkable infectious disease mortality decline achieved in the West from the 19th century onward resulted from improved living and working conditions. McKeown estimated that in England between 1854 and 1971, 74% of the mortality gains resulted from a decrease in infectious disease–related mortality. His premise was that most of the decrease occurred even before effective drugs like sulfamides and antibiotics (developed in the 1940s) came in and therefore can only be attributable to other causes, such as improved hygiene and diet. Only 3% of tuberculosis deaths were prevented by treatments between 1848 and 1971, argues McKeown, but the proportion rises to 51% in the period from 1948 to 1971. In the health conditions that now prevail in the developed world, medical treatments are decisive in keeping mortality low, but their role in the rapid mortality decline since the turn of the 18th century was much more limited.

Infectious diseases aside, what share of current mortality improvements can be attributed to medical treatments? Indirect estimates have been done from available health statistics, especially in the Scandinavian countries. In Finland, for example, Kari Poikolainen and Jaakko Eskola (1986) estimated the health services impact on curable diseases between 1969 and 1981. The authors of the study selected the curable diseases (i.e., "amenable to intervention") from a list that included enteritis, tuberculosis, syphilis, other infectious and parasitic diseases, diabetes, hypertension or high blood pressure, and duodenal ulcer. They excluded all malignant tumors (including cervical carcinomas, which are treatable if diagnosed at a sufficiently early stage), the most serious and most frequent cardiovascular diseases (e.g. myocardial infarction, strokes), and sublethal accidents. Mortality from these curable diseases was compared with other-disease mortality. The study showed that mortality for curable diseases fell by approximately two-thirds in 12 years, compared with just one-third for incurable-disease mortality. Most of the decrease in incurable-disease mortality results from the blurred dividing line between curable and incurable diseases and, to a lesser extent, from effective primary prevention measures for nontreatable diseases. The main curable diseases that registered mortality improvements are tuberculosis, hypertension, gallstones, and causes of neonatal deaths. The Finnish authors conclude from this that health services were probably responsible for approximately one-half the curable-disease mortality decline over the review period. Other studies reached highly comparable conclusions. In an international comparison of 10 curable diseases (almost identical to those of the Finnish study), very similar decreases were observed in a set of industrialized countries, including Italy and France (Charlton and Velez, 1986). Table 52–1 shows a marked drop in curable-disease mortality across all countries, from 51% in England to 72% in Japan. Japan is a case apart because curable-disease mortality started from a much higher level, especially for tuberculosis, maternal mortality, and cerebrovas-

TABLE 52–1 Mortality Improvements from Avoidable Causes and Mortality from All Other Causes in Six Industrialized Countries, 1956 to 1978: Comparative Annual Rates per 100,000 People

Country	Curable causes		All other causes	
	1956	1978	1956	1978
England	68.6	33.4	327.2	313.0
United States	74.7	24.6	398.6	364.0
France	69.4	25.0	385.1	311.6
Japan	177.6	49.2	355.3	201.9
Italy	81.4	34.9	331.5	275.3
Sweden	50.4	19.7	267.2	259.4

From Charlton and Velez, 1986.

cular diseases. That notwithstanding, incurable diseases were the primary cause of death in the review period, with annual rates of approximately 300 to 400 per 100,000 people compared with 60 to 80 per 100,000 for curable diseases (except in Japan).

II. RISKS OF CLINICAL PRACTICE AND MEDICAL UNCERTAINTIES

Clinical practice remains beset by risks that to some extent highlight the shortcomings in mechanisms for *transferring research outcomes into practice*. At a workshop held in the San Giovanni Hospital in Turin, a group of doctors were asked in a series of simple questions about current clinical situations: whether they thought vitamin C helped heal scabs; whether they gave vitamin C to prevent or treat a cold; what drug they thought most effective in treating rheumatoid arthritis and which had least side effects; and whether they thought that surgical patients with serious trauma (e.g., femoral fracture) should be preventively treated for deep vein thrombosis (DVT) with heparin or a physical method (e.g., lower limb compression) (Vineis and Canavese, 1999). The respondent doctors were not a representative sample of the entire hospital, but they were not specially selected either. As expected, the doctors' replies were not entirely uniform. For rheumatoid arthritis, for example, 48% gave corticosteroids as the treatment of choice, and 34% opted for antimalarials. For 41%, their choice was based on information about randomized trials or other scientific evidence, and for 50%, the choice reflected their knowledge of the mechanisms of action (i.e. physiopathologic considerations). Although 19% reported using drug treatment alone to prevent DVT, 75% preferred combination therapies. The effectiveness of heparin was significantly overestimated; 88% of respondents thought it cut the risk of thrombosis by 5 to 10 times, whereas reduction established in trials is about twice; 95% thought that heparin cut the risk of pulmonary embolism by more than 25% and reduced mortality.

Having canvassed the clinical practitioners' views, we jointly consulted on each question the *Cochrane Library*, a database of systematic reviews of studies of clinical trials. The aim was not to look for answers but to see what relationship might obtain between what practitioners thought and a good source of updated knowledge. For example, we checked the number of results referenced for prevention of DVT. We found approximately 90 randomized trials to be accessible on the prevention of different forms of thromboembolism in patients with fractured femurs. Of these, however, only 26 were considered relevant and acceptable by the Cochrane Library review groups' criteria. The final outcome of the consultation was that heparin reduces the incidence of DVT, but it increases the risk of nonfatal pulmonary embolism and, paradoxically, the probability of dying. Physical methods (e.g., lower limb compression) present a somewhat different case. For an equal effect on DVT, there is also a preventative effect on pulmonary embolism deaths. Surprisingly, far fewer studies were found on the effectiveness of physical methods (4) than on heparin (19), perhaps because of the different commercial interests involved.

Rheumatoid arthritis is a much more clear-cut example. Treatment of acute cases with injectable gold salts and treatment with corticosteroids are about equally effective (measured as the difference between the average number of inflamed joints in treated patients and a control group), but gold salts have clearer and much more severe side effects.

The broad findings from these few examples could be demonstrated for many other clinical practices, including widely varying opinions and attitudes among doctors and scientific results that overall are hard to evaluate and to condense and that are often anything but compelling and clearly transferable into practice. The risks of clinical practice do not result from doctors' ignorance or inefficiency. Rather, choosing the best treatment is often a complex procedure of sifting through mountains of information that includes research outcomes and subjective aspects such as patient acceptance of treatment.

The need to address this problem was behind the founding of the *Cochrane Collaboration* (named after one of the inventors of randomized clinical trials) some years ago and the more general movement of EBM. Its main purpose is to facilitate transfer of information into practice through the preparation and dissemination of systematic literature reviews (e.g., the Cochrane Library). EBM has the merit of signaling which medical treatments have proved valid and which are useless or harmful. However, it has been criticized on the ground that not all medical activity is fitted to randomized, controlled trials, especially complex ones such as prevention practices; EBM also disregards the interpersonal and subjective aspects of the doctor-patient relationship.

III. BRIDGING NATURAL SCIENCE AND ANTHROPOLOGY

The relationship between biomedical science and public expectations has always been difficult (Turney, 1998). The role of biomedical technology often has been overestimated at the macro level (e.g., mortality decline in the West from the 19th century and onward) and the micro level (e.g., daily clinical practice). Broadly, the issue is one of transferring research outcomes into clinical practice. The main emphasis here has been on the objective aspects, such as how to facilitate the preparation and dissemination of clear and usable abstracts of medical research to update clinical practices (e.g., Cochrane Library), but the subjective aspects of the doctor-patient relationship—or more generally, those between biomedical research and public expectations and images—should not be disregarded. The doctor-patient private consultation is often a fragmentary and frustrating dialogue, but is it right to class patients' beliefs as superstitions or as short of the scientifically evidenced mark? Evidence shows that the problem is much more complex, as suggested, for example, by the work of the leading anthropologist Byron Good, who demonstrates that the doctor-patient dialogue must be looked at in the whole context. This means looking at the substance of the communication, as well as turns of phrase, the form of the discourse, and even the punctuation. It is to Byron Good's credit that he has illuminated how the doctor-patient narratives are distinguished not only by "high" versus "low" knowledge, but also by "knowledge" versus "belief" (Good, 1994). It is highly valid and significant that a trend toward dissemination of proven scientific knowledge (i.e., EBM) has developed within clinical medicine, but we also hope that doctors assimilate at least two lessons of anthropology. First, the significant improvements in Western public health in the 19th century were more the result of demographic and anthropologic than technologic factors. Second, the transfer of medical knowledge into clinical practice must deal with patients' own cultures, which should be respected.

Acknowledgment

This chapter was translated from the French version by Glenn D. Robertson.

References

Berman Marshall, 1983. *All that is solid melts into air*. London, Verso, 101 p.

Callahan Daniel, 1999. *False hopes: overcoming the obstacles to a sustainable, affordable medicine*. Rutgers, NJ, Rutgers University Press, 330 p.

Charlton John R. H. and Velez Ramon, 1986. Some international comparisons of mortality amenable to medical intervention, *British Medical Journal*, vol. 292, p. 295–301.

Collins Harry and Pinch Trevor, 1988. *The Golem at large*. Cambridge, MA, Cambridge University Press, 163 p.

Goethe Johann Wolfgang von, 1790 (English translation 1990). *Faust*. New Haven, CT, Yale University Press, 272 p.

Good Byron, 1994. *Medicine, rationality and experience*. Cambridge, MA, Cambridge University Press, 242 p.

Jonas Hans, 1985. *Technik, Medizina und Ethik*. Frankfort, Insel Verlag, 323 p.

Loeb Jacques, 1908. La dynamique des phénomènes de la vie. Paris, F. Alcan, 439 p.

Loeb Jacques, 1912 (reprinted 1964). *The mechanistic conception of life*. Cambridge, MA, Harvard University Press, 232 p.

McKeown Thomas, 1990. *The origins of human disease*. Oxford, New York, Blackwell, 238 p.

Poikolainen Kari, and Eskola Jaakko, 1986. The effect of health services on mortality: decline in death rates from amenable and non-amenable causes in Finland, 1969–81. *Lancet*, 1(8474), p. 199–202.

Shelley Mary, 1818 (reprinted 1996). *Frankestein*. London, Norton, 352 p.

Turney John, 1998. *Frankestein's footsteps*. New Haven, CT, Yale University, 320 p.

Venter Craig, 2000. *The human genome*. Rockville, MD, Celera Genomics. (Paper presented at the conference Biotechnology: The Science and the Impact, Netherland Congress Center, The Hague, January 20–21, 2000.)

Vineis Paolo and Canavese Caterina, 1999. Presentazione dell'edizione italiana, *in*: Alfredo Morabia (ed.), *L'epidemiologia clinica*, p. IX–XIII. Rome, Il Pensiero Scientifico Editore, 105 p.

P A R T

III

THE INTRICACY OF DIFFERENTIAL MORTALITY FACTORS

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Mortality varies widely (as will be seen more in detail in Volume III, on the history of population settlements), not least by level of national economic, social, medical or cultural development. Patterns differ not just between countries, but also between national subpopulations—be they regional or local—as well as subgroups within a population defined by membership criteria like gender, social status, marital status, educational level, occupation, etc. In fact, any criterion of differentiation used, provided it has a social, economic, psychological, cultural, or other significance for health and mortality, is highly likely to reveal mortality differentials.

These inter-group mortality differentials may have three very different but complementary origins. First and most obvious is belonging to a group. If, for example, femaleness is biologically propitious to survival, then femaleness guarantees a longer life than maleness. Likewise, being married, educated, or in a higher social status group generally entails category-specific behaviors that improve probabilities of survival. Secondly, where the criterion is not based on a genetic, unchangeable characteristic like sex, for example, the change of state may itself add a further differential to the category-specific one. So, widowhood, unemployment, loss of social status, etc., may be a time-specific factor of excess mortality. Finally, health-specific entry into a category may create a selective inequality. Married people's lower mortality than single people may in part be due to less healthy individuals' greater difficulty in finding a marriage partner and tendency to remain never-married. Likewise, at the normal economically active ages, there is a significant excess of nonworking over working population mortality, often due to ill-health keeping or driving them out of the labor market.

The factors that predicate these three grounds of differentiation are themselves varied and complex. There is an intricate number of explanations for each difference observed. An examination of all possible dimensions of differential mortality is clearly outside the scope of this treatise, so it was thought preferable to confine the analysis to three significant examples, striving for as comprehensive as possible an outline of each.

In Chapter 53, Jacques Vallin discusses the relationship between mortality, sex and gender. In Chapter 54, Tapani Valkonen will consider the mortality gap by social status. Finally, in Chapter 55, Graziella Caselli and Jacques Vallin will examine geographical mortality variations.

C H A P T E R

53

Mortality, Sex, and Gender

JACQUES VALLIN
Institut national d'études démographiques (INED), Paris, France

Girls are fewer in number at birth and live longer than boys. It appears that these two "laws of nature" biology impose on the human race complete one another to ensure the presence of men and women in almost identical numbers at reproductive ages. However, these are two laws that psychological attitudes and social behaviors can profoundly modify to the detriment of one or the other sex. The topic of the sex ratio of births is discussed in Chapter 4. This chapter analyzes the causes of excess male mortality, which is prevalent in most countries.

Life expectancy among women is higher than among men almost everywhere today (Table 53–1). In the United States in 1995–2000, the difference was 6.7 years; in France, it was 7.8 years. In some cases, the difference is much higher, as in Russia, with a difference of more than 12 years, but in other cases, it is much smaller, as in Iran (1.5 years), India (0.6 year), or Bangladesh (0.1 year), to mention only countries with more than 50 million inhabitants (systematically included in Table 53–1).

This diversity indicates that the difference is not merely a biologic effect. It is unlikely that the genetic distance between large human populations is enough to bring about such differences in a purely biologic parameter. The difference in mortality between men and women is not only a question of biologic sex; it is mainly a question of "socially constructed sex," or a question of gender.

The range of mortality situations is the result of historical evolution, which brings into play an even greater diversity, the examination of which makes it possible to understand the factors determining the difference in life expectancy between men and women. From the outset, it seems plausible that women have a biologic advantage. In the past and in certain cases today, the social conditions imposed on women deprived them of the benefits of this advantage, which women eventually regained through improved social status. Regaining this lost advantage led to widening the gap in life expectancy between men and women far beyond the strictly biologic difference. Is this because of the development among men of specific behaviors harmful to health? Alternatively, does it have more to do with a greater ability on the part of women to make the most of improved health conditions? Both successively made a difference.

I. A PLAUSIBLE NATURAL ADVANTAGE

Part of excess male mortality, which according to different investigators varies in degree, has always been considered as the expression of the biologic difference between the sexes. Some authorities even consider the XY pair of chromosomes, which determines the male sex, a lesser version of the XX pair of the female sex, perceiving males as having got the short end of the stick (Federici, 1950; Lenz, 1940).

1. An Obvious Genetic Advantage

If a mutation affects one of the genes of the X chromosome in women, it does not cause a disease or disorder in most cases because the second X chromosome

TABLE 53–1 Differences in Life Expectancy among Men and Women in 1995–2000 by Regions and Main Countries of More Than 50 Million Inhabitants

Regions and countries	Life expectancy		Difference Women – Men
	Men	Women	
Whole world	63.2	67.6	4.4
Africa	50.0	52.8	2.8
North Africa	63.3	66.4	3.1
Egypt	64.7	67.9	3.2
Central Africa	47.1	49.9	2.7
Nigeria	48.7	51.5	2.8
Ethiopia	42.4	44.3	1.9
Southern Africa	51.5	57.5	6.0
Eastern Asia	68.8	73.4	4.6
China	67.9	72.0	4.1
Japan	76.8	82.9	6.1
Southern Asia	62.6	64.7	2.1
Southeast Asia	63.7	67.8	4.1
Indonesia	63.3	67.0	3.7
Vietnam	64.9	69.6	4.7
Philippines	66.5	70.2	3.7
Thailand	65.8	72.0	6.2
Central South Asia	61.8	62.9	1.1
India	62.3	62.9	0.6
Pakistan	62.9	65.1	2.2
Bangladesh	58.1	58.2	0.1
Iran	68.5	70.0	1.5
Southwest Asia	65.9	70.2	4.3
Turkey	66.5	71.7	5.2
Latin America	66.1	72.6	6.5
Central America	67.9	73.3	5.4
Mexico	69.5	75.5	6.0
South America	65.3	72.3	7.0
Brazil	63.1	71.0	7.9
North America	73.6	80.2	6.6
United States	73.4	80.1	6.7
Europe	69.2	77.4	8.2
Western Europe	73.8	80.3	6.4
Germany	73.9	80.2	6.3
United Kingdom	74.5	79.8	5.3
France	74.2	82.0	7.8
Italy	75.0	81.2	6.2
Eastern Europe	63.3	73.8	10.5
Russia	60.6	72.8	12.2
Ukraine	63.8	73.7	9.9
Oceania	71.4	76.3	4.9

From United Nations, 2001.

compensates for the loss of function (see Chapter 49); in men, all genes on the single X chromosome express themselves, even if they are deleterious. This is why genetic diseases such as hemophilia or myopathy mostly affect males. More than 1000 diseases are linked to the X chromosome. Not all of them specifically affect males, because the Y chromosome can sometimes compensate for a defective gene on the X chromosome, but in the end, a female advantage is evident. The female advantage is derived from more than protection from genetic diseases. Certain specifically female functions also are associated with greater life expectancy, and the genetic difference between the sexes generally seems to be associated with better resistance to biologic aging. In this chapter, only a few selected factors and their possible link with life expectancy are discussed.[1]

The female body must have food reserves for pregnancy and breast-feeding. According to some biologists, this capacity in women to store and eliminate food reserves without damage to the body allows them to deal with changes in living conditions with greater ease, to endure overfeeding, and to more efficiently eliminate excess food (Seely, 1990).

Folliculine-stimulating hormone (FSH) is secreted until menopause. FSH facilitates dilatation of veins and protects against circulatory diseases. Male and female hormone profiles are very different. The estrogen-to-testosterone ratio is approximately 10 in women and 0.1 in men. Estrogen in women enables them to eliminate more easily the bad low-density lipoprotein cholesterol and to afford them greater immunity against cerebrovascular diseases. At the same time, testosterone, the hormone of aggressiveness, tends to incite men to violence and to taking risks.

In addition to the effect of FSH, the female circulatory system has other advantages. For equal physical effort, women have lower increases in arterial pressure and greater elimination of lipids, which protects them against ischemic heart diseases (Nygaard *et al.*, 1990). Female cardiac muscle, which functions in a more complex way, is more flexible than male cardiac muscle (Ryan *et al.*, 1994).

The pair of X chromosomes plays an important role in defending the body against oxidizing lesions caused by free radicals by ensuring the replication of repair enzymes, a function that a single X chromosome is much less able to fulfill. If a gene on the X chromosome is defective and unable to code for the correct repair enzyme, men are unprotected, whereas women can retain the repair capacity because the corresponding gene on their second X chromosome is functional (Hollyday, 1987). Jean-Claude Chesnais (1998) described other genetic mechanisms that may underlie the greater life expectancy of women. There are many other examples, but apart from the fact that these phenomena and their effects on life expectancy

[1] The factors are among those mentioned by Jean-Claude Chesnais (1998) in a review on the subject presented at an Association internationale des démographes de langue française conference.

have not yet been entirely elucidated (Magnami and Accorsi, 1993), the part played by all of the genetic factors in the difference in life expectancy between men and women is very difficult to isolate from the effects of nonbiologic factors.

2. An Advantage That Is Difficult to Measure

It is difficult and perhaps impossible to measure the part played by genetic factors and that played by sociocultural factors in the differences in mortality observed between groups of human beings. How can we isolate the effect of the genotype from that of the sociocultural context when, from a genetic standpoint, differences between individuals are often much greater than differences between groups? How can a demographer interpret the differences observed between groups when, in many cases, the constitution of these groups itself has an influence on mortality through selection? These are only two of many difficult questions to solve for the analysis of differential mortality. We are at least certain that sex divides the population into two distinct groups on the basis of a genetic difference, the famous Y chromosome, and because it constitutes for each individual a constant characteristic from conception until death, it can induce no selection effect.

Is there an "admissible reference that will make it possible to sort out what is attributable to biologic factors from what is linked to social and behavioral factors in excess male mortality"[2] (Pressat, 1973)? It is very unlikely, because sex in itself is a support for sociocultural differences, the effects of which are extremely difficult to isolate to evaluate those of biologic difference.

Some demographers have presented very different arguments to demonstrate that the difference in life expectancy and its recent evolution almost entirely resulted from women's purely biologic advantage over men. For example, in trying to understand the reasons for the constant increase in excess male mortality, which at the time gave an advantage of up to 6 years in life expectancy to American women, Francis Madigan wrote in 1957: "Not only sociocultural pressures *[are]* less important than biologic factors in relation to the mortality differentials of sexes, but they are of comparatively small importance in this respect." He reached such an extreme conclusion by making a comparison between two groups of religious communities of different sexes that obeyed exactly the same monastic rule and that he therefore assumed had the same living conditions. The slight excess female mortality observed at the beginning of the century had in the space of 50 years given way to increasing excess male mortality, similar to that prevailing in the total population. Considering that between these two religious populations, the biologic difference linked to sex was the only intervening element, he therefore concluded that for the whole of the population, as for the religious communities, the evolution of excess male mortality could be explained almost entirely by the sharp decline in infectious diseases, against some of which (i.e., tuberculosis) women supposedly were less well armed than men, and by the increasing weight of cancers and cardiovascular diseases to which they were less susceptible than men. This did not take into consideration a major difference, admitted by the author himself, that existed between the two religious groups: the consumption of tobacco.[3] This approach also denied the fact that sociologic sex, or gender, could more generally induce different sociocultural behaviors, even within the framework of similar monastic rules. While avoiding the thesis of the total prevalence of biologic factors that was very much in vogue at the time, other investigators attempted to measure the proportion of excess male mortality that might be attributed to it.

In an attempt to generalize the distinction he had made between *endogenous* and *exogenous* mortality in infancy (Bourgeois-Pichat, 1951a, 1951b), Jean Bourgeois-Pichat (1952) built a *table de mortalité biologique limite* (i.e., table of maximum life expectancy), attributing to women a life expectancy greater than that of men by 1.9 years: 78.2 years compared with 76.3. The concept of *endogenous* mortality is not, however, understood here in the strictly biologic sense (and even less in the genetic sense of the term). It covers more of the morbid processes that, at a given stage of knowledge, medical science is more or less powerless to counteract (see Chapter 48). Jean Bourgeois-Pichat took into account future progress in medical science that would push this limit back. Twenty-five years later, in making the same calculations with the same concepts, he obtained a considerably greater difference in mean life expectancy between the two life tables:

[2] "Référence admissible pour démêler ce qui dans la surmortalité masculine tient aux facteurs biologiques et ce qui revient aux facteurs sociaux et de comportement."

[3] It is true that at the time the harmful effects of tobacco were not as well known as they are today, but as Roland Pressat stressed at the time (1973) in a criticism of Francis Madigan's work, according to a contemporary study by the American Cancer Society, male excess-mortality beyond the age of 35 in 54% of cases among cigarette smokers was due to smoking, with the harmfulness of this habit principally increasing mortality from cancer and cardiovascular diseases: exactly the diseases to which Francis Madigan thought that women were more biologically resistant.

80.3 years for women and 73.8 years for men, corresponding to a difference of 6.5 years (Bourgeois-Pichat, 1978). During this period, what Jean Bourgeois-Pichat referred to as the "hard rock" was largely eroded in the case of women, whereas for men, it was reinforced. This cannot have been the result of increased biologic differentiation between the sexes; it must instead reflect improved health measures or changes in behavior that may have evolved differently according to sex. The main reason for the apparent decline in maximum life expectancy among men is linked to increased smoking, which provoked during that period an explosion in the number of lung cancers, which Jean Bourgeois-Pichat included with all other cancers in his calculations of endogenous mortality. Measured in this way, maximum biologic mortality gives only an approximation by excess of the biologic difference between the sexes.

In a different way, Roland Pressat calculated the biologic advantage of women to be 2 years of life expectancy at birth. In the absence of efficient health practices, a difference of this amount seems to be almost totally attributable to biologic factors. This assumption seems to be confirmed by the fact that "currently, in the first year of life, when no outside influence is likely to cause a mortality differential according to sex, at least in the West, mortality among boys exceeds by 25% to 30% that of girls" and that "the continued difference during the whole of life would be such as to result in a moderate difference of approximately [. . .] 2 years [in life expectancy at birth]"[4] (Pressat, 1973). Although a modern demographer may have some doubts concerning the historical argument (i.e., the biologic advantage of women was masked by the negative consequences of their social status), the second argument based on infant mortality is one of the most convincing (Pressat, 1987). Extrapolating this information to other ages remains risky. Although there is an advantage for females at the start, the advantage most likely lessens with age. However, certain elements in the biologic difference may lead a person to believe that it increases at very old ages through improved resistance to aging. At this stage, there is a certain convergence of opinions concerning the existence of a female biologic advantage, probably a rather small one, even if it must be acknowledged that it cannot be measured accurately. Let us assume that God made women a little stronger than men.

[4] "Actuellement, dans la première année de vie, quand ne joue aucune influence extérieure de nature à différencier la mortalité selon le sexe, au moins dans l'univers culturel occidental, la mortalité des garçons surpasse de quelques 25 à 30% celle des filles . . . le maintien de cet écart tout au long de la vie serait de nature à entraîner une différence modérée, de l'ordre de [. . .] deux ans [espérance de vie à la naissance]."

II. ADVANTAGE THAT IS REDUCED OR INVERSED BY UNFAVORABLE SOCIAL STATUS

The first in a series of paradoxes in the analysis of mortality differentials between the sexes in past populations or more recently in populations with high mortality is the fact that an almost equal chance of survival (i.e., combining excess female mortality in adolescence and the beginning of adulthood with excess male mortality at older ages) or even a more global excess female mortality, which leaves women with a life expectancy inferior to that of men, usually is observed when measurement is possible. In certain circumstances, corresponding in all likelihood to a given state in social and economic development, nature's generosity to women has been thwarted by the actions of men (having for a long time confined women to an inferior social status) or of nature itself (reserving for women most of the risks linked to reproduction). Despite of the biologic advantage of women, past populations and some modern populations in developing countries have been marked by certain forms of excess female mortality.

1. Maternal Mortality

Excess female mortality at reproductive ages is well known. It has been abundantly documented in Europe in the past (Tabutin, 1978) and measured in France, thanks to the analysis of the representative sample of parish registers conducted by Louis Henry in the 1960s. In the 18th century and at the beginning of the 19th century, between the ages of 25 and 45 years, the probability of women dying was higher than that of men, between 5% and 20% according to the age group and the period (Blayo, 1975). In developing countries, the same pattern was seen until very recently (Potter and Volpp, 1993).

Excess female mortality is linked essentially to maternal mortality. It could therefore be qualified as biologic because only women are exposed to this risk because of pregnancy, childbirth, and postpartum complications. It is therefore partly nature that took back from women some of the advantage it had conferred on them. However, beyond the purely biologic risks, maternal risks also depended on the economic and social conditions of women's reproductive lives, which were closely linked to their status and level of fertility, which in itself is closely linked to the cultural and social context of females. It is therefore also (male) society that partly eroded women's natural advantage. In 18th century France, women's overall life expectancy at 25 years of age was slightly less than that of men, because male excess mortality observed after the age of 50 years was not enough to compensate for excess female mortality between the ages of 25 and 45 years.

2. Mortality at Young Ages

The role of women's status is even more obvious concerning excess female mortality at young ages, which can be explained only by differences in behavior toward children according to their sex. In the past, female excess mortality was manifest in Europe, where it seemed to have been especially pronounced in the 1- to 4- and 5- to 9-year-old age groups. The degree of female excess childhood mortality seemed to worsen in the 19th century (Tabutin, 1978). In France, more precise statistics from 1899 show that the issue of excess mortality was still a concern for all girls between 4 and 17 at the beginning of the 20th century and that it disappeared only during the Second World War after gradually becoming limited to a declining number of age groups (Vallin, 1983; Tabutin, 1978). Dominique Tabutin stressed the essential role played by infectious diseases, notably tuberculosis, in this excess mortality among little girls, which was sufficiently high to overtake excess male mortality due to accidents or violence, which was already high in these age groups. For Tabutin, the excess female mortality due to infectious diseases was directly linked to the unfavorable conditions imposed on girls at the time, who were more often deprived of health care and food, under the influence of an antifeminist ideology[5] that attributed little value to girls compared with boys.[6]

This lack of value of the female sex is at the origin of excess mortality among girls today in many developing countries. The most glaring examples come from the Indian subcontinent (Das Gupta, 1987; Chen *et al*., 1981). In Bangladesh, mortality among girls between the ages of 1 and 14 years was 50% higher than that of boys at the end of the 1970s. Inequality between food rations and health care for boys and girls was the cause (Chen *et al*., 1981). In India, the burden on the parents of female children imposed by the dowry system often means that the birth of a female is a catastrophe for the family. The phenomenon has been observed in most cultural regions. In addition to Bangladesh, these practices continued until recently in most Moslem countries (Adlakha and Suchindran, 1985), including in the Maghreb (Haffad, 1984). In Algeria, for example, excess mortality among little girls from the third month of life constitutes exclusively the *exogenous* component of infant mortality, whereas the *endogenous* component is marked by high excess male mortality (Vallin, 1979). As in Bangladesh, the phenomenon seems, to be closely linked to the fact that less attention is given to the female sex (Ferry, 1979; Vallin, 1979). Even in sub-Saharan Africa, for which statistics are too imprecise to isolate the different components of infant mortality, it may be that despite a cultural context that is less unfavorable to women, the absence of an infant mortality differential between the sexes is linked to unfavorable treatment of little girls, who should otherwise have a lower infant mortality rate than boys (Gbenyon and Locoh, 1989, 1992).

The relation between women's social status, the value of the female sex, and excess mortality among little girls stands out in China, where there is evidence of the survival of the old practice of female infanticide, which the governmental policy of the single child has probably revived (Calot and Caselli, 1988, 1989). In the province of Anhui, where the probability of dying before the age of 1 year for females is more than 12% higher than for males, Wangsong Zhang (1983) estimates that infanticide, which may responsible for almost 60% of infant deaths, is the main cause of this exceptional excess female mortality. This prejudice against female babies also is found in the practice of discriminatory *feticide*, which consists of eliminating female embryos after having an ultrasound scan (Miller, 1996).

The biologic advantage women theoretically have in terms of life expectancy has been all or partly masked by the unfavorable influence of women's social status for a long time in Europe and in many developing countries, where prejudice against females continues to this day. In India (Ragavashari *et al*., 1974), in Bangladesh (Bean and Khan, 1967; Yusuf, 1967), in Pakistan (Afzal, 1973; Yusuf, 1967), and in Sri Lanka (Sri Lanka, 1974), the accumulation of the different aspects of excess female mortality from young ages to the ages of maternity was so high that up until the end of the 1960s, it led to a much shorter mean life expectancy for women than for men. In Algeria at the beginning of the 1970s, excess male mortality in the

[5] Developed, as Dominique Tabutin (1978) reminds us, by great minds such as Jean-Jacques Rousseau, August Comte, Balzac, and even the Socialist Proud'hon, who had "scientifically calculated" that the value of a woman was only $^{8}/_{27}$ of that of a man!

[6] To illustrate to what extent this ideology prevailed in the past, André Armengaud (1973) mentioned two anecdotes drawn from *Histoire Morale des Femmes* by Legouvé (1842): "Question a peasant about his family, and he will answer: 'I have no children; I only have daughters.' And the Breton farmer, whose wife gives birth to a daughter, still says today: 'my wife had a miscarriage.' "

first months of life and beyond the age of 50 years just counterbalanced life expectancy at birth for both sexes, and the phenomenon persists today. In France in the 18th century, at a time when the absence of health care made mortality less sensitive to social inequalities, low excess male mortality in the first year applied to a very high overall rate was enough to produce a shorter mean life expectancy for the male sex. However, from the age of 1 year, life expectancy, combining excess female mortality before the age of 45 and excess male mortality beyond that age, was equivalent for the two sexes or slightly lower for women. Although "God made women" a little more resistant than men, for a long time, men took away from them the benefits of this natural advantage.

III. FROM REGAINING THE ORIGINAL ADVANTAGE TO OBTAINING A NEW PRIVILEGE

Everything changes. Extraordinary economic and social progress from the 18th century was accompanied by a reduction in social differences between the sexes and a radical lightening of the weight of maternities. Early on, demographic transition led to greater life expectancy among women. In improving their status, women might have been merely regaining the lost benefit of their natural advantage.

If so, it would not be enough to account entirely for the comparative evolution of women's status and the mortality differential between the sexes. The life expectancy of women depends on many social factors beyond the biologic difference that is responsible for the substantial mortality differences between the sexes. While pursuing improved social status, women have continued to benefit from the protection resulting from their biology. This partially explains why everywhere in Europe since the beginning of the 19th century, excess male mortality has continually increased.

In the past 2 centuries in Europe and more recently (and much faster) in developing countries, women have acquired a new social status that will soon compare favorably with that of men. Moreover, the attitudes of adults toward children have changed completely during this period, attributing to them value that is increasingly independent of their sex (Vallin and Lery, 1975, 1978).

Having obtained equality in food, health, and cultural terms, there no longer was any reason for little girls to die younger than little boys, and they could recover the full benefit of their biologic advantage. With the decline in fertility rates and with the improvement of conditions for women during pregnancy and childbirth, the only specific health disadvantage for women disappeared. However, the recent change in mortality has gone far beyond that of recovering women's original advantage to creating a new and even greater advantage for them.

1. Disappearance of Excess Female Mortality and the Development of Excess Male Mortality

In France, even under the *Ancien Régime*, women's life expectancy at birth has never been lower than that of men. According to research by Louis Henry, from the 1740s to the 1860s, the female advantage remained small, shifting rather irregularly between 0.6 and 2.4 years, and generally remaining less than 2 years, which according to Roland Pressat, corresponded to the female biologic advantage. It was therefore not entirely canceled by the unfavorable status given to women, unless we also take into account the nonbiologic factors of excess male mortality that at that time were added to the biologic factors: violence and accidents. Since then, in 1 century, the difference between life expectancy at birth has increased substantially, going from 1.7 years between 1865 and 1869 to 8 years between 1984 and 1986 and reaching a level four times higher than the assumed 2-year female biologic advantage. However imprecise estimating their biologic advantage may prove to be, it is evident that women have done much more than simply recover it, or men have lost much more than strict equality of treatment would have produced for them in terms of life expectancy. These facts are not limited to France. Even if there are great disparities today between industrial countries concerning excess male mortality, everywhere the tendency has been the same. Although it was very limited in the 19th century, the difference in life expectancy increased and even accelerated in the 1950s and 1980s (Table 53–2). Although the difference remains smaller in the most advanced developing countries, it already exceeds 2 years.

An investigator becomes even more clearly aware of the extent of women's progress when examining the evolution in ratios of male excess mortality by age. Figure 53–1 provides an illustration for France. To evaluate this ratio, a demographer may use, for exploratory purposes, Roland Pressat's two assumptions. First, excess male infant mortality essentially results from biologic factors. Second, without the intervention of other factors, the same excess male mortality is observed at other ages. At less than 1 year, since the beginning of the 19th century, excess male mortal-

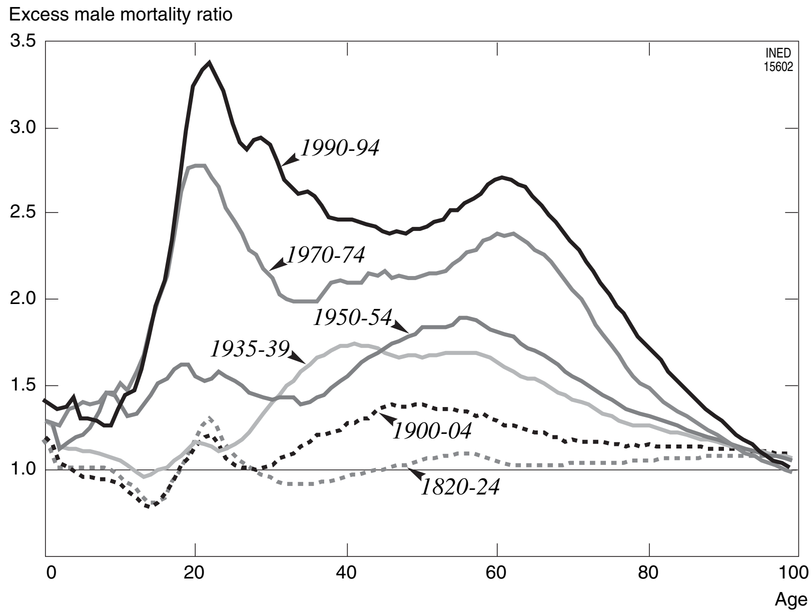


FIGURE 53–1 Evolution in excess male mortality ratios by age in France since the beginning of the 19th century.

TABLE 53–2 Life Expectancy by Sex in France since the 18th Century and in Russia, the United States, and Japan in the 20th Century

Period	Life expectancy		Difference between sexes
	Men	Women	
France			
1770–1779	28.2	29.6	1.4
1820–1824	38.2	39.5	1.4
1865–1869	39.9	41.6	1.7
1909–1913	48.6	52.5	3.9
1950–1954	64.0	69.9	5.9
1990–1994	73.1	81.3	8.2
1997	74.2	82.1	7.9
Russia			
1896	30.9	33.0	2.1
1938	40.1	46.0	5.9
1958	63.0	71.3	8.3
1994	57.7	71.3	13.6
1997	61.0	73.0	12.0
United States			
1900	46.3	48.3	2.0
1950	65.6	71.1	5.5
1979	70.0	77.8	7.8
1990	72.0	78.8	6.8
1995	72.5	79.3	7.2
Japan			
1900	37.5	38.0	0.5
1950	57.5	60.8	3.3
1994	76.6	83.0	6.4
1996	77.0	83.6	6.6

ity has remained almost perfectly constant, ranging between 25% and 30%, and this excess can be attributed to the male biologic disadvantage. Today, there is no difference in attitudes toward children according to sex,[7] and despite the marked prejudice in the past, this difference has had only a small impact on mortality in the first year of life. However, the situation is very different at other ages for which radical changes have occurred.

At the beginning of the 19th century, excess female mortality was clearly demonstrated in childhood and adolescence (5 to 18 years) and in at reproductive ages. Around 1900, there remained excess female mortality at young ages, but excess male mortality was already high after the age of 35 years, exceeding the level reached for men approximately 50 years of age at less than 1 year of age. Just before the Second World War, excess female mortality at young ages had almost disappeared (except at 13 and 14 years), but the ratio of excess male mortality remained at all ages clearly below that observed in the first year of life, suggesting that unequal treatment of young girls probably still had a negative effect on their survival. At the same time, excess male mortality considerably increased in adulthood, reaching 60% between the ages of 40 and

[7] A precise analysis of substantial excess fertility after the loss of a child shows that in the 1960s, the desire to replace the child expressed by the parents bears no relation to the sex of the deceased child (Vallin and Lery, 1975, 1978).

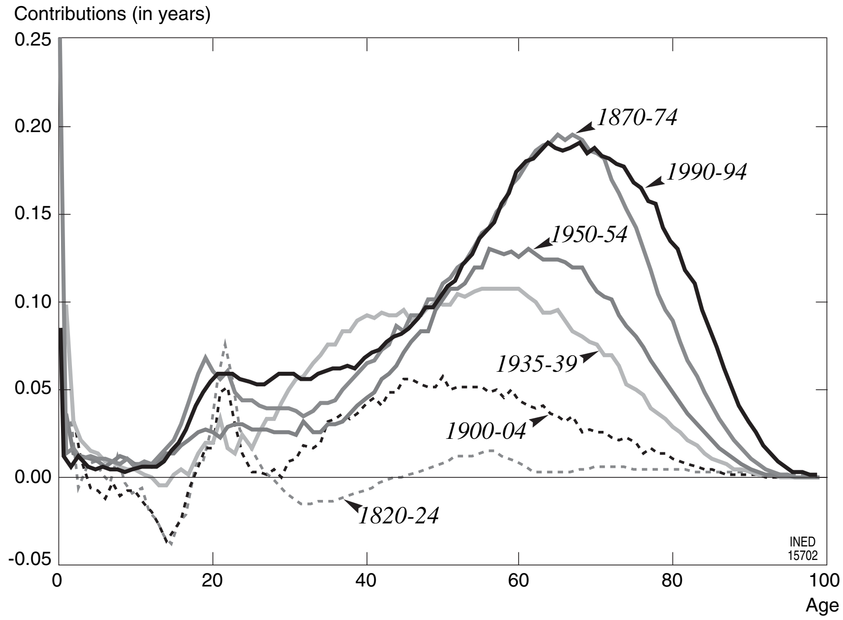


FIGURE 53–2 Evolution of the contribution of the differences in mortality by age to the difference in life expectancy at the beginning of the 20th century.

60. Just after the war, excess male mortality was almost the same as at birth until the age of 35 years, suggesting that women had probably fully recovered their biologic advantage at these ages, whereas at older ages, men were still losing ground. From the 1950s, excess male mortality continued to worsen and even accelerated at older ages, and a second upsurge in excess male mortality harshly affected a few years around 20 years of age.

At the beginning of the 1970s, the excess male mortality ratio exceeded 200% from 18 to 70 years, reaching as high as 240% between the ages of 60 and 70 and even 270% at 20. Since then, the relative situation of men has worsened again, with a peak of 350% at 20 to 22 years and very wide coverage, with a rate of more than 250% at adult ages. At all ages, this is 10 times the reference level of 25% to 30%!

This vision of the mortality difference between the sexes given by the ratio of the age-specific death rates does not, however, take into account the very great variation in death rates according to age and their weight in the difference in life expectancy at birth. Using a method of decomposition of the difference in life expectancy between the sexes (Pollard, 1988, 1990) (see Chapter 16), it is possible to measure the contribution of each year of age to this difference (Vallin, 1993a). Although excess male mortality at less than 1 year of age has remained almost constant, its weight in the difference in life expectancy between the sexes has greatly declined. At the beginning of the 19th century, it alone explained almost all of the difference in life expectancy between the sexes (i.e., 1.4 years of a total of 1.5 years between 1820 and 1824). Between 1900 and 1904, it still accounted for almost one-half of the difference (i.e., 1.4 of 3.4 years). However, in the middle of the 1970s, it accounted only for 3% (i.e., 0.26 of 7.9 years). The decline in absolute figures (from 1.4 to 0.26 years) stems entirely from the decline in infant mortality, but the marginalization of its weight (from more than 90% to 3%) is accounted for by the great increase in excess male mortality at the other ages (Fig. 53–2).

Even reduced in this manner, the contribution of excess male mortality at less than 1 year of age remains higher than the contributions by the other age groups taken individually. The case of infant mortality is isolated, and it is the combination of the contributions of all of the contiguous age groups from roughly 50 to 75 years that now make up the greater part of the difference between mean life expectancies. The exceptional peak in the ratios of excess male mortality observed at approximately the age of 20 carries a weighting factor of only secondary importance because of the very low mortality at these ages.

2. Increasing Predominance of Environmental and Behavioral Factors

It is possible to measure with the same decomposition method (Pollard, 1988, 1990) the weight of the different medical causes of death in the difference in life expectancy between the sexes and in its worsening.

This was done 15 years ago for France, based on a reconstitution of mortality by cause that made it possible to follow its evolution from 1925 to 1978 (Vallin and Meslé, 1988). This approach allowed several assessments that, without being entirely conclusive, showed interesting trends (Vallin, 1993a). The fact that a large fraction of current excess male infant mortality is closely linked to hereditary or congenital processes substantiates the theory of a genetic origin for this difference. The greater susceptibility of boys to infectious pediatric diseases, remarkably stable for the whole period, does not contradict this idea. It is only because of its collapse that infectious pediatric mortality plays a moderating role in the aggravation in the difference in life expectancy between the sexes. Today and since the 1920s, the fight against infectious diseases seems to have been well managed in France, independently of the child's sex, and little boys are therefore almost constantly at a disadvantage due to their genetic handicap.

Findings were not the same at other ages. Although for a long time the role of infection in young girls masked their biologic advantage, the statistical disadvantage results more from their unfavorable social status than from any specific susceptibility to certain diseases. As soon as women's social status improved for any length of time and at all ages, infectious diseases resulted in excess male mortality. The latter situation goes well beyond the male fragility observed among babies. A more detailed analysis shows that at adult ages, susceptibility concerns mainly infectious respiratory diseases that are strongly linked to the consumption of tobacco and to industrial pollution, two exogenous factors that, until recently, affected men much more than women.

It might be said of degenerative diseases, which are the consequence of the body wearing out, that they affect men more because they are less resistant. However, it would be just as plausible to attribute the effect to a lifestyle that is more harmful to men's health. The second interpretation is more realistic insofar as excess mortality linked to this type of cause has greatly increased in the past 50 years.

The case of tumors, which play an extremely important role in increasing excess male mortality, is even clearer because tumors of the respiratory system closely linked to smoking and industrial pollution and tumors of the esophagus linked to alcoholism are first among the causes for excess male mortality. Behavioral and environmental factors clearly play a role.

The same factors lead to very strong excess male mortality rates linked to accidents, suicide, alcoholism, and cirrhosis of the liver, which are sufficiently strong that, despite the secondary part they play in overall mortality, they contribute substantially to the difference in life expectancy between the sexes and in its increase over time.

IV. FROM THE NEGATIVE EFFECTS OF MALE BEHAVIOR TO THE POSITIVE EFFECTS OF FEMALE BEHAVIOR

All of these observations perfectly coincided, at least at first sight, with the idea that was prevalent for a long time: Beyond a slight biologic advantage for females, excess male mortality and its spectacular increase in industrial countries resulted from the emergence of typically male "man-made diseases."[8] Work-related risks in industrial activity, alcoholism, smoking, and car accidents were the main factors contributing to excess male mortality. Their unprecedented expansion from the beginning of the 20th century, which was barely interrupted during World War II and which has continued to grow since the end of the 1940s, provided an obvious explanation for the continued increase in excess male mortality. However, this explanation has a limit.

1. They Smoke, They Drive, They Work—But Women Still Die at Older Ages

Although differences in behavior or living conditions are the main factors of excess male mortality, what can explain the continued increase at a time when, especially since the end of the 1960s, these differences have considerably declined?

In recent decades, most male and female behaviors have become greatly similar. Whereas smoking has been greatly reduced among men—the success is such that in England, for example, bronchial and lung cancer mortality has begun a spectacular decline (Meslé and Vallin, 1992)—it has greatly increased among women. In the same way, alcoholism, previously thought to be a very male condition, has declined in France since the beginning of the 1960s (Vallin and Meslé, 1988). Road accidents have been declining since the oil crisis in the 1970s, but they are increasingly less specifically associated with men because women now drive in almost equal numbers. Women's participation in economic activity has increased to a level close to that of men, which at the same time has perceptibly declined.

These changes already seemed rather likely to occur or were already clearly taking shape in the 1960s for the authors of life tables to assume in the construction

[8] "Man-made diseases" as described by Omran (1971).

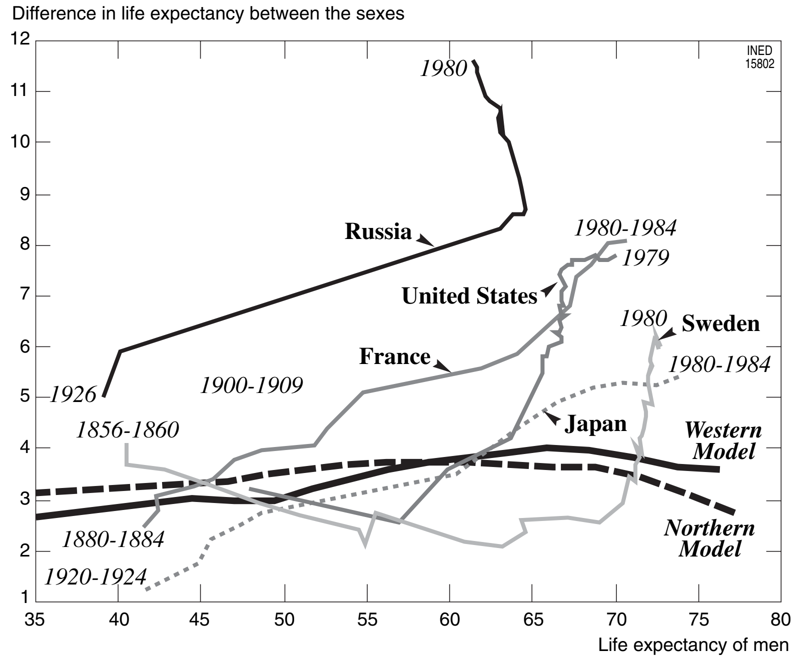


FIGURE 53–3 Correlation between the level of life expectancy and the difference between the sexes according to the models of Coale and Demeny and according to the real situation in some industrial countries from the end of the 19th century to the 1980s.

of their models that there would soon be a decline in the difference in life expectancy between the sexes and therefore a reversal in the constant trend toward an increase that had been observed until then (Coale and Demeny, 1966). The subsequent evolution entirely contradicted a projection that seemed certain at the time. The difference between the sexes continued to increase everywhere while life expectancy was reaching higher levels, and instead of slowing down, this trend tended to accelerate, as cruelly demonstrated in Figure 53–3 (Vallin, 1983). Despite contradiction of the facts, the authors of the tables did not modify their models when they republished them in the 1980s (Coale and Demeny, 1983).[9]

Although behaviors have become more similar for men and women, they are not identical. Men continue to smoke and drink more than women, they drive more often, and they have higher rates of economic activity. Although this may be enough to explain continued excess male mortality beyond the initial biologic advantage of women, it in no way explains the increase observed.

There has always been a tendency, concerning excess male mortality, to highlight the causes of death likely to increase mortality, in the belief that they are the result of social vices men are more likely to engage in. This approach does not acknowledge the fact that the exceptional increase in excess male mortality has emerged at the same time as the equally exceptional development of health care. Beyond the negative behavioral and environmental factors that have more adverse effects on men's health than on women's health, is it not likely that there is a more fundamental difference in lifestyle that enables women to benefit more from improved health care than men?

2. Being Female: An Art of Living?

a. Inherited Aspects

From the socially inferior status that was their lot for so long, women reaped more than disadvantages. Although they were treated as slaves and as means of production and reproduction, they were also reputed to be weak and fragile, and women benefited from a certain degree of protection. Machismo and politeness toward women have often gone hand in hand. Denied access to power by a patriarchal system, women were also preserved from the most dangerous tasks involved in war and hunting. Is it not customary to cry out in case of danger, "Women and children first!"? During the industrial age, to such customs were added many laws forbidding women access to certain dangerous jobs (in mines) or arduous work (night shifts),

[9] A correction has been published since, but it is only for the Western and Northern models (Coale and Guo, 1989, 1990).

limiting the length of their working day and protecting their fertility (e.g., monitoring of pregnancies, maternity leave). At the beginning of the century, they were almost excluded from many factory jobs and relegated to activities more "in keeping with their nature," such as domestic work. In the same way as for war and military service, they were for the most part spared from the inhuman working conditions imposed by heavy industry. It is also essentially because of their status and social pressures that women were denied the dangerous male privileges of consuming alcohol and tobacco or of driving automobiles, not to mention that of sexual freedom and its associated health risks (formerly syphilis and more recently AIDS).

Although mainly a thing of the past, this heritage of protection has not completely disappeared. In addition to having gained greater social equality and substantially reducing the risks of reproduction, women have managed to preserve some of the positive aspects of their former status in the form of a different approach to life and to societal interaction.

b. Female Form of Development

The fact that women are attaining equal status with men does not mean that female behavior is becoming increasingly male, nor does it signify that women's relation to modern society increasingly resembles that of men. Fundamental differences remain, some of which may explain why, until now, women have gained greater benefit from improved health care than men.

It is true that modern women have access to alcohol, tobacco, and cars and that excess male mortality linked to the causes of death directly related to these factors has tended to decline in recent years. Many of the consequences of closing this gap between behaviors remain to come insofar as some harmful effects emerge only later (e.g., those linked to smoking). It seems, however, that the way women behave in these areas is less dangerous than that of men. For example, among younger generations, the proportion of women who smoke is almost equal to and sometimes greater than that of men; however, the average number of cigarettes smoked each day by regular smokers is lower today (Blanc, 1985a, 1985b). In the same way, women drivers are on average more careful, and they drive more slowly and take fewer risks. The first aspect is probably minor compared with the two other characteristics.

Female participation in economic activity has been massive during the last three generations (Lévy and Labourie-Racapé, 1983), but it remains fundamentally different from that of men (Vallin, 1995). It is often pointed out that women have less access to the positions of highest responsibility. Less often mentioned is the fact that they are generally more qualified and that they often hold less degrading jobs. In France from the beginning of the 1980s, the census showed that 40% of jobs were held by women, and this proportion has remained stable since, as the 1996 employment survey revealed (INSEE, 1996). Women now participate almost as much as men in economic activity. However, their participation in the different professions is very unequal (INSEE, 1996; Vallin, 1995).

Jobs with the greatest share in decision-making are seldom held by women (in 1996, only 16% of corporate chief executive officers were women), and this is even more the case for technical jobs (female engineers: 13%), but the same applies for manual jobs (i.e., industry, crafts, and agriculture), whether qualified or unqualified (20%), especially if the jobs involve the responsibility of management (female foremen: 9%). Intermediate-level jobs are considered feminine (77%) in the ancillary professions of health and social care, in civil service (80%), and in companies and the service industry (83%). The proportion of these jobs compared with the total number of jobs held by women is huge (70%), whereas among men, manual jobs are prevalent (more than 40%).

I cannot help but relate this difference in the distribution of men and women in the workplace with the even greater difference observed concerning social inequality in death (Vallin, 1995). Among men, risks of death vary greatly according to socioprofessional category. In France, between senior executives and the professions on the one hand and manual workers on the other, the difference in life expectancy at 35 years of age is 9 years (Desplanques, 1976, 1984). Among women, the variation is very slight (Desplanques, 1984). Not only is female economic activity essentially concentrated in professions in which health risks are low (e.g., schoolteachers) or average (e.g., office workers), but even when they work in sectors that are more dangerous for men, women seem to enjoy better protection. Detailed analysis clearly reveals this trend (Vallin, 1995). The two most dangerous jobs in peacetime (i.e., policemen and drivers) are among the least often held by women (8% and 6%, respectively). Almost as active as men and even more so if domestic activity is taken into account, women take part in activities that generally are less harmful to health, and they manage their affairs in a fashion that is kinder to their health.

A third aspect to be highlighted here is that the attitude women generally have concerning their body, their health, and their lifestyle is very different from that of men. Here again we find that initially, this is a consequence of the reproductive function and of its evolution. At the same time that control of fertility lightened the weight of maternities, maternal and infant protection, gynecology, and obstetrics have mobilized an increasing share of medical resources, an area exclusively reserved for women. Development of this specific area of health care has done much more than compensate for the risks of pregnancy. It has contributed to improving overall female health, making women more accustomed than men to contacts with medical services and more concerned about health issues. Screening for tumors of the uterus during visits to gynecologists, the principal reason for the reduction in mortality from uterine cancer, is just one example of improved health care for women (Hatton *et al.*, 1985a, 1985b).

Women's special relation to health is not just a consequence of their reproductive function. For the human body, the culture of femininity is a very different constraint than that imposed by the exaltation of virility. At the risk of caricaturizing, it can be said that the search for beauty is in opposition with the search for power. In the first instance, the body must remain young and healthy as long as possible; in the second, the body is subjected to risk very early on. The fact remains that women seem to be much more attentive than men to their body and to their health needs. They use health services more often, and they often talk with the practitioner with greater ease and openness.

More inclined to temperance, practicing less harmful activities in a manner probably more respectful of their health, more inclined to care for their bodies and to preserve their lives, women have quite naturally drawn greater benefit from medical and social progress. Since the turning point of the 1960s, they have reinforced their advantage in all of these areas thanks to education. In France, for example, during the 1960s, the proportion of girls in a same birth cohort succeeding at the *baccalauréat* (end of secondary examination) exceeded that of boys (Lévy, 1981). Although fewer women than men have access to the most advanced levels of higher education, on average they have a higher level of general education—another major asset in their favor.

3. Women: Champions in the Second Phase of the Health Transition

Women had all the described assets in hand at the time of the health transition. During the 1960s, at the end of a spectacular convergence of all industrial countries toward the life expectancy of countries formerly more advanced (i.e., Japan, Russia, or the Mediterranean countries catching up with the countries of Northern and Western Europe), progress slowed down almost everywhere. The maximum benefits from the fight against infectious diseases had been attained thanks to the distribution of vaccines and antibiotics. Mortality due to these diseases, especially at very young ages, had become so low that a continuation of its decline was no longer likely to produce a significant increase in life expectancy. Cardiovascular diseases and cancer, against which medical science remained essentially powerless, ranked first among causes of death, far ahead of infectious diseases, and economic and social changes also had brought about an unprecedented increase in negative phenomena such as alcoholism, smoking, and road accidents. This explains the stagnation or even reversal in life expectancy observed in industrial countries during the 1960s, spawning pessimism about returning to improvement in the trend.

It is in such a context that Abdel Omran (1971) formulated his theory of epidemiologic transition, which brought humanity from "the age of pestilence and famine" to that of "degenerative diseases and man-made diseases" (see Chapter 57). Because the huge benefit of the victory over infectious diseases was a thing of the past, there was no more hope of a return to sustained improvement in life expectancy. However, this improvement is a reality. The 1960s were nothing more than a transition crisis from a first phase of improvement, which Abdel Oman called *epidemiologic transition*, to the second phase of the *health transition* (Vallin, 1993b). The health transition was founded on the success made in treating and preventing man-made diseases and cardiovascular diseases and cancers. In France, improvements in life expectancy resumed in the 1970s (Meslé and Vallin, 1993). One after the other, increases in man-made diseases were curbed. The trend began to reverse from 1965 for alcoholism, from 1974 for road accidents, and near the end of the 1980s for lung cancer, but it was the spectacular decline in cardiovascular diseases that was key to this new phase of improvement. What happened in France also happened more or less rapidly in most Western industrialized countries. In Japan, the transition was so rapid and so successful that the improvement in life expectancy did not even experience a slowdown in the 1960s. Eastern countries, particularly Russia, have failed to introduce this transition, and life expectancy has been in stagnation or decline since the 1960s (Shkolnikov *et al.*, 1995, 1996).

This East-West divergence has been much written about, and the causes of the failure in the former Communist countries are as complex as the reasons for the success of Western countries. However, one of the fundamental reasons for the difference lies in the fact that whereas the fight against infectious diseases might have resulted essentially from a policy of dissemination for the benefit of the greatest number of simple and efficient techniques such as vaccines and antibiotics, success in the area of cardiovascular diseases or cancer requires much more sophisticated medical techniques and the active participation of each individual in the preservation of his or her own health. Man-made diseases can be checked only by individual awareness and changes in behavior. The transition to this new health strategy is naturally easier for societies founded on individual responsibility than for countries where for many decades the state was omnipresent and discouraged individual initiatives.

Mutatis mutandis, the acceleration of the divergence observed during the 1960s and 1970s between female and male life expectancies, can therefore be explained. For all of the reasons previously given, women were initially much better prepared than men to succeed in the second phase of the health transition. Whereas the crisis of the 1960s in Western countries was marked in almost all of the industrialized nations by a substantial slowdown, stagnation, or even decline in male life expectancy, women's life expectancy everywhere continued to increase, sometimes without the rate even slowing down. In many cases, the decline in cardiovascular mortality for women seamlessly took over from the decline in infectious diseases, whereas men needed a much longer period of adaptation. Subsequently, the 1970s and 1980s were marked by a much faster decline in cardiovascular mortality among women than among men (Meslé and Vallin, 1992). Women took better advantage of this new phase in the transition to improved health much more quickly than men.

With this background, it is easier to understand how the difference in life expectancy between men and women could continue to grow at the same time the behaviors of men and women came to resemble one another as they increasingly shared "masculine vices." The more fundamental virtues of women enabled them to embark much faster and much more effectively on the second phase of the health transition.

V. WHEN GEOGRAPHY CONTRADICTS HISTORY BUT CONFIRMS THE PREVIOUS HYPOTHESIS

It is the dual aspect of the role of behaviors in mortality differences (i.e., the harmfulness of certain practices and the capacity to take advantage of improvements in health care) that makes it possible to elucidate a third paradox concerning excess male mortality. Over the centuries, as life expectancy has increased, so too has the mortality difference between the sexes. However, when considering the geographic variations in a country such as France (and many other countries), it is clear one observes that the greater the life expectancy, the smaller is the difference between the sexes (Vallin, 1990).

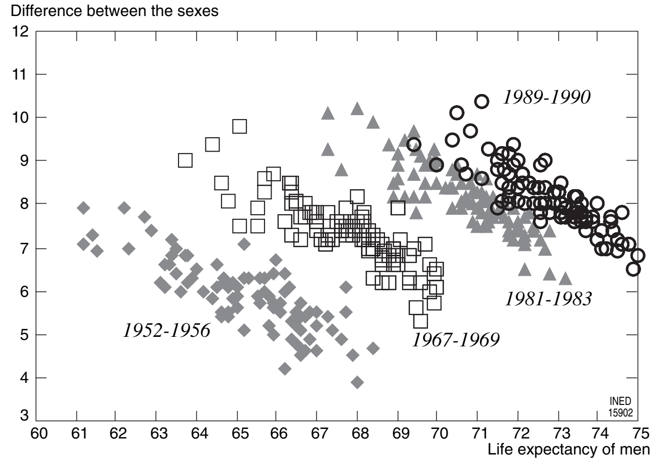


FIGURE 53–4 Correlation between the variations in male life expectancy and those of the difference in life expectancy observed between the French departments in 1952–1956, 1967–1969, 1981–1983, and 1989–1990.

1. The More Life Expectancy Increases, the More the Difference Increases, but the Greater the Life Expectancy, the Smaller the Difference

Figures 53–4 and 53–5 illustrate the paradox as it may be observed in France: Variations between *départements* show a very strong negative relation between life expectancy at birth among men and the difference in life expectancy between the sexes[10] (see Fig. 53–4), whereas the evolution over a century (see Fig. 53–5) reveals a positive correlation between the two indicators that is just as clear.

This phenomenon, which also is observed in many other countries (Vallin, 1990), is based on the fact that geographic variations are mainly linked to the action of different types of male behavior harmful to health (e.g., alcoholism, especially in the case of France) that shorten male life expectancy and increase the difference between men and women, whereas the evolution over time is essentially linked to improved health that,

[10] As demonstrated in Figure 53–4, this phenomenon is constant over time. The increase in life expectancy shifts the cluster of points without modifying its orientation.

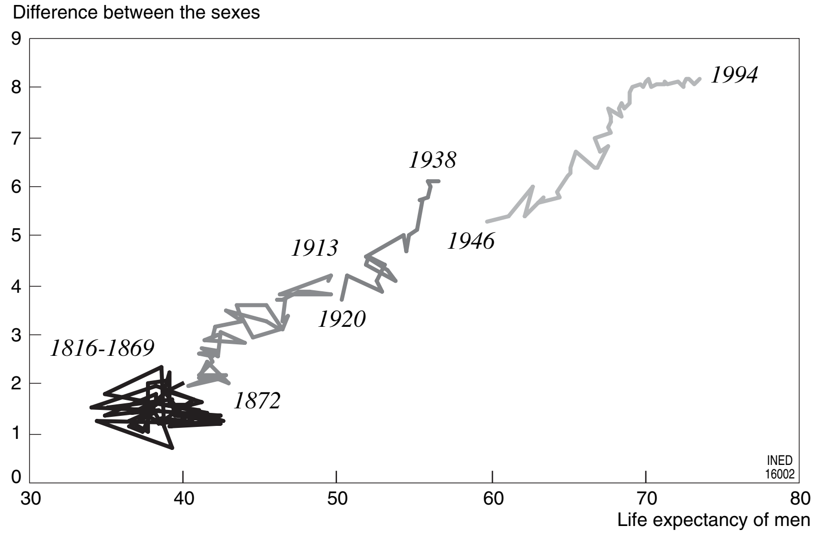


FIGURE 53–5 Correlation between annual variations in male life expectancy and those of the difference in life expectancy between the sexes observed in France since the beginning of the 19th century.

although it increases life expectancy of both sexes, increases the life expectancy of women more because they are able to draw greater benefit from it.

Social inequality in mortality exhibits the same sort of paradox. In past decades, the difference between manual workers and senior executives increased as life expectancy increased. Regional variations also show an inverse relationship because regions where excess mortality among manual workers was greatest were those where life expectancy was lowest (Vallin, 1995). The reason for this is that for social differences and for differences between the sexes, the same causes produce the same effects. Whereas regional differences are very sensitive to behaviors harmful to health (i.e., more widespread among manual workers than among executives), evolution over time is much more sensitive to the capacity to benefit from improved health (i.e., more widespread among senior executives).

This third paradox takes on a special characteristic in Eastern countries, especially in Russia (Meslé and Vallin, 1998). Although geographic variations show that in these countries, as almost everywhere else, a negative correlation exists between the life expectancy of men and the difference between the sexes, in Russia, Hungary, and Poland, the historical evolution only demonstrates a positive correlation until the 1960s. Since then, the correlation has been reversed, with the difference between the sexes continuing to widen while increased life expectancy has stalled or declined. In Russia and elsewhere before the crisis of the 1960s, geographic variations dominated by man-made diseases specifically affecting men contrasted with the historical effects of rapid progress in health conditions, which were more favorable to women. Eastern countries—Russia in particular—completely missed passage to the second phase of the health transition, and their inability to curb the spread of man-made diseases or to fight cardiovascular diseases has resulted in a stagnation or decline in life expectancy and in an increase in the difference in life expectancies between men and women. In Russia, for example, the difference in life expectancy between men and women went from 8.7 years in 1964 to 13.6 years in 1994, whereas male life expectancy fell in the same period from 64.6 years to 57.7 years.

2. Different Roles Played by Different Causes of Excess Male Mortality

The third paradox, far from confusing the issue, substantiates the explanations and examples previously presented. Beyond the biologic differences between the sexes, excess male mortality correlates with two types of causes of death. Certain diseases that may be designated as man-made diseases, whether they are related to individual behaviors harmful to health (e.g., alcoholism, lung cancer, road accidents, suicide), or to risks associated with certain economic activities (e.g., accidents and diseases linked to work), affect in a highly discriminatory manner certain categories of population (i.e., men, when distinguishing the sex, and manual jobs when distinguishing economic activities), and these diseases make up the major part of the geographic variations in mortality within a given country. The evolution of mortality over time is affected more by other causes; in the past, these deaths were caused by infectious diseases, and currently, mortality is affected more by cardiovascular diseases and cancers. Because infectious diseases were initially more unfavorable to women, especially at younger ages and at reproductive ages, their decline since the end of the 19th century had already increased the difference in life expectancy between the sexes. Success since the 1970s in the fight against cardiovascular diseases and to a more modest degree against cancers has mainly benefited women, resulting in an unprecedented difference in life expectancy between the sexes despite a homogenization of behaviors and a reduction in the differences linked to man-made diseases. In countries such as Russia, where man-made diseases (i.e., alcoholism and violence) have a greater impact than anywhere else and where passage to the second phase of the health transition has not yet occurred, both phenomena penalize men much more than women and push excess male mortality to new heights, resulting in a negative reconciliation of history and geography.

VI. THE FOURTH PARADOX: WILL EXACERBATION OF EXCESS MALE MORTALITY ACCOMPANY A REDUCTION IN DIFFERENCES IN LIFE EXPECTANCY?

It would be rash to conclude that an endless increase in the difference in life expectancy between the sexes is inevitable. A fourth paradox suggests that the phenomenon will soon halt. The fourth paradox is that an extreme exacerbation of excess male mortality may accompany a reduction in the difference in life expectancy between the sexes. In France, this is what was brought to light by extrapolating over a period of 100 years recent mortality trends by age and by sex (Vallin and Meslé, 1989). The decline in female mortality was at all ages faster during the last 20 years than that in male mortality, and it is not surprising that the extrapolation led to higher and sometimes quite extravagant excess male mortality ratios at all ages. In this analysis, the ratio of male rates to female rates would exceed 10 against 1 at approximately 50 years of age and would be close to 15 against 1 at approximately 20 years of age, compared with the current figures of 2.5 against 1 and 3.5 against 1, respectively. Despite this incredible increase in excess male mortality, the difference in life expectancy between the sexes would quite quickly reach a peak of 9 years (without ever reaching the current real level in Russia) and drop back at the end of the projection to 7 years, definitely less than the current 8 years. This paradox is explained by the fact that the extrapolation brings the female rates to levels so much lower than male rates that they would require a stronger relative reduction to produce the same effect on life expectancy. Male rates can therefore continue to decline less quickly than female rates (and to worsen excess male mortality) while producing a progression of life expectancy markedly higher than that obtained by females. Except if the extreme age limit were to be pushed back (an assumption that is rarely made in this type of projection) and even if female mortality continued to decline faster than male mortality, the difference in life expectancy between the sexes could no longer increase, and it would tend to decline. The same projection in a country like Russia leads to a completely different result. With increasing male mortality, there is no close limit to the increase in the difference in life expectancies. The facts speak for themselves because the difference is already more than 13 years.

This fourth paradox is not the only reason to hope for a decline in the difference in life expectancy between the sexes, nor is it the main reason. English-speaking countries and Scandinavian countries, which are often ahead in terms of improved health, especially when health status relies on a massive effort in terms of individual prevention, have very clearly shown for a number of years that the difference in life expectancy between the sexes can be reduced by new increases in life expectancy (Fig. 53–6).

In Australia, Sweden, and Norway, the increase in male life expectancy from 70 to almost 75 years during the 1980s and 1990s led to a reduction in the difference between the sexes of approximately 1 year of life expectancy. The same phenomenon has been observed in the United States, albeit at a lower level of life expectancy. Male behaviors harmful to health play a diminishing role, and in these countries, men have partially succeeded in copying the positive behaviors of women in terms of disease prevention and accessing health care. There is no reason why the experience of English-speaking countries and Scandinavian countries should not become generalized, enabling men in most places to eventually regain some of the ground they lost during the 20th century.

At the beginning of the 1960s, to build their model life tables, Ansley Coale and Paul Demeny, made (and lost) the bet that women would increasingly imitate men. Maybe it would be enough today to bet the other way to win and say that men will increasingly imitate women. Will they go so far as to reduce their disadvantage to the strict minimum imposed by biology? That is more difficult to say.

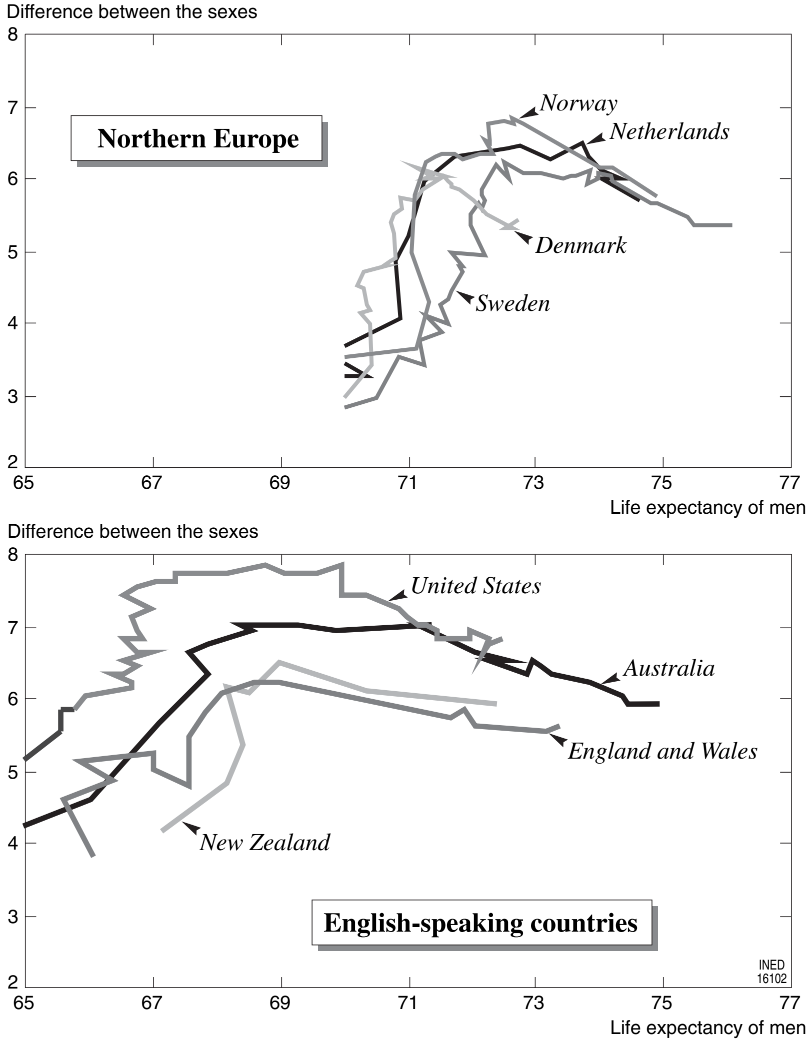


FIGURE 53–6 Relation between the level of life expectancy and the difference in life expectancy between the sexes during recent decades in some English-speaking and Northern European countries.

Acknowledgment

This chapter was translated from the French version by Paul Belle.

References

Adlakha Arjun and Suchindran C. M., 1985. Infant and child mortality in Middle Eastern countries, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference, Florence* 1985, vol. 2, p. 367–376. Liége, Ordina éditions, 456 p.

Afzal Mohammad, 1973. 1972 census: population, expected and actual, *The Pakistan Development Review*, vol. 12(2), p. 123–133.

Armengaud André, 1973. L'attitude de la société à l'égard de l'enfant au XIX[e] siècle, *Annales de démographie historique*, p. 310–311.

Bean Lee and Khan Mazihur Rahman, 1967. *Mortality patterns in Pakistan*. Karachi, Pakistan Institute of Development Economics, 8 p.

Blanc Michel, 1985a. Les effets à long terme des programmes d'intervention contre le tabagisme: application à la France, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 238–256. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 542 p. (Travaux et documents, cahier no. 108.)

Blanc Michel, 1985b. The long-term effect of anti-smoking programmes, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 241–260. Liège, Ordina édition, 558 p.

Blayo Yves, 1975. La mortalité en France de 1740 à 1829, *Population*, vol. 30, special issue *Historical demography*, p. 123–142.

Bourgeois-Pichat Jean, 1951a. La mesure de la mortalité infantile. I. Principes et méthodes, *Population*, vol. 6(2), p. 233–248.

Bourgeois-Pichat Jean, 1951b. La mesure de la mortalité infantile. II. Les causes de décès, *Population*, vol. 6(3), p. 459–480.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Bourgeois-Pichat Jean, 1978. Future outlook for mortality decline in the world, *Population Bulletin of the United Nations*, vol. 11, p. 12–41.

CALOT Gérard and CASELLI Graziella, 1988. *La mortalité en Chine d'après le recensement de 1982*. I. *Analyse selon le sexe et l'âge au niveau national et provincial*. II. *Tables de mortalité par provinces*. Paris, Institut national d'etudes demographiques, Union internationale pour l'etude scientifique de la population, 72 + 112 p. (Dossiers et recherches no. 16.)

CALOT Gérard and CASELLI Graziella, 1989. La mortalité en Chine d'après le recensement de 1982. Analyse selon le sexe et l'âge au niveau national et provincial, *Population*, vol. 44(4–5), p. 841–872.

CHEN Lincoln C., HUQ Embadul, and D'SOUZA Stan, 1981. Sex bias in the allocation of food and health care in rural Bangladesh, *Population and Development Review*, vol. 7(1), p. 55–70.

CHESNAIS Jean-Claude, 1998. La sous-mortalité féminine: de la sociologie à la biologie, *in*: Association internationale des démographes de langue française (ed.), *Morbidité, mortalité: problèmes de mesure, facteurs d'évolution, essai de prospective*, p. 11–18. Paris, Presses universitaires de France, 737 p. (Proceedings of the international seminar in Sinaia, September 3–6, 1996.)

COALE Ansley and DEMENY Paul, 1966. *Regional model life tables and stable populations*. Princeton, Princeton University Press, 871 p.

COALE Ansley and DEMENY Paul, 1983. *Regional model life tables and stable populations*, 2nd ed. New York, Academic Press, 496 p.

COALE Ansley and GUO Guang, 1989. Revised regional life tables at very low levels of mortality, *Population Index*, vol. 55(4), p. 26–41.

COALE Ansley and GUO Guang, 1990. New regional life tables at high expectation of life, *Population Index*, vol. 56(1), p. 613–643.

CONTI Susanna, FARCHI Gino, MASOCCO Maria, MINELLI Giada, TOCCACELI Virgilia, and VICHI Monica, 2003. Gender differentials in life expectancy in Italy, *European Journal of Epidemiology*, vol. 18(2), p. 107–112.

DAS GUPTA Monica, 1987. Selective determination against female children in rural Bangladesh, *Population and Development Review*, vol. 13(1), p. 77–100.

DESPLANQUES Guy, 1976. *La mortalité des adultes suivant le milieu social 1955–1971*. Paris, National Institute of Statistics and Economic Studies, 148 p. (Les collections de l'INSEE, série D, no. 44.)

DESPLANQUES Guy, 1984. L'inégalité sociale devant la mort. *Économie et Statistique*, January (162), p. 29–50.

DESPLANQUES Guy, 1985. *La mortalité des adultes. Résultats de deux études longitudinales (période 1955–1980)*. Paris, National Institute of Statistics and Economic Studies, 212 p. (Les collections de l'INSEE, Série D, no. 102.)

FEDERICI Nora, 1950. La mortalità differenziale dei due sessi e le sue possibili cause, *Statistica*, vol. 10(3), p. 274–320.

FEDERICI Nora, OPPENHEIM-MASON Karen, and SOGNER Solvi (eds.), 1993. *Women's position and demographic change*. Oxford, Clarendon Press, 368 p.

FERRY Nicole, 1979. *La femme et l'enfant en milieu rural algérien: étude sociologique et médicale de la maternité et du premier âge*. Doctoral thesis in medicine. Lille, Université de Lille, 120 p.

GBENYON Kuakuvi and LOCOH Thérèse, 1989. Différences de mortalité entre garçons et filles, *in*: Gilles Pison, Étienne van de Walle, and Mpembélé Sala-Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara*, p. 221–244. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 446 p. (Travaux et documents, cahier no. 124.)

GBENYON Kuakuvi and LOCOH Thérèse, 1992. Mortality differences in childhood by sex in sub-Saharan Africa, *in*: Étienne van de Walle, Gilles Pison, and Mpembélé Sala-Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 230–252. International Studies in Demography. Oxford, Clarendon Press, 450 p.

HAFFAD Tahar, 1984. *Les différences de mortalité selon le sexe et leurs conséquences*. Doctoral thesis. Paris, l'Ecole des Hautes études en Sciences Sociales, 494 p.

HATTON Françoise, FLAMANT Robert, BOUVIER-COLLE Marie-Hélène, and MAUJOL Léone, 1985. La lutte contre la mortalité cancéreuse, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 217–238. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 562 p. (Travaux et documents, cahier no. 108.)

HATTON Françoise, FLAMENT Robert, BOUVIER-COLLE Marie-Hélène and MAUJOL Léone, 1985. The fight against cancer, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 217–240. Liège, Ordina Edition, 558 p.

HERDAN Gustave, 1952. Causes of excess male mortality in man, *Acta Genetica et Statistica Medica*, vol. 3(4), p. 351–375.

HOLLIDAY Robin, 1987. Ageing X-chromosome reactivation, *Nature*, vol. 327(6124), p. 661–662.

INSEE, 1996. Enquête sur l'emploi de 1996. Résultats détaillés. *INSEE Résultats*, vol. 107–108. Paris, National Institute of Statistics and Economic Studies, 206 p.

LEGOUVÉ Ernest, 1842. *Histoire morale des femmes*. Paris, Librairie Académique Didier, 460 p.

LENZ Fritz, 1940. Die Ubersterblichkeit der Knaben im lichte Erblichkeitslehre, *Archiv für Hygiene*, vol. 93, p. 126–150.

LEVY Michel-Louis, 1981. Garçons et filles à l'école, *Population et Sociétés*, no. 151, p. 1–3.

LEVY Michel-Louis and LABOURIE-RACAPÉ Annie, 1983. Le salariat féminin en perspective. *Population et Sociétés*, no. 165, p. 1–3.

LOPEZ Alan and RUZICKA Lado (ed.), 1983. *Sex differentials in mortality: trends, determinants and consequences*. Canberra, Australian National University, Demography Department, 498 p.

MADIGAN Francis C., 1957. Are sex mortality differentials biologically caused? *The Milbank Memorial Fund Quarterly*, vol. 35(2), p. 202–223.

MAGNANI Mauro and ACCORSI Augusto, 1993. The female longevity phenomenon. Hypotheses on some molecular and cell biology aspects, *Mechanisms of Ageing Development*, vol. 72(2), p. 89–95.

MELSÉ France, 1995. La mortalité en France: le recul se poursuit, *Population*, vol. 50(3), p. 743–776.

MELSÉ France, 2004. Écart d'espérance de vie entre les sexes: les causes du recul de l'avantage féminin, *Revue d'épidémiologie et de santé publique*, vol. 52(4), p. 333–352.

MELSÉ France and VALLIN Jacques, 1992. *Évolution de la mortalité par cancer et par maladies cardio-vasculaires en Europe depuis 1950*. Paris, Institut national d'etudes demographiques, 46 p. (Dossiers et recherches no. 34.)

MELSÉ France and VALLIN Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference. Montreal 1993*, vol. 2, p. 365–382. Liège, Ordina éditions, 493 p.

MELSÉ France and VALLIN Jacques, 1998. Évolution et variations géographiques de la surmortalité masculine: du paradoxe français à la logique russe, *Population*, vol. 53(4–5), p. 1079–1102.

MILLER Barbara D., 1996. Changing equality: the politics of sex selective abortion in Asia, *in*: International Union for the Scientific Study of Population Committee on Anthropological Demography and Centre for Developement Studies (ed.), *Seminar on socio-cultural and political aspects of abortion in a changing world, Kovalam, Trivandrum, India, March 25–28, 1996*, p. 1–15. Liège, Union Internationale pour l'etude scientifique de la population.

Nygaard Eric, Madsen-Gleerup Annette, and Christensen H., 1990. Endurance capacity and longevity in women, *Health Care Women International*, vol. 11(1), p. 1–10.

Omran Abdel R., 1971. The epidemiologic transition: a theory of the epidemiology of population change, *The Milbank Memorial Fund Quarterly*, vol. 49(4), p. 509–538.

Pollard John H., 1988. Causes de décès et espérance de vie: quelques comparaisons internationales, *in*: Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Mesure et analyse de la mortalité: nouvelles approches*, p. 291–313. Paris, Institut National d'Etudes Demographiques, Presses Universitaires de France, 458 p. (Travaux et documents, cahier no. 119.)

Pollard John H., 1990. Cause of death and expectation of life: some international comparisons,*in*: Jacques Vallin, Stan D'Souza, and Alberto Palloni (eds.), *Measurement and analysis of mortality: new approaches*, p. 269–292. International Studies in Demography. Oxford, Clarendon Press, 416 p.

Potter Joseph E. and Volpp Letitia P., 1993. Sex differentials in adult mortality in less developed countries: the evidence and its explanation, *in*: Nora Federici, Karen Oppenheim-Mason, and Solvi Sogner (eds.), *Women's position and demographic change*, p. 140–161. Oxford, Clarendon Press, 1993, 368 p.

Pressat Roland, 1973. Surmortalité biologique et surmortalité sociale, *Revue Française de Sociologie*, vol. 14, special issue, p. 103–110.

Pressat Roland, 1981. *Perspectives de réduction de la surmortalité masculine dans les pays ayant une faible mortalité.* Canberra, The Australian National University, 15 p. (Paper presented at the meeting on Sex Differentials in Mortality: Trends, Determinants, and Consequences.)

Raghavachari S., Natarajan K. S., Biswas A. K., and Bawa S. S., 1974. *The population of India.* New Delhi, Office of Registrar General and Census Commisioner. Paris, Commission for International Cooperation in National Research in Demography, 152 p.

Ryan Sheila M., Goldberger Ary L., Pincus Steven M., Mietus Joseph, and Lipsitz Lewis A., 1994. Gender and age related differences in heart rate dynamics, *Journal of the American College of Cardiology*, vol. 24(7), p. 1700–1707.

Seely Stephen, 1990. The gender gap: why do women live longer than men? *International Journal of Cardiology*, vol. 29(2), p. 113–119.

Shapiro Sam, 1954. The influence of weight, sex, and plurality on neonatal loss in the United States, *American Journal of Public Health and the Nation's Health*, vol. 44, p. 1142–1153.

Shkolnikov Vladimir, Melsé France, and Vallin Jacques, 1995. La crise sanitaire en Russie. I. Tendances récentes de l'espérance de vie et des causes de décès de 1970 à 1993. II. évolution des causes de décès: comparaison avec la France et l'Angleterre (1970–1993), *Population*, vol. 50(4–5), p. 907–943, 945–982. (Published in English in *Population, an English Selection*, 1995, vol. 8.)

Shkolnikov Vladimir, Melsé France, and Vallin Jacques, 1996. Health crisis in Russia. I. Recent trends in life expectancy and causes of death from 1970 to 1993. II. Changes in causes of death: a comparison with France and England and Wales (1970 to 1993), *Population, an English Selection*, vol. 8, p. 123–154, 155–190.

Sri Lanka, Department of Census and Statistics, 1974. *The population of Sri Lanka*. Colombo, Department of census and statistics, 128 p.

Tabutin Dominique, 1978. La surmortalité féminine en Europe avant 1940, *Population*, vol. 34(1), p. 121–148.

Trovato Frank and Lalu N. M., 2003. Contribution of cause-specific mortality to changing sex differences in life expectancy: seven nations case study, *Social Biology*, vol. 45(1–2), p. 1–20.

United Nations, 2001. *World population prospects: the 2000 revision, vol. 1. Comprehensive tables.* New York, United Nations, Population Division, 746 p.

Vallin Jacques, 1979. Un fait social, la surmortalité des petites filles en Algérie, *in*: AMEP (ed.), *Actes du III^e colloque de démographie maghrébine, Tunis, Avril 24–28 , 1978*, vol. 1, p. 189–215. Tunis, Association Maghrebine pour l'étude de la Population, 2 vol., 625 + 641 p.

Vallin Jacques, 1983. Sex patterns of mortality: a comparative study of model life tables and actual situations with special reference to the cases of Algeria and France, *in*: Alan Lopez and Lado Ruzicka (eds.), *Sex differentials in mortality. Trends, determinants and consequences*, p. 443–476. Canberra, Australian National University, 498 p.

Vallin Jacques, 1990. Quand les variations géographiques de la surmortalité masculine contredisent son évolution dans le temps, *Espace, Populations, Sociétés*, vol. 3, p. 467–478.

Vallin Jacques, 1993a. Social change and mortality decline: women's advantage regained or achieved? *in*: Nora Federici, Karen Oppenheim-Mason, and Solvi Sogner (eds.), *Women's position and demographic change*, p. 190–212. International Studies in Population. Oxford, Clarendon Press, 368 p.

Vallin Jacques, 1993b. Life expectancy: past, present, and future possibilities, *in*: Jean-Marie Robine, Colin Mathers, Margaret Bone, and Isabelle Romieu (eds.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*, p. 63–77. Paris, John Libbey, Institut National de la Santé et de la Recherche Médicale, 388 p. (Colloque INSERM, no. 226.)

Vallin Jacques, 1995. Can sex differentials in mortality be explained by socioeconomic differentials? *in*: Alan Lopez, Graziella Caselli, and Tapani Valkonen (eds.), *Adult mortality in developed countries, from description to explanation*, p. 179–200. International Studies in Population. Oxford, Clarendon Press, 362 p.

Vallin Jacques and Lery Alain, 1975. *Essai d'estimation de la surfécondité consécutive au décès d'un enfant en bas âge*. Paper presented at the Commission for International Cooperation in National Research in Demography seminar on infant mortality and its relation to fertility, Bangkok, May 1975. Paris, CICRED, 1975, p. 119–137.yy

Vallin Jacques and Lery Alain, 1978. Estimating the increase in fertility consecutive to the death of a young child, *in*: Samuel H. Preston (ed.), *The effects of infant and child mortality on fertility*, p. 69–90. New York, Academic Press, 262 p.

Vallin Jacques and Melsé France, 1988. *Les causes de décès en France de 1925 à 1978.* Paris, Institut National d'Etudes Demographiques, Presses Universitaires de France, 608 p + 7 volumes of annexes. (Travaux et documents, cahier no. 115.)

Vallin Jacques and Meslé France, 1989. À long terme, l'écart d'espérance de vie entre hommes et femmes devrait diminuer, *Population*, vol. 44(6), p. 1244–1251.

Yusuf Farhat, 1967. Abridged life table for the Pakistan and its provinces 1962–1964, *in*: International Union for the Scientific Study of Population (ed.), *Contributed papers. Sydney conference, August 21–25, 1967*, p. 533–541. Liège, International Union for the Scientific Study of Population, 1100 p.

Zhang Wansong *et al*., 1983. Yinger xingbili shitiao yao qieshi jiuzheng [Effective measures have to be taken against the anomalously high sex ratio at birth]. *Shesshui* [Society], 1983, no. 2.

C H A P T E R

54

Social Inequalities in Mortality

TAPANI VALKONEN

Département de Sociologie, Université de Helsinki, Helsinki, Finland

Persons in lower socioeconomic positions die on average younger than those in higher positions. This social inequality has been observed in almost all studies of different populations and using different indicators of socioeconomic position, such as social or occupational class, socioeconomic status, educational attainment, income, and housing characteristics.

Research on social inequalities in mortality has long traditions. In his early 20th century monumental work, *Die Lehre von der Mortalität und Morbidität*, Harald Westergaard (1901) reviewed a large number of studies on mortality by wealth, social class, and occupation carried out in Europe in the 19th century. The classic study on the length of life by Louis Dublin, Alfred Lotka, and Mortimer Spiegelman (1949) dealt with social differences in mortality on the basis of British and North American studies. Antonovski (1967) wrote a review of the studies carried out on social classes, life expectancy, and mortality. According to him:

> Despite the multiplicity of methods and indices used in the 30-odd studies cited, and despite the variegated populations surveyed, the inescapable conclusion is that class influences one's chance of staying alive. Almost without exception, the evidence shows that classes differ on mortality rates.

Research on social inequalities in mortality attempts to answer several questions. To what extent are there differences in the length of life and in death rates between social groups? What causes the existence of such differences? Are these differences lessening or rather increasing? How and why does the size of the differences change? Does the size of the differences vary from one country to another?

There are several reasons for studying these questions. They are important from the point of view of demographic research, because socioeconomic differences are often more discriminatory than other differences between sub-groups of the population. Studies contribute to a better understanding of the determinants of mortality levels and trends. For epidemiologists, social mortality differences provide clues to the causes of diseases. Mortality differences are also important from a social and health policy perspective, because they indicate inequalities in health and well-being.

I. NATIONAL STUDIES ON SOCIAL MORTALITY DIFFERENCES AND METHODS USED

Statistics on mortality by social category do not usually feature in regular statistics. The regular exploitation since 1911 of the British data published by the Registrar General since 1851 provide the only regular time series of mortality by occupation and social class (Fox, 1979; OPCS, 1978). The death rates in these statistics are calculated by relating the deaths, classified according to the last occupation recorded on the death certificates, to the population classified in the same way on the individual census forms. Deaths are usually grouped in a period of 3 to 5 years focused on the census. Social class is based on the classification of occupations into six categories. The three classes ranking highest in the social scale include nonmanual occupations at different levels, and the others include manual occupations.

Figure 54–1 shows age-standardized death rates for men between the ages of 15 and 64 by social class in England and Wales between 1970 and 1972. The death rates are based on two sources; deaths registered between 1970 and 1972 by the age and social class of the deceased person provide the numerator, and a 10% random sample of the population recorded in the 1971 census (mean population) by age and social class provides the denominator.

Mortality increases with decreasing social class. The age-standardized mortality of the 15- to 64-year-old class V (i.e., unskilled manual workers) is 80% higher than that of class I (i.e., professional).

This classic method has been used in other countries as well as in the United Kingdom (Mielck and do Rosário Giraldes, 1993). In many countries, however, it cannot be used, because the death certificate does not indicate the social category of the deceased person. Another serious limitation is that the results are likely to be biased, because the information on the occupation of decedents and persons in the risk category comes from different sources. Even though the same question is asked in the census and on the death registration form, the accuracy of the information may differ. There is evidence that this bias (often called numerator/denominator bias) can be substantial (Lévy and Vallin, 1981; Fox 1979). Because of this increased bias, the results on social mortality presented in the Registrar General's Decennial Supplement for mortality in the 1981 census (OPCS, 1986) are considered to be less reliable than those preceding them (Thatcher, 1986).

To avoid problems associated with the use of unlinked records, other methods are employed. These are usually based on the linkage of death records registered by the civil authority with population records. As early as 1865, Chapin (Antonovski, 1967) had compared death rates among taxpayers and nontaxpayers using linked records. He assigned every person who died in 1865 in Providence, Rhode Island, to the taxpayer or nontaxpayer group on the basis of census and tax records, and he calculated the death rates for both groups. He found the rates for the nontaxpayers to be more than double those of the taxpayers.

The idea of linking individual death records to census records was applied on a large scale for the study on differential mortality in the United States carried out by Evelyn Kitagawa and Philip Hauser (1973). About 340,000 of the deaths that occurred during the period of May to the end of August 1960 were matched with the 1960 census records to obtain social and economic characteristics of decedents as reported in the 1960 census. Death rates were calculated by occupation, level of education, and income. According to the investigators, the collected education data are more reliable indicators than those concerning occupation and income. The approach of death does change the occupational system and income

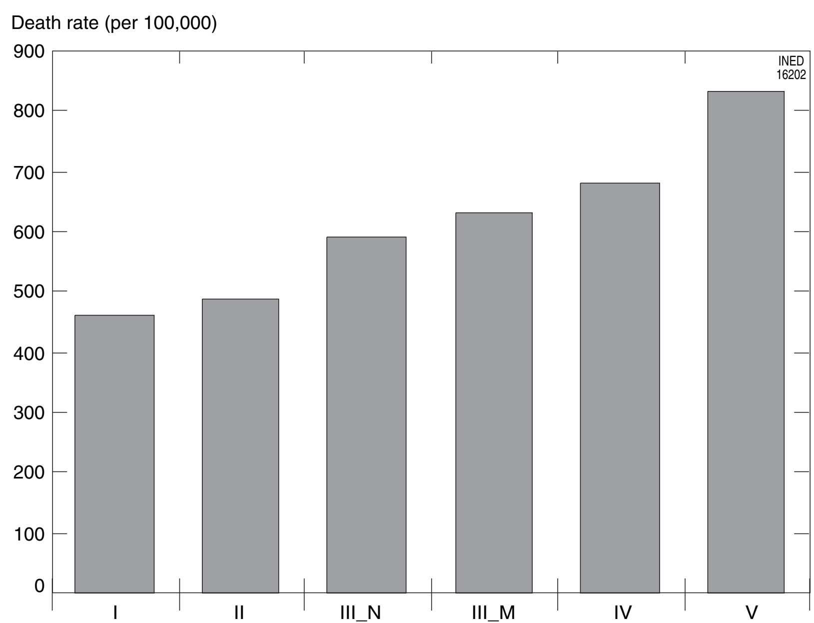


FIGURE 54–1 Age-standardized death rates (per 100 000) by social class of men between the ages of 15 and 64 years in England and Wales, 1970 to 1972.

situation, but the educational level is unaffected by it. As anticipated by the authors of the study, mortality varied inversely with level of education. For example, among white persons between the ages of 25 and 64 years, men with 4 years or more of college education had 30% lower mortality than the whole age group. Among women, the difference was 22%.

Another version of linking records back to census records involves following a sample of persons from the census, as in the longitudinal prospective studies common in epidemiology. An early example of this approach was the study carried out in France by the National Institute of Statistics and Economic Studies (INSEE), in which a sample of 800,000 persons from the 1954 census was followed for mortality (Desplanques 1976). The data made it possible to measure mortality by social-occupational group, level of education, and marital status. Another longitudinal study of 800,000 persons from the census was undertaken in 1975 (Desplanques, 1984). Table 54–1 shows the probabilities of death for men between the ages of 35 and 60 years during first 5 years after each of these two studies.

Results for the two periods show the expected gradients: lowest mortality in the group for administrators and professionals and highest mortality among unskilled workers. The ratio between the two extreme social categories was 2.8 : 1 between 1975 and 1980.

The ranking of the social-occupational categories by mortality remained almost constant during the 20-year period separating these two studies. Nevertheless,

TABLE 54–1 Probabilities of Dying in France for Men between the Ages of 35 and 60 Years by the Main Social-Occupational Categories for 1955 to 1959 and for 1975 to 1980

Social-occupational category	Probability of dying (%)	
	1955–1959	1975–1980
Administrative and professional groups	12.7	9.1
Teachers	12.0	10.1
Foremen	15.9	11.6
Technicians	15.0	11.8
Farmers	15.3	12.0
Middle-level administrative groups	15.7	12.4
Employers in industry and commerce	18.2	13.6
Non-manual workers	19.4	13.9
Skilled workers	20.8	17.2
Semi-skilled workers	21.3	18.7
Agricultural workers	21.8	20.2
Unskilled workers	27.7	25.4
All economically active	19.0	15.0

From Desplanques, 1984.

although mortality declined in all categories, the effect was most evident in the most favored social groups. Differences in social mortality increased. Between 1975 and 1980, life expectancy at the age of 35 for male teachers was 43.2 years, but for unskilled male workers, it was 34.3 years (i.e., a difference of 8.9 years), whereas between 1955 and 1959, the difference was 8 years.

Although France had been one of the first countries to produce reliable statistics on social inequalities in mortality, work on measuring and following the development has remained sporadic (Leclerc *et al*., 2000).

Another example of a longitudinal study based on linked records is the longitudinal study for England and Wales, based on a 1% population sample (about 500,000) from the 1971 census. Data on deaths, as well as some other demographic events occurring in this population, have since been linked to the sample records. Detailed studies on different aspects of socioeconomic mortality differences have been carried out using this data set (Harding *et al*., 1997; Goldblatt 1990).

Studies based on linked records offer better possibilities for analyzing social differences in mortality than cross-sectional unlinked studies, but they still have limitations. Manual linking of census and individual death records is a time-consuming and very expensive process. The costs involved make the study of large populations impossible, and samples must be used. Although the samples used in the three studies described here were large enough for a statistically satisfactory analysis of differences in total mortality, the possibilities for detailed analysis by cause of death and in subgroups of the population are limited. Moreover, because of problems in linking the census and death records, the quality of data is not always very good. In the Kitagawa and Hauser study (1973), for example, 23% of the decedents could not be linked to any census record.

Longitudinal linked records of a population sample can best be carried out in countries in which a system of personal identification numbers makes the computerized linkage of records possible. Mortality studies (Borgan, 1996; Vågerö and Lundberg, 1995; Valkonen *et al*., 1993; Andersen, 1985) have been carried out in Denmark, Finland, Norway, and Sweden. Practically all registered death records can be linked to census records. In Finland, for example, only 0.3% of all deaths between 1986 and 1990 could not be linked to the 1985 census records (Valkonen *et al*., 1993). An additional advantage of the system of personal identification numbers is that information from other registers, such as tax and pension registers, can also be used.

Table 54–2 shows comparative death rates by occupation-based social class in Finland, on the basis of death records for 1986 to 1990, linked to the records of the 1985 census (Valkonen and Martikainen, 1997). For the purposes of calculating death rates by social class for the economically active population, only those who were employed in 1985 were assigned to social classes. All economically inactive persons and persons who were unemployed were excluded. For the purpose of calculating the death rates by social-occupational category for the whole population, retired, disabled, and unemployed persons were classified on the basis of their former occupation obtained by record linkage to the 1970 census. Nonworking family members (mainly housewives) were classified on the basis of the occupational category of the head of the household. The *All* column in Table 54–2 shows mortality differences for the whole population, not just for those who were employed in 1985.

The results for the economically active and the whole population are clearly different, particularly among women. Mortality in each social-occupational class is lower for the economically active, because the mortality of retired and other non-employed groups is, at an equal age, much higher than that of employed persons. More importantly, the differences between classes are clearly smaller among the economically active than in the whole population, because the proportion of economically inactive persons is higher in the manual classes. Among economically active women, class differences are much smaller than those among men.

This difference between the economically active population and the whole population is significant. Findings on social mortality have often only represented the economically active population, because access to information on the former occupations of non-employed persons is more difficult than for the employed. Such findings, however, underestimate the real magnitude of social differences. The amount of bias diminishes when the follow-up time becomes longer (Lévy and Vallin, 1981; Valkonen and Martikainen, 1979).

Different approaches have been used in countries where, for legislative or other reasons, studies based on the linkage of records on individuals were not possible. One example is ecologic research, in which death rates in particular geographic areas are correlated with

TABLE 54–2 Age-Standardized Mortality and Relative Age-Standardized Mortality for 1981 to 1985 by Social Class for the Economically Active Population in 1980 and for the Whole Population, with Distributions of Persons by Social Class for Men and Women between 35 and 64 Years Old in Finland

Social class	Age-standardized mortality rates (per 100,000)		Relative age-standardized mortality (upper white-collar employees = 1)		Population distribution (%)	
	Econ. active	All	Econ. active	All	Econ. active	All
Men						
Upper white-collar employees	595	768	1.00	1.00	15.3	13.0
Lower white-collar employees	709	1062	1.19	1.38	17.0	16.3
Skilled workers	781	1229	1.31	1.60	40.1	39.8
Unskilled workers	944	1744	1.59	2.27	7.1	9.8
Farmers	702	1068	1.18	1.39	13.1	13.8
Other self-employed persons	843	898	1.42	1.17	7.4	6.2
Others	391	2724	0.66	3.55	0.0	1.2
All	749	1176	1.26	1.53	100.0	100.0
Population (in 1000s)					2990	3839
Women						
Upper white-collar employees	242	344	1.00	1.00	9.0	9.3
Lower white-collar employees	267	393	1.10	1.14	38.4	32.5
Skilled workers	280	444	1.16	1.29	24.6	27.0
Unskilled workers	268	490	1.11	1.42	12.0	12.8
Farmers	276	431	1.14	1.25	11.7	13.1
Other self-employed persons	300	310	1.24	0.90	4.3	4.1
Others	256	1148	1.06	3.34	0.0	1.1
All	273	428	1.13	1.24	100.0	100.0
Population (in 1000s)					2768	4193

From Valkonen and Martikainen, 1995.

socioeconomic indicators and other characteristics of those areas (Wilkins *et al.*, 1989). This method is inexpensive and generally feasible, but it suffers from the usual weaknesses of ecologic analyses. The results of ecologic studies are therefore not comparable with those of studies in which data on individuals are used. The aggregate-level relationship between the socioeconomic status and mortality of populations in specific areas may be quite different from individual-level association between the same variables.

II. MORTALITY DIFFERENCES BY AGE

All examples of social mortality differences presented have dealt only with the working-age population. The extent of differences between social classes varies considerably with sex and age. Figure 54–2 shows excess mortality by age of men and women of different social-occupational groups in relation to that of upper-class nonmanual employees in Finland (Valkonen *et al.*, 1993). For the youngest age groups, the social classification is based on the social-occupational category of the head of the household.

With few exceptions, the order of occupational classes is the same in all age groups for both sexes. Mortality is highest among manual workers, second highest among farmers, and third highest in the lower nonmanual class (Table 54–3). It is always lowest in the upper nonmanual class, which is set at 100 in Figure 54–2. The relative differences are the smallest among children and elderly people and most pronounced among young adults 25 to 45 years old. The differences even exist in the much older groups. Among females, the age pattern of mortality differences is similar to that among males, but the relative differences are clearly smaller in all age groups, as shown in Table 54–2 for women between the ages of 35 and 64 years.

Age-specific death rates can be used to calculate life tables by social class. The difference in life expectancy at birth between upper nonmanual employees and manual workers was about 6 years for males and 3 years for females in Finland in between 1986 and 1990 (Valkonen *et al.*, 1993). The extent of the differences depends on the classification used. A more detailed classification of manual workers would show larger differences than the 6 years in male life expectancy at birth.

In England and Wales, the difference in life expectancy at birth between the social classes I/II (i.e.,

TABLE 54–3 Life Expectancy at Birth and at 30 Years of Age by Social Class and Sex in Finland, 1986 to 1990

Social class	Men		Women	
	e_0	e_{30}	e_0	e_{30}
Upper nonmanual employees	75.3	46.6	81.1	51.9
Lower nonmanual employees	73.0	44.4	79.9	50.8
Manual workers	69.2	40.9	78.1	49.1
Farmers	71.6	43.3	78.5	49.6
Total	70.8	42.4	78.8	49.8
Difference between the extremes	6.1	5.7	3.0	2.8

From Valkonen et al., 1993.

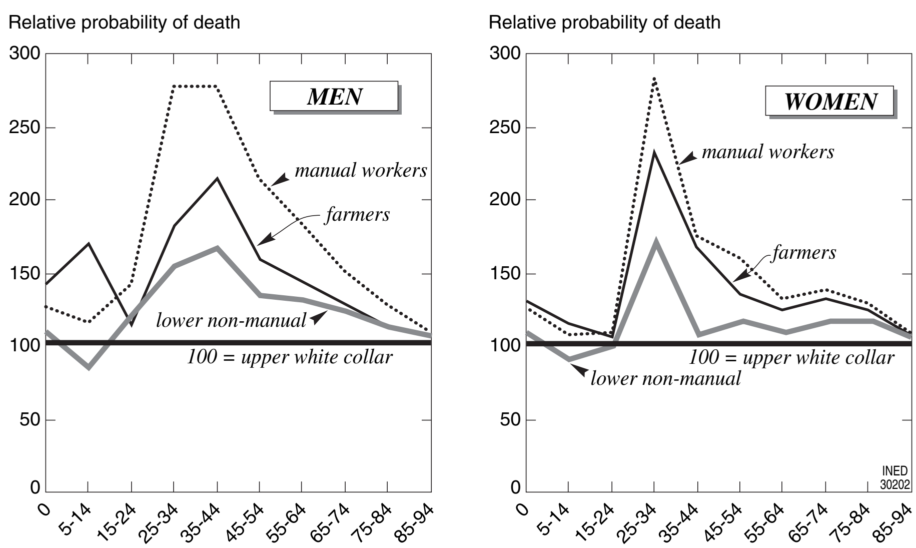


FIGURE 54–2 Relative probability of death by age (10-year age groups) and occupational class between 1986 and 1990 for men and women (upper white-collar employees taken as base 100).

upper nonmanual employees or self-employed technicians) and classes IV/V (i.e., manual workers) was estimated at 5 years for men and 3 years for women in between 1987 and 1991 (Hattersley, 1997). No recent estimate of life expectancy by social category is available for other countries.

III. MORTALITY BY CAUSE OF DEATH

The first step toward a better understanding of social inequalities in mortality is to study them by cause of death. As an example, Table 54–4 shows the higher mortality ratios of manual workers to upper white-collar employees of 35 to 64 years old in Finland in the 1980s for selected causes of death.

Among men, mortality is higher among manual workers than nonmanual workers for all causes of death except cancer of the intestines and rectum. Although the direction of the difference between the social classes is practically the same for all causes of death, the extent varies. Male manual workers have at least 2.5 times higher mortality than upper nonmanual employees due to alcohol poisoning, diseases of the respiratory system, and lung cancer. As for males, female manual workers also have higher mortality for almost all causes of death, except for cancers of the intestine and rectum and, in this case, breast cancer.

Table 54–5 shows the contribution of the most common causes of death to the difference in life expectancy at the age of 35 years between workers and upper nonmanual employees in Finland at the end of the 1980s. The total difference is 5.5 years for males and 2.7 years for females. For both sexes, the main causes of differences are ischemic heart disease and other cardiovascular diseases. These account for 46% of the male difference and 68% of the female difference in life expectancy. Among males, lung cancer, respiratory diseases, and accidents provide marked contributions to the inequality of life expectancy between social classes.

TABLE 54–4 Ratio of Age-Standardized Mortality Rates by Cause of Death of Manual Workers to that of Upper Nonmanual Employees by Sex for Ages 35 to 64 Years in Finland, 1981 to 1990

Cause of death	Men	Women
Malignant neoplasms	1.54	1.00
Stomach cancer	1.54	1.48
Cancer of intestine and rectum	0.97	0.88
Lung cancer	2.93	1.82
Breast cancer	—	0.68
Other cancers	1.14	1.09
Circulatory diseases	2.01	2.10
Ischemic heart disease	2.02	2.37
Cerebrovascular diseases	1.90	1.73
Other diseases	2.15	1.81
Respiratory diseases	3.58	2.06
Cirrhosis of the liver	1.83	1.33
Accidents and violence	2.53	1.44
Road accidents	1.68	1.31
Alcohol poisoning	3.89	—
Suicide	2.33	1.15
All causes	2.01	1.40

From Valkonen *et al*., 1993.

IV. WHY ARE THE SOCIAL DIFFERENCES IN MORTALITY LESS AMONG WOMEN THAN AMONG MEN?

In all countries for which reliable data exist, the social differences in mortality are less among women than men. Seppo Koskinen and Tuija Marelin (1994) tried to understand this phenomenon by using data for men and women 35 to 64 years of age in Finland. They showed that the results did not depend on the indicator chosen to define the social-occupational status. For each of the four social-occupational indicators used in the study, the inequalities between women were much less than those between men.

Analysis of the differential mortality by cause of death showed that for cardiovascular diseases, women's relative socioeconomic differences were slightly higher than those of men. For other diseases affecting both sexes, the relative social mortality differences are practically the same, except for respiratory diseases, for which the mortality differences are higher

TABLE 54–5 Contribution in Years of Selected Causes of Death to the Difference in Life Expectancy between Upper Nonmanual Employees and Manual Workers by Sex in Finland, 1986 to 1990

Cause of death	Men	Women
Lung cancer	0.68	0.05
Breast cancer		–0.12
Other cancers	0.19	0.16
Ischemic heart disease	1.75	0.91
Other cardiovascular diseases	0.79	0.91
Respiratory diseases	0.50	0.18
Other diseases	0.50	0.34
Accidents	0.55	0.15
Alcohol poisoning	0.21	0.04
Suicide	0.36	0.08
All causes	5.53	2.73

From Valkonen *et al*., 1993; for method, see United Nations, 1988.

among men. The differences are also more marked among men for accidents and violence.

The fact that the social range of total mortality is much greater among men than among women is partly explained by this steeper social gradient for violent death among men. The main reason, however, is that the distribution of deaths by causes varies according to sex. Cancer of the breast, which according to the data employed by Seppo Koskinen and Tuija Martelin is responsible for 9% of female deaths, is practically nonexistent among men. Deaths from breast cancer also fall with social-occupational status, and this significantly reduces the social range of female mortality. Cancers other than those of the stomach, intestines, lungs and breast, whose social gradient is very slight, account for at least one-fourth of the deaths of women and only one-tenth of those of men. If the same distribution of deaths by cause among men is applied to women, the social gradient of mortality among women becomes quite similar to that of men. According to these Finnish data, the big difference in social range between men and women mainly results from the difference in structure by cause of death. We do not know to what extent this result can be applied to other populations and age groups.

V. EXPLANATIONS OF SOCIAL INEQUALITIES IN MORTALITY

The most frequently used explanatory framework for social mortality differences is that which appeared in the *Black Report*, set up in Great Britain in 1980 by a task force presided over by Sir Douglas Black (Townsend and Davidson, 1982). This framework distinguishes four main types of explanation: statistical artefact, natural and social selection, material and structural causes, and cultural and behavioral factors. These explanations, which had already been put forward by other authors (Behm and Vallin, 1982; Vallin, 1979), were also the object of further developments (MacIntyre, 1997; Vågerö and Illsley, 1995; Davey Smith *et al*., 1990).

Social and mortality differences could only be a *statistical artifact* if they were viewed as mere measurement errors. Admittedly, such errors may effectively bias the extent of the observed differences, but there is considerable proof that the differences are quite real. Problems of measurement tend more frequently to underestimate rather than overestimate them.

According to the *selection hypothesis*, the association between social position and mortality is in line with the reasoning that personal health determines social position and not *vice versa*. People are not in poorer health because they belong to disadvantaged social categories; they belong to these social categories because their health is less good than the health of people in other categories. Young people in poor health have greater difficulty in completing their studies, and adults with health problems move down the social scale. Although direct measurement of the role of health in this selection bias is difficult, most investigators have concluded that social mobility due to health reasons is not a highly determinant factor in social mortality differences (Davey Smith *et al*., 1994; Wilkinson, 1986; Fox *et al*., 1985).

Another type of selection, sometimes called *indirect selection*, may contribute to social mortality differences because some *confounding factors* are capable of affecting the social-occupational status of persons and their health and mortality. High social status of the parents, good school performance, and certain psychological and physical characteristics, for example, may favor access in adulthood to a socially privileged status. The inherited lifestyle model (especially behavior in health matters) may also contribute to health and chances of survival. Consequently, a part of the relation observed between social status and adult mortality may not depend on the status itself. Unfortunately, no recent empirical study enables us to appreciate these distinctions.

The materialistic/structural and behavioral/cultural explanations are two additional elements in an interpretive outline that stresses—through the risk factors involved—the effects of social position on a person's health. It explains why those benefiting from a higher social status are also less exposed to the risks of death than those of a lower status.

In the *materialistic/structural* view, persons of a disadvantaged status suffer from health problems inherent in material living conditions, such as precarious living conditions, difficult working conditions, and insufficient income for ensuring correct nutrition and health care. This type of explanation seems plausible for the past and for less-developed countries, but it is less certain that poverty and precarious living conditions are pertinent factors today in developed countries. Results for all countries indicate that differences in mortality are not solely between the poor and the rest of the population, but that there are also stepwise differences within the middle and upper classes. Working conditions are more discriminatory and, even though they tend to become homogeneous, their role in social mortality differences persists in the most advanced societies.

According to the *cultural/behavioral explanation*, socioeconomic mortality differences result from differences in the prevalence of risk behaviors, such as smoking, alcohol abuse, lack of exercise, and unhealthy dietary habits. For example, class differences in the prevalence of smoking may account for a substantial part of the socioeconomic differences in mortality from lung cancer and from coronary heart disease, because smoking is a risk factor for both diseases and because there are differences in smoking habits among social classes in most countries.

Literature on the subject has suggested hypotheses to explain social mortality differences other than those put forward by the *Black Report*. One of these concerns the effects of biologic programming that is set up *in utero* or during infancy (Barker, 1991). According to this hypothesis, differences in present social mortality differences may correlate with conditions prevailing in a previous generation. These conditions would have influenced maternal health, which would have influenced the growth and development of the fetus and child and the probability of suffering from a disease or of dying prematurely at an adult age. Because adults in a disadvantaged social category have experienced precarious living conditions since their childhood, their higher mortality could be attributed to the conditions prevailing during their first years of life. The validity of this hypothesis has yet to be established.

Another hypothesis is that the higher mortality of persons forming part of a disadvantaged social category reflects psychosocial factors, particularly those relating to work, and is more frequent in the categories from the lower part of the social spectrum (Marmot *et al*., 1997). Differences in access to health services and their use also contribute to social inequalities in mortality (Keskimäki *et al*., 1997).

The theories and hypotheses about the causes of social differences in mortality are numerous, but few empirical investigations have verified the validity and measured the role of the risk factors evoked. Of all the studies aiming to explain the social differences in mortality, the one most frequently cited is the so-called Whitehall investigation, carried out with about 17,000 British civil servants (Marmot *et al*., 1984). According to this study, only one-fourth of the considerable differences in mortality by coronary diseases observed among five grades of civil servants could be attributable to those involving smoking, physical activity, blood pressure, or cholesterol level. Of all the risk factors, only smoking varied significantly from one category to the other, and the explication for differences in mortality must essentially be attributed to other risk factors and behavior.

The authors of the study also suggested another possible explanation. Exposure to specific risks favoring specific causes of mortality, such as those of the Whitehall study, cannot take account of all the observed differences in mortality. The great similarity of social differences for most of the causes of death in many national studies cannot be explained by employing a list of specific risk factors. Michael Marmot and his colleagues (1984) think that social inequalities in mortality are largely attributable to "general risk factors" that concurrently influence most causes of death. Psychosocial stress could be one of these factors (Williams, 1990). A new follow-up study (Whitehall II) is being carried out in an attempt to find a better explanation of the differences in morbidity and mortality observed between different categories of civil servants in Britain (Marmot, 1996).

According to some other studies, a greater share of the social inequalities of mortality by coronary diseases appears attributable to some disparities in the prevalence of risk factors. In most cases, a large part of the difference remains unexplained (Kaplan and Kiel, 1993).

The empirical studies seeking an explanation of the social-occupational differences are generally inconsistent and do not reach any firm conclusion. The poverty of these results may reflect the nature of the problem—that the explanation of the differences varies from one cause of death to another, from one country to another, and according to the era.

The persistence and even worsening of social-occupational mortality differences have preoccupied a number of governments and the World Health Organization (WHO). A very complete British report (Independent Inquiry into Inequalities in Health, 1998) included 10 pages of recommendations for health policies and programs aimed at reducing the inequalities of health and mortality. Some of these recommendations are aimed at a global level of society (e.g., reduction in the inequality of incomes), but others are very specific (e.g., improvement in access to therapy for replacing nicotine). No study exists, however, for measuring the effectiveness of the policies for reducing the social differences in mortality.

VI. INTERNATIONAL COMPARISONS

Although the level of mortality is everywhere higher among persons in the most disadvantaged social positions, it is not clear to what extent the size of this difference varies from country to country. International comparisons of social-occupational mortality differences are difficult to carry out. Reliable data exist only for a few countries. Even when data do exist, there are differences between studies in terms of time periods and age bands covered, social-occupational indicators used, coverage (e.g., exclusion of economically inactive persons), measures of mortality, and quality of data. There is no generally agreed method for measuring the magnitude of the differences even if the data problems are resolved. A large number of inequality measures have been constructed, and the choice of these may substantially influence the results. It is therefore not surprising that the results of attempts to compare the sizes of socioeconomic mortality differentials have been inconsistent (Kunst and Mackenbach, 1994c).

The diversity of the results observed from one country to another was studied in an ambitious project covering 11 countries in Western Europe, the United States, and three socialist countries in Central and Eastern Europe, for which representative mortality data were available at national level, according to educational and social-occupational categories (Kunst, 1997; Mackenbach *et al*., 1997). These data were collected between 1980 and 1990, and generally were for male adults between 30 and 64 years old.

In the 14 countries where this information was available, the death rate among manual workers was higher than that for nonmanual workers, and the relative excess mortality of the manual workers was remarkably similar from one country to another. Nevertheless, some larger differences were observed in France, the Czech Republic, Finland, and (especially) Hungary than in other countries (Kunst, 1997).

It has been possible to measure these differences in the cause of death in 11 West European countries and in the United States. The excess mortality of manual workers is manifest for each of the nine groups of mortality causes employed, except for one notable exception: ischemic heart diseases. No excess mortality can be observed for this group of causes in France, Switzerland, Italy, and Spain. In Portugal, the mortality rate for ischemic heart diseases was even lower among the manual than the nonmanual workers. It is possible, however, that this exception is only an apparent one. Many of the deaths from ischemic diseases are confused with other cardiac causes, or they remain ill defined (Meslé and Vallin, 1988). This problem of data quality is often more acute for the most disadvantaged social categories. The social differences in causes of death other than ischemic heart diseases were more marked in those of the southern countries than in those of the north of Europe.

In an attempt to explain the weak social discrimination of ischemic heart diseases in southern Europe, Anton Kunst (1997) has suggested that some specific factors have contributed to protecting people there against these diseases, such as a traditional diet comprising a high consumption of fish and vegetable oil and a traditionally moderate consumption of alcohol. These factors particularly have favored the most disadvantaged social categories.

Differences in mortality by social class are higher in Finland and France than in other countries of Western Europe. In both countries, alcohol-related mortality is a contributing factor. In France, death by cancer of the digestive system or cirrhosis of the liver—for which the consumption of alcohol is the main risk factor—is seven times higher among unskilled workers than in the higher social-occupational categories (Desplanques, 1984). Similarly, accidental and violent deaths in Finland, which play an important part in the mortality differences for all causes, are closely linked to the Finnish drinking habits; according to the information available on death certificates, one-half the Finnish adult men who died a violent death were inebriated (Mäkelä, 1998). In Finland, alcohol-related deaths account for one-fourth of the difference in life expectancy between manual and nonmanual workers (Mäkelä *et al*., 1997).

VII. TRENDS IN SOCIAL MORTALITY DIFFERENCES

The problems encountered in studying trends in mortality differences are similar to those in making international comparisons. It is difficult to obtain comparable data, and even when data exist, the conclusions may depend on the population groups covered and the indicators of inequality used.

Anton Kunst (1997) compared some evaluations of the extent of differences in mortality by education level and social class for men between the ages of 50 and 59 years in Finland, Norway, France, and Italy (Turin) during the first and second halves of the 1980s. In all these countries, the differences have tended to increase. An increase in the differences for men 40 to 54 years old in the United States has also been observed. The Kunst results do not only apply to a limited age group. Other data clearly show that, since the 1980s, the relative differences in social mortality have increased in all developed countries where measurement was possible. This increase in the social range of male mortality at working ages was already apparent in Sweden and Spain in the 1970s (Diderichsen and Hallqvist, 1997; Regidor *et al*., 1995) and since the 1960s in France (Desplanques, 1984).

Figure 54–3 illustrates the increase in differences between educational levels observed in Finland since 1971 for life expectancy according to sex at the age of 35. Men with a higher education had a life expectancy in 1970 that was 4.3 years more than those with a lower one. The difference remained almost the same until the

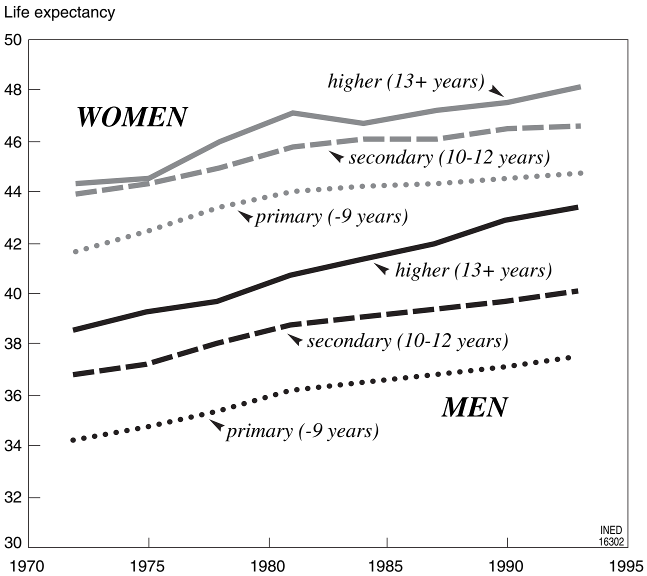


FIGURE 54–3 Life expectancy (years) at age 35 according to sex and level of education in Finland between 1971 and 1993.

end of the decade, but since the early 1980s, it has grown rapidly, reaching up to 5.9 years in the early 1990s. Among women, the difference grew in the same period from 2.7 to 3.4 years. Differences by occupational category have followed similar trends.

Some data for the trend of life expectancy by social class are available for England and Wales. They show that the difference in life expectancy at birth between the classes I/II (i.e., professional, managerial, and technical occupations) and IV/V (i.e., partly skilled and unskilled occupations) increased by almost a year between 1979 and 1981 and between 1987 and 1991 among men and women (Hattersley, 1997).

This increase in social mortality differences contradicts the objective advocated in 1985 by the Regional Bureau for Europe of the WHO: "By the year 2000, the actual differences in health status between and within countries should be reduced by at least 25%" (WHO, 1985). The reasons for this failure have not been analyzed in detail, but it has been suggested that the increase might reflect the general increase in social inequalities and income differences (Carrol and Davey Smith, 1997). Although this explanation is valid in certain cases, it does not provide a sufficient explanation. Differences in mortality have increased in countries where income differentials have increased (e.g., Great Britain) and in countries where they have decreased (e.g., Finland).

The principal common factor in the increased social mortality differences is the decline in cardiovascular deaths and especially in deaths from ischemic heart diseases—a decline that has been more rapid among persons with the highest education or employed in nonmanual occupations. In the Nordic countries and the United States, death rates from heart diseases probably have not always been higher among manual than nonmanual workers. In the United States and in England and Wales, this has been recognized only since the 1950s or 1960s (Kunst, 1997; Marmot *et al.*, 1978), and it has been true for Sweden and Norway only since the 1960s and 1970s (Diderichsen and Hallqvist, 1997; Borgan, 1996). Similarly, scarcely any social differences in mortality due to ischemic diseases were observable in France, Switzerland, Italy, or Spain in the 1980s, and even the reverse is the case in Portugal. In all the countries for which data are available, a more rapid decline in deaths from cardiovascular or ischemic heart disease has been observed among men of working age from the upper, rather than the lower, socioeconomic groups. This is the major reason for the increase in social inequalities in mortality.

This increase in social mortality differences, largely due to greater deviations for cardiovascular diseases, does not necessarily imply that social-occupational inequalities in mortality will increase in the future. The rapid decline in deaths from cardiovascular diseases in the socially privileged classes shows that this type of death has become avoidable and that substantial advances are still possible. Although the most favored social categories have benefited most from this advance, this is probably because they have changed their health behavior more rapidly (e.g., diet, smoking, physical activity) and because they have had better access to new medical treatments, such as coronary bypass and angioplasty (Keskimäki *et al.*, 1997). Doubtlessly, the most disadvantaged classes will eventually adopt the same behavior in health matters and benefit from the new medical techniques when these have become routine treatments.

Cardiovascular diseases are not the only conditions that have contributed to an increase in the differences. The data concerning the part played by other causes are not always available, and it is difficult to generalize. It appears, however, that there is no uniformity between countries in the trend of the mortality profiles other than for cardiovascular diseases. For example, although death through lung cancer has contributed to accentuation of the social mortality differences in certain countries, this was not the case in Finland between 1980 and 1990. Although death through alcoholism was one of the major causes for the increase in mortality differences in Finland during the 1980s and, to a lesser extent, during the 1990s (Valkonen and Jalovaara, 1996), there is no information concerning such processes in other countries. Changes in differences in mortality result from a complex combination of different and even opposite trends for various causes of death, which may vary by sex, class and period. It is most unlikely that a single explanation could account for this.

References

ANDERSEN Otto, 1985. *Dodelighed og erhverv 1970–80*. Copenhagen, Danmarks Statistik, 251 p. (Statistiske undersogelser, no. 41.)

ANTONOVSKY Aaron, 1967. Social class, life expectancy and overall mortality, *Milbank Memorial Fund Quarterly*, vol. 45, p. 31–73.

BARKER David J.P., 1991. The foetal and infant origins of inequalities in health in Britain, *Journal of Public Health Medicine*, vol. 13, p. 64–68.

BEHM Hugo and VALLIN Jacques, 1982. Mortality differentials among human groups, *in*: Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life*, p. 11–37. Liège, Ordina éditions.

BORGAN Jens-Kristian, 1996. Socioeconomic trends in differential mortality among middle-aged men in Norway 1960–1990, *Scandinavian Population Studies*, vol. 11, p. 73–81.

CARROL Doug and DAVEY-SMITH George, 1997. Health and socioeconomic position. A commentary, *Journal of Health Psychology*, vol. 2, p. 275–282.

DAVEY-SMITH George, BARTLEY Mel, and BLANE David, 1990. The Black Report on socioeconomic inequalities 10 years on, *British Medical Journal*, vol. 301, p. 491–493.

DAVEY-SMITH George, BLANE David, and BARTLEY Mel, 1994. Explanations for socioeconomic differentials in mortality. Evidence from Britain and elsewhere, *European Journal of Public Health*, vol. 4, p. 131–144.

DESPLANQUES Guy, 1976. *La mortalité des adultes suivant le milieu social 1955–1971*. Paris, National Institute of Statistics and Economic Studies, 148 p. (Les collections de l'INSEE, série D, no. 44.)

DESPLANQUES Guy, 1984. L'inégalité sociale devant la mort, *Économie et statistique*, janvier (162), p. 29–50.

DIDERICHSEN Finn and HALLQVIST Johan, 1997. Trends in occupational mortality among middle-aged men in Sweden 1961–1990, *International Journal of Epidemiology*, vol. 28(4), p. 782–787.

DUBLIN Louis I., LOTKA Alfred J., and SPIEGELMAN Mortimer, 1949. *Length of life. A study of the life table, Revised edition*. New York, The Ronald Press Company, 321 p.

FOX A. John., 1979. Prospects for change in differential mortality, *in*: World Health Organization *et al*. (eds.), *Proceedings of the meeting on socioeconomic determinants and consequences of mortality, Mexico City 1979*, p. 515–561. New York, Geneva, World Health Organization, 622 p.

FOX A. John, GOLDBLATT Peter O., and JONES David R., 1985. Social class mortality differentials: artefact, selection or life circumstances, *Journal of Epidemiology and Community Health*, vol. 39(1), p. 1–8.

GOLDBLATT Peter O. (ed.), 1990. *Longitudinal study. Mortality and social organisation 1971–1981*. London, Office of Population Censuses and Surveys, 193 p. (Series LS, no. 6.)

HARDING Seeromanie, BETHUNE Ann, MAXWELL Roy, and BROWN Joanna, 1997. Mortality trends using the Longitudinal Study, *in*: Frances Drever and Margaret Whitehead (eds.), *Health inequalities. Decennial supplement*, p. 143–155. London, The Stationery Office, 257 p. (Series DS no. 15.)

HATTERSLEY Lin, 1997. Expectation of life by social class, *in*: Frances Drever and Margaret Whitehead (eds.) *Health inequalities. Decennial supplement*, p. 73–82. London, The Stationery Office, 257 p. (Series DS no. 15.)

KAPLAN George A. and KEIL Julian E., 1993. Socioeconomic factors and cardiovascular disease: a review of the literature, *Circulation*, vol. 88, p. 1973–1998.

KESKIMÄKI Ilmo, KOSKINEN Seppo, SALINTO Marjo, and ARO Seppo, 1997. Socioeconomic and gender inequities in access to coronary artery bypass grafting in Finland, *European Journal of Public Health*, vol. 7, p. 392–397.

KITAGAWA Evelyn M. and HAUSER Philip M., 1973. *Differential mortality in the United States: a study in socioeconomic epidemiology*. Cambridge, MA, Harvard University Press, 205 p.

KOSKINEN Seppo and MARTELIN Tuija, 1994. Why are socioeconomic mortality differences smaller among women than men? *Social Science & Medicine*, vol. 38, p. 1385–1396.

KUNST Anton E., 1997. *Cross-national comparisons of socioeconomic differences in mortality*. Doctoral thesis. Rotterdam, Erasmus University, 264 p.

KUNST Anton E. and MACKENBACH Johan, 1994c. *Measuring socioeconomic inequalities in health*. Copenhagen, World Health Organization, 115 p.

LECLERC Annette, FASSIN Didier, GRANDJEAN Hélène, KAMINSKI Monique, and LANG Thierry (eds.), 2000. *Les inégalités sociales de santé*. Paris, La Découverte et Institut National de la Santé et de la Recherche Médicale, 448 p.

LÉVY Claude and VALLIN Jacques, 1981. La mortalité par catégorie socioprofessionnelle, un essai de calcul direct, *Population*, vol. 36(4–5), p. 938–946.

MACINTYRE Sally, 1997. The Black Report and beyond: what are the issues? *Social Science & Medicine*, vol. 44, p. 723–745.

MACKENBACH Johan P., KUNST Anton E., CAVELAARS Adriënne E.J.M., GROENHOF Feijke, GEURTS José J.M., and EU Working Group on Socioeconomic Inequalities in Health, 1997. Socioeconomic inequalities in morbidity and mortality in western Europe: a comparative study, *Lancet*, vol. 349, p. 1655–1659.

MÄKELÄ Pia, 1998. Alcohol-related mortality by age and sex and its impact on life-expectancy: estimates based on the Finnish death register, *European Journal of Public Health*, vol. 8, p. 43–51.

MÄKELÄ Pia, VALKONEN Tapani, and MARTELIN Tuija, 1997. Contribution of deaths related to alcohol use to socioeconomic variation in mortality: register based follow up study, *The British Medical Journal*, vol. 315, p. 211–216.

MARMOT Michael G., 1996. The social pattern of health and disease, *in*: David Blane, Eric Brunner, and Richard Wilkinson (eds.), *Health and social organization. Towards a health policy for the 21st century*, p. 42–67. London, Routledge, 326 p.

MARMOT Michael G., ADELSTEIN Abel M., ROBINSON Nicola, and ROSE Geoffrey A., 1978. Changing social-class distribution of heart disease, *British Medical Journal*, vol. 2, p. 1109–1112.

MARMOT Michael G., BOSMA Hans, HEMINGWAY Harry, BRUNNER Eric, and STANSFELD Stephen, 1997. Contribution of job control and other risk factors to social variations in coronary heart disease incidence, *Lancet*, vol. 350, p. 235–239.

MARMOT Michael G., SHIPLEY Martin J., and ROSE Geoffrey A., 1984. Inequalities in death—-specific explanations of a general pattern, *Lancet*, no. 8384, p. 1003–1006.

MESLÉ France and VALLIN Jacques, 1988. Les composantes de la mortalité cardio-vasculaire en France depuis 1925: résultats d'une reconstitution historique, *Population*, vol. 43(2), p. 391–425.

MIELCK Andreas and DO ROSÁRIO Giraldes Maria (eds.), 1993. *Inequalities in health and health care. Review of selected publications from 18 Western European countries*. Münster, Waxman, 238 p.

OPCS, 1978. *Occupational mortality. The Registrar General's decennial supplement for England and Wales, 1970–72*. London, Her Majesty's Stationery Office, Office of Population Censuses and Surveys, 224 p. (Series DS, no. 1.)

OPCS, 1986. *Occupational mortality. The Registrar General's decennial supplement for Great Britain, 1979–80, 1982–83. Part I. Commentary*. London, Her Majesty's Stationery Office, Office of Population Censuses and Surveys, 128 p. (Series DS, no. 6.)

Regidor Enrique, Gutiérrez-Fisac Juan L., and Rodrigue Carmen, 1995. Increased socioeconomic differences in mortality in eight Spanish provinces, *Social Science & Medicine*, vol. 41(6), p. 801–807.

Thatcher A. Roger, 1986. Preface, *in*: Office of Population Censuses and Surveys (ed.), *Occupational mortality. The Registrar General decennial supplement for Great Britain, 1979–80, 1982–83. Part I. Commentary*, p. I—IX. London, Her Majestys's Stationery Office, 128 p. (Series DS, no. 6.)

Townsend Peter and Davidson Nick, 1982. *Inequalities in health: The Black report*. Harmondsworth, UK, Penguin Books, 240 p.

Vågerö Denny and Illsley Raymond, 1995. Explaining health inequalities: beyond Black and Barker, *European Sociological Review*, vol. 11, p. 219–241.

Vågerö Denny and Lundberg Olle, 1995. Socioeconomic mortality differentials among adults in Sweden, *in*: Alan Lopez, Graziella Caselli, and Tapani Valkonen (eds.), *Adult mortality in developed countries. From description to explanation*, p. 223–242. Oxford, Clarendon press, 361 p.

Valkonen Tapani and Jalovaara Marika, 1996. Kuolleisuus koulutusasteen mukaan 1987–95. [Mortality according to educational level 1987–95], *Statistics Finland, Causes of Death 1995, Health 1996:5*, p. 17–19 and 50–55.

Valkonen Tapani and Martikainen Pekka, 1997. Problems of comparability resulting from the exclusion of economically inactive persons in studies on social class differences in mortality, *in*: Guillaume Wunsch and Attila Hancioglu (ed.), *Morbidity and mortality data: problems of comparability. Proceedings of the European Association for Populations Studies and the Hacettepe Institute of Populaton Studies Workshop*, p. 97–114. Ankara, Hacettepe University, Institute of Population Studies, 233 p.

Valkonen Tapani, Martelin Tuija, Rimpelä Arja, Notkola Veijo, and Savela Soili, 1993. *Socioeconomic mortality differences in Finland 1981–90*. Helsinki, Statistics Finland, 100 p. (*Population* 1993, no. 1.)

Vallin Jacques, 1979. Facteurs socio-économiques de la mortalité dans les pays développés, *in*: Organisation Mondiale de la Santé, Pan American Health Organization, and Commission for International Cooperation in National Research in Demography (eds.), *Délibération de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité. El Colegio de Mexico, Mexico, 19–25 juin 1979*, p. 266–302. Geneva, WHO; New York, PAHO, 622 p.

Westergaard Harald, 1901. *Die Lehre von der Mortalität und Morbilität. Anthropologisch-statistische Untersuchungen*. Jena, Verlag von Gustav Fischer, 702 p.

WHO, 1985. *Targets for health for all. Targets in support for the European strategy for health for all*. Copenhagen, World Health Organization, Regional Office for Europe, 201 p.

Wilkins Russell, Adams Owen B., and Brancker Anna, 1989. Changes in mortality in urban Canada from 1971 to 1986, *Health Reports*, vol. 1(2), p. 137–174.

Wilkinson Richard G., 1986. Socioeconomic differences in mortality: interpreting the data on their size and trends, *in*: Richard G. Wilkinson (ed.), *Class and health. Research and longitudinal data*, p. 1–20. London, Tavistock, p. 223.

Williams David R. 1990. Socioeconomic differentials in health: A review and redirection, *Social Psychology Quarterly*, vol. 53, p. 81–99.

C H A P T E R

55

Geographic Variations of Mortality

GRAZIELLA CASELLI AND JACQUES VALLIN

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

Mortality varies according to sex (see Chapter 53), to social background (e.g., social class, social-occupational category, level of education; see Chapter 54), to a variety of individual characteristics (e.g., weight, height, marital status), or to collective criteria (e.g., religion, ethnic origin). Mortality also varies according to geographic location. This statement may seem obvious, given the differences between situations observed throughout the world, but such differences also can be observed within a single country. The finer the territorial division, the more we find differences; geographic diversity exists within a single region or a single province and at all levels in the hierarchy of administrative territorial divisions.

In this chapter, we first provide an overview of this diversity.[1] Later, we demonstrate that geographic variations are not necessarily the same for different components of mortality. For example, the patterns of child mortality may be different from those of adult mortality, whereas cause-specific mortality patterns may sometimes show striking contrasts. In the same way, geographic variations of mortality may depend on sex, social class, or other factors.

Geographic variations of mortality are partly linked to the particular history of each country, region, or other geographic entity. For this reason, geographic patterns of mortality necessarily change over time, even if they do show some constant features. In some cases, we observe the persistence of specific patterns over long periods, whereas in other places, the situation may undergo a complete reversal.

We examine the different factors that can explain geographic variations of mortality and see to what extent they can be attributed to variations in the composition of the populations of different geographic units, to geographic factors in the strict sense of the term (e.g., climate, altitude), or to socioeconomic, cultural, and environmental factors (e.g., occupation, eating habits, smoking, industrial pollution).

I. A GREAT DIVERSITY OF SITUATIONS

1. Inequalities on a World Scale

The most noticeable geographic difference in mortality is that which can be established between different world populations. This can be observed by referring to differences between countries. Over the past several decades, the United Nations has regularly published estimates of life expectancy at birth for country.[2] Figure 55–1 and Table 55–1 provide data for the 1995–2000 period; the life expectancy data are for both sexes taken together. The map was obtained

[1] A more detailed description of geographic variations of mortality and health can be found in numerous specialized atlases (Britton, 1999a, 1999b; Salem *et al.*, 1999; Jozan and Prokhorskas, 1997; Merenne *et al.*, 1997; Dorling, 1995; Potrykowska and Clarke, 1995; Plane and Rogerson, 1994; Peters et Larkin, 1993; Holland, 1988, 1991, 1993; Bähr *et al.*, 1992; Decroly and Vanlaer, 1991; Union géographique internationale, 1990, 1991; Clieff and Haggett, 1988; Yue Lu, 1987; ISTAT and ISS, no date).

[2] Some of these estimates may be criticized. Nonetheless, we provide them here as they are given, because this set of data has the advantage of being complete and is sufficient to support our point in this chapter.

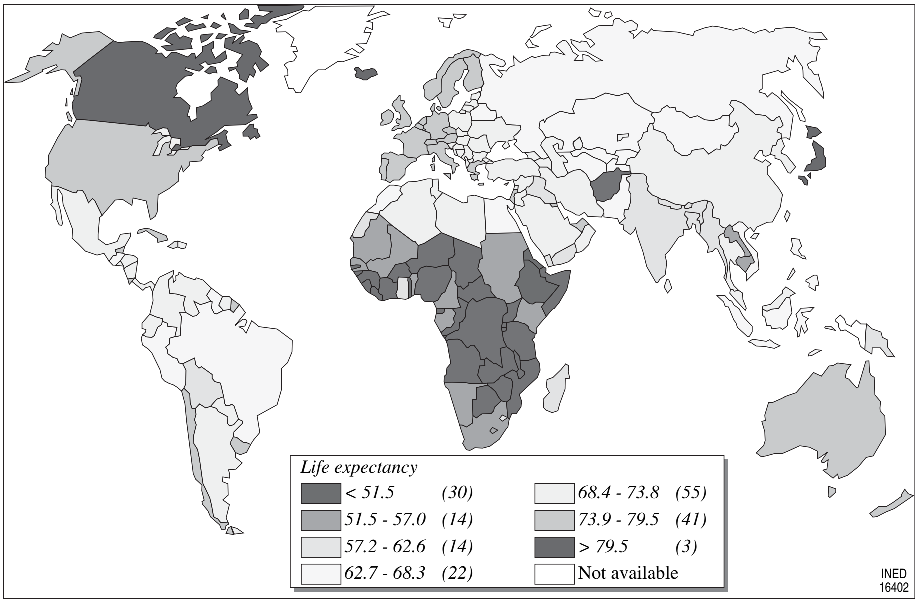


FIGURE 55–1 Life expectancy at birth for both sexes between 1995 and 2000: inequalities between different countries of the world. (From United Nations, 2001.) (Figure 55–1 also reproduced in color plate section.)

by distinguishing seven classes of life expectancy defined on the basis of the simple average, *m*, and the standard deviation, σ, of the 189 values observed.[3] The central class covers $m \pm \frac{\sigma}{4}$, and on either side, each class represents one-half of a standard deviation.

According to these estimates, at the end of the 20th century, life expectancy ranged from 37.2 years in Sierra Leone to 80.0 years in Japan. On the map, we can see a clear opposition between a region of low life expectancy covering all the countries of sub-Saharan Africa and (in a less homogenous way) a substantial part of Southern or Southeast Asia and Oceania (from Afghanistan to Papua New Guinea) and a region of very high life expectancy, occurring mainly in the most developed countries (i.e., Japan, North America, Europe, Australia, and New Zealand) but also extending toward less developed countries (i.e., Central America, northern and southern parts of South America, Middle-East, China, and neighboring countries). Each of these two regions forms a relatively continuous expanse, highlighting the clear-cut division of the planet in terms of health status. Between these regions, for different reasons, we find the center of Latin America (e.g., Brazil, Peru), which is less advanced in the health transition, and most of the territory of the old USSR, whose health status has been declining for more than 30 years.

[3] Whereas the life expectancy calculated for the world population taken as a whole is 65.4 years, the simple average of life expectancies per country is 65.5 years. This may be surprising because of the wide diversity of population counts (i.e., several thousand for some microstates to more than a billion for China). This finding probably means that the distribution of the diversity of population counts is not related to life expectancy levels. The variability is quite high because the standard deviation represents 11.2 years, almost 20% of the average.

2. A Matter of Scale

The world map of life expectancy raises a few problems of interpretation, particularly in terms of scale, prompting comments about heterogeneity of geographic distribution and disproportion of population numbers.

a. Heterogeneity of the Geographic Division of the Planet

The political division of the planet into separate states is far from ideal for the observation of geographic variations of mortality. Indeed each state has its own policies and observes a certain consistency in health regulations and behavior, and from this point of view, international comparisons are extremely useful. However, due to differences in size among countries, such a world map can easily mask the internal discrepancies existing within certain countries; internal lines of separation are as important as national borders, which sometimes have little impact. This type of situation is common when the focus shifts from

TABLE 55–1 Variations in Life Expectancy in the World according to United Nations' Estimates for 1995 to 2000

Country	e_0	Country	e_0	Country	e_0	Country	e_0
Africa	51.4	**North America**	76.7	**Asia**	65.8	**Europe**	73.2
Northern Africa	64.6	Canada	78.5	*East Asia*	70.9	*Northern Europe*	76.7
Algeria	68.9	United States	76.5	China	69.8	Denmark	75.9
Egypt	66.3			Hong Kong	79.1	Estonia	70.0
Libya	70.0	**Latin America**	69.3	North Korea	63.1	Finland	77.2
Morocco	66.6	*Central America*	71.0	Japan	80.5	Iceland	78.9
Sudan	55.0	Belize	73.6	Macao	78.5	Ireland	76.1
Tunisia	69.5	Costa Rica	76.0	Mongolia	61.9	Latvia	69.6
West Sahara	61.4	Salvador	69.1	South Korea	74.3	Lithuania	71.4
Western Africa	50.0	Guatemala	64.0	*South Central Asia*	61.5	Norway	78.1
Benin	53.5	Honduras	65.6	Afghanistan	42.5	Sweden	79.3
Burkina Faso	45.3	Mexico	72.2	Bangladesh	58.1	United Kingdom	77.2
Cape Verde	68.9	Nicaragua	67.7	Bhutan	60.7	*Western Europe*	77.7
Côte d'Ivoire	47.7	Panama	73.6	India	62.3	Austria	77.7
Gambia	45.4	*Caribbean*	67.5	Iran	68.0	Belgium	77.9
Ghana	56.3	Bahamas	69.1	Kazakhstan	64.1	France	78.1
Guinea	46.5	Barbuda	76.4	Kyrgyzstan	66.9	Germany	77.3
Guinea-Bissau	44.1	Cuba	75.7	Maldives	65.4	Luxembourg	77.0
Liberia	48.1	Dominican Republic	67.3	Nepal	57.3	Netherlands	77.9
Mali	50.9	Guadeloupe	77.3	Pakistan	59.0	Switzerland	78.6
Mauritania	50.5	Haiti	52.0	Sri Lanka	71.6	*Eastern Europe*	68.2
Niger	44.2	Jamaica	74.8	Tajikistan	67.2	Belarus	68.5
Nigeria	51.3	Martinique	78.8	Turkmenistan	65.4	Bulgaria	70.8
Senegal	52.3	Netherlands Antilles	75.5	Uzbekistan	68.3	Czech Republic	74.3
Sierra Leone	37.3	Puerto Rico	74.9	*Southeast Asia*	65.3	Hungary	70.7
Togo	51.3	Trinidad and Tobago	73.8	Brunei	75.5	Poland	72.8
Middle Africa	48.9	*South America*	68.9	Cambodia	56.5	Moldavia	66.6
Angola	44.6	Argentina	72.9	East Timor	47.5	Rumania	69.8
Cameroon	50.0	Bolivia	61.4	Indonesia	65.1	Russia	66.1
Central African Republic	44.3	Brazil	67.2	Laos	52.5	Slovakia	72.8
Chad	45.2	Chile	74.9	Malaysia	71.9	Ukraine	68.1
Congo	50.9	Colombia	70.4	Myanmar	55.8	*Southern Europe*	77.0
Congo, Democratic Republic	50.5	Ecuador	69.5	Philippines	68.6	Albania	72.8
Equatorial Guinea	50.0	Guyana	63.7	Singapore	77.1	Bosnia Herzegovina	73.3
Gabon	52.4	Paraguay	69.6	Thailand	69.6	Croatia	73.3
Eastern Africa	45.7	Peru	68.0	Vietnam	67.2	Greece	78.0
Burundi	40.6	Suriname	70.1	*Southwest Asia*	67.9	Italy	78.2
Comoros	58.8	Uruguay	73.9	Armenia	72.4	Malta	77.6
Djibouti	45.5	Venezuela	72.4	Azerbaijan	71.0	Portugal	75.2
Eritrea	51.5			Bahrain	72.9	Slovenia	75.0
Ethiopia	44.5	**Oceania**	73.5	Cyprus	77.8	Spain	78.1
Kenya	52.2	*Australia and New Zealand*	78.4	Palestine	71.4	Macedonia (former Republic of Yugoslavia)	72.7
Madagascar	51.6	Australia	78.7	Georgia	72.7	Yugoslavia, Federal Republic of	72.2
Malawi	40.7	New Zealand	77.2	Iraq	58.7		
Mauritius	70.7	*Melanesia*	58.7	Israel	78.3		
Mozambique	40.6	Fiji	68.4	Jordan	69.7	Total for world	65.4
Reunion	73.8	New Caledonia	74.0	Kuwait	75.9		
Rwanda	39.4	Papua-New Guinea	55.6	Lebanon	72.6		
Somalia	46.9	Solomon Islands	67.4	Oman	70.5		
Tanzania	51.1	Vanuatu	67.2	Qatar	68.9		
Uganda	41.9	*Micronesia*	71.8	Saudi Arabia	70.9		
Zambia	40.5	Guam	73.5	Syria	70.5		
Zimbabwe	42.9	Polynesia	70.3	Turkey	69.0		
Southern Africa	55.4	French Polynesia	71.7	United Arab Emirates	74.6		
Botswana	44.4	Western Samoa	68.5	Yemen	59.4		
Lesotho	51.2						
Namibia	45.1						
South Africa	56.7						
Swaziland	50.8						

From United Nations, 2001.

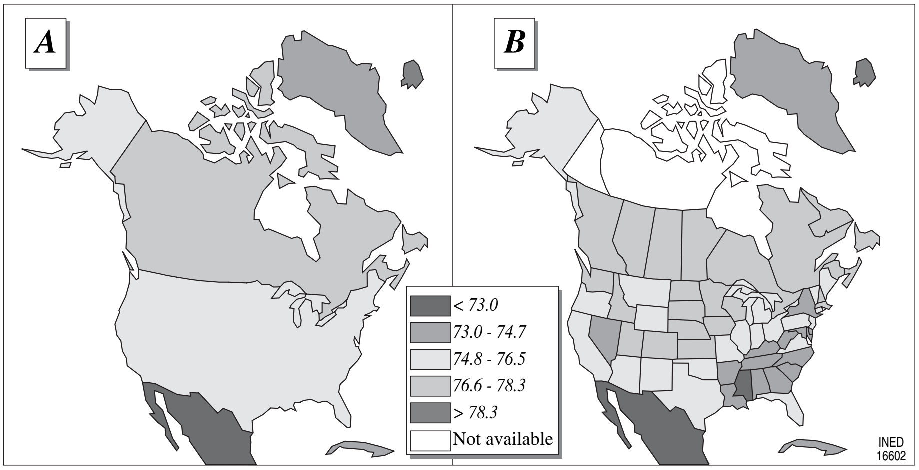


FIGURE 55–2 Life expectancy at birth for both sexes in North America in 1990, without **(A)** and with **(B)** state or province borders within the United States and Canada. (Figure 55–2 also reproduced in color plate section.)

large countries such as the United States or China to smaller neighboring countries.

Figure 55–2 illustrates the case of North America compared with its closest neighbors in Latin America (e.g., Mexico, Cuba) and in Europe (e.g., Iceland). If we follow national borders, nothing particular can be observed, except that Canada does better globally than the United States (see Fig. 55–2A). However, if we establish distinctions between the different states or provinces that make up these two federations, we see that the situation of the north-central part of the United States resembles that of the Canadian provinces, whereas that of the southeastern part of the United States bears a closer resemblance to Cuba or Mexico (see Fig. 55–2B). The fact that these states have large African-American or Latin American minorities probably plays a role. In any case, it appears that neither the border between the United States and Canada nor the Rio Grande have any more significance in terms of health than the line separating the U.S. southern states from the northern states.

Another striking example is that of China compared with its neighbors (Fig. 55–3). Like the United States, China is an enormous expanse, apparently homogenous, in the middle of a diversity of smaller countries, among which is Japan, which holds the record of highest life expectancy, and Burma, whose life expectancy is one of the lowest in the world (see Fig. 55–3A). We refer to figures from the early 1980s to be able to use the life expectancy estimates for every Chinese province established on the basis of the 1982 census (Calot and Caselli, 1989).

If we look separately at the Chinese provinces, about 10 of which have populations larger than those of France, Italy, or the United Kingdom, we can see how they are divided in relation to the surrounding countries. The situation of the far-western provinces resembles that of South Asia (e.g., Pakistan, India, Nepal, Bhutan, Bangladesh, Burma, Laos, Vietnam), whereas the maritime provinces (e.g., Formosa) are colored blue and are closer to Japan (see Fig. 55–3b).

This problem, which is obvious in the case of the United States or China, exists at all levels, even within countries or smaller geographic units (Decroly and Grimmaud, 1991; ISTAT and ISS, 1988). Heterogeneity is common, and the finer the territorial division, the more precisely this heterogeneity can be defined.

In the case of Italy, as we move down in the territorial hierarchy from the 20 regions to the 95 provinces or 8100 communes,[4] we obtain a better picture of the true situation, although a complete description of geographic variations can never be obtained. In this case, to perform such a comparison, given the insufficient data, we had to replace life expectancy by a comparative index of relative mortality, but the result is about the same.

The three maps shown in Figure 55–4, which represent the variations of male mortality, reproduce globally the same well-known profile of excess mortality in the northern region and low mortality in the southern region. However, this picture is considerably qualified when we divide the regional map into smaller areas. On the regional map, we observe a striking contrast between the regions situated north of the Po River (i.e., Aosta Valley, Piedmont, Liguria, Lombardia, Veneto, Trentino-Alto-Adige, and Friuli-

[4] According to the communal boarders in force in 1991, the year of the census, on which the analysis of mortality for 1991 to 1993 is based.

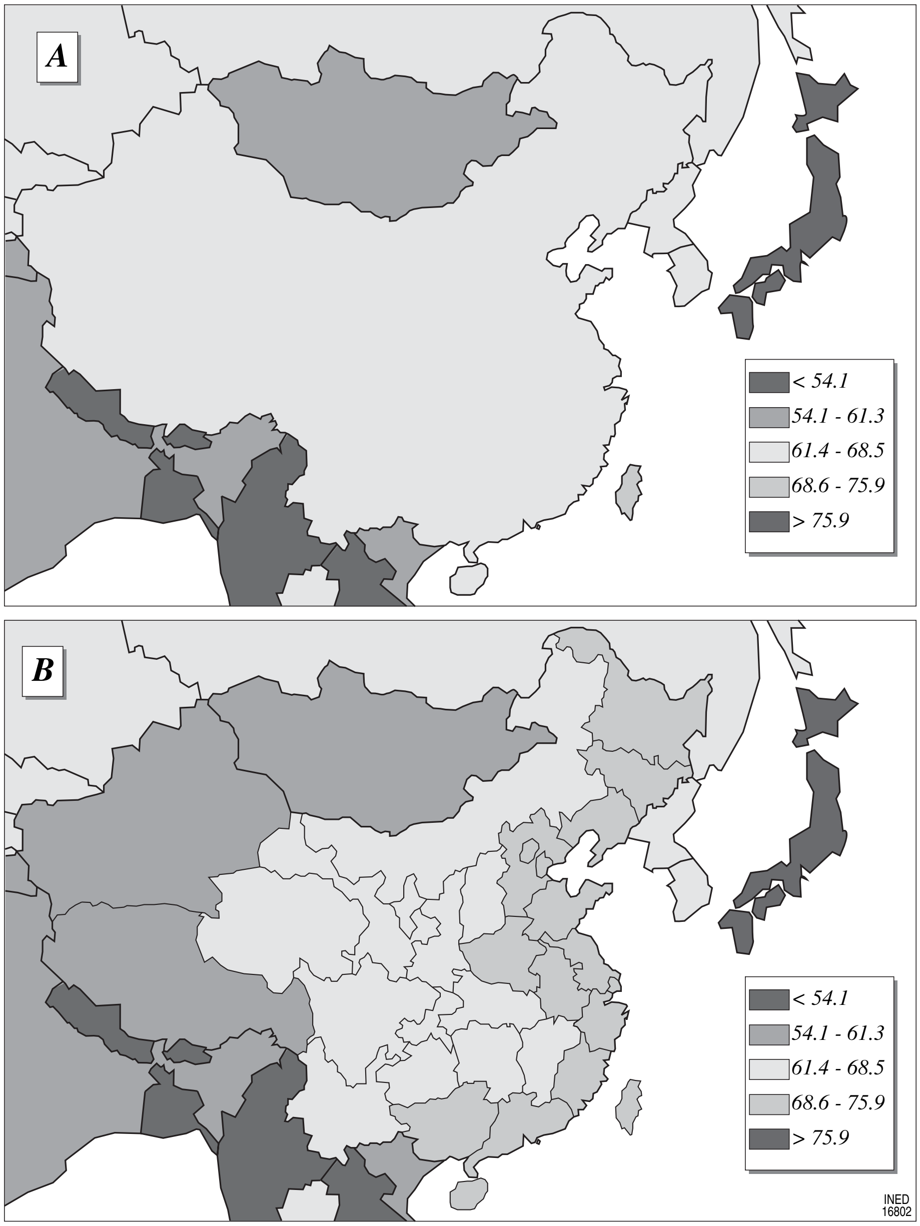


FIGURE 55–3 Life expectancy at birth for both sexes in eastern Asia in the early 1980s, with **(A)** and without **(B)** borders between Chinese provinces. (Map based on data from Calot and Caselli [1989].) (Figure 55–3 also reproduced in color plate section.)

Venezia-Giulia), which have a higher mortality, and the regions closer to the Adriatic Sea (i.e., Marches, Umbria, Abruzzi e Molise, and Puglia) and the Ionian Sea (i.e., Basilicata and Calabria), where mortality is lower. Between these two regions, there is an intermediate zone with mortality rates close to the national average, which includes the islands (e.g., Sardinia, Sicily), Emilia-Romagna, Tuscany, and Latium. The only exception to this apparently clear-cut division is Campania, colored red on the map, in the middle of blue or yellow regions.

If we look at the map of provinces, the situation appears more complex. We observe that excess mortality does not affect the area north of the Po as homogeneously as it first seemed. Piedmont and Liguria, for example, contrast with the provinces close to Lombardia, especially in the northeast Piedmont (i.e., Novara and Vercelli) and in the eastern part of Liguria (i.e., Genova), where male mortality is high, and all the western provinces, where mortality is close to the national average. Similarly, the region of Veneto includes only several high-mortality provinces (e.g.,

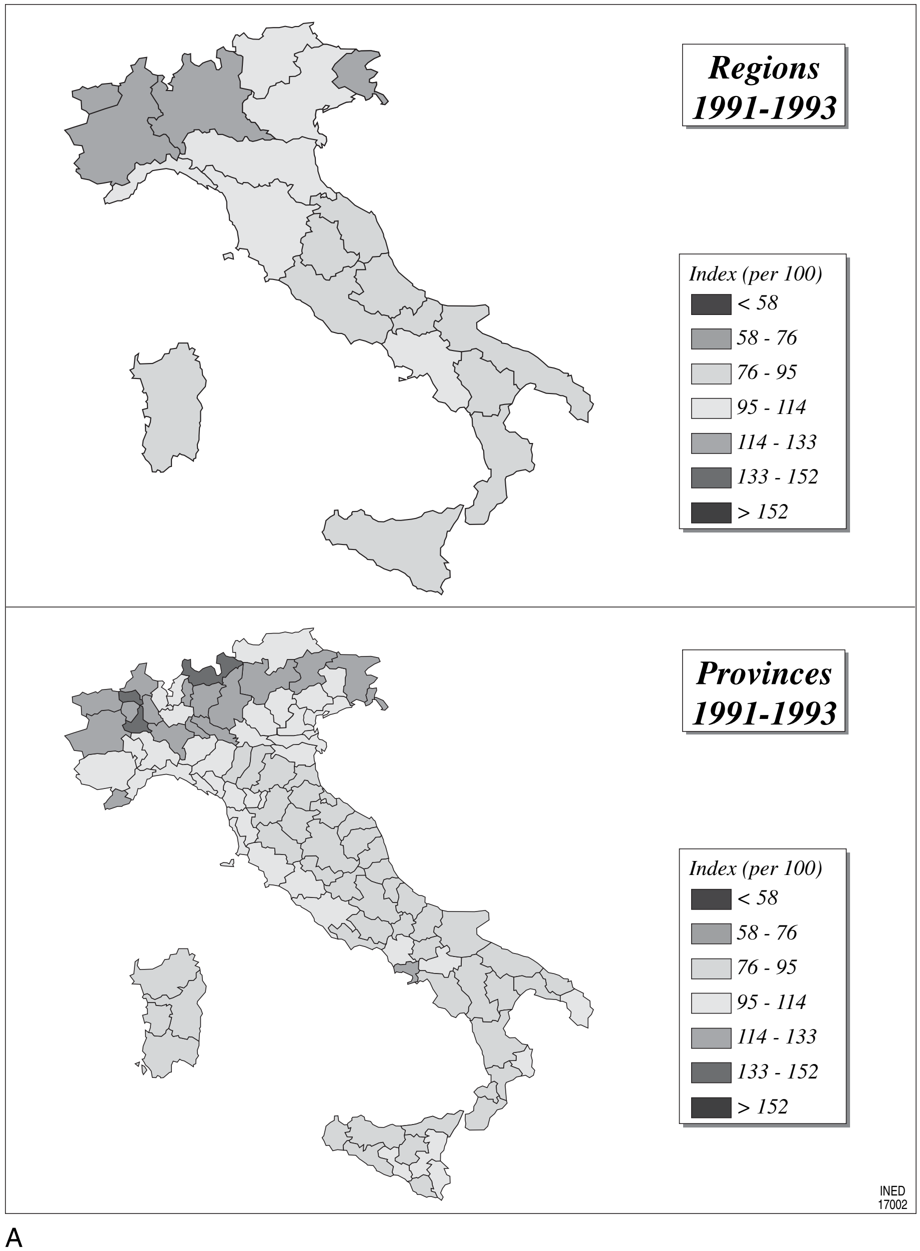


FIGURE 55–4 Geographic variations of Italian male mortality according to three levels of administrative territorial division: regions, provinces, and communes, in 1991–1993. **A:** Relative mortality index (per 100) by region and province. (Maps based on data from ISTAT [2000], Caselli [2002], and Caselli *et al*. [2003].) (Figure 55–4 also reproduced in color plate section.)

Belluno, Vicenza, Rovigo), whereas mortality in the others is about average.

In Tuscany, where mortality seemed to be average for the country, only three provinces (i.e., Grosseto, Livorno, and Lucca) are on average level, whereas the northernmost province (i.e., Massa-Carrare) has the excess mortality of northern Italy; all the other provinces, which are more eastern, have low mortality rates.

However, the most interesting case is that of Campania, identified as an exception. Whereas this region, taken as a whole, presents moderate excess mortality, it is made up of four provinces, two of which have very high mortality (i.e., Naples and Casertae), two of which have very low mortality (i.e., Benevento and Avellino), and a fifth province (i.e., Salerno) that has average mortality.

When we shift from the regional to the provincial level, excess mortality rates are located in smaller zones, whereas the lower mortality rates are spread over wider expanses, overflowing into Emilia, Tuscany, Latium, and Campania, as a continuation

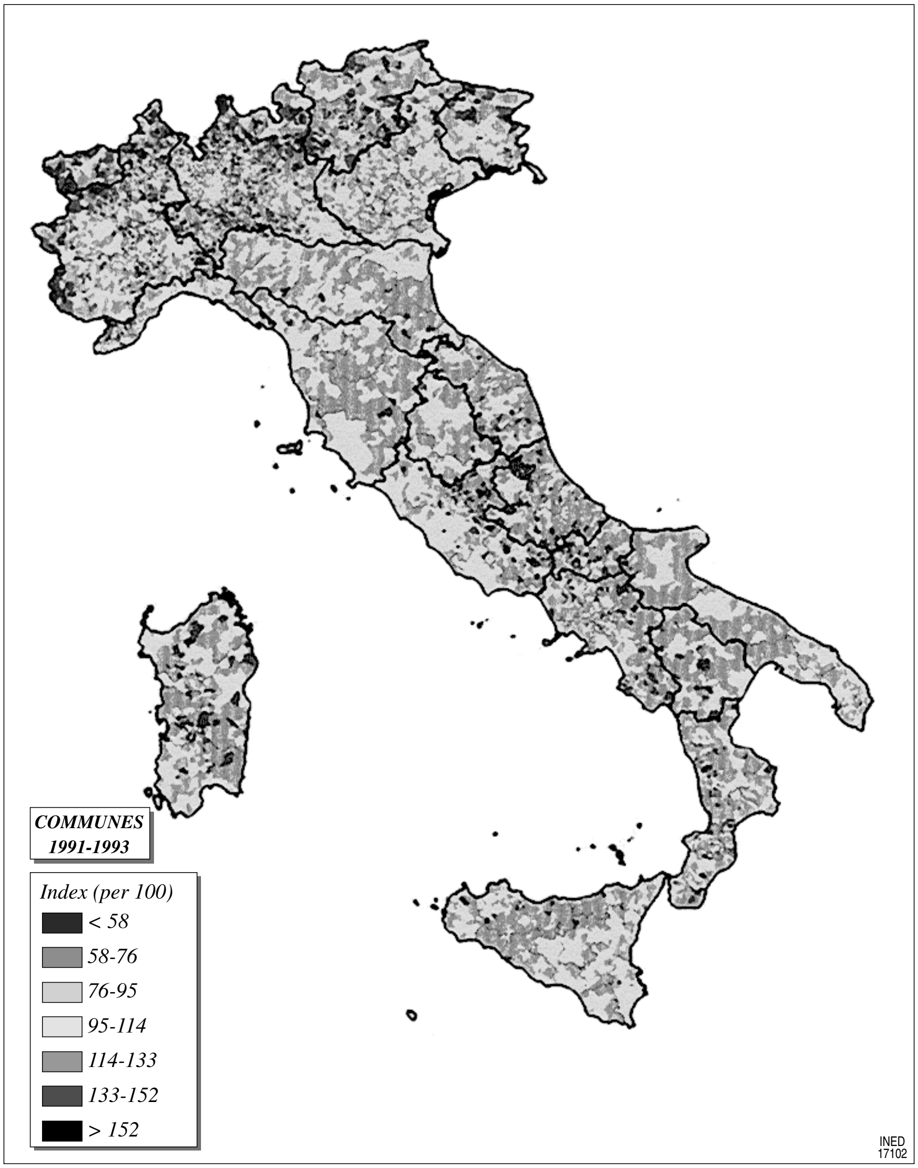


FIGURE 55–4 (*Continued*) **B:** Relative mortality index (per 100) by commune in 1991–1993. (Maps based on data from Cislaghi [2000].) (Figure 55–4 also reproduced in color plate section.)

of the regions where levels are low everywhere. Conversely, only four provinces (i.e., Tarente, Brindisi, Reggio-Calabria, and Aquila) are excluded from the low-mortality area, but some provinces of Sicily and Sardinia can be included. In shifting from the regional to the provincial level, the weight of some cities draws some provinces toward higher mortality rates, whereas the more rural provinces tend to have lower mortality levels.

The next step, the shift to the *commune* or municipal level (see Fig. 55–4B), allows still greater precision in defining the complexity of the division between high and low mortality. Some municipalities located in northern regions with high mortality have very low mortality, such as some communes of the Aosta Valley or other Alpine provinces and some communes of Liguria. Conversely, a few communes in low-mortality regions of central or southern Italy show excess mortality levels.

Figure 55–5 is a close-up view of the Campania region, showing the contrast between the communes that are the cause of the region's global excess mortality rates and those whose health conditions are no worse than elsewhere in the country. We see how the city of Naples contrasts with the communes located at the region's periphery, close to surrounding regions

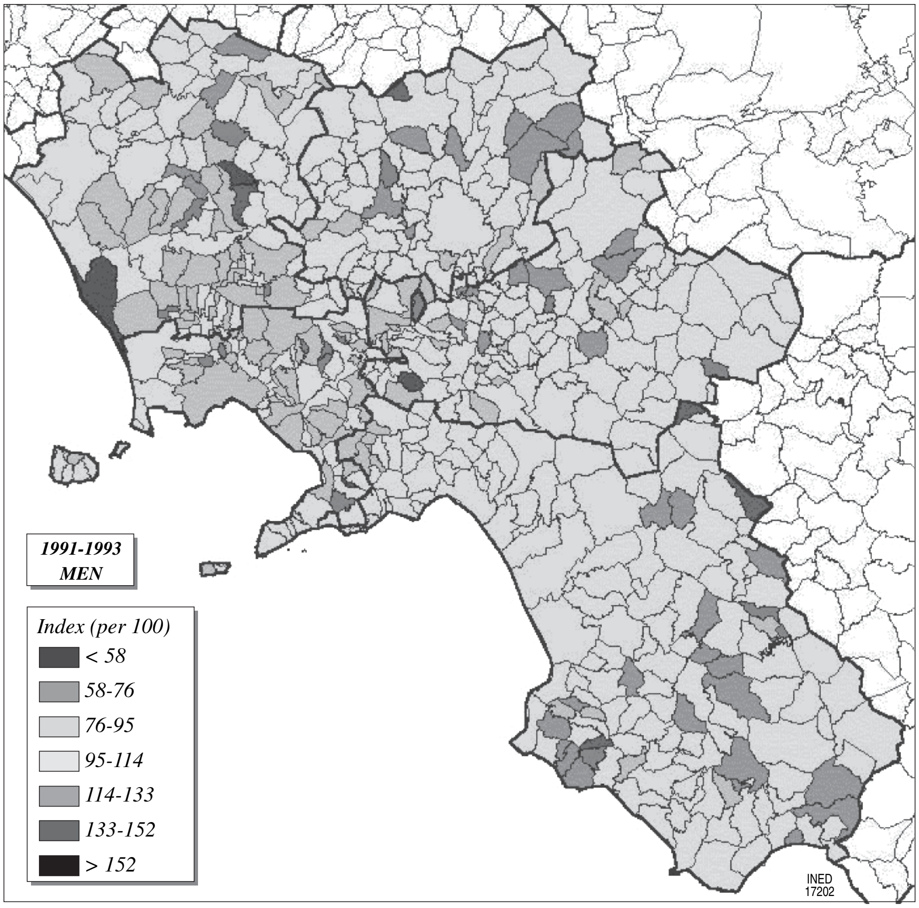


FIGURE 55–5 Closer view of the geography of mortality in Campania: provinces and communes. (Maps based on data from Cislaghi [2000].) (Figure 55–5 also reproduced in color plate section.)

that have low mortality rates. Besides the province of Naples itself, which is made up almost exclusively of communes with high mortality, the four other provinces of Campania are divided, in various proportions, into an area of average to high mortality closer to Naples and an area of low mortality closer to the neighboring regions.

If the division into smaller units plays an important role in the image we get of reality, that image also depends on the lines along which divisions are made. This was demonstrated in the cases of North America and China, where the difference in size of the areas can be a problem; however, defining borderlines is not a simple matter. The administrative limits of a city, for example, do not necessarily separate zones that suffer more or less from urban pollution. For this reason, geographers suggest dividing a given territory into geographic units of constant shape and size (see Chapter 70). The available data cannot be applied to such an operation. An analyst would have to use complex adjustment methods. In this case, such adjustments would be all the more relevant if the basic data were obtained on the smallest possible scale. To study mortality in a country such as Italy, the demographer would have to reach below the communal level.

b. Disproportion in Population Numbers

However, even analysis of populations below the communal level is not enough to solve the problem of unequal territorial division (i.e., inequality in the size of the areas covered or in the determination of area limits). It is difficult, when analyzing geographic variations of mortality not to address the issue of the size of the populations concerned.

In the geographic representations provided earlier, the same visual importance was given to two territories covering the same area, even if one of them had a population of a hundred million and the other had a population of only several million. In Figure 55–3, for example, the Chinese province of Szechwan, which according to the 1982 census had a population of 100 million, covers a smaller area than the Xinjiang

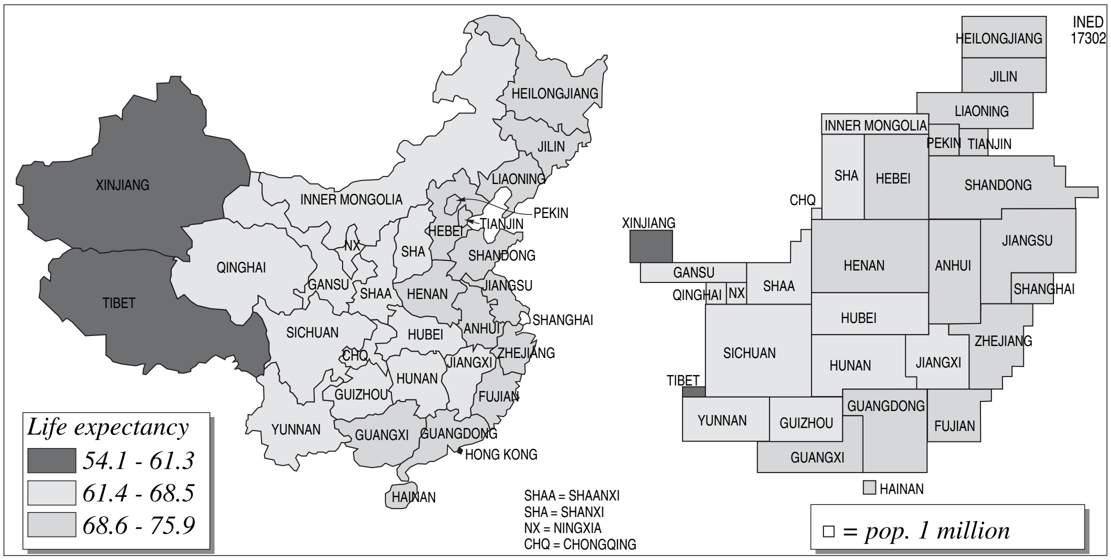


FIGURE 55–6 Geographic variations of life expectancy at birth in China in 1981: traditional representation and representation in proportion with population numbers. (Maps based on data from Calot and Caselli [1989] and from the United Nations [1999b].) (Figure 55–6 also reproduced in color plate section.)

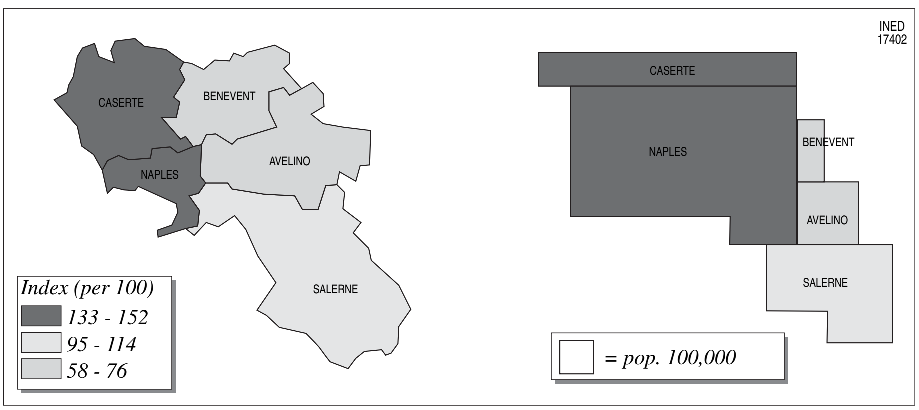


FIGURE 55–7 Geographic variations of the relative mortality index for males in the five provinces of Campania: traditional representation and representation in proportion with population numbers. (Figure 55–7 also reproduced in color plate section.)

province or Tibet, which had populations of 4 and 2 million, respectively. This means that the weight of the mortality levels observed in the latter provinces is exceedingly high compared with that observed in Szechwan.

To overcome this obstacle, we can draw schematic maps in which the area of each geographic unit is proportional to its population. Figure 55–6 isolates China's provinces as they appear on Figure 55–3 and compares them with a map of this type. The result is striking: The blue zone of low mortality is much larger than the rest; the orange zone of excess mortality is quite limited in size.

What is obvious for Chinese provinces is also true for more ordinary situations and much more widely so than often thought. There are no deserts in Italy, but we can observe the same phenomenon. For example, of the 5.6 million people of Campania, more than 3 million live in the province of Naples and 800,000 in the province of Caserte, the two provinces with high mortality, whereas only 400,000 live in Avellino and 200,000 in Benevento, both provinces with low mortality. The extension of the low-mortality zone, which was the result of the shift from the regional to the provincial level, was in part an optical illusion (Fig. 55–7). From a demographic point of view, what is most important is that the most-populated area is a high-mortality zone.

c. The Cities

A similar problem of scale occurs systematically when dealing with cities. Most often, the apparent map of mortality changes when a demographer calculates separately the mortality index of cities and that of the administrative units they belong to. One example is the case of mortality due to cancer in Baltic countries (Krumins *et al*., 1999). In these countries, the mortality due to cancer observed in cities is almost systematically higher than that of the corresponding administrative territorial unit. It is possible that this difference may be an observational problem (i.e., deaths due to cancer happen more often in cities because that is where the hospitals are located). Whether the phenomenon is real or artificial, the chosen system of representation changes our perception of the phenomenon. For most geographic units that include a city, the mortality index for cancer decreases if the city is accounted for separately (Fig. 55–8).

For greater precision, the cities could be represented by areas proportional to their population, which is not the case in the maps of Figure 55–8. However, this would be no more than an unsatisfactory half-measure if the other administrative units were still represented according to their real borders, without any indication of population size.

II. COMPONENTS OF GEOGRAPHIC INEQUALITY IN MATTERS OF MORTALITY

The previous example concerning mortality due to cancer prompts us to take a new look at the whole of our analysis of geographic variations of mortality. The differences in life expectancy or standardized mortality rates that we have accepted until now provide only an overall measurement. Within the same area of geographic observation, the mortality map can vary greatly depending on sex, age, causes of death, or other components. With the use of Italian data, we assess the roles played by these three basic components.

1. Reversal of the Geography of Mortality from Male to Female

Figure 55–9 shows to what extent the sex-specific mortality maps differ for Italy. The classic opposition between northern and southern areas, already illustrated in several ways in Figure 55–4, concerns mainly

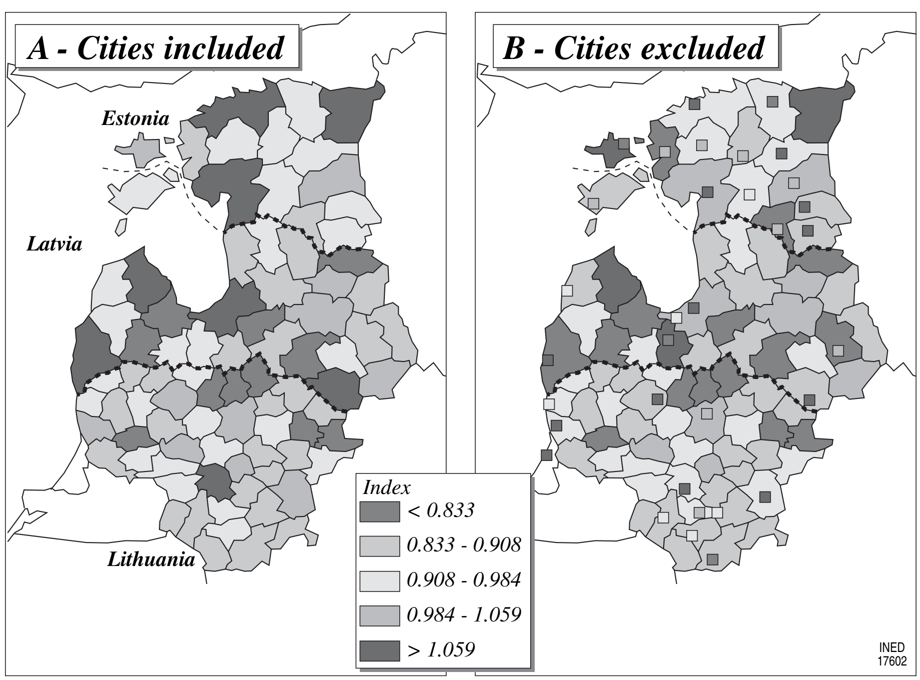


FIGURE 55–8 Variation of the index of comparative mortality due to cancer in Baltic countries for 1987–1990, depending on whether the cities are included **(A)** or considered separately **(B)** from the administrative territorial division to which they belong. (From Krumins *et al*., 1999.) (Figure 55–8 also reproduced in color plate section.)

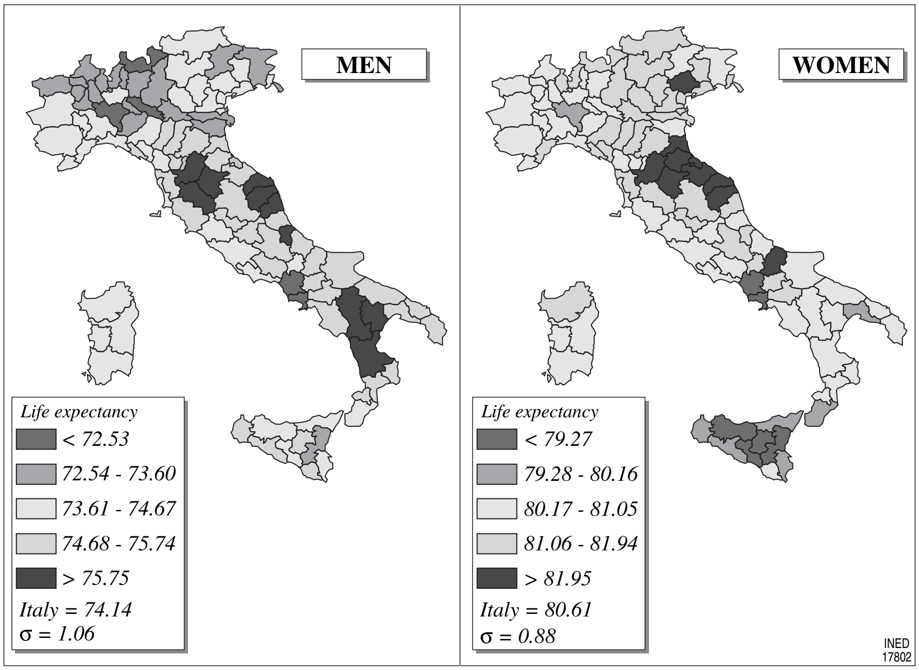


FIGURE 55–9 Geographic variations of sex-specific life expectancy at birth in Italy for 1991–1993. (Maps based on data from Caselli [2002].) (Figure 55–9 also reproduced in color plate section.)

male mortality. The situation for female mortality is completely different. We do observe a privileged area where the life expectancy at birth of both sexes is much higher than the national average; this area includes mainly the Marches, Umbria, and parts of Tuscany and Emilia. Similarly, both sexes have high mortality rates in Naples and Caserte. For the remaining areas, almost everything points to an opposite case. In southern Italy (e.g., Calabria, Puglia, Basilicate), male life expectancy is very high, whereas it usually is less than average for females. Sicily has the lowest female life expectancy, whereas male life expectancy figures are somewhat higher than average (at least in the southern and western provinces of the island). Conversely, in the northern provinces of Italy, where male life expectancy is usually low, women's life expectancy is at least equal to the national average and often higher (with the exception of Pavia). This major contrast between the distribution of male and female mortality in Italy is the result of complex combinations of multiple factors (Caselli *et al.*, 2003; Caselli, 2002). However, we may say, even though it is a slight simplification, that men suffer more than women from industrial pollution in the north of the country, whereas women suffer more than men from the state of underdevelopment in the southern area, particularly as concerns health care services (Vallin, 1981). The geographic map underscores the differences in life expectancy between the sexes and illustrates the changes that occur over time (Meslé and Vallin, 1998) (see Chapter 53).

2. Reversal of the Geography of Mortality from Infants to Adults

The reversal of the geographic map is even more spectacular when we shift from small children to adults. Although, as concerns the mortality probability between the ages of 40 and 65 years for males (Fig. 55–10, second map), the map is almost an exact replica of the map of life expectancy variations, the map of infant mortality is its complete opposite (see Fig. 55–10, first map). The northern region almost systematically has the lowest probability, whereas the southern region has the highest. Even more clearly than for the male-female difference, this reflect the fact that infant mortality is closely linked to the quality of health care, which is better in the northern region than in the southern region, whereas male adult mortality is much more affected by industrial pollution in the north. Geographic and other variations are most sensitive at adult ages (Van Houte-Minet and Wunsch, 1978).

The geography of mortality at older ages (65 to 84 years) is a little more complex. It bears more of a resemblance to adult mortality than to infant mortality. Similar to the rate for persons between the ages of

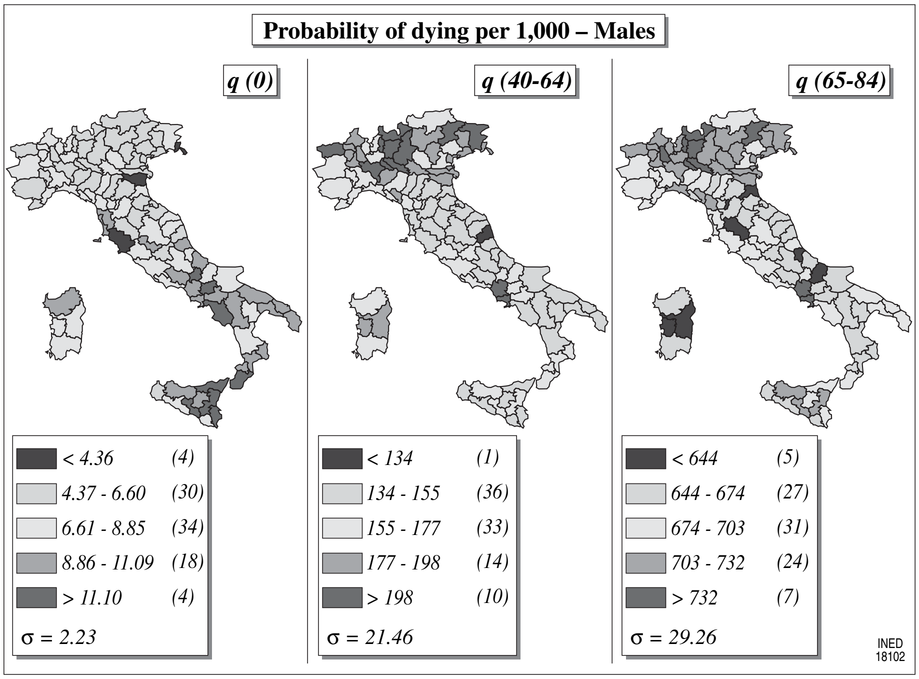


FIGURE 55–10 Geographic variations of the male probability of dying before the age of 1 year, between 40 and 64 years, and between 65 and 84 years in Italy for 1991–1993. (Maps based on data from Caselli [2002].) (Figure 55–10 also reproduced in color plate section.)

40 and 64 years, excess mortality can be found by far mostly in the northern part of the country, but low mortality is less frequently observed in other regions in the country than for the previous age groups. Almost all of the provinces of Italy or Sicily that are near the Tyrrhenian Sea, along with Puglia, are at or above the national average. Conversely, mortality is low in all of Sardinia, even though the rate is relatively poor at adult ages. The factors behind these specific aspects of mortality at advanced ages are complex. In the northern region, industrial causes play a negative role, but there are other causes in certain central provinces (e.g., urban pollution) and in the southern region (e.g., poor health care services, emergency care in particular).

3. Reversal of the Geography of Mortality from One Social Category to Another

The weight of social class in geographic variations of mortality is high. In a country such as France, for example, regional differences in mortality are for a large part explained by differences in the distribution of the population per social-occupational category. Social inequalities in terms of mortality are more or less pronounced, depending on the region. Whereas the Paris region has one of the highest life expectancy levels in France, the mortality of the poorer social groups is higher there than elsewhere. In other words, it often happens that the geography of mortality is different, even reversed, for different social classes (Desplanques and Nizard, 1978).

4. Reversal of the Geography of Mortality from One Cause to Another

The various regional contrasts reflect geographic variations in the cause of death, and we can observe complete reversals in this area as well. In Italy, the pattern of male mortality due to diabetes at age 40 to 64 years is almost the exact opposite of that of mortality due to bronchial and lung cancer (Fig. 55–11). Although bronchial and lung cancer affects mainly the northern part of the country, diabetes affects the southern part. The northern regions suffer from high levels of air pollution, with a combination of industrial emissions and frequent fogs, whereas the southern area is spared these two factors. Even more important is smoking, which is much more widespread in the northern region than in the southern region of Italy. Conversely, there are more factors favoring diabetes in the southern area (particularly sugar consumption), along with insufficient health services to control mortality due to this cause.

The geographic opposition between lung cancer and diabetes can be found in almost identical form for

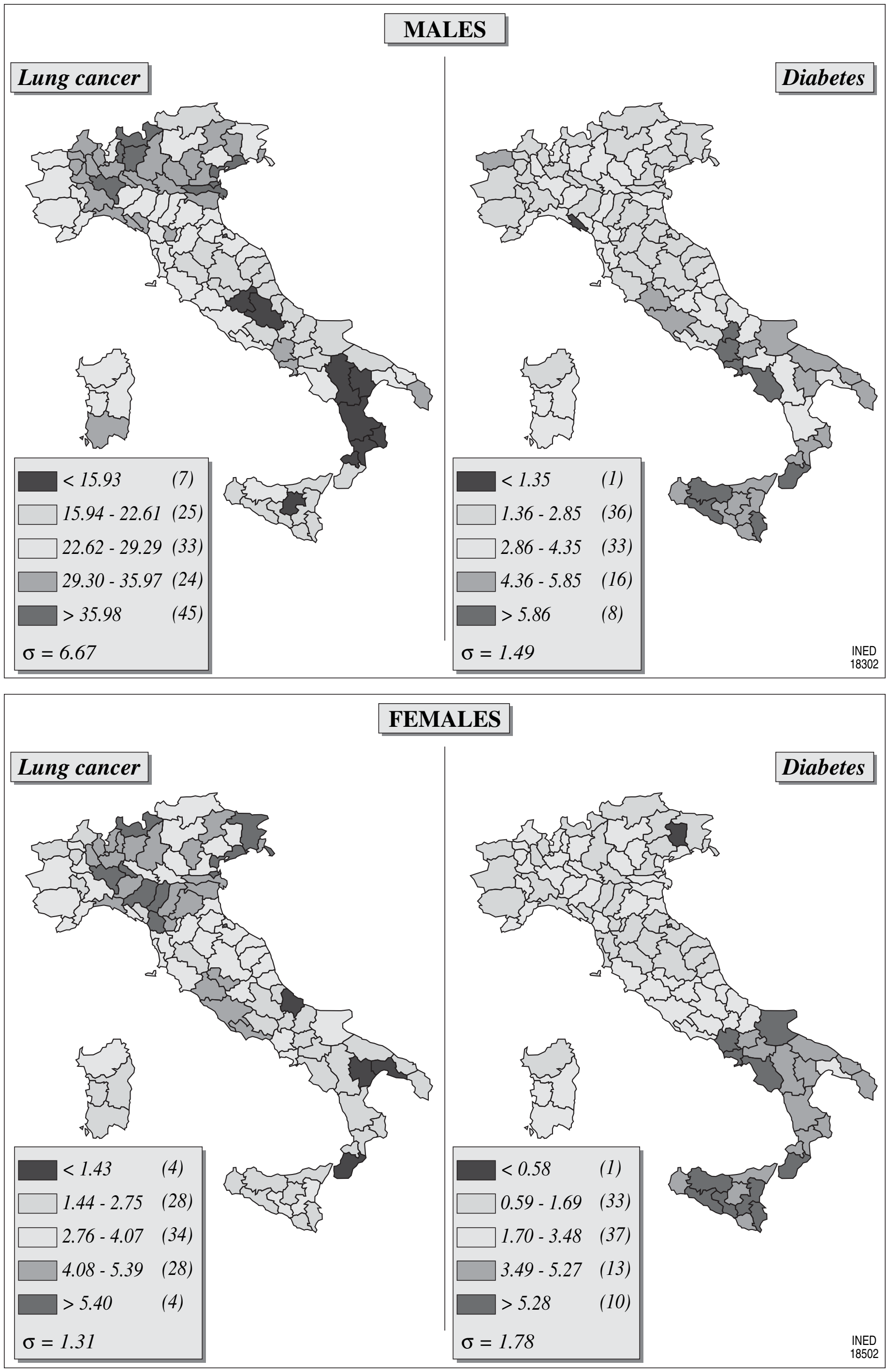


FIGURE 55–11 Geographic variations of the probability of dying of lung cancer or diabetes between 40 and 64 years of age in Italy for 1991 to 1993. (Maps based on data from Caselli [2002].) (Figure 55–11 also reproduced in color plate section.)

females (see Fig. 55–11, lower part), which is not surprising. Nonetheless, the difference in mortality for these two causes plays a role in the larger geographic opposition in mortality for males and females. Whereas diabetes kills about equal numbers of men and women, bronchial and lung cancers kill almost 10 times as many men than women. If these were the only causes, they would be enough to account for the male-female geographic contrast, but there are many other causes of mortality. Most of them, including ischemic heart disease (Fig. 55–12), have similar geographic profiles for both sexes, and the geographic contrast between males and females for mortality due to all causes is mainly the result of intensity differences.

However, for some causes, the geographic profile changes according to sex. This is particularly true for mortality due to "other diseases of the circulatory system" (mainly cerebrovascular diseases). For both sexes, there is excess mortality for this cause in the south, whereas most of the central regions have low mortality. However, if this mortality is very high for males in the northwestern regions, it is much less so for females, who enjoy rather low mortality rates due to this cause in the north. The negative situation in the south may be caused by certain eating habits that may favor cerebrovascular disease, but it is probably also affected by the lack of adequate emergency health care. This phenomenon can be observed for both sexes, but the negative factors prevalent in the north, which are mainly of industrial origin, affect mostly men. This contrast observed for the other diseases of the circulatory system is all the more important in the sex-specific mortality pattern in that this group of causes affects women much more than ischemic heart diseases, whereas the latter cause of death is more frequent among men.

III. PERMANENCE AND CHANGE

An analysis[5] of mortality trends in Europe since 1950, conducted under the aegis of the Council of Europe, has made it possible to bring together all the existing data and to provide estimates of life expectancy for all the European countries according to their current borders (Vallin and Meslé, 2001). We compared the situations observed in 1950, 1965, and 1995. In the mid-1960s, large gains in life expectancy were made thanks to the reduction of mortality due to infectious diseases and to advances in the treatment of cardiovascular diseases (Meslé and Vallin, 1993). What were the consequences for European mortality? To answer this question, Figure 55–13 was composed on the basis of statistical classes common to the three maps.[6] We can see how the maps are modified while keeping track of trends from one date to the next. What we observe is a radical transformation. In the early 1950s, there was a small group of Nordic countries (e.g., Scandinavian countries, Iceland, Netherlands) with life expectancies considerably above average. These countries, along with the United Kingdom, Belgium, Germany, and Switzerland, which were also above average, represented a north-northeastern corner spreading toward the center of Europe and bringing together the more prosperous countries. In striking contrast with this group, there was Eastern Europe (i.e., the entire Communist bloc except for the Czech Republic and Bulgaria) and a large part of southern Europe, the Iberian Peninsula, Yugoslavia, and Turkey. France, Italy, and Greece held an intermediate position. These different positions on the map show the stage reached at the time by each country in the first phase of the health transition, the phase during which most infectious diseases are eradicated (see Chapter 57). The northern countries were far ahead of those in the east and south, for which the transition occurred much later and which remained behind.

By 1965, the picture had completely changed. Progress had slowed down in the most advanced countries and accelerated in those that were behind, and in 15 years, the gap had narrowed. Only Turkey, Portugal, and two or three Yugoslav republics remained a little behind. The homogenization process was considerable, and most of the contrasts prevalent in 1950 had disappeared in a context of overall improvement of the situation.

In 1995, however, a new contrast emerged, more striking than that of 1950, with a clear opposition between East and West. All the western countries are colored blue. All those of the former USSR are colored brown or red, with the exception of the Transcaucasian republics, where we know that mortality is underestimated.[7] Between these regions, there are the former

[5] Analysis was based on previous work (Caselli *et al.*, 1999; Vallin, 1998; Caselli and Egidi, 1981a; van Poppel, 1979; Vallin and Chesnais, 1974).

[6] To do this, we calculated the average and the standard deviation for all the figures used for these three periods, and we created seven classes from a central class of one-half standard deviation centered on the average.

[7] This is not so much due to the under-registration of deaths, which is relatively moderate, as to the overestimation of the population. After the collapse of the USSR, the massive departure of nonlocal populations (especially Russians) was only partially taken into account in official population estimates (Badurashvili *et al.*, 2002; Yeganyan *et al.*, 2001).

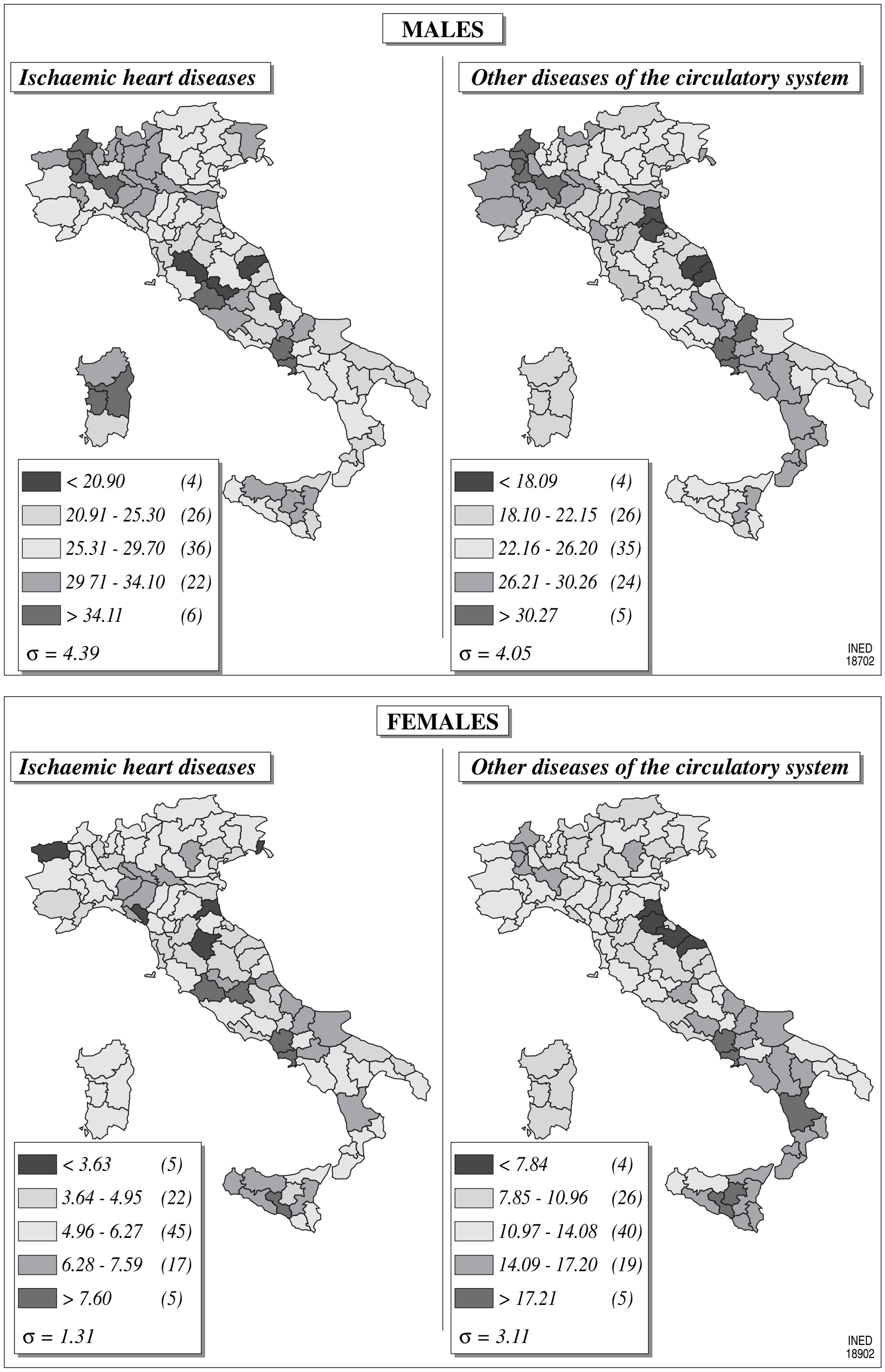


FIGURE 55–12 Geographic variations of the probability of dying of ischemic heart disease or other diseases of the circulatory system between 40 and 64 years of age in Italy for 1991 to 1993. (Maps based on data from Caselli [2002].) (Figure 55–12 also reproduced in color plate section.)

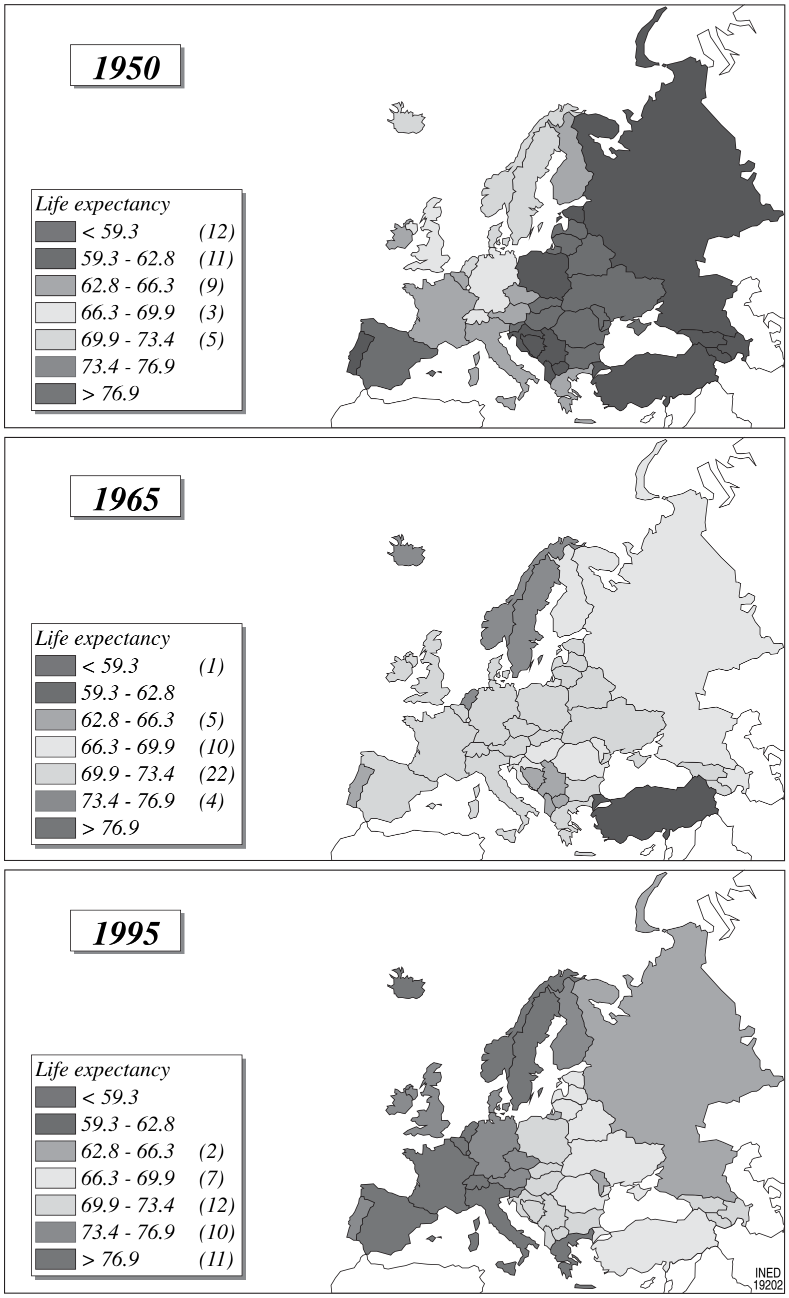
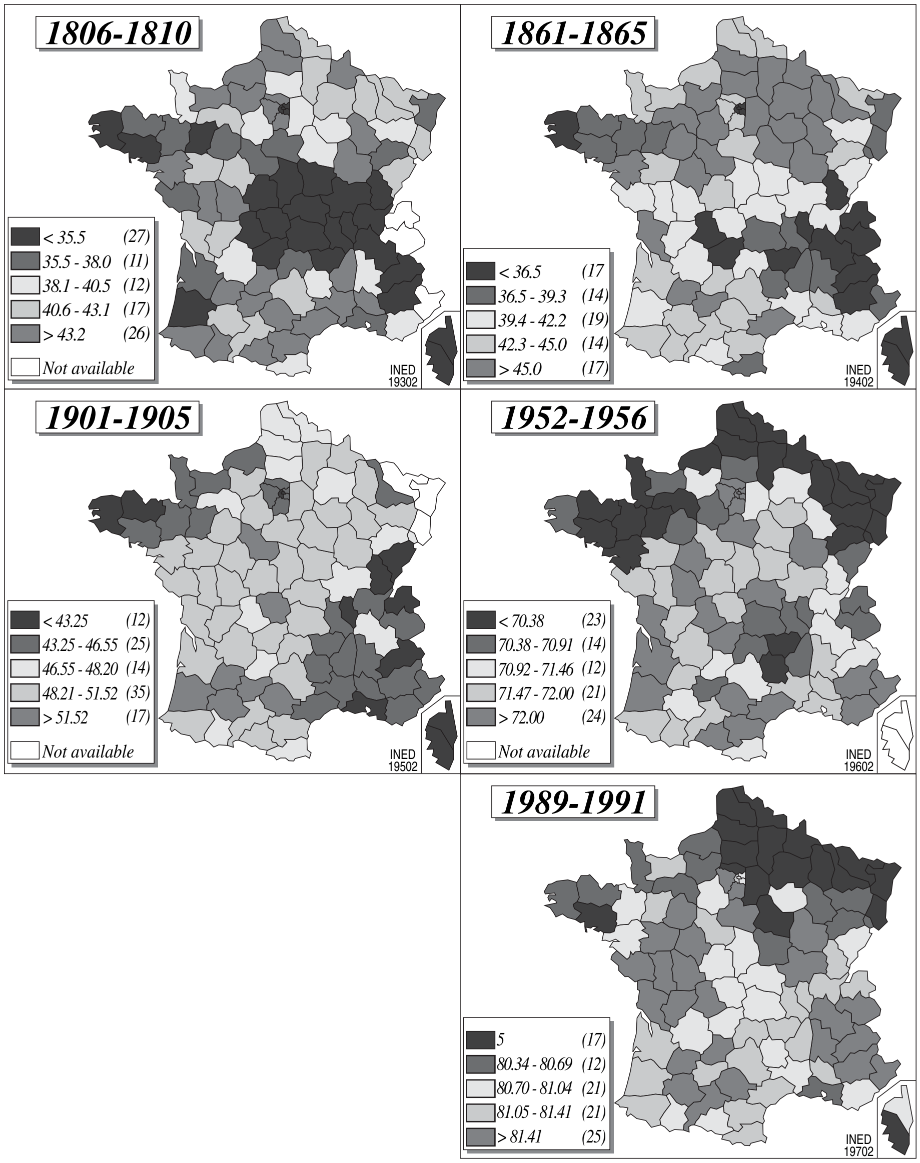


FIGURE 55–13 Life expectancy at birth for both sexes in Europe for 1950, 1965, and 1995. (Maps based on data from Vallin and Meslé [2001].) (Figure 55–13 also reproduced in color plate section.)

satellite countries of the USSR, the countries of the former Yugoslavia and of Turkey. Since the end of the 1960s, Western countries have succeeded in gaining more years of life expectancy thanks to advances made in the field of cardiovascular disease and some forms of cancer (i.e., the second phase of the health transition). However, the formerly Communist countries are experiencing a crisis. They have not succeeded in entering the second phase of the health transition, and certain social ills, such as alcoholism or violent deaths, have increased while health services have deteriorated, leading to an increase in the number of causes of death, which previously had been almost under control (Meslé *et al*., 1998; Meslé, 1991).

The example of Europe shows that strong political differences can have a great impact on the overall state of health in a country. Once the question of infectious



A

FIGURE 55–14 **A:** Trends by department in variations of life expectancy at birth of French women between 1806 and 1810 to between 1989 and 1991. (Maps based on data from Bonneuil [1997] for the 19th century and INSEE for the 20th century and on data from Isnard and Lavertu [1995] and INSEE [1964].) (Figure 55–14 also reproduced in color plate section.)

diseases—mainly a question of simple technical means—has been solved, overall progress in health issues depends on economic and social progress and on cultural changes leading to a modification in behavior and lifestyles that are detrimental to health (Vallin and Meslé, 2001; Caselli and Egidi, 1981d). Such factors are powerful enough to create a health gap between countries undergoing different political and economic trends, while leading to greater homogeneity between countries belonging to the same geopolitical group.

What can be said about internal variations within a single country? Usually, the trends move at a slower pace (Tonnelier, 1991), but they are no less significant. The map of life expectancy at birth for French women, for example, has completely changed in 2 centuries[8] (Fig. 55–14A).

[8] The comparison over 2 centuries is possible only for females. This was made possible by reconstitution of data for each département by Noël Bonneuil (1997) for the 19th century, and this research was carried out only for females.

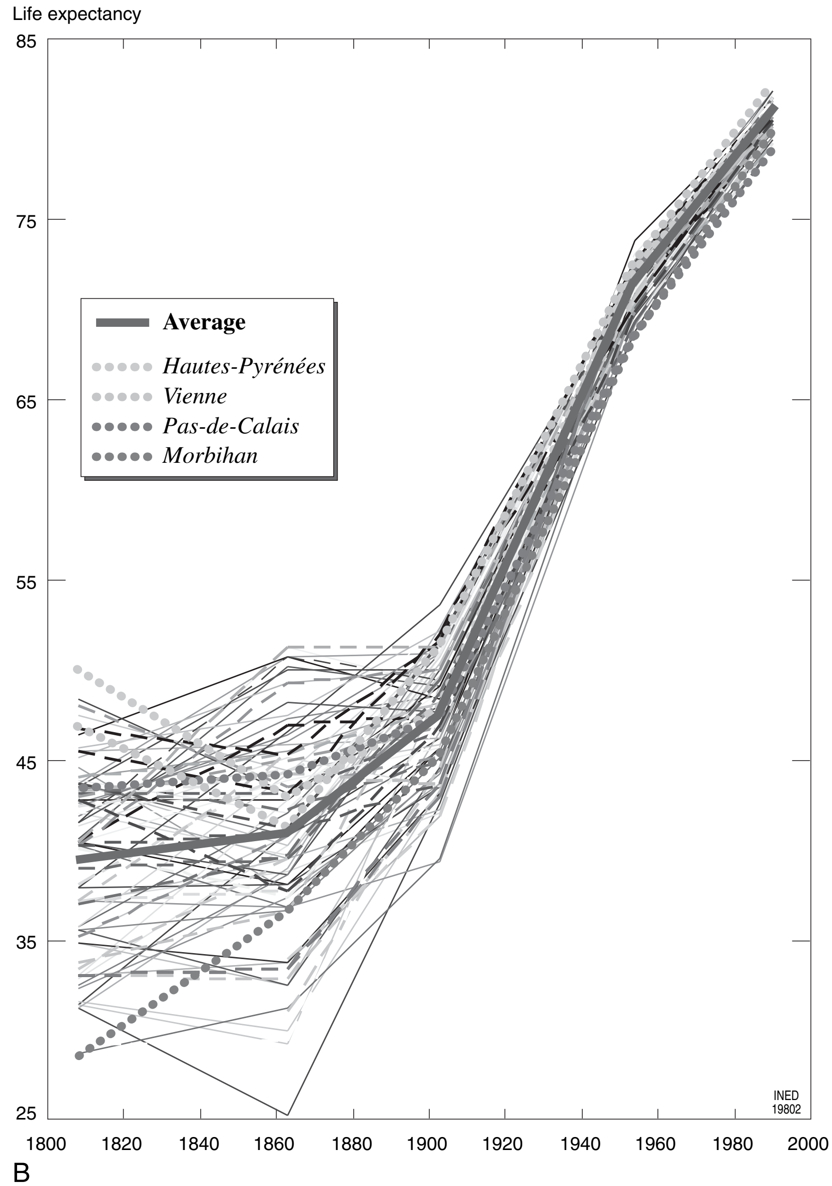


FIGURE 55–14 (*Continued*) **B:** Trends in female life expectancy by department in France since the beginning of the 19th century. (Data from Bonneuil [1997], INSEE [1964], and Isnard and Lavertu [1995].) (Figure 55–14 also reproduced in color plate section.)

In France at the beginning of the 19th century, in 1806–1810, there were two vast zones of low mortality in the northeast (i.e., Normandy, Nord-Pas-de-Calais, Picardy, Ile-de-France, Champagne-Ardennes, Lorraine, Alsace, Burgundy, and Franche-Comté) and in the southwest (i.e., Poitou-Charentes, Lot, Tarn, Pyrenees, and Languedoc), separated by a vast area of excess mortality in the center (i.e., Berry, Limousin, and Auvergne), with extensions into the Alps on one side and into Brittany on the other.

Fifty years later, in 1861–1865, excess mortality had declined in some parts of the center and become specific to the Alps and Brittany; in the north, however, the low mortality zone extended over the Paris basin (except for Paris itself and the suburbs), but it no longer included the Alsace region. In the early 20th century, the excess-mortality zone covered the entire southeastern part of France (from the Jura mountains to the Cévennes and Provence) and the Brittany-Normandy region, whereas mortality was low in the center and the southwest.

In 1952–1956, excess mortality had for a large part disappeared in the southeast, but had returned to the north and northeast areas. In those years, there appeared a pattern that was to remain for the next 50 years, with an opposition between the north, which had poor health conditions (from Brittany to Alsace, through Normandy and Pas-de-Calais, Picardy and Lorraine) and nearly all of the remainder of France (except Savoie and two or three *départements* of the Massif Central region). At that time, the dominant factor in geographic mortality variations was alcoholism (Meslé, 1990; Nizard and Prioux, 1975; Ledermann, 1956, 1964), even when taking into account specific regional factors, such as infant mortality in the Nord-Pas-de-Calais or industrial pollution in the north and northeast (Caselli and Egidi, 1988; Caselli 1984).

The years 1989–1991 show a change in this familiar pattern. Normandy in particular is no longer included in the excess mortality zone, which then ceases to be a continuous expanse. The zone of excess mortality covers several *départements* of the central region (i.e., from Yonne to Creuse, through Nièvre and Allier).

If we compare the first map with the last, the geography of mortality completely changed over 2 centuries, in large part because of economic and social changes. In the past, the areas with very high mortality were situated in the traditionally poor regions of Brittany, the Massif Central, and the Alps; the southwest was rich farming country, whereas the north and east areas were entering the economic development process tied to the industrial revolution. At the end of the 1980s, excess mortality was still a fact in Brittany, but it was especially marked in the north and the east of France due to the consequences of heavy industrialization.

At the same time, the geography of mortality also changed in terms of the inequality between regions. In the early 19th century, the average life expectancy of French women was 39.4 years; it was only 28.4 years in the Morbihan (Brittany), but it reached 50.0 years in the Hautes-Pyrénées, a difference of nearly 22 years and a relative gap of 76% between both extremes (Table 55–2). This difference does not result from extreme values, because the standard deviation was 5.1 years. In the mid-19th century, inequality increased still more; with just about the same life expectancy (41 years), the absolute gap was 26 years, from 25.2 years in the Hautes-Alpes to 51.2 in Yonne (i.e., standard deviation of 5.6 years). In the early 20th century, inequality strongly decreased. For an average life expectancy of 47 years, the absolute difference was only 14.2 years, from the Finistère (39.3 years) to the Loir-et-Cher (53.4 years). However, the differences were dramatically reduced during the 20th century, and by 1952 to 1956, the average life expectancy had greatly increased, reaching 71.2 years; the gap between extremes had

TABLE 55–2 Statistical Characteristics of the Variations in Life Expectancy at Birth of French Women at Different Times in History

Characteristics	Periods				
	1806–1810	1861–1865	1901–1905	1952–1956	1989–1991
Average	39.35	40.77	47.22	71.24	80.92
Standard deviation	5.14	5.63	3.33	1.08	0.68
Minimum	28.40	25.20	39.30	68.30	78.70
Maximum	50.00	51.20	53.50	73.20	82.40
Amplitude	21.60	26.00	14.20	4.90	3.70
Relative amplitude (%)	76.06	103.17	36.13	7.17	4.70

dropped to only 4.9 years (between Morbihan, with 68.3, and Alpes-Maritimes, with 73.2 years), with a standard deviation of 1.08. Between 1989 and 1991, the average life expectancy reached almost 81 years. There was a difference of only 3.7 years between Pas-de-Calais (78.7 years) and Vienne (82.4 years), and the standard deviation had dropped to 0.68 years.

Figure 55–14B illustrates this spectacular decrease in the geographic inequalities of French mortality. The trends observed in the different *départements* are extremely varied in the 19th century, but their differences suddenly fade during the 20th century, and life expectancy rapidly increases. To make our table easier to read, we have indicated the average value in bold red type, whereas the two *départements* with extreme values between 1805 and 1810 are in bold blue type, and those with extreme values between 1989 and 1991 are in bold green.

Changes in mortality patterns are even more striking in Italy. Figure 55–15 shows that the current pattern of excess mortality for males north of the Po River compared with low mortality in the center and south (with the remarkable exception of Campania) is the opposite of the situation observed in the beginning of the 20th century, when excess mortality was specific to the south, and the situation was better in the north (the center, however, already had a very privileged situation). The explanation is simple. In the past, when mortality mainly was attributed to infectious diseases, the low level of development of the south represented a greater disadvantage than the industrial pollution in the north. Today, the main causes of mortality are cardiovascular diseases and cancer, and industrial pollution in the north has a greater impact than the persistent underdevelopment of the south. The advantage of the central region, which has been spared the industrial pollution of the north but is very developed,

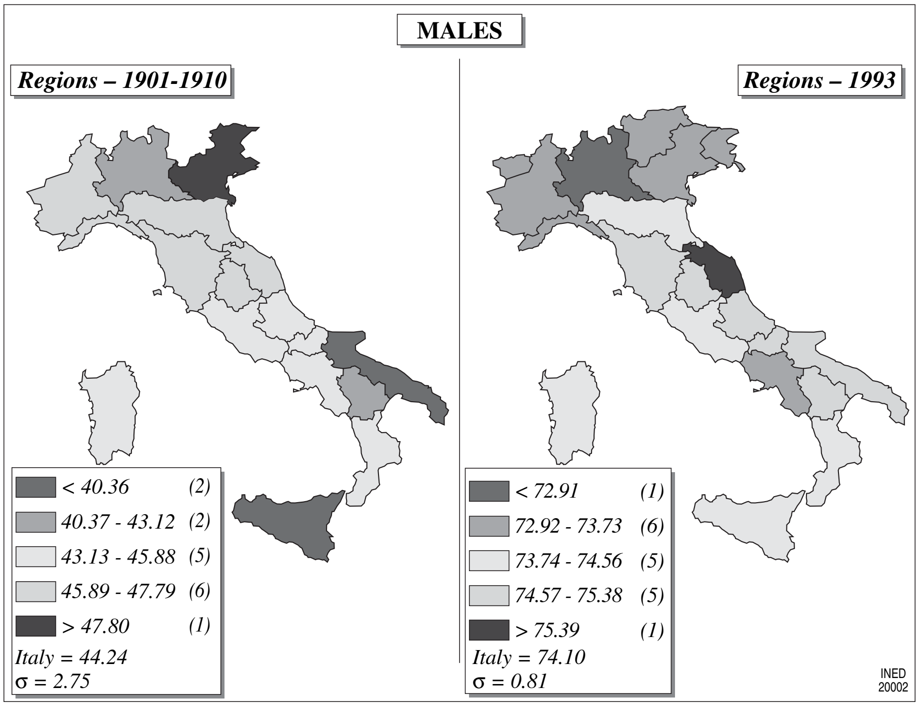


FIGURE 55–15 Trends in regional variations of male life expectancy at birth in Italy from 1901 to 1910 and in 1993. (From ISTAT, 1996a; Caselli, 1990.) (Figure 55–15 also reproduced in color plate section.)

remains permanent (Caselli *et al*., 1993; Caselli and Egidi, 1981b, 1981c).

This complete change of pattern mainly results from the near disappearance of child mortality. Child mortality had always been high in the southern region, whereas very early on, adult male mortality was higher in the north than in the south. At the beginning of the century, child mortality was still high in Italy (more than 100 deaths per 1000 children) and had a large impact on life expectancy, with a considerable advantage for the north. Today, it has decreased 10-fold (less than 10 deaths per 1000 children), and although it remains higher in the south than in the north, the consequences in terms of life expectancy are negligible. The difference in life expectancy now reflects adult excess mortality in the north (Caselli and Reale, 1999).

IV. FACTORS BEHIND GEOGRAPHIC INEQUALITIES

What can be said of these maps of mortality and causes of death? Overall mortality maps can be compared with of this or that variable that is thought to play a role in mortality differences. One example is provided by the two maps shown in Figure 55–16. The first is a map of male mortality at ages 25 to 54 years in each Italian province, and the second illustrates the proportion of the working population employed in mining and manufacturing industries. They are extremely similar, which suggests a strong correlation between male adult mortality and the harmful effects of these industries. There is also a surprising coincidence in France between the geographic distribution of olive oil consumption and that of the lowest rate of mortality due to colon cancer (Meslé, 1983). Although we could provide many such illustrations, this type of comparison remains very limited due to the number of factors involved, the necessity of comparing a large number of maps, and the difficulty of drawing meaningful conclusions on the basis of an intuitive reading of them.

Different methods of statistical analysis, such as principal components analysis, correspondence analysis, canonical analysis, hierarchical analysis, and regression models, can lead to more objective interpretations. With geographic interpretations based on areas that are small enough to be homogenous in terms of mortality, it is possible to identify more precisely high and low mortality zones and to verify or determine more precisely the relationships that exist between different mortality levels (as well as profiles by age, sex, and cause) and the environmental factors

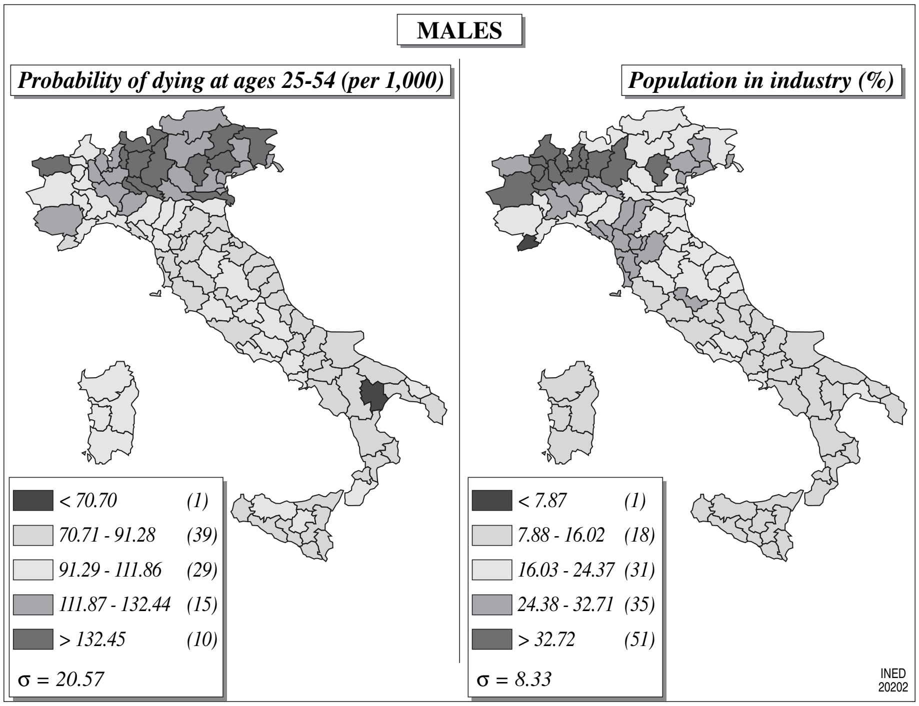


FIGURE 55–16 Geographic variations of male mortality at ages 25 to 54 years and of the proportion of the working population employed in mining and manufacturing industries in Italian provinces between 1971 and 1973. (From Caselli, 2002.) (Figure 55–16 also reproduced in color plate section.)

that may explain these differences. This approach, based on observations that were aggregated for a geographic area and not for individual data, does not enable us to establish any causal relationships between external variables and mortality, but it can suggest their existence and provide assumptions to explain observed differences; such assumptions can be verified only through in-depth epidemiologic studies based on individual data.

To illustrate how this type of analysis can contribute to the interpretation of geographic variations of mortality, we refer to the main conclusions of a principal components analysis carried out on the geographic variation of male mortality in Italy between 1971 and 1972 for two age groups representative of the adult (25 to 54 years) and elderly population (55 to 79 years) (Caselli and Egidi, 1981c). At that time and for these two age groups, the mortality maps were about the same as those in Figure 55–10 for the 40 to 64 and 55 to 84 age groups between 1991 and 1993, with the same striking contrast between northern and southern parts of Italy. In this study, the cause-specific geographic differences in mortality (Table 55–3, 16 groups) observed in the 94 Italian provinces were compared with 38 indicators, including socioeconomic and environmental indicators, indicators pertaining to eating behavior, and other behavior indicators than can be linked to health[9] (Table 55–4).

Thanks to the image obtained by projection of the 94 province-points on the factorial plane, the proximity of the provinces, in terms of analogy of behavior pertaining to the variables, becomes significant. The projection of the 16 + 38 variables makes it possible to attribute a significance to the proximity of the

[9] Briefly recall that the principal components analysis is a method of factorial analysis that makes it possible to reduce the dimensions of the data matrix with the use of synthetic variables (factors) obtained through linear combination of the initial variables. This analysis is based on a geometric interpretation of the data matrix, according to which each of the n (94 for Italy) provinces is assimilated to a point whose coordinates are the p (16 + 38) variables (causes of death and external variables). Similarly, each variable can be interpreted as a point whose coordinates are the n values taken from the provinces. The provinces and the variables are elements belonging to two Euclidian vectorial spaces whose dimensions are, respectively, p and n. The procedure consists in projecting the cloud of 94 points-provinces, as well as the 16 + 38 variables, on a subspace of fewer dimensions but that guarantees the lowest possible distortion of the initial image (see Bouroche and Saporta, 1980). The first two or three factors almost always suffice—especially in the field of demography—to explain most of the total dispersion of points, and the first factorial plane (formed through the first two factors) reflects most of the dispersion.

TABLE 55–3 Groups of Causes of Death Used for the Principal Components Analysis of the Geography of Italian Mortality

No.	Code	Group of causes	ICD code[a]
1	MIP	Infectious and parasitic diseases	000–139.9
2	TMD	Malignant neoplasm of digestive organs and peritoneum	150–159.9
3	TMR	Malignant neoplasm of respiratory system	160–163.9
4	TMU	Malignant neoplasm of genito-urinary organs	180–189.9
5	AT	Other malignant neoplasms	140–149.90; 164–179.9; 190–239.9
6	D	Diabetes mellitus	250
7	AGN	Other diseases of endocrine glands, nutritional and metabolism disorders	240–249.9; 251–279.9
8	CI	Ischemic heart disease	410–414.9
9	AAC	Other diseases of circulatory system	290–409.9; 415–458.9
10	B	Acute bronchitis and bronchiolitis; bronchopneumonia, unspecified; chronic bronchitis; emphysema and asthma	466; 485; 490–493.9
11	AAR	Other diseases of respiratory system	460–465.9; 467–484.9; 486–489.9; 494–519.9
12	MAD	Diseases of the digestive system	520–577.9
13	MAU	Diseases of genitourinary organs	580–629.9
14	SEM	Symptoms and ill-defined conditions	780–796.9
15	AET	Accidents, poisoning, and traumas	E800–E999.9
16	AM	Other conditions	280–289.9

[a] International Classification of Diseases: detailed list of the eighth revision.
From Caselli and Egidi, 1981c.

provinces in terms of correlation: The nearer each other on the plane, the better correlated are the distributions of two variables, all the more so as they are distant from the origin.

The proximity of the points (provinces or variables) shows their homogeneity, but in this analysis, the demographer cannot project simultaneously the province-points and the variable-points, because they do not belong to the same cloud or to the same space. However, it is possible to compare the respective positions of two or several provinces with the total group of variables and *vice versa*.

To interpret the results, it is particularly important to attribute a meaning to factors, because they are linear combinations of the initial variables. To this end, we used the coordinates of the variable-points that express correlations between variables (i.e., cause-specific death probability and external variables) and the factorial axes.

An initial analysis, conducted with only the external variables, enabled us to show the relationships between variables and to divide the territory into geographic areas according to their dominant characteristics. We observed that the first factorial axis opposes variables linked to economic zones at a lower stage of development to those pertaining to the most developed zones. The provinces all follow this axis, from south to north according to the well-known social economic division of Italy. This dichotomy, as we have seen, can be found in the pattern of adult and elderly mortality, although in a form contrary to what might have been expected, because adult and elderly mortality is higher in the wealthier regions of the north and lower in the poorer regions of the south. The second factorial axis, which is tied mainly to the geographic distribution of the population, establishes an opposition between the populations of the sparsely populated rural and mountainous regions (i.e., Alpine and Apennine regions) and the large urban centers such as Rome, Milan, and Naples.

If we apply the principal components analysis to the entire group of 16 + 38 variables, the joint study of the territorial distributions of causes of death and external variables suggests interesting considerations for the research of factors underlying the observed geographic differences. For the sake of brevity, we do not show the projection of the provinces on the first factorial plane; we instead concentrate on the projection of the variable-points (Figs. 55–17 and 55–18), while calling attention to the specificities of the geography of different causes of death, as well as the meaning given to the two main axes of the analysis of external variables.

Between the ages of 25 and 54 years, the first factorial axis of this joint analysis (which explains 27% of the variance) offers two pieces of information: one concerning economic development, especially in the industrial sphere, and the other the high rate of mortality, closely tied to the high probability of dying of

TABLE 55–4 External Variables Used for the Principal Components Analysis of the Geography of Italian Mortality

Number	Code	Variable
Use		
1	Activity	Rate of activity
2	Salaried	Proportion of salaried workers
3	Construction	Proportion of the working population salaried in construction and civil engineering
4	Indupop	Proportion of the working population salaried in mining and manufacturing industries
5	Induprov	Proportion of working population working in mining and manufacturing industries among those working in the province
6	Dimind	Average dimension of the industry
Geographic mobility of workers		
7	+1 h	Proportion of persons commuting 1 hour or more to get to work
8	Commute	Proportion of commuters who drive to work every day
Risks at work		
9	Malprof	Incidence of diseases contracted at the workplace in industry
10	Accind	Incidence of accidents at the workplace leading to death or permanent invalidity in industry
Environment and demographic density		
11	Density	Population density per km 2
12	Size/household	Average number of persons per household
13	Urban	Proportion of persons living in communes with population >20,000
14	Agglom	Proportion of persons living in an urban environment
15	Mountain	Proportion of persons living in the mountains
16	SO_2	Emission of SO_2 air pollution indicator
17	Pollwater	Estimate of degree of water pollution
18	Radiac	Basic radioactivity
Climate		
19	Precip	Annual average precipitation in millimeters
20	Excuther	Thermal excursion
Housing		
21	<1919	Old housing (proportion of private dwellings built before 1919)
22	Bath	Proportion of private dwellings with running water, bathroom, or water closet (WC)
23	Heating	Proportion of private dwellings with central heating
24	Persons/room	Average number of persons per room
Medical and hospital infrastructure		
25	Physicians	Average number of practicing physicians for 1000 inhabitants
26	Beds	Number of beds for adults available for 1000 inhabitants
27	Assistance	Public expenditure for assistance
Consumption		
28	Bread	Consumption of bread and pasta
29	Sugar	Consumption of sugar
30	Fat	Consumption of vegetable and animal fats
31	Meat	Consumption of meat and fish
32	Milk	Consumption of milk and cheese
33	Non-food con.	Non-food consumption
34	Smoking	Consumption of tobacco
Income and standard of living		
35	Income	Net income per capita
36	Car	Average number of privately owned cars per capita
Education		
37	Instruction	Proportion of persons 15 years old or older having had at least 8 years of education
Marital status		
38	Marital status	Ratio of unmarried and widowed persons 15 years old or older to married persons

From Caselli et Egidi, 1981c.

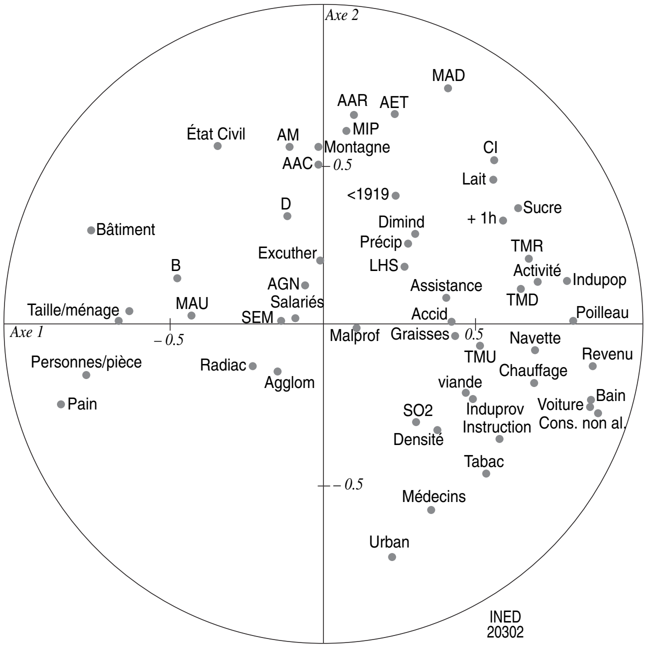


FIGURE 55–17 First factorial plane in the principal components analysis of male mortality between the ages of 25 and 54 years that is associated with the external variable matrix for Italian provinces between 1971 and 1972. See Tables 55–1 and 55–4 for causes of death and external variables. (From Caselli and Egidi, 1981c.)

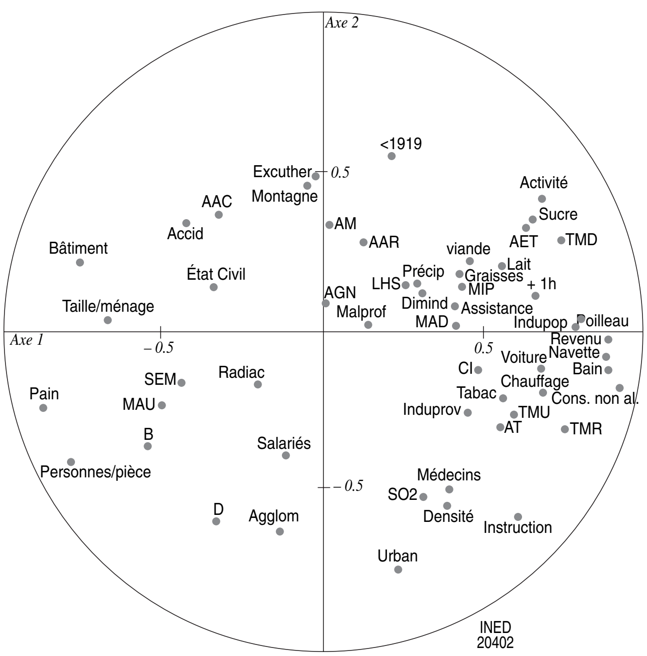


FIGURE 55–18 First factorial plane in the principal components analysis of male mortality between the ages of 55 and 79 years that is associated with the external variable matrix for Italian provinces between 1971 and 1972. See Tables 55–1 and 55–4 for causes of death and external variables. (From Caselli and Egidi, 1981c.)

cancer or ischemic heart disease. The second factorial axis (12% of the variance), which tells us about the spatial distribution of the population, does not play a discriminating role in terms of cause-specific mortality structure. Overall, the correlation of mortality causes with the first factorial plane is low.

The external variables taken into account have little impact on the explanation of the territorial differences in mortality due to diseases of the endocrine glands and the metabolism (D and AGN), diseases of the genitourinary organs (MAU), or poorly defined causes (SEM). We also observe the low level of relevance of the variables linked to climate conditions and the variables taken as indicators of environmental degradation, most often typical of urban zones. The latter variables (i.e., density and SO_2) are rather correlated with the third factor (not represented here), which establishes an opposition between malignant tumors of the respiratory system (TMR) and malignant tumors of the digestive system (TMD). The latter are frequently associated with high-protein and high-fat diets. The main results, listed in Table 55–5, demonstrate several of the relationships between cause-specific mortality and these variables. We see that ischemic heart diseases and cancers are common in the most highly developed and wealthiest areas of the country. Diseases such as diabetes (which, thanks to medical advances, are rarely fatal today) continue to cause death in regions that are economically underdeveloped and where health services are still deficient.

Between the ages of 55 and 79 years (see Fig. 55–18), the general structure of the projection of points on the first factorial plane (29% of the variance) is in large part similar, but some interesting differences are revealed. The correlation between external variables and factors seems generally higher. As a result, the variables that seem to have an impact on the territorial distribution of mortality at that age are more numerous.

The first axis, the development indicator, correlates more with mortality due to malignant tumors of the respiratory system (TMR) or the digestive system (TMD) and mortality due to violent causes (AET). This axis also correlates, negatively this time, with mortality due to acute infectious diseases of the respiratory system (B) and to diabetes (D).

The second axis (11% of the variance), linked to urbanization, takes on a significant role in the description of mortality between the ages of 55 and 79 years. This axis establishes an opposition between the characteristics of mortality in urbanized regions, with a greater frequency of diabetes (D), acute diseases of the respiratory system (B), and cancers, and the characteristics of mortality in the mountainous

TABLE 55–5 Correlations between the Geographic Distribution of Causes of Death and Selected External Variables, with Reference to the First Factorial Plane of the Principal Components Analysis

Ischemic heart diseases	Malignant tumors of the respiratory system	Malignant tumors of the genitourinary organs	Malignant tumors of the digestive system	Diseases of the digestive system	Accidents, poisonings, and trauma	Bronchitis, bronchopneumonias, and similar diseases
For age 25 to 54 years						
+1 h	Indupop	Pollwater	Meat	Milk sugar	+1 h	Construction
Indpop	Induprov	Indupop	Indpop	+1 h	Commute	Persons/room
Pollwater	+1 h	Income(*)	Pollwater	Mountains	Marital status	Size/household
Milk	Commute		Commute	Marital status	Sugar	Bread
Sugar	Pollwater		Income[a]			
	Income[a]					
For age 55 to 79 years						
Indupop	Indupop	Pollwater	Meat	Milk	Activity	Persons/room
Induprov	+1 h	Indupop	Sugar	Sugar	Indupop	Size/household
Income[a]	Pollwater	Induprov	Fat	Fat	Income[a]	Bread
Milk	Density	Density	Pollwater	+1 h	Sugar	
Smoking	SO_2	SO_2	Indupop			
	Smoking	Smoking	Revenue[a]			
	Income[a]	Income[a]				

[a] Significant of the entire group of variables pertaining to well-being and standard of living, which are highly correlated among one another, and with income.

See Table 55–4 for an explanation of the external variables.

regions of the Apennines, where other diseases of the digestive system (AAR), violent deaths (AET), and malignant tumors of the digestive system (TMD) are prevalent.

The influence of urbanization on ischemic heart disease seems entirely different from what we observed in the previous age group. Between the ages of 25 and 54 years, mortality due to ischemic heart disease is higher in the northern provinces, which have a lower density (even if they are more industrialized), whereas between 55 and 79 years, this type of mortality is high mainly in the large, industrial cities.

Several strong relationships between variables and cause-specific mortality are modified by age. For example, the influence of certain variables pertaining to individual behavior (e.g., smoking) on mortality due to malignant tumors of the respiratory system or the genitourinary system and on mortality due to ischemic heart disease grows stronger after the age of 55 years (see Table 55–5). The fourth factorial axis (not shown here) reveals a specific dimension, because it positively correlates with other diseases of the respiratory system and with public health expenditures, the number of physicians per capita, and altitude. This factor appears as the axis illustrating the situation of some provinces of the Alps and the Ligurian Riviera, where the climate and the availability of adequate medical services attracts elderly persons suffering from respiratory diseases.

Among young and older adults, there is a coincidence between the geography of mortality due to malignant tumors of the respiratory system or genitourinary organs and the geography of air and water pollution, as well as the level of industrialization. However, a significant correlation with smoking appears only at older ages. At these ages, the variables that play a part in the geography of cancers also play a considerable role in the geography of ischemic heart disease, whose geographic distribution at younger ages seems very much linked to that of high-protein and high sugar diets, as well as stressful lifestyles.

The differences between the death probability due to malignant tumors and diseases of the digestive system are also linked to eating habits, but in this case, especially among older persons, the high consumption of fatty foods and animal protein seems to have a very negative impact.

Beyond permanent patterns, the specific trends of certain causes of death are responsible for changing the Italian geography of mortality. Another study (Caselli, 2002), which focuses precisely on the more sensitive causes, shows that such changes are still taking place, particularly in relation to modifications in eating behavior (which are increasingly homogenous), in the nature of industrial activities (which are increasingly diversified), or in health prevention practices (for which the north-south divide has grown deeper).

CONCLUSIONS

Like all variables that divide a population into different subgroups, geography brings to light mortality variations that underscore one aspect of the population's heterogeneous nature. The concept of geography in itself requires explanations, which partly explains the global heterogeneity of the population in terms of risks of dying. However, geographic differences in mortality can also help show the relationship between mortality and other socioeconomic or cultural variables, to the extent that the latter also vary according to geographic location. The study of multiple correlations between variables in the realm of geographic differences can be very useful in this respect.

If studies on geographic variations of mortality more often focus on developed countries than on developing countries, it does not mean that the phenomenon is less important in the developing countries (Verhasselt and Pyle, 1993). The paucity of studies indicates only that the necessary data are much harder to obtain in some countries.

Acknowledgment

This chapter was written with the collaboration of Rosa-Maria Lipsi of the Dipartmento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Italy, and it was translated from the French version by Zoé Andreyev.

References

Badurashvili Irina, McKee Martin, Tsuladzé Giorgi, Meslé France, Vallin Jacques, and Shkolnikov Vladimir, 2001. Where there are no data: what happened to life expectancy in Georgia since 1990? *Public Health*, vol. 115(6), p. 394–400.

Bähr Jürgen, Jentsch Christoph, and Kuls Wolfgang, 1992. Bevölkerungsgeographie, *in*: *Lehrbuch der Allgemeinen geographie*, vol. 9. Berlin, Walter de Gruyter, 158 p.

Bonneuil Noël, 1997. *Transformation of the French demographic landscape 1806–1906*. Oxford, UK, Clarendon Press, 218 p.

Bouroche Jean-Marie and Saporta Gilbert, 1980. *L'analyse des données*. Paris, Presses Universitaires de France, 128 p.

Britton Malcom, 1989a. *Mortality and geography: a review in the mid-1980s. The Registrar General's decennial supplement for England and Wales*. London, Office of Population Census and Surveys, 223 p. (Series DS, no. 9.)

Britton Malcom, 1989b. Mortality and geography, *Population Trends*, vol. 56, p. 16–23.

Calot Gérard and Caselli Graziella, 1989. La mortalité en Chine d'après le recensement de 1982. Analyse selon le sexe et l'âge au niveau national et provincial, *Population*, vol. 44(4–5), p. 841–872.

Caselli Graziella, 1984. Les causes de décès en France. III. Un essai d'interprétation des différences géographiques: application à la période 1974–1976, *Population*, vol. 39(6), p. 1011–1044.

Caselli Graziella, 1990. Mortalità e sopravvivenza in Italia dall'Unità algi anni '30. Temi di demografia storica italiana (secc XVII–XIX), *in*: Eugenio sonnino (ed.), *Popolazione, società e ambiente*, p. 275–309. Bologne, Cooperativa Libraria Universitaria Editrice Bologna, 270 p. (Proceedings of the Conference Hispano Luso Italià de Demografia Historica, Barcelone, April 22–25, 1987.)

Caselli Graziella, 2002. *How environmental factors cause differences in the Italian geography of mortality by cause*. Rome, University La Sapienza, 34 p. (Paper presented at the meeting of the Population Association of America, Atlanta, May 9–11, 2002).

Caselli Graziella, Cerbara Loredana, and Leti Giuseppe, 1993. The geography of adult mortality: results from the fuzzy clumping method, *Genus*, vol. 40(1–2), p. 1–24.

Caselli Graziella and Egidi Viviana, 1981a. *Nouvelles tendances de la mortalité en Europe*. Strasbourg, Conseil de l'Europe, 101 p. (études démographiques no. 5.)

Caselli Graziella and Egidi Viviana, 1981b. La géographie de la mortalité italienne: différences territoriales et milieu, *Genus*, vol. 37(1–2), p. 1–153.

Caselli Graziella and Egidi Viviana, 1981c. L'analyse des données multidimensionnelles dans l'étude des relations entre mortalité et variables socio-économiques d'environnement et de comportement individuel, *Genus*, vol. 37(3–4), p. 57–91.

Caselli Graziella and Egidi Viviana, 1981d. Géographie de la mortalité en Europe: influence de l'environnement et de certains aspects du comportement, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference, Manilla, 1981*, vol. 2, p. 165–205. Liège, International Union for the Scientific Study of Population, 556 p.

Caselli Graziella and Egidi Viviana, 1988. Les variations géographiques de la mortalité, *in*: Jacques Vallin and France Meslé (eds.), *Les causes de décès en France de 1925 à 1978*, p. 507–607. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 607 p. (Travaux et documents, cahier no. 115.)

Caselli Graziella, Meslé France and Vallin Jacques, 1999. Le triomphe de le médecine, *in*: Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998*, p. 125–181. Paris, Librairie Arthème Fayard, 792 p.

Caselli Graziella and Reale Alessandra, 1999. Does cohort analysis contribute to the study of the geography of mortality? *Genus*, vol. 55(1–2), p. 27–59.

Caselli Graziella, Cerbara Loredana, Heins Frank, and Lipsi Rosa Maria, 2003. What impact do contextual variables have on the changing geography of mortality in Italy? *European Journal of Population*, vol. 19, p. 339–373

Cislaghi Cesare, 2000. *GIS Mortalità 1981–1994*. Milan, Verbatim, CD-Rom. (Versione aggiornata con Atlante provinciale e elaborazioni batch, version 6.0.)

Clieff Andrew D. and Haggett Peter, 1988. *Atlas of disease distributions. Analytic approaches to epidemiological data*. Oxford, UK, Blackwell, 300 p.

Decroly Jean-Michel and Grimmeau Jean-Pierre, 1991. Variations intercommunales de la mortalité en Belgique, *Espace, populations, sociétés*, no. 1, p. 75–73.

Decroly Jean-Michel and Vanlaer Jean, 1991. *Atlas de la population européenne*. Brussels, éditions de l'Université de Bruxelles, 172 p.

Desplanques Guy and Nizard Alfred, 1978. Mortalité sociale et mortalité régionale, *in*: René Duchac (ed.), *Les disparités démographiques régionales*, p. 318–331. Paris, éditions du CNRS, 614 p. (Proceedings of the fifth National Demographic Seminar of the National Center of Scientific Research (CNRS), Nice, April 14–16, 1975, colloque no. 935.)

Dorling, Daniel, 1995. *A new social atlas of Britain*. Chichester, John Wiley & Sons, 247 p.

Grosclaude Anne, Lux Baudouin, van Houte-Minet Michèle, and Wunsch Guillaume, 1979. Mortalité régionale et comportements différentiels, *Population et Famille*, vol. 48, p. 1–43.

Holland Walter W. (ed.), 1988. *European Community atlas of avoidable death*. Oxford, UK, Oxford University Press, 356 p.

Holland Walter W. (ed.), 1991. *European Community atlas of avoidable death*, vol. 1, 2nd ed. Oxford, UK, Oxford University Press, 350 p.

Holland Walter W. (ed.), 1993. *European Community atlas of avoidable death*, vol. 2, 2nd ed. Oxford, UK, Oxford University Press, 178 p.

INSEE, 1964. *Données de démographie régionale 1954 déduites du recensement de 1954 et des statistiques de l'état civil des années voisines*. Paris, Imprimerie nationale, 234 p.

Isnard Michel and Lavertu Jacques, 1995. *Données de démographie régionale 1990*. Paris, National Institute of Statistics and Economic Studies, 233 p. (INSEE-Résultats, no. 430–431, Démographie-Société, no. 45–46.)

ISTAT and ISS, 1988. *Mortalità per causa e unità sanitaria locale 1980–1982*. Rome, Istituto Centrale di Statistica and Istituto Superiore di Sanità, 750 p.

ISTAT, 1996a. *Tavole di Mortalità della Popolazione Italiana per Regione. 1989–93*. Rome, Istituto Nazionale di Statistica, 54 p.

ISTAT, 1996b. Ricostruzione delle populazione residente per età nelle province italiane. Anni 1982–1991. Rome, Istituto Nazionale di Statistica, 414 p. (Speciali informazioni.)

ISTAT, 2000. *Tavole provinciali di Mortalità. Anno 1995*. Rome, Istituto Nazionale di Statistica, 212 p. (Informazioni, no. 12.)

ISTAT and ISS, no date. *La mortalità in Italia nel periodo 1970–1992: evoluzione e geografia*. Rome, Istituto Nazionale di Statistica, 436 p.

Jozan Peter and Prokhorskas Remigijus (eds.), 1997. *Atlas of leading and avoidable causes of death in countries of Central Eastern Europe*. Budapest, Hungarian CSO Publishing House, Hungarian Central Statistical Office, World Health Organisation, 324 p.

Krumins Juris, Vallin Jacques, Jasilionis Domantas, Meslé France, and Stankuniéne Vlada, 1999. *Do geographical variations of mortality overpass national frontiers? The case of the three Baltic countries and their neighborhood*. Paris, Institut National d'Etudes Demographiques, 38 p. (Paper presented at the European Population Conference, The Hague, August 30–September 3, 1999.)

Ledermann Sully, 1956. *Alcool, alcoolisme, alcoolisation*, vol. 1. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 314 p. (Travaux et documents, cahier no. 29.)

Ledermann Sully, 1964. *Alcool, alcoolisme, alcoolisation*, vol. 2. Paris, Institut national d'etudes demographiques, Presses universitaires de France, 613 p. (Travaux et documents, cahier no. 41.)

Mérenne Bernadette, van der Haegen Hernan, and van Hecke Étienne (eds.), 1997. *La Belgique: diversité territoriale*. Brussels, Service fédéraux des affaires scientifiques techniques et culturelles (SSTC) and Crédit Communal de Belgique, 144 p.

Meslé France, 1983. Cancer et alimentation: le cas des cancers de l'intestin et du rectum, *Population*, vol. 38(4–5), p. 733–762.

Meslé France, 1990. Géographie de la mortalité liée à l'alcoolisme: évolution 1962–1982, *Espace, populations, sociétés*, no. 3, p. 521–531.

Meslé France, 1991. La mortalité dans les pays d'Europe de l'Est, *Population*, vol. 46(3), p. 599–650.

Meslé France and Vallin Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in*: International Union for the Scientific Study of Population (ed.), *International Population Conference, Montreal, 1993*, vol. 2, p. 365–382. Liège, International Union for the Scientific Study of Population, 493 p.

Meslé France and Vallin Jacques, 1998. Évolution et variations géographiques de la surmortalité masculine: du paradoxe français à la logique russe, *Population*, vol. 53(4–5), p. 1079–1102.

Meslé France, Vallin Jacques, and Shkolnikov Vladimir, 1998. Reversal of mortality decline: the case of contemporary Russia, *World Health Statistics Quarterly*, vol. 51(2–4), p. 191–206. (Historical epidemiology: mortality decline, and old and new transitions in health, special issue edited by Odile Frank.)

Nizard Alfred and Prioux France, 1975. La mortalité départementale en France, *Population*, vol. 30(4–5), p. 781–824.

Peters Gary L. and Larkin Robert P., 1993. *Population geography: problems, concepts, and prospects*. Dubuque, IA, Kendall & Hut, 365 p.

Plane David A. and Rogerson Peter A., 1994. *The geographical analysis of population with application to planning and business*. New York, John Wiley & Sons, 417 p.

Potrykowska Alina and Clarke John, 1995. Population and environment, *in*: *Geographia Polonica*, vol. 64. Warsaw, Polish National Academy of Sciences, Institute of Geography and Spatial Organization, 300 p.

Salem Gérard, Rican Stéphan, and Jougla Éric, 1999. *Atlas de la santé en France*, vol. 1. *Les causes de décès*. Paris, John Libbey, 190 p. (DREES, MIRE, Ministère de l'emploi et de la solidarité.)

Tonnellier François, 1991. Évolution des inégalités géographiques de mortalité en France depuis 1911, *Espace, Populations, Sociétés*, no. 1, p. 29–36.

Union Géographique Internationale, 1990. Les inégalités géographiques de la mortalité. Actes du colloque UGI, Lille, Avril 1990, vol. 1, *Espace, Populations, Sociétés*, no. 3, p. 343–576.

Union Géographique Internationale, 1991. Les inégalités géographiques de la mortalité. Actes du colloque UGI, Lille, Avril 1990, vol. 2, *Espace, Populations, Sociétés*, no. 1, p. 1–244.

United Nations, 1999a. *World population prospects: the 1998 revision. Vol. 1: Comprehensive tables*. New York, United Nations, Population Division, 627 p.

United Nations, 1999b. *World population prospects: the 1998 revision*, vol. 2. *The sex and age distribution of the world population*. New York, United Nations, Population Division, 883 p.

United Nations, 1999c. *World population prospects: the 1998 revision*, vol. 3. *Analytical report*. New York, United Nations, Population Division, 247 p.

United Nations, 2001. *World population prospects: the 2000 revision*. New York, United Nations, Population Division, CD-ROM.

Valkonen Tapani, 2001. *Trends in differential mortality in Europe*, p. 185–321. Strasbourg, Council of Europe Publishing, 332 p.

Vallin Jacques, 1981. Le développement économique est-il devenu nuisible à la santé des hommes? *Population*, vol. 36(4–5), p. 929–933.

Vallin Jacques, 1998. Politiques de santé et contrastes internationaux de mortalité, *Revue de géographie*, vol. 122(4), p. 305–320.

Vallin Jacques and Chesnais Jean-Claude, 1974. Évolution récente de la mortalité en Europe, dans les pays anglo-saxons et en Union soviétique, 1960–1970, *Population*, vol. 29(4–5), p. 871–898.

Vallin Jacques and Meslé France, 2001. Évolution de la mortalité en Europe depuis 1950: mortalité par âge, sexe et causes de décès, *in*: *Tendances en matière de mortalité et mortalité différentielle*, p. 33–188. Strasbourg, Éditions du Conseil de l'Europe, 344 p. (Études démographiques no. 36, also published in English: *Trends* *in Mortality and Differential Ortality*, p. 31–186. Strasbourg, Council of Europe Publishing, 334 p.)

van Houte-Minet Michèle and Wunsch Guillaume, 1978. La mortalité masculine aux âges adultes, *Population et Famille*, no. 43, p. 37–68.

van Poppel Frans, 1979. Regional differences in mortality in Western and Northern Europe: a review of the situation in the seventies, *in*: World Health Organization (ed.), *Délibération de la réunion de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité*, p. 351–372. Geneva, New York, United Nations/WHO, 622 p.

Verhasselt Yola and Pyle Gérald F., 1993. Geographical inequalities of mortality in developing countries, *Social Science and Medicine*, vol. 36(10–44), p. 1239–1382.

Yeganyan Ruben, Badurashvili Irina, Andreev Evgueni, Meslé France Shkolnikov Vladimir, and Vallin Jacques, 2001. Life expectancy in two Caucasian countries: how much due to overestimated population? *Demographic Research*, vol. 5(7), p. 217–243. (Electronic journal of the Max Planck Institute for Demographic Research.)

Yue Liu (ed.), 1987. *The population atlas of China*. Oxford, UK, Oxford University Press, 216 p.

PART

IV

THEORIES AND FRAMEWORKS OF MORTALITY

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Morbidity and mortality have changed beyond all recognition since man first trod the Earth. The conditions of humankind's survival changed with their ability to control the natural world, and after millennia of hesitant progress, a sweeping revolution began in the 18th century. Nevertheless, vast differences remain between the situations of populations and between the constituent groups of populations, as was stressed in earlier chapters.

The efforts to account for the dynamics of these changes and the diversity of situations are reviewed in the next two chapters from two very different but complementary angles. Many tentative theories have been formulated about the very high infant and child mortality in the least-developed countries. Godelieve Masuy-Stroobant critically reviews the main frameworks of infant and child mortality trends and differentials in Chapter 56.

In Chapter 57, France Meslé and Jacques Vallin look at attempts to find a paradigmatic explanation for health gains since the 18th century, from Abdel Omran's *epidemiologic transition theory* to the broader concept of *health transition*.

C H A P T E R

56

Theories and Frameworks for the Study of Child Mortality

GODELIEVE MASUY-STROOBANT

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Thus although we may know in how many studies has infant mortality been found to slope downwards with increasing education of the mother, and in how many cases the relation is presumably not affected by disturbing factors, we know little about the theories and frameworks that led to the study of the association. . . . It is no wonder then that one of the most common findings in the literature—that mother's education exerts a negative effect on infant mortality—is also one of the least understood.

(Palloni, 1987, p. 31)

The importance given to the mother's education at the individual level and the development of horizontal health policies in developing countries—of which access to education is one of the main components—at the collective level, were of central importance in the development of theories and frameworks on infant and child mortality in the demographic literature in the late 1970s.

This chapter does not discuss the respective advantages and limits of theories and frameworks; the former are usually considered as more rigorous but too limited, and the latter are often too ambitious in trying to integrate all known risk factors, whatever the level of observation or the nature of the relation (direct, indirect, or only contributing through other factors) with infant and child mortality. Nor should the testability of the proposed hypotheses be evaluated. Theories and frameworks are presented here as part of a theorization process in the sense defined by Hubert Gérard: "a process of critical systematization of acquired 'knowledge,' in an appointed body of qualified proposals resulting in hypotheses likely to be subjected to empirical testing verification" (Gérard, 1989, p. 271).

The theoretical hypotheses concerning mortality in the general process of demographic transition are not considered here because they are presented in later chapters. Such developments usually refer to the historical decrease of overall mortality (and incidentally, of infant and child mortality) and consider it a social process, observed and measured at a macrosocietal level and analyzed jointly with the dynamics of fertility decline.

This synthesis is limited to the tentative theorization and understanding of the mechanisms of infant mortality (i.e., from birth to the first birthday) and child mortality (i.e., between the first and the fifth birthday) observed at the individual level. Because of the diversity of mortality levels, research focuses on infant mortality (including fetal mortality) in the developed countries and on infant and child mortality in the developing countries. Because they are still closely dependent on the cultural, socioeconomic, political, and temporal contexts they refer to, such theorizing attempts have not yet resulted in a general theory, although a number of potentially common elements can be derived from them. The theorizing process is under way but still poorly developed: “Mortality is the theoretically under-developed area of demography which will probably incur the greatest evolution in the next 30 years” (Crimmins, 1993, p. 589).

I. FROM HYPOTHESES TO THEORIES

Several attempts have been made, in developed and in developing countries, to understand and produce frameworks or even theories on the universal or quasi-universal (negative) relationship observed between mortality in early infancy and the mother’s education. A similar relationship is observed at the aggregate level (i.e., country, region, or other geographically defined unit of analysis) between the level of infant or child mortality and women’s education and, at the individual level, between a child’s risk of dying and its mother’s educational level.

1. In the Developed Countries

Long before demographers, sociologists have tried to uncover the mechanisms behind this relationship by proposing hypotheses very similar to John Caldwell’s theories for Nigeria (1979).

When trying to understand the biologic and social factors of infant mortality in England and Wales from 1911 to 1950, J. Morris and J. Heady (1955, p. 558) suggested three ways of accounting for persisting social inequality: “The first is concerned with time-lag in the improvement of health as a result of social improvements; the second approach deals with the changing composition of the social groups; the third emphasizes that the whole question—a crucial one for contemporary social medicine—is still open.”

The time lag in the dissemination of the effects of social improvements may be cultural, because the influence of grandmothers and traditions are a socially variable obstacle to change and to the adoption of new health behaviors. The investigators further suggest that long-term capital assets such as education could well be more effective than income as far as maternal and child health are concerned.

The second approach deals with the hypothesis of selective social mobility upheld by Raymond Illsley, based on the 1954 Aberdeen survey of women having their first child. Illsley said, “The moves from one social class to another are frequent between generations. . . . It is possible to show that moves of this type are selective not only on the social and the educational level, but also where health is concerned. . . . Disparities will usually be all the greater as the (social) mobility was more significant and/or linked to the level of education, the child’s development, growth or death” (Illsley, 1987, p. 51).

Luc Boltanski (1969) develops the same hypothesis of a superposition of the social and temporal distances in the acquisition of the knowledge and of the health and childcare behaviors:

> The dissemination of medical knowledge occurs faster and more easily among the upper classes than in the lower ones. This is first of all because the patient-doctor . . . communication pattern varies according to the patient’s social class—he is more familiar with them when they belong to the upper classes, and more authoritarian when they are working class. Another reason is that members of working classes, having a low educational level, do not have a critical view on their knowledge or “prejudices” because they lack that “examination mind” which is a consequence of the formal education. . . . The implementation of the newest rules, those that best comply with (childcare) norms, requires less the acquisition of any particular knowledge than taking up an overall attitude to knowledge (Boltanski, 1969, p.12, 72–79).

The attempts to describe the mechanisms of the action of a mother’s education on the health (and survival) of her child were all developed in reference to the Western European context. At that time, child mortality was very low in those countries, with infant mortality levels ranging from 27 deaths for every 1000 live births (England-Wales, 1951 to 1955) to 32 per 1000 (France, 1956 to 1960). Compulsory education was already introduced for several decades, and all Western European countries benefitted from national maternal and child health systems. Social security systems, including compulsory health care insurance, had been organized immediately after the Second World War. Although infant mortality today reaches well levels well below 10 per 1000, the gap between social groups still persists (Masuy-Stroobant and Gourbin, 1995); such inequalities are also observed in fetal mortality, even in countries, such as Sweden, experiencing the lowest mortality rates (Cnattingius and Haglund, 1992).

2. In the Developing Countries

John Caldwell’s research on Nigeria (1979) played a leading role in the renewed interest demographers took in a more explanatory approach to child mortality in Third World countries. He first establishes the fact that the level of education (specifically, the mother’s education) is the most discriminating socioeconomic characteristic as far as children’s survival is concerned: “Maternal education is the single most significant determinant of these marked differences in child mortality . . . maternal education cannot be employed as a proxy for general social and economic change but must be examined as an important force in its own right.” He adds, “Furthermore, in Nigeria, as doubtless in much of the Third World, education serves two roles: it increases skills and knowledge as well as the ability to deal with new ideas, and provides a vehicle for the import of a new culture" (Caldwell, 1979, p. 408–409).

John Caldwell then suggests three hypotheses of growing importance:

> The first explanation is usually given as the only reason. That the mothers and other persons involved break with tradition or become less "fatalistic" about illness, and adopt many of the alternatives in childcare and therapeutics that become available in the rapidly changing society. The second explanation is that an educated mother is more capable of manipulating the modern world. She is more likely to be listened by doctors and nurses. . . . She is more likely to know where the right facilities are and to regard them as part of her world and to regard their use as a right and not as a boon. There is a third explanation, which may be more important than the other two combined. . . . That is, that the education of women greatly changes the traditional balance of familial relationships with profound effects on child care.

Most of the studies subsequently published, including the explanatory frameworks on child mortality, refer to this series of hypotheses, even though they were not systematically verified on the field. The tentative theory centered on the mother's education was subsequently refined or even expanded by considering it within a more general context of generalization of the schooling of the population (Palloni, 1985a, 1985b) and in its dynamics from one generation to another (Ware, 1984).

To Alberto Palloni, the effect of a mother's educational level should be considered with reference to the population's general education level: The higher it is at collective level, the less significant its effect will be at individual level. For him, this is a sign of greater equity in the distribution of national resources, and he suggests the following hypothesis: "In countries with socially integrated populations and higher levels of participation, the beneficial effects of an individual's education are less important since access to the basic resources and services is less constrained by individual assets and characteristics" (Palloni, 1985b, p. 474).

The effects mentioned here are measured in absolute terms, but relative differences persist in very-low-mortality countries.

With reference to Bangladesh, Helen Ware mentions an transgenerational dimension of the health-education interaction: "Educated mothers may be taller and fitter, not only in cases where their parents were also educated and lived in comfortable circumstances, but also in cases where poor parents selected the healthiest and brightest of their children to go to school and then strengthened their investment with preferential treatment in the allocation of food and medical services" (Ware, 1984, p. 195).

3. Similarities of Hypotheses

There are many similarities between the hypotheses concerning developing countries and those on developed ones. When Luc Boltanski argues in 1961 that the effect of education has proved to be more significant than any specific training in childcare, he is later echoed in observations made in Africa by Helen Ware (1984, p. 194). The hypotheses made by Luc Boltanski and John Caldwell on how education results in an increase in maternal autonomy are very much alike, whether such autonomy is shown in mothers breaking with tradition or in their feeling more comfortable in using modern medicine and communicating with it. The hypothesis of selective social mobility and of health-education transmission of girls from one generation to another, such as made by Raymond Illsley, turns out to be relevant in observations made in Bangladesh and Nepal (Ware, 1984). The idea that the acquisition of a certain level of education by the general population, especially by mothers, act as a "long-term investment" (Morris and Heady, 1955) in health is further expanded on by people who believe in social, rather than economic, development and in horizontal and basic policies, rather than in vertical health interventions, to secure the mortality decline and the health of Third World children.

This information shows that a theory of child mortality focused on mothers' education is still being developed. When taking the different levels of action (i.e., individual, family, and societal) into account, its temporal dimension, through the transgenerational transmission, may make it possible to turn it into a more universal theory. The formal education mentioned here, like medical technologies and preventive health behaviors, was exported from the developed into the developing countries. The former and the latter are probably inextricably linked. But how?

II. FRAMEWORKS FOR CHILD MORTALITY

Frameworks for child mortality appear by the early 1980s in the demographic literature, several decades after the well-known fertility framework suggested by Kingsley Davis and Judith Blake in 1956. Although theories on demographic transition, of which mortality is an essential component, were developed in the meantime, few or no attempts at systematic approach to mortality have been made. How can such delay be accounted for, and what triggered the emergence of the numerous frameworks on child mortality in the early 1980s?

The understanding of mortality seems more complex than that of other demographic phenomenon, partly because of its biologic component; its variable, sometimes unknown, latent period between exposure to or origin of the disease; its clinical manifestations; and its fatal issue. Setting aside the identification of what epidemiologists call risk factors of infant or even child mortality, multivariate analysis of mortality is strongly limited by the fact that deaths are a statistically rare event (even in situations of high mortality), whereas it is widely acknowledged that the analysis of its causality is multifactorial.

Demographers, as social scientists, and epidemiologists, belonging to the biomedical sciences, share a common interest in the study of infant and child mortality. Social sciences, including demography, focus mainly on its measurement, the social and economic determinants of its historical decline at a collective level, and in showing how mortality differences are an indicator of persisting social and economic inequalities at the individual level. The biomedical sciences have analyzed the biologic process of diseases, attempting to identify how to prevent diseases (mainly by vaccination), or to cure them, according to the classic epidemiologic model of infectious diseases, by intervening on one of the three elements of the agent-exposure-host triad. In doing so, demographers found it difficult to access the mechanisms (which were partly biologic, but also related to health behaviors) that would enable them to establish a link between social class and early death in childhood, whereas physicians and epidemiologists often neglect the social factors about which they felt helpless. There was also the widespread feeling that mortality, especially mortality in childhood, was supposed to continue decreasing as the poorer countries experienced process of economic and social development. This belief was reinforced by the fact that medical technology was then considered efficient and able to control infectious diseases by means of mass vaccination, securing a clean environment (e.g., in the case of malaria), or educating mothers to practice oral rehydration when their infant suffer from diarrhea. However, such beliefs were completely upset by the fact that the rate of decrease reportedly slowed down in high-mortality countries and that health conditions worsened in others (Gwatkin, 1980). Theses linked to the very conception of development were then confronted, resulting in the idea that vertical interventions, usually punctual health interventions, such as mass vaccination programs and targeting specific diseases, could only have a limited effect. Children who escape dying of the targeted disease would die of another one. Reaching a certain level of economic development, measured, for example, by means of the gross national product growth is not enough to improve the population's health; the wealth acquired must benefit the majority, which implies that social policies are set up to ensure that the results of economic development are redistributed: "Major barriers to good health are social and economic; more attention should therefore be given to such activities as education of women and community development" (Mosley, 1985, p. 189). Alberto Palloni, referring to Latin America, gives primary importance to making chances equal: "Equally relevant is the presence of broad social developments as reflected in equalization of access to material resources and health services, active participation in community and national life by all groups in society, and suppression of traditional, locally based norms of behavior which are detrimental to the health of some individuals (excess fertility, short birth intervals, sex discrimination in the treatment of children, child rearing practices, traditional resistance to adoption of modern medical services)" (Palloni, 1985a, p. 449). To vertical interventions meant to have an effect over specific diseases using specific technologic means are opposed horizontal interventions that "are designed to deal with generic conditions producing high prevalence of a wide variety of diseases by providing egalitarian and better access to medical services, improving sanitation and water supply, supporting dietary and nutritional campaigns and engaging individuals and families in campaigns for the prevention and treatment of common diseases" (Palloni, 1985a, p. 449). How successful such horizontal interventions can be depends on the development of social policies such as previously mentioned.

The emergence of the nutritionist movement resulted on the medical side in the development of a more general approach to children's health conditions, using simple anthropologic measures in which their age, weight, and size are combined. The underlying logic was based on the observation that, except in cases of strictly accidental causes, children do not die of any single disease, but rather of the combined effect of repeated infectious attacks, from which a child recovers with even greater difficulty because it was suffering from malnutrition from the start or was weakened by a previous disease. Various measurements of nutritional status have been developed with a view to evaluate some transitory or chronic deficit, but what matters is that the concept makes it possible to establish a link between the social and biomedical approaches to mortality, because the child's nutritional status (including its weight at birth) is closely linked to its parents' social and economic conditions and because measuring nutritional status is not specifically linked to any particular disease, as is the measurement of overall mortality, with which demographers are more familiar.

1. In Developing Countries

The frameworks of child mortality have been developed by referring to contexts in which infant and child mortality rates are still very high (along with very high fertility rates). They were also largely inspired by the fertility framework developed by Kingsley Davis and Judith Blake (1956), drawing clear distinction between intermediate and proximate variables of mortality, which have direct effect on a child's chances of survival, and other variables whose indirect effect is felt through intermediate variables.

Beyond aiming at giving some order and coherence to the series of risk factors and to the alleged causes of premature death in early infancy or during childhood hitherto identified in scientific literature, such frameworks also aim at reconciling or even bringing together the social and biomedical approaches by "identifying a set of proximate determinants that can link social processes in populations to biological outcomes in individuals" (Mosley, 1985, p. 202). At the same time, the different levels of influence (and analysis) were deliberately taken into account. In some cases, these range from the macrosocietal political system to the genetic factors at the individual level, the ecological environment, the community infrastructure, household composition, and other levels.

What matters about these frameworks is the identification and enumeration of determinants, as well as their classification into broad categories, rather than the complex relations between each of the elements within each category or with those of other categories. Along with the frameworks, a presentation of and justification for the consideration of each of their constituents are usually given but no hypotheses are offered that could be confronted with reality. That was not part of the authors' objectives. For some, "the task . . . is the identification of those characteristics of individuals, households, and communities that need to be maintained, altered or eliminated by means of health interventions and social development" (Palloni, 1985a, p. 457) and consequently to help developing an efficient health policy, whereas for others, it was "to provide a construct that can assist researchers in reexamining the assumptions underlying their work in order to better identify the relevant questions to ask" (Mosley, 1985, p. 202).

The frameworks suggested by Sriniwasa Meegama (1980), Michel Garenne and Patrice Vimard (1984), Henry Mosley and Lincoln Chen (1984), and Alberto Palloni (1985a, 1985b) are presented here as illustrations. The field experiences of the authors cover the three major developing regions—Asia (Meegama, Mosley, and Chen), Africa (Garenne and Vimard), and Latin America (Palloni)—and they allow us to trace their thinking process and to identify what was at stake in the health and development policies with regard to children's health and mortality.

Instead of suggesting a single pattern that would be applicable *mutatis mutandis* to the deaths of all children before their fifth birthday, Sriniwasa Meegama draws radical distinction between neonatal deaths and post-neonatal and children's deaths. Rather than integrating the paradigms of social and medical sciences, his approach consists of enlarging the traditional epidemiologic model of infectious diseases. For example, post-neonatal and children's mortality is broken down into groups of major causes of death (e.g., respiratory diseases, digestive system disorders, illnesses due to malnutrition, anemia); in the next stage, the exposure vectors (e.g., the air, water, environment, inadequate nutrition) and the contributing causes are identified. For example, inadequate clothing or housing may lead to exposure to adverse weather conditions, which may cause respiratory diseases that may lead to death; the lack of prenatal or postnatal monitoring, insufficient information, or the mother's economic activity may result in early weaning or insufficient feeding of the infant, which may cause malnutrition or anemia that, in interaction with infectious disease, may be direct or indirect causes of a child's death.

The same reasoning applies to neonatal deaths. Sriniwasa Meegama's approach offers a synthetic vision of the biologic determinants of mortality in childhood, grouping together infectious diseases and identifying their causes, which are mainly linked to social and economic conditions. The proposed causal process, however, remains specific to each group of causes of death and to age at death. Malnutrition appears as a direct cause of death and as contributing to it when combined with infectious diseases. The framework, however, is strictly limited to the individual level, whereas most of the frameworks subsequently published achieved a much larger scope.

The framework designed by Michel Garenne and Patrice Vimard (1984) goes beyond the individual's social and economic conditions to the political and ecological contexts (Fig. 56–1). Infant and child mortality is considered as a whole, as is the case in most frameworks designed for developing countries. Their framework has a rather modest objective: "It is not meant to

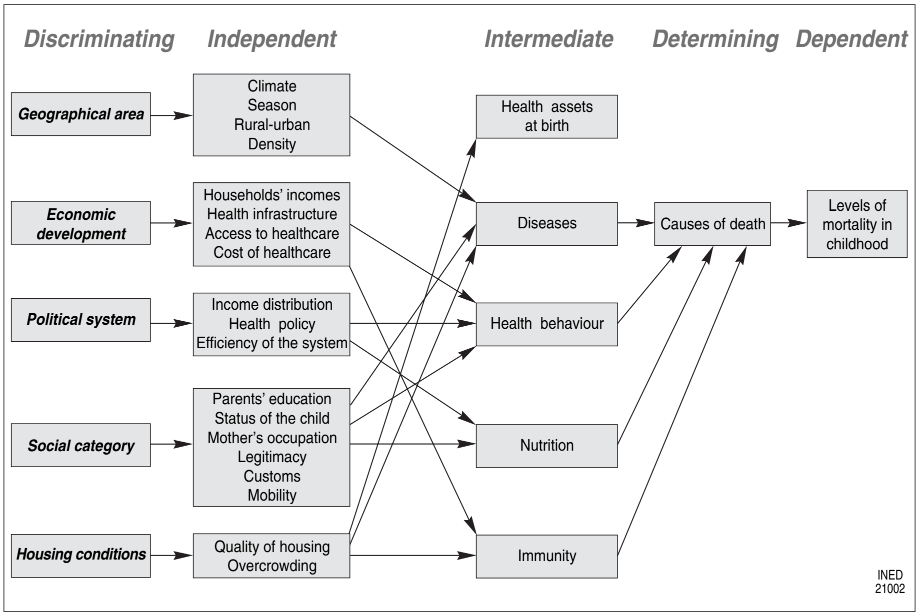


FIGURE 56–1 Analytic framework for the study of child survival, first example. Derived from the framework developed by Michel Garenne and Patrice Vimard (1984).

be exhaustive in any way, only to situate things so as to clarify the discussion" (Garenne and Vimard, 1984, p. 306). The authors distinguish five groups of variables, corresponding to different levels of causality:

1. The discriminating variables "should not have any direct effect over mortality, but refer to the classical studied differential mortality: geographic characteristics, economic development level, political system, social categories, and housing conditions.

2. The independent variables describe in detail the different dimensions of the variables collectively defined at the previous level; they are "likely to have an effect on mortality through one or several intermediate variables." Geographic characteristics therefore are broken down into categories of climate, season, residence (rural or urban), and population density and the political system into "distribution of income," "health policy," "efficiency of the system," and other categories. Maternal education is mentioned as one dimension of the social category, but the status of the child, the mother's occupation, the customs, and other factors are also included.

3. The intermediate variables are those through which most of the independent variables are likely to exert their influence and "their action over the survival of the child or the cause of its death"; here, the chain of causes is much more complex than Sriniwasa Meegama's (there is no specific path to each cause of death), and the authors have proposed five groups of variables (i.e., the health status at birth, diseases, health behaviors, nutrition, and immunity).

4. The determining variables are medical causes of death. No cause or disease is explicitly mentioned; the authors mention that it is not the result of any simplistic medical diagnosis "because the death of a child is often the result of a complex process that can seldom be summarized in any single cause of death."

5. The dependent variable is the overall level of mortality.

Michel Garenne and Patrice Vimard provide a classification of the variables by the nature of their relationship with the risk of death. What they refer to as *levels* do not necessarily cover the level of observation or analysis of the variables taken into account. The levels are defined from a functional point of view. From an operational point of view, the variables pertaining to a specific category are not necessarily observed at the same levels—independent variables that correspond to such aspects of economic development such as "households' income," "health infrastructure," "access to health care," and "cost of health care" should be observed at the household's or at the community level. The independent variables belong to the "social category," and discriminating variables are observed at the individual level (e.g., the mother's and the father's education) or at the household level (e.g., status of the child) or a more aggregate level (e.g., mobility of the population).

Henry Mosley and Lincoln Chen (1984) keep the idea of a distinction between independent and intermediate variables, but they classify the independent variables by level of observation or analysis into household-related individual variables and community variables. The authors state that they mean to suggest a framework integrating the social and biomedical sciences approaches. They do it in two ways. First, they combine the nutritional level of growth faltering (nutritional status) of the survivors with the level of mortality of the respective birth cohort into a more general health index for the dependent variable. In doing this, the authors go beyond the usual deceased-survivor dichotomy, classifying the survivors by their health levels, with death being considered an extreme situation of ill health. The point was to consider child mortality "more as a chronic disease process with multifactorial origins than as an acute, single-cause phenomenon" (Mosley and Chen, 1984, p. 41). Second, they select proximate determinants (or intermediate variables) in which "clinical" relevance with regard to the risk of malnutrition or death is combined with their being observable or measurable in studies made at the population level. The intermediate variables or proximate determinants are categorized into five major groups: maternal factors linked to reproductive behavior, environmental contamination, nutrient deficiency, injury, and personal illness control (Fig. 56–2).

Independent (or underlying or socioeconomic) determinants, which are operating through the proximate determinants, are grouped into three broad categories corresponding to three levels of observation:

1. Individual-level factors include productivity or the capacity of producing a surviving, healthy child, such as can be measured by the father and mother's education, their health condition, and their time availability; the norms, traditions, and attitudes likely to change economic choices and practices related to health (as measured by the power relations within the household); the value of the child; beliefs about disease causation; food preferences; and taboos.
2. Household-level factors include availability of food (including water), clothing, bedding, housing, fuel or energy availability, transportation, access to health care (both preventive and curative), and access to information.
3. Community-level factors are related to the characteristics of the context, such as the ecologic setting and political economy, including the stability of the political institutions and the degree to which individuals participate in local and national decision-making, and are related to health system variables.

The studies by John Caldwell (1979) concerning individually measured independent variables are used as references, and most importantly, Henry Mosley considers the mother's educational level as a key factor: "Because of her responsibility for her own care during pregnancy and the care of her child through the most vulnerable stages of its life, her educational level can affect child survival by influencing her choices and

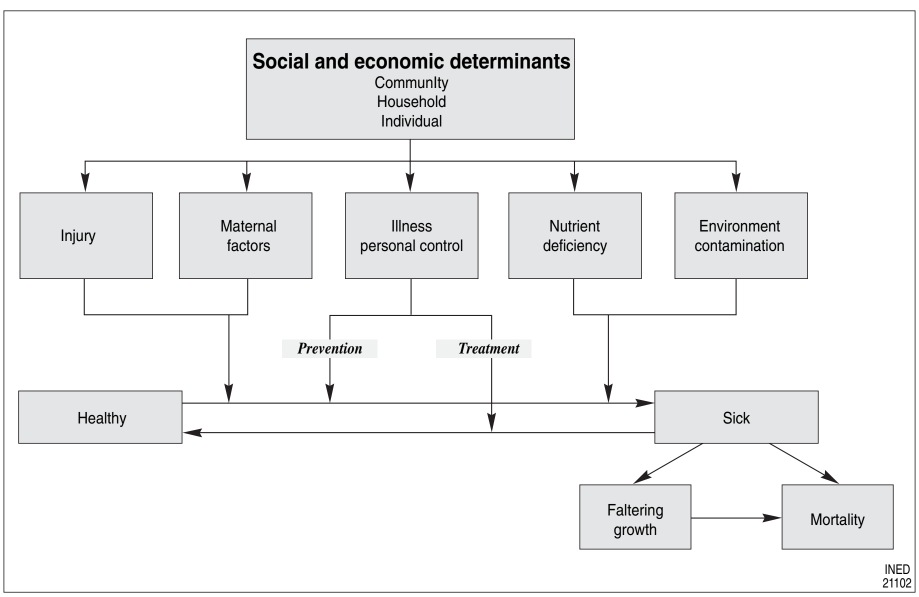


FIGURE 56–2 Analytic framework for the study of child survival, second example. Derived from the framework developed by Henri Mosley and Lincoln Chen (1984).

increasing her skills in health care practices related to contraception, nutrition, hygiene, preventive care, and disease treatment. In fact, so many proximate determinants may be directly influenced by a mother's education to radically alter chances for child survival, that one of the authors [Mosley] was prompted to label the process 'social synergy.'" (Mosley and Chen, 1984, p. 35).

Whereas Henry Mosley and Lincoln Chen's objectives were to suggest a framework for research and action, Alberto Palloni (1985a, 1985b) is more oriented toward action. He strives to prove the necessity to develop social policies and the so-called horizontal health interventions (Fig. 56–3).

Palloni states that "the excess of infant and child mortality and the anomalous structure of causes of death are clues about the nature of the comparative failure of the health policies implemented up to then" (Palloni, 1985a, p. 455). His framework was intended to help formulate a health policy, and its elements were selected with reference to their vulnerability to a social policy action. It includes three levels of action:

1. The policy level emphasizes social policies and horizontal interventions (i.e., education campaigns and socialization of health care).

2. The individual or collective level distinguishes among individual (i.e., biologic, behavioral, and social characteristics, including maternal education), household (e.g., resources, overcrowding, family structure), and community (e.g., ecology, health system, accessibility to services) characteristics.

3. The third level, which is closest to morbidity, refers to the biomedical paradigm (i.e., exposure to infectious agents, vulnerability or susceptibility of the individual, and capacity to resist the disease).

Action may be conducted at any of these levels, but for Alberto Palloni, any vertical action related only to the third level cannot result in any sustainable improvement of children's health or in any long-term increase of their chances of survival. If the causes of individual differences (second level) in mortality can be identified, substantial reduction of mortality among the less-privileged social group may lead to a radical decrease in total mortality, but that is unlikely to occur without reinforcing social policies and horizontal interventions (first level).

2. In Developed Countries

In most developed countries, the stagnation of infant mortality is recent, and it is probably artificial, at least partly, and linked to the evolution of birth notification practices (Masuy-Stroobant and Gourbin, 1995), as well as to the fact that very low levels have

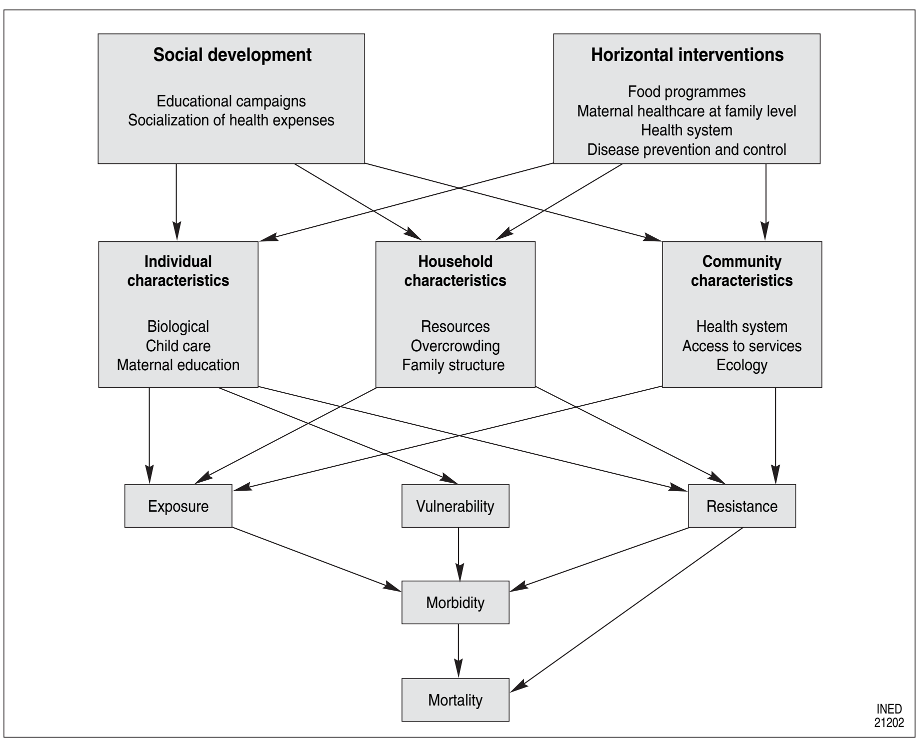


FIGURE 56–3 Analytic framework for the study of child survival, third example. Derived from the framework developed by Alberto Palloni (1985).

been achieved. As a social phenomenon, however, infant mortality in developed countries is not a solved problem—significant social inequalities are still observed, even when fetal mortality is considered. The believe that causes of fetal or neonatal mortality are mainly endogenous should be revised.

Whereas demographers are traditionally more interested in identifying mortality differentials (i.e., by maternal age, birth order, legitimacy, or the mother's social class and education), epidemiologists have repeatedly sought *factors* and *markers* of preterm birth or low birth weight. When risk factors are vulnerable to medical intervention, markers, among which most social and economic variables are included, are useful in identifying at risk groups, but they cannot be changed through medical intervention. The cause and effect relationship between preterm birth and those factors has seldom, if ever, been proved: "Many factors—more than 100—have been brought out. But the question of whether those factors play any causal role has not yet been investigated. . . . Efficient measures then might well be taken for high risk women, though no action was taken against the risk factors, since their causal role is ignored, and some of them cannot even be changed" (Schwartz, 1989, p. 138). The fact that infant mortality is a statistically rare event has also contributed to particular attention being paid to measures of a child's vulnerability at birth, such as preterm birth or the risk of low birth weight—which, like malnutrition, are not specific to any given disease—as dependent variables.

Beyond identifying risk criteria, a few attempts have been made to classify variables, as was done for developing countries. However, such attempts were definitely situated at the individual level; the social, health, and ecologic circumstances in developed countries are very different from those in developing ones. Concerning the United States, Charles Nam and colleagues (1989) draw a clear distinction between independent or basic variables (i.e., ethnic group, maternal education, marital status, and place of residence), intermediate variables (i.e., weight at birth, prenatal care, last birth interval, age of the mother, previous experience of death, sex of the child, and twin births), and immediate factors (i.e., age at death and medical cause). For Belgium, I have established a classification of social identification variables (i.e., occupation, education, nationality, and marital status of the child's parents) and intermediate variables (i.e., past fertility behavior, behaviors and preventive attitudes about the health system and in daily life) "in order to gain a better understanding of the mechanisms through which the social and cultural class determines unequal survival in early infancy" (Masuy-Stroobant, 1989, p. 48).

The process of systematizing available knowledge is ongoing, and the stage of developing one or several infant and childhood mortality theories has not been reached for developing or developed countries. Similar approaches are applied; social and cultural factors admittedly play a more extensive part, although they are nonspecific and must be used in combination with biologic factors to account for infant or child mortality. However, the persistent social component in children's health and mortality, even in the most developed countries, remains largely unaccounted for. The time or process dimension also needs to be inserted into the theorization attempts, which are still exceedingly static, so that that this essential aspect of the child mortality process can be considered.

Acknowledgment

This chapter was translated from the French version by Mireille Rabenoro.

References

Boltanski Luc, 1969. *Prime éducation et morale de classe*. Paris, Mouton, 154 p.

Caldwell John C., 1979. Education as a factor in mortality decline: an examination of Nigerian data, *Population Studies*, vol. 33(3), p. 395–415.

Cnattignius Sven and Haglund Bengt, 1992. Effects of socioeconomic factors on late foetal and infant mortality in Sweden, *in: Proceedings of the international collaborative effort on perinatal and infant mortality*, vol. III, p. IV38–IV42. Hyattsville, NCHS, 370 p.

Crimmins Eileen M., 1993. Demography: the past 30 years, the present, and the future, *Demography*, vol. 30(4), p. 579–591.

Davis Kinsgley and Blake Judith, 1956. Social structure and fertility: an analytic framework, *Economic development and cultural change*, vol. 4, p. 211–235.

Garenne Michel and Vimard Patrice, 1984. Un cadre pour l'analyse des facteurs de la mortalité des enfants, *Cahiers de l'ORSTOM, Série sciences humaines*, vol. 20(2), p. 305–310.

Gérard Hubert, 1989. Théories et théorisation, *in*: Josiane Duchêne and Guillaume Wunsch (eds.), *L'explication en sciences sociales. La recherche des causes en démographie*, *Chaire Quetelet 1987*, p. 267–281. Louvain-la-Neuve, CIACO, 478 p.

Gwatkin Davidson, 1980. Indication of change in developing country mortality trends. The end of an era? *Population and development review*, vol. 6(4), p. 615–644.

Illsley Raymond, 1987. Les inégalités sociales de santé en Grande-Bretagne, *in*: Pierre Aïach, Roy Carr-Hill, Sarah Curtis, and Raymond Illsley (eds.), *Les inégalités sociales de santé en France et en Grande Bretagne*, p. 43–63. Paris, INSERM, La Documentation Française, 206 p.

Masuy-Stroobant Godelieve, 1988. *Santé de l'enfant et inégalités sociales*. Louvain-la-Neuve, CIACO, 283 p.

Masuy-Stroobant Godelieve, 1989. Social inequalities in perinatal health, *Biology of the Neonate*, vol. 55(1), p. 42–49.

Masuy-Stroobant Godelieve and Gourbin Catherine, 1995. Infant health and mortality indicators. Their accuracy for monitoring the socio-economic development in the Europe of 1994, *European Journal of Population*, vol. 11, p. 63–84.

Meegama Sriniwasa A., 1980. Socio-economic determinants of infant and child mortality in Sri Lanka: an analysis of post-war experience. London, World Fertility Survey, 55 p. (WFS scientific report, no. 8.)

Morris J. N. and Heady J. A., 1955. Social and biological factors in infant mortality. V. Mortality in relation to the father's occupation 1911–1950, *The Lancet*, March 12, p. 554–559.

Mosley Henry W. and Chen Lincoln C., 1984. An analytical framework for the study of child survival in developing countries, *Population and development review*, vol. 10, suppl., p. 25–45.

Mosley Henry W., 1985. Biological and socio-economic determinants of child survival. A proximate determinants framework integrating fertility and mortality variables, *in*: IUSSP (ed.), *International Population Conference, Florence, 1985*, vol. 2, p. 189–208. Liège, IUSSP, 456 p.

Nam Charles B., Eberstein Isaac W., Deeb Larry C., and Terrie E. Walter, 1989. Infant mortality by cause: a comparison of underlying and multiple cause designation, *European Journal of Population*, vol. 5, p. 45–70.

Palloni Alberto, 1985a. Santé et lutte contre la mortalité en Amérique Latine, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 447–473. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Palloni Alberto, 1985b. Health conditions in Latin America and policies for mortality change, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 465–492. Liege, Ordina Edition, 558 p. (Proceedings of an IUSSP seminar, Paris, February 28–March 4, 1983.)

Palloni Alberto, 1987. Theory, analytical frameworks and causal approach in the study of mortality at young ages in developing countries, *Annales de la société belge de médecine tropicale*, vol. 67, suppl. 1, p. 31–45.

Schwartz Daniel, 1989. L'explication en épidémiologie, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales. La recherche des causes en démographie*, p. 127–140. Bruxelles, CIACO, 478 p. (Chaire Quetelet, 1987.)

Tabutin Dominique, 1995. Transitions et théories de mortalité, *in*: Hubert Gérard and Victor Piché (eds.), *Sociologie des populations*, p. 257–288. Montreal, Presse de l'Université de Montréal, 518 p.

Ware Helen, 1984. Effects of maternal education, women's roles and child care on child mortality, *Population and development review*, vol. 10, suppl., p. 191–214.

CHAPTER

57

The Health Transition: Trends and Prospects

FRANCE MESLÉ AND JACQUES VALLIN

Institut national d'études démographiques (INED), Paris, France

For thousands of years and up until the mid-18th century, human life expectancy rarely exceeded 25 or 30 years, with limited variations according to time or place. This does not mean that the epidemiologic profile did not change during that time. On the contrary, research has shown that since prehistoric times, there has been a succession of *pathocenoses*, periods marked by a specific epidemiologic development involving specific pathologic combinations (Biraben, 1999). However, from the mid-18th century on, a new era began in Europe during which the shift from one pathocenosis to the next also went along with a decisive and lasting lengthening of life expectancy. This marked the beginning of the *demographic transition*, a term coined to designate the historical process that enabled mankind to shift from an older demographic regime, in which high fertility rates balanced the losses due to high mortality, to a new regime, in which mortality between birth and the end of reproductive life dropped to such an extent that generation replacement could be ensured with barely more than two children per woman (Notestein, 1945; Kirk, 1944; Landry, 1934).

Mortality and fertility declines have their roots in the economic, social, and political transformations that led to the Industrial Revolution.[1] However, to each of these declines corresponds a specific set of factors. The dynamics of the factors leading to these long-term advances in life expectancy were given the name of *health transition* (Frenk *et al*., 1991).

These advances resulted from a decline in the spread of certain diseases, which deeply modified the epidemiologic profile of the population; however, unlike the other changes involved in pathocenosis, for the first time, the decline affected all infectious diseases, and these were no longer replaced by others in such a way as to maintain mortality globally at the same level at a given age. Age at death has risen, and life expectancy has increased. The term *epidemiologic transition* is used to designate the shift from one dominant pathologic structure to another and a radical transformation process in ages at death.

The epidemiologic transition was first theorized at the end of the 1960s (Omran, 1971). This research elucidated the reasons why life expectancy had begun to increase, how this increase accelerated, and how, beginning in the 1960s, it reached a threshold in the most advanced countries.

Omran's theory was soon undermined by the dramatic decrease in cardiovascular diseases in Western countries. The concept of epidemiologic transition, which focuses on trends in the spread of diseases, turned out to be too narrow, and as a result, in the first issue of the *Health Transition Review*, founded in 1991 by John Caldwell, Julio Frenk, and others, the term *health transition* was chosen. Health transition refers to changes in the epidemiologic characteristics of the health situation, and it reflects society's response to this state of health.

[1] This question is dealt with in Chapters 68 and 69, devoted, respectively, to Europe's demographic transition and to the globalization of the transition.

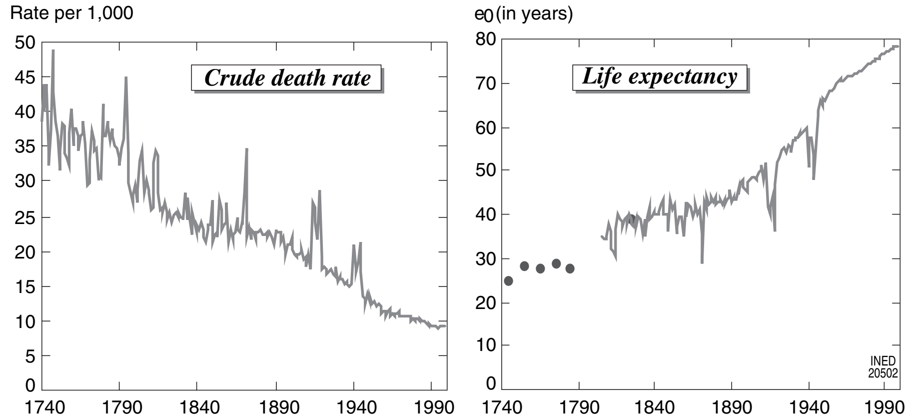


FIGURE 57–1 Crude death rate and life expectancy at birth in France: trends since the middle of the 18th century.

We face two important issues for the future. After 30 years of decline in mortality due to cardiovascular diseases, can we expect any further advances in life expectancy? It is probable that the new epidemiologic situation will change, because causes of death will be changing, and these developments probably carry new potential in terms of life expectancy gains. However, one question remains: Will these advances soon reach their limit, or is this limit still remote or uncertain?

Regardless of possible life expectancy gains in the future, existing gains have already and necessarily entailed the aging of the population. What is the relationship between aging and the state of health of populations? Our task is to determine for a given age whether the state of health has remained the same, is improving, or is deteriorating.

I. FROM THE EPIDEMIOLOGIC TRANSITION TO THE CARDIOVASCULAR REVOLUTION

In the demographic regimes of the past, mortality was very high, and there were strong fluctuations due to epidemics and famines. In a country such as France in the middle of the 18th century, when national statistical series began to be drawn up on a more or less regular basis, the crude death rate was still about 40 per 1000, hovering between 30 and 50 per 1000, as it had been in the past. By then, these fluctuations were much lower than they had been in previous times. The plague, for instance, which devastated Europe between 1347 and 1352, killed up to one-half of the population in some regions (Biraben, 1975). In Finland during the 1868 famine, the crude death rate rose to 80 per 1000 (Pitkänen, 1993).[2] In France in the space of 2 centuries, the crude rate dropped to less than 10 per 1000, and the fluctuations completely disappeared (Fig. 57–1). In the 20th century, such wide-scale disasters were caused only by the two World Wars, in violent contrast to the steadiness of the generally declining trend. The annual variations of life expectancy can be observed only since the beginning of the 19th century (Meslé and Vallin, 1989), but 10-year estimations are also available from the middle of the 18th century (Blayo, 1975). The average life span was about 25 years around 1740 and is about 80 years today. In Figure 57–1, we can see only the more recent fluctuations of the 19th century, but it is easy to imagine their scope in earlier times.

1. Epidemiologic Transition or Health Transition?

Abdel Omran (1971, 1983) attempted to synthesize the different factors at play in these secular mortality trends, in his theory of *epidemiologic transition*, based on three epidemiologic ages, with the second age corresponding to the *transition* between the first age, the age of high mortality, and the third age, that of the new epidemiologic regime. He determined the following stages:

[2] In Europe during peace time, the only country that reached a crude death rate of that scope during the 20th century was Ukraine, during the famine of 1933; that year, 2.5 million persons died in a population of about 30 million (Vallin *et al.*, 2002). Contrary to the Finnish famine, this famine was not caused by climatic conditions but was the consequence of Stalin's policy of forced collectivization of agriculture.

1. The *age of pestilence and famine*, in which mortality was high and fluctuating and life expectancy was less than 30 years on average
2. The *age of receding pandemics*, during which life expectancy increased strongly, rising from less than 30 to more than 50 years
3. The *age of degenerative and man-made diseases*, in which mortality, after a continuous decline, tends to stabilize at a very low level

The first age was marked by the predominance of infectious diseases; large-scale endemics determined the average level of the death rate, and epidemics and famines were responsible for the fluctuations. John Graunt, who studied death certificates in the city of London, showed that in the 17th century, three-fourths of the mortality in London was caused by infectious diseases, malnutrition, or complications of childbirth. In France, the first age lasted until the mid-18th century.

During the second age, epidemics became less frequent, and endemic infectious diseases began to decline; mortality decreased, and the fluctuations diminished. The "transition" began in the 18th century in northwestern Europe (around 1750 in France) and gradually spread to the rest of Europe, reaching the south in the 20th century. In Europe, cholera replaced the plague epidemic, although with far less violence, and this new disease was curbed much faster than previous ones. Among major endemic diseases, tuberculosis, which first spread considerably due to industrialization and urban crowding, began to recede near the end of the 19th century. In France, the development of industrial capitalism during the Second Empire went along with a pause in the decline of mortality, but beginning in the 1880s, mortality resumed a sharp decline, with a massive and widespread decline in infectious diseases. This decline continued throughout the 20th century, with two interruptions due to the two World Wars.

According to Omran, during the third age, the progress of life expectancy slows down and tends to reach a threshold. Numerous traditional causes of death continue to decrease, but the benefits of this new situation are countered by the rise of new endemic diseases, whether they are degenerative diseases (e.g., cardiovascular diseases, cancers, diabetes, metabolic disorders) or man-made diseases. Omran includes in this category "radiation injury, accidents, occupational hazards, carcinogens in the environment and in industry, and food additives." It is debatable whether mortality due to degenerative diseases has really increased globally. In France, the mortality increase attributed to cardiovascular diseases is deceptive, because if deaths due to ill-defined causes are realistically distributed going as far back in time as possible (1925), cardiovascular mortality is receding (Vallin and Meslé, 1988; Meslé and Vallin, 1988). There is no doubt that these diseases proportionally weigh more and more in calculating the total death rate. For man-made diseases, mortality has genuinely increased—or rather skyrocketed. However, the elements mentioned by Omran are not necessarily the determining factors. Smoking and alcoholism, which he does not even mention, are prime factors; most accidents are traffic accidents, and domestic or workplace accidents are declining.

Notwithstanding this criticism, Omran's division into three ages accurately reflects reality up until the end of the 1960s, when he wrote his first article. At that time, in all the industrialized countries, the progress of life expectancy was beginning to slow down, even stagnate, and in some cases (Eastern Europe), it began to decrease (Vallin and Chesnais, 1974). In France, a distinct pause can be observed among males, and progress in life expectancy was slowing among females (Meslé and Vallin, 1993). Many investigators believed at the time that life expectancy in the most advanced countries was close to the maximum limit imposed by human nature (see Chapter 48). Jean Bourgeois-Pichat (1952) calculated in the early 1950s a *maximum biologic life table*, showing that the maximum life expectancy was 76 for men and 78 for women. Similarly, to establish the first world population forecasts, the United Nations stated in the early 1980s that all populations were converging toward a life expectancy of 75 years (i.e., a little less for males and a little more for females) (United Nations, 1985).

However, events very soon belied this model. In all the Western countries, the advances of life expectancy accelerated again in the early 1970s because of a massive decline in cardiovascular diseases. At that time, Jay Olshansky and Brian Ault (1986), followed by Richard Rogers and Robert Hackenberg (1987), began to speak of a "fourth stage of the epidemiologic transition."[3] For the former, the diseases remained the same; only the ages at death had grown older. For the latter, we were dealing with a new era, in which great advances were and would be occurring because individuals were acquiring healthier habits. This notion of a fourth stage, which can be added on to Omran's epidemiologic transition, is rather unconvincing. Omran's idea was based mainly on the shift

[3] This was a "fourth stage of the epidemiologic transition" for Jay Olshansky and Brian Ault; it was a "new stage," also qualified as "hybristic," for Richard Rogers and Robert Hackenberg.

from an old permanent regime to a new permanent regime, after a transition toward lower mortality. It was therefore possible to say that the transition phase has grown longer and still retain the same three-phase explanatory model. However, this would not account for the fact that the new era of progress beginning in the 1970s was based on a major epidemiologic change, the cardiovascular revolution, which is not linked in any way with the end of the era of infectious diseases, even though it did follow closely. For this reason, it seemed preferable to us (Vallin, 1993) to adopt the semantic shift proposed by Julio Frenk *et al*. (1991) and include within the wider concept of health transition an initial phase (that described by Omran) of life expectancy gains, attributed mainly to the decline in mortality due to infectious diseases, followed by a second phase dominated by the decline in cardiovascular diseases, leaving open the possibility of identifying later phases, such as that suggested by Shiro Horiuchi (1999) for cancer and perhaps senescence.

2. Changes in Causes and Ages at Death

Two important phases in health advances can be distinguished based on the control of two very different types of diseases that have had an impact on different ages at death. If we take the example of France, we can divide the life expectancy gains obtained during a given period between those due to the changes in the different causes of death and those due to the probability of dying at different ages.[4] It is difficult to closely follow cause-specific mortality trends due to the breaks periodically introduced into statistical series by successive changes in the nomenclature of diseases. France, like most developed countries, uses the International Classification of Diseases (ICD), which was adopted by the International Institute of Statistics (IIS) in 1893 and by the World Health Organization (WHO) after World War II. This classification was revised 10 times within the space of a century, sometimes with wide-ranging modifications (see Chapter 42). The main reason for taking France as an example is that we have at our disposal reconstructions of historical series of deaths, classified by cause of death, on the basis of a constant definition (Vallin and Meslé, 1988, 1998). Nonetheless, this series only goes back to 1925, and we can therefore measure with precision the role of epidemiologic changes in the life expectancy trends only over the past 75 years of the period of health transition. However, we can obtain a good idea of it by identifying three periods: 1925 to 1948, when the decline of infectious diseases was still the dominant pattern; 1948 to 1968, during which this decline was less productive and partly countered by the rise of man-made diseases; and 1968 to 1996, when the advances in life expectancy gained new impetus with the control of man-made diseases and the decline of cardiovascular diseases.

Figure 57–2 represents a breakdown of life expectancy gains obtained in France during these three periods, between the different effects due to the trends of six large groups of causes:[5]

1. Infectious and respiratory diseases (i.e., Chapters 1 and 8 of the ICD, 9th revision)
2. Other cancers (i.e., Chapter 2 of the ICD, except bronchial and lung cancer)
3. Cardiovascular diseases (i.e., Chapter 7 of the ICD)
4. Man-made diseases (i.e., bronchial and lung cancers, alcoholism,[6] car accidents, suicide, and homicide)
5. Other diseases (i.e., all other items except those in Chapters 16 and 17 of the ICD)
6. Other violent deaths (i.e., other items of Chapter 17 of the ICD).

From 1925 to 1948 in France, of the 10.8 years of life expectancy gained by males, almost three-fourths (7.4 years) of the years were gained because of the decline in infectious and respiratory diseases. Among females, the proportion (7.5 of 12.2 years) is a little smaller because of the significant decline in cardiovascular diseases. From 1948 to 1968, the total life expectancy gain is twice as low (5.1 and 6.6 years, respectively), and the fact that the time interval is shorter (20 years instead of 23 years) does not alone explain it. There are probably two reasons. The impact of the decline of infectious diseases is decreasing, the impact of man-made diseases is decisively negative. During the following period (1968 to 1996), life expectancy gained once again—6.5 years for males and 6.9 years for females—but for entirely different reasons. The decline in mortality due to infectious diseases, which was still continuing, produced almost nothing in terms of life

[4] Different methodologies are available, including those developed by Evgueni Andreev (1982) and John Pollard (1982), which are based on the same principle and provide almost identical results (see Chapter 16).

[5] We first performed a proportional distribution of deaths due to ill-defined causes (i.e., Chapter 16 of the ICD) to avoid the considerable bias from the fact that the proportion of deaths due to ill-defined causes has dropped from more than 35% in the 1920s to less than 6% today.

[6] This category includes items 291 (alcoholic psychoses), 303 (alcohol dependence syndrome), 305.0 (alcohol abuse) and 571.0 to 571.3, and 571.5 (cirrhosis of the liver) in the 9th revision of the ICD.

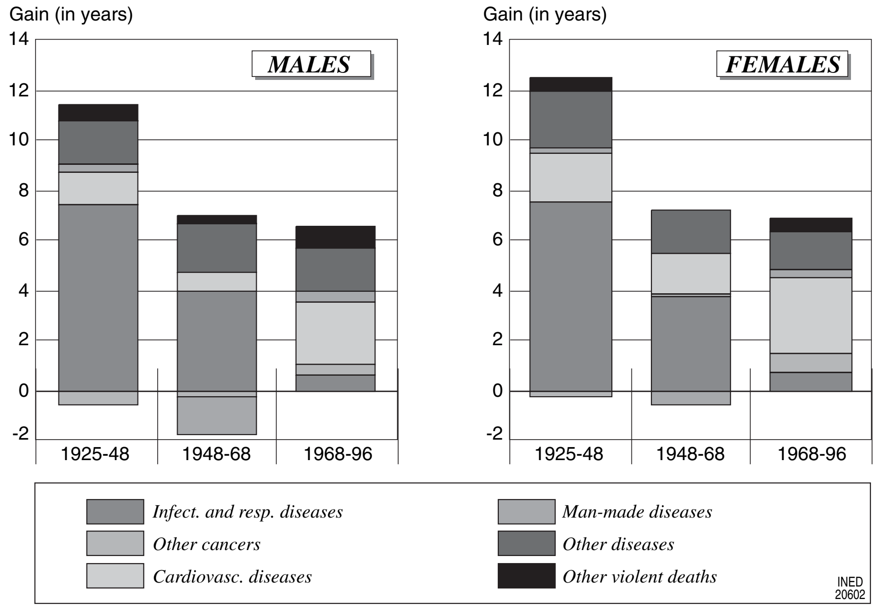


FIGURE 57–2 Contribution of the six large groups of causes of death in life expectancy gains according to sex during three periods in France.

expectancy gains because the level of this type of mortality was already very low. However, the decline in man-made diseases reversed the impact of this new trend on life expectancy and especially the decisive decline of cardiovascular mortality increased life expectancy by 2.6 years among males and 3.1 years among females. Moreover, for the first time, the category "other cancers" played a role in the lengthening of life expectancy.

The 1960s were marked by an important turning point in the health transition, when cardiovascular diseases replaced infectious diseases as the main source of life expectancy gains. The stage where Omran's epidemiologic transition left off is now far behind. However, there is another point concerning which Omran's pessimism is outdated: Not only has the rise in man-made diseases been curbed, but these causes of death have to a large extent been reduced. This is particularly clear among males, who are more affected by these diseases than females. Figure 57–3 shows how in France, from the 1960s on, these causes of death were gradually overcome.

The reversal of the tendency occurred first for alcoholism, beginning in the mid-1960s. For this cause of death, the losses in life expectancy turned into gains beginning in 1968–1978. The phenomenon occurred again a little later for traffic accidents. We know that the reversal of mortality trends due to traffic accidents began in 1974; for this reason, in Figure 57–3, life expectancy gains appear substantial only as of 1978–1988. The negative impact on mortality due to bronchial and lung cancer ceased only in the more recent period. This decline began in the early 1990s and should lead to life expectancy gains among males in later decades.

This extensive redistribution of the impact of different death causes on life expectancy trends went along with a radical transformation of the distribution of different age-specific death rates. From this point of view, available statistics make it possible to follow the effects of the health transition since it began. Figure 57–4 represents life expectancy gains due to the mortality decline of six age groups during six periods since the mid-18th century, in France. The three most recent periods are the same as those used in Figure 57–2 for causes of death, and we simply divided the two previous centuries into three periods.

For 2 centuries and up to the 1920s, the increase in life expectancy was almost entirely accounted for by the decline of infant and child mortality. From 1740–1749 to 1925, the life expectancy of males gained 29 years and that of women gained 32 years. In both cases, 23 years can be ascribed to the mortality decline of children younger than age 15 years. Between 1817 and 1885, the life expectancy gains are much lower than during the preceding and following periods, and this reflect the fact that the decline of child mortality slowed down at that time. A temporary increase in child mortality was observed in France during the Second Empire. The decisive impact of infant and child

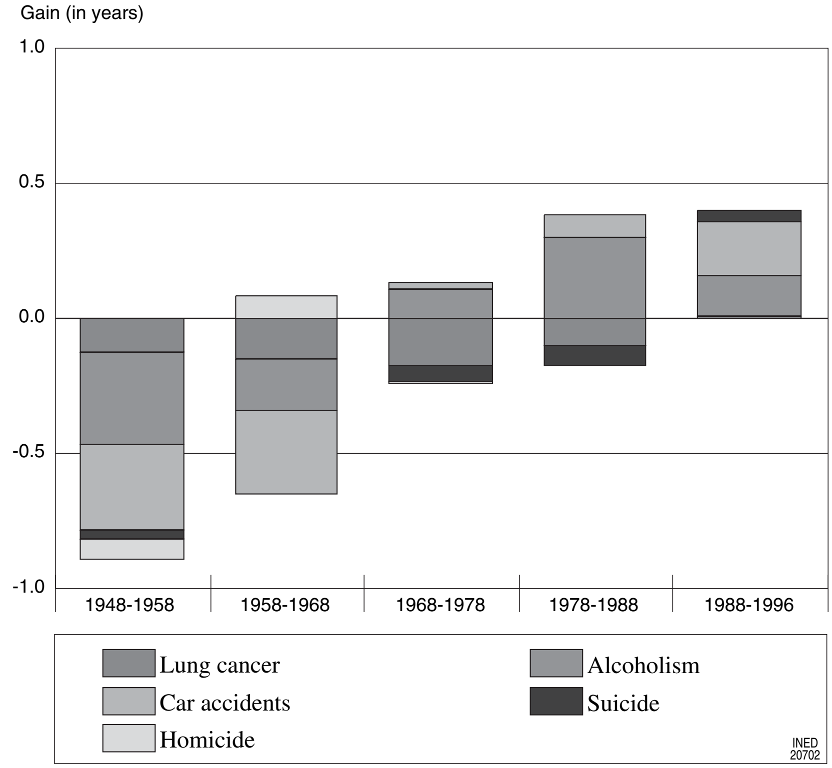


FIGURE 57–3 Contribution of five types of man-made diseases to gains or losses in life expectancy for males during three periods in France.

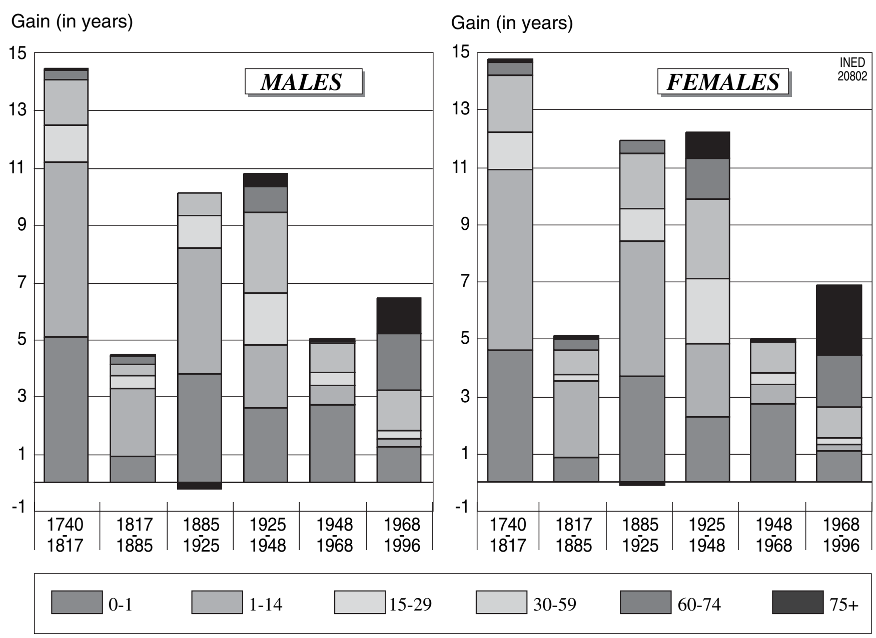


FIGURE 57–4 Contributions of six age groups to the gains or losses of life expectancy for each sex during three periods in France.

mortality trends corresponds to that of childhood infectious diseases.

The period between 1925 and 1948 appears to be quite different. The decline of child mortality led to still more gains in life expectancy, but substantial gains were also caused by the decline of adult mortality, particularly between at age 30–59 years. The decline of infectious diseases led to gains at all ages. Tuberculosis in particular began to recede, and maternal mortality from infectious causes began a steep decline with the arrival of sulfamides. Very soon, thanks to vaccines and antibiotics, the decline in child mortality led again to most of the life expectancy advances after World War II, whereas the rise of man-made diseases deprived adults of most of the advantages resulting from these medical advances.

The situation radically changes when we move to the last period (1968 to 1996). The decline of child mortality produced only marginal effects, whereas that of elderly persons began to play an unprecedented role. Among females in particular, the decline in mortality after age 60 represents a gain of 4.2 years, for a total of 6.9 years or more than 60%. Among males, the proportion is 50%. The decline in mortality at age 75+ represents on its own a 2.4-year gain among females and 1.2-year gain among males.

3. Health Transition Factors

There has been much debate concerning the factors responsible for the health transition, particularly about the respective roles of medical advances and social and economic development (Vallin, 1989). Other points of view have also stressed the importance of cultural and social changes (Tabutin, 1995; Caldwell, 1986, 1990).

In the 1960s, most researchers agreed that medical advances were mainly responsible for the health transition. From Pasteur to antibiotics, effective weapons against infectious diseases had been discovered, and after World War II, the effectiveness of these techniques were tested among the populations of the developing countries, still very much affected by infectious diseases. Paul Demeny (1965) went so far as to deny the impact of economic and social progress in the decline of world mortality during the 3 decades that followed the war, whereas Colin Clark (1967) believed that medical progress was the only factor responsible for the increase of life expectancy in England in the 18th and 19th centuries. At that time, however, Thomas McKeown (McKeown, 1965, 1976; McKeown and Record, 1962) held an opposite view, because he believed that the mortality decline observed in England in the 19th century was for the most part attributable to advances in nutrition; this opposing view led to violent disputes. Today, these disputes have lost their relevance, because it appears that these two factors do not suffice alone or together to account for health advances. Reality is far more complex, and factors have also varied to a large extent in time and space.

In Europe, during the 18th century, mortality began to recede thanks to the double impact of the control of the widespread epidemics that were so frequent in the past (not so much through medical advances but through measures taken to protect populations) and the increase of available food, which was attributed to the development of agriculture and the improvement of transport and storage conditions. In France, life expectancy rose from less than 30 years in the mid-18th century to a little more than 40 years in the 1840s. During the Second Empire, economic progress based on industrialization and the development of uncontrolled capitalism led to an interruption of the life expectancy increase during 3 decades (from 1850 to 1880) because of the concentration of an ever-larger proportion of the population in insalubrious urban zones and unhealthy working and living conditions, particularly for children.

Life expectancy began to increase again because of the discoveries of Louis Pasteur and their implementation through new health policies. However, even if medical progress played a very important role, it was not the only factor involved. Pasteur's ideas had to be concretely applied in daily life, particularly through rules of hygiene and food preserving techniques. Medical and health advances were spread mainly through educational progress, because education became free and compulsory at that time thanks to Jules Ferry. This double progress, both medical and sociocultural, largely contributed to the control of infectious diseases in a general context of economic and social development. More progress in medical techniques was made in the 1930s with the invention of sulfamides, and the decisive moment occurred in the 1940s with the invention of antibiotics and new vaccines. At that time, these advances made their way to the population thanks to the gradual implementation of generalized health insurance systems after World War II. In France, life expectancy rose from 43 years in the early 1880s to almost 60 years on the eve of World War II and 70 years in the early 1960s.

This considerable increase in life expectancy, due to two centuries of progress in nutrition and to the victorious struggle against infectious diseases, although first limited to Europe and North America, rapidly spread during the 20th century to the rest of the world. Significant advances were made in some regions of

Latin America and Asia in the first half of the century, but it is mainly after World War II that these advances reached the developing countries. The question of the spread of health advances to these countries led to the development of numerous theories and explanatory schemes of the mortality decline. These naturally involved multiple factors (Palloni, 1985; Mosley and Chen, 1984). The medical techniques developed in Europe and North America played a determining role, enabling some poor countries (e.g., Sri Lanka, Cuba, Costa Rica, Kerala) and even the largest among them, China, to reach life expectancy levels in the 1970s or 1980s that were quite close to those of the wealthy countries, even though these countries had not experienced major economic development (Preston, 1975). Nonetheless, efficient health policies still remained to be implemented so that the majority could benefit from these advances. After World War II, the success of vaccines and antibiotics led some to believe that it would be enough to organize eradication campaigns targeted at the main endemic diseases (Coale and Hoover, 1958). Indeed, because of the efforts of the WHO, some campaigns were very successful, the most famous case being the eradication of malaria in Sri Lanka whereby life expectancy in that country gained 12 years within 2 years (Molineaux, 1985a, 1985b; Meegama, 1967; Frederiksen, 1960). However, it was soon found that to spread health progress to the population in a lasting and effective way, health care policies had to account for these numerous factors. The WHO completely revised its action by adopting in 1978, at the Alma Ata conference, the strategy of "primary health care," which involves the dissemination at the local level of a number of simple and efficient medical techniques; the development of food production, drinking water, and environmental measures; and educational progress and improvement of the status of women. This strategy of primary health care has often been criticized, but the reason why it has often been unsuccessful is that its implementation requires political reforms that few governments are willing to implement (Mosley, 1985a, 1985b; Suarez-Ojeda and Yunes, 1985a, 1985b; Rosera-Bixby, 1985a, 1985b). The factors that determine health progress are doubtless a complex combination of technical, socioeconomic, cultural, and political factors in which each element interacts with the others.

What was true for the first phase of the health transition based on the control of infectious diseases is even truer for the next phase, in which the increase of life expectancy depends more and more on the decline of other diseases, particularly cardiovascular diseases. The issue is much more complex in terms of medical aspects. There are no miracle drugs such as vaccines and antibiotics to treat cardiovascular diseases, only various forms of treatment such as anticoagulants, beta-blockers, pacemakers, and coronary bypass surgery; treatment also includes the organization of medical emergency services. Medical prevention is more difficult to organize, because in this case it is not possible to vaccinate once and for all, but only to monitor certain risk indicators (e.g., blood pressure, cholesterol levels). In the case of infectious diseases, global policies are enough to improve environmental factors (e.g., environmental clean-up, drinking water distribution), but the context favoring cardiovascular disease can be modified only through the complex interaction of global policies and changes in individual behavior. In the case of smoking, for example, the decision to stop smoking is an individual matter, but this decision will more likely not be taken if smoking remains valued in the social culture. Political action and individual awareness are interdependent. Such changes are even more difficult in the realm of eating habits. In this case, the problem is not the consumption of one specific, clearly identified food item, but the subtle balance of quantity, quality, and the type of ingredients that make up daily meals. At the political level, the message is much harder to convey, and on the individual level, it is very difficult to change a person's eating habits. Beyond the obvious impact of advances in medical technology, as concerns the cardiovascular revolution and the initial successes in the struggle against cancer, public awareness of the fundamental importance of behavioral risk factors is a crucial element. It is the driving force of the second phase of the health transition, on a par with the role of Pasteur's discoveries and the use of antibiotics during the first phase.

This combination of factors came together quite late in the history of economic and social development in industrial countries. True, infectious diseases had to be dealt with before cardiovascular diseases could be addressed, but a high degree of economic and social development was also needed to ensure the necessary material and cultural means for these advances. This is what happened in the 1970s in most developed countries, and as a result, life expectancy in France rose from 72 years in 1970 to almost 79 years in 1999, crossing the 75-year barrier that the United Nations in the early 1980s still considered to be the upper limit of life expectancy. However, this new stage of health progress did not occur in all industrialized countries (Bourgeois-Pichat, 1985a, 1985b). The most notable exception is Eastern Europe, which instead of following the path of decline in cardiovascular disease, combined all the disadvantages of the third age of Omran's epidemiologic transition: pandemic of degenerative diseases and expansion of man-made diseases (Shkolnikov *et al*., 1995, 1996; Meslé, 1991).

II. THE HEALTH TRANSITION AND FUTURE PROSPECTS

The future developments of the health transition increasingly depend on the relationship between life expectancy and longevity, which so far has not been studied in depth. Until now, researchers have focused on life expectancy, with the implicit idea that longevity is an intangible fact; however, as demonstrated in Chapter 48, the question of longevity has become extremely important. Depending on whether the upper limit of human life can be extended, the potential increase of life expectancy can vary considerably. If it is intangible, life expectancy can still significantly increase, but if it is not, it can increase indefinitely. In any case, we must raise the question of the consequences of the health transition on the quality of the years gained.

1. Rectangularization of the Survival Curve?

Since the middle of the 18th century, the increase in life expectancy has mostly resulted from the fact that deaths that occurred at earlier ages began to occur at older ages, even if this extension does not necessarily go beyond the age limit biologically imposed on human life. As described in Chapter 48, as mortality declined, the survival curve of the population became increasingly "rectangularized" (Fig. 57–5).

The life tables for the years 1740–1749 show that the proportion of survivors dropped suddenly at very early ages due to the high level of child mortality. At the age of 1 year, it was down to 70%, and before the age of 10, one-half of the cohort had already died. Some exceptional individuals did come near or even exceeded 100 years. The life table of French women in 1996 showed that there were still 50% of survivors at 85, but even if living more than 100 years is becoming less and less unusual, it is still an exceptional feat.

On the basis of this observation, it can be concluded that there is a limit to the human life span. This limit is well over 100 years, although it cannot be precisely defined because our only reference is the empirical value of a record that will always remain temporary and whose validity depends on the probability of its happening again and the effectiveness of the administrative and statistical system. This record is held by Jeanne Calment, who was a little older than 122 years when she died in 1997. Observation shows that only very few individuals have come anywhere near such a record and that most people die before or very little after 100 years. The available space for such an increase in life expectancy is limited overall to the area situated between the present survival curve and the top right angle of the rectangle defined by a vertical line at 100 years or a little more than 100 years (see Fig. 57–5).

If we take into account the heterogeneous nature of the population and individual life stories, it would be unlikely that all individuals would live until 100 and that afterward all or almost all immediately die. We can easily conceive that the limit survival curve should

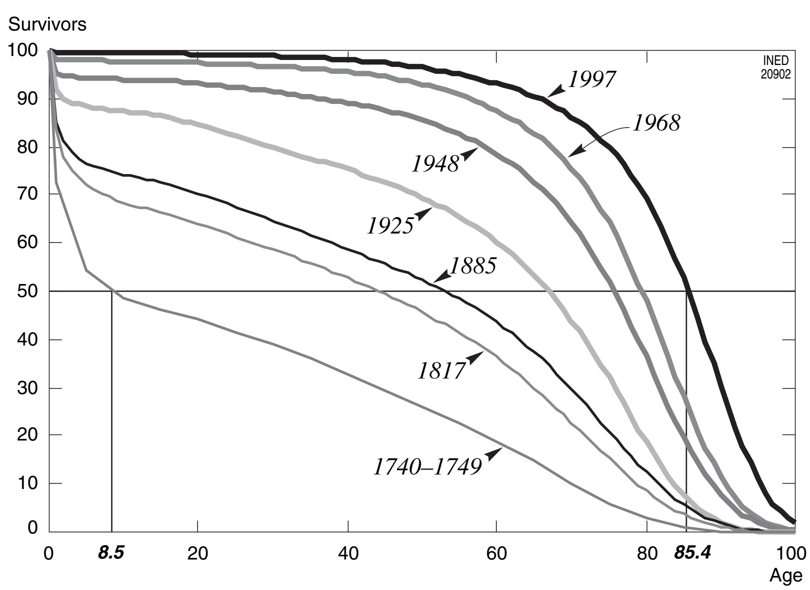


FIGURE 57–5 Rectangularization of the survival curve of French females.

be situated somewhere between the present curve and a perfect rectangularization. However, opinions diverge about the location of this "somewhere."

Certain demographers believe that we are very close to this limit. Jay Olshansky (1990), drawing his inspiration from the very pessimistic views of biologist James Fries (1980), shows that, on the basis of the 1985 life table of the United States, to reach a life expectancy of 85 years, it would be necessary to reduce by 43% the age-specific death risk of females and by 65% that of males. Because, according to this researcher, it would be very improbable that higher reductions might be obtained, he believed it most unlikely that life expectancy could ever exceed 85 years. Other researchers are more optimistic. Josianne Duchêne and Guillaume Wunsch (1990), for example, using the age limit of 115, on the basis of research carried out by several gerontologists (McGinnis, 1985; Medvedev, 1975; Korenchevsky, 1961) and taking a modal age at death of 95 years, suggest, after having conducted different types of modeling, a limit life table yielding a life expectancy of 91.4 years. In a more empirical fashion, we extrapolated the particularly stable and regular trends of the different age-specific death risks observed in France over the last 20 years and observed that within the next century, life expectancy could reach 91 years for males and 95 years for females (Vallin and Meslé, 2001). It seems quite probable, and in any case possible, that life expectancy will continue to increase well beyond the 85-year age limit suggested by Jay Olshansky. Nonetheless, if human longevity is finite, it would be very difficult to reach a life expectancy of 100 years.

2. Or a New Expansion of Ages at Death?

The second issue is to find out whether the limit of the human life span, which is today a little more than 120 years, is really intangible. In the early 1950s, Paul Vincent (1951) had suggested the idea of determining the upper limit of human life by extrapolating the trends of exponential mortality increase with age until the death risk reaches 100%. He then found, on the basis of French data, that the maximum duration of human life was 107 years. Twenty years later, using the same method but applying it to more recent data, Françoise Depoid (1973) estimated this maximum age to be 115 years. Did this mean that the age limit had really increased? As seen in Chapter 48, the data of five countries that have long had vital statistics enabled John Wilmoth and Hans Lundström to show that since statistics have become reliable, particularly in recent decades, the highest age at death has increased every year. The fact that the population is growing increases the likelihood that some will reach a very advanced age, especially given the increase in the proportion of persons reaching old age due to the decline of mortality among children and adults. Nonetheless, it is probable that part of the observed phenomenon actually reflect an extension of the human life span. This idea is also suggested by the recent acceleration of the mortality decline at more advanced ages. In France, this acceleration occurred in the early 1970s for mortality between 65 and 75 and in the early 1980s for mortality between 75 and 85 (Meslé, 1995). Even more significant, life expectancy at the age of 100 has also been increasing in the past decades. Among French females, it rose from 1.5 years in the mid-1950s to 2.2 years in the early 1990s (Allard *et al.*, 1996; Meslé and Vallin, 1998). The more reliable data on mortality at very advanced ages we have, the more it seems that the mortality increase with age slows down, contrary to the Gompertz model, which until recently was universally accepted. The explanation seems to be the very strong selection due to high mortality at these ages among a population that is very heterogeneous to begin with but becomes more and more homogeneous, being made up of the more resistant individuals (Barbi *et al.*, 2003; Vaupel *et al.*, 1979, 1998). Given these facts, the idea that recent health advances may partly reflect the slowing of biologic aging becomes more credible. This would obviously represent a new phase in the health transition, radically different from the previous ones. Without reaching the extreme conclusions of André Klarsfeld and Frédéric Revah (1999), who suggest that if immortality does not exist, the reason is not biologic necessity as always believed, but only the fact that it was not selected by evolution because it had no biologic purpose, we may nevertheless recall that in the early 1980s, Roy Walford (1984) did come to believe that a life expectancy of 150 years could be reached in the 21st century.

If, after the cardiovascular revolution, advances in cancer research should bring about the end of the second phase of the health transition—that of the decline of degenerative diseases, mainly thanks to the prevention of risk factors—it is too soon to find out whether a third phase will ever occur, such as through the control of genetic factors responsible for biologic aging.

Whether the third phase will ever happen, it is now certain that the second phase of the health transition—which, unlike the first, is no longer based on the decline of mortality at young ages but on that of mortality at advanced ages—will lead to new societal issues tied to the aging of the population, with mortality at advanced ages representing a major factor in the latter phenomenon. At the heart of this issue, paradoxically, lies the question of the health of the population.

3. Aging of the Population, Biologic Aging, and State of Health of the Population

In the past, during the phase when the health transition concerned mainly the control of infectious diseases of infants and children, the gains in terms of life expectancy at birth were almost exclusively gains in years of good health. However, because most of the years gained have resulted from the struggle against degenerative diseases, we may wonder whether health progress does not result in large part in making it possible for people with health problems to live longer. This question is more extensively discussed in later chapters,[7] but it is useful to say a few words about it at this stage.

Numerous surveys have been devoted to this issue during the past 2 decades in developed countries. The first results, mainly from predominantly Anglo-Saxon countries (e.g., Australia, Canada, United States), seem to indicate that the years of life that go along with disabilities and deficiencies have increased at a faster pace than total life expectancy. However, initial French surveys have partly contradicted these pessimistic results. If we make a distinction between moderate or slight disabilities and severe or very severe disabilities, we observe that years of life without severe disability everywhere increase as fast as the total life expectancy (Mormiche, 1999; Robine, 1993). These apparently contradictory results reflect the fact that in this field, there is no really objective measuring system to assess health and that as health improves globally, individual demands also rise, and along with these demands, the level of discomfort felt or declared. This phenomenon is particularly noticeable in cases of slight or moderate disabilities, which used to be considered normal. It is much less visible in cases of severe disabilities. For a given age, it would seem that health is improving, which is the logical conclusion, to the extent that the aim of health care is first and foremost to fight disease and suffering, and the fact that people die later than they used to is only a consequence of this effort.

However, if, at a given age, health improves while life expectancy increases, this does not necessarily mean that the average health situation of the population has improved. On the contrary, it may be deteriorating. In the current stage of the health transition, the life expectancy increase depends mainly on the decrease of mortality at advanced ages, which itself has become a major factor in the aging of the population. However, even if health improves at each age, there is no person whose health does not deteriorate with age. The question is to find out whether the improvement of health at a given age is enough to offset the deterioration of the global health situation caused by the aging of the population. If we retain present assessments of severe disability, we could be led to believe that for the time being, a certain balance has been created between these two phenomena, because the proportion of years spent in good health compared with the total number of years of life that have been gained remains about constant. However, this reasoning supposes that the population count does not change, and this is not the case. In the real context of most developed countries, we must expect in the next decades a significant increase of the disabled population, both in absolute figures and in proportion to the rest of the population.

[7] See Chapter 80 by Jean-Marie Robine and Carol Jagger, which is devoted to the consequences of demographic change.

Acknowledgment

This chapter was translated from the French version by Zoé Andreyev.

References

Allard Michel, Vallin Jacques, Andrieux Jean-Michel, and Robine Jean-Marie, 1996. In the search of the secret of centenarians: a French demographic and medical survey about centenarians, *in*: Graziella Caselli and Alan Lopez (eds.), *Health and mortality among elderly populations*, p. 61–86. International Studies in Demography. Oxford, Clarendon Press, 360 p.

Andreev Evgueny, 1982. Analysis of life expectancy through the components method [in Russian], *Vestnik Statistiki*, no. 3, p. 42–47.

Bah Sulaiman and Rajulton Fernando, 1991. Has Canadian mortality entered the fourth stage of epidemiologic transition? *Canadian Studies in Population*, vol. 18(2), p. 18–41.

Barbi Elisabetta, Caselli Graziella, and Vallin Jacques, 2003. *Trajectories of extreme survival in heterogeneous population*. *Population E*, vol. 58(1), p. 43–66.

Biraben Jean-Noël, 1975. *Les hommes et la peste en France et dans les pays européens et méditerranéens. Tome I: la peste dans l'histoire*. Paris, Mouton, 459 p.

Biraben Jean-Noël, 1996. *Les pathocénoses en Europe*. *in*: Mirko Grmek (ed.) *Histoire des maladies*, p. 1–36. Paris, Association Économie et Santé.

Blayo Yves, 1975. La mortalité en France de 1740 à 1829, *Population*, vol. 30 (special issue: *Démographie historique*), p. 123–142.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Bourgeois-Pichat Jean, 1985. L'évolution récente de la mortalité dans les pays industrialisés, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 489–522. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Bourgeois-Pichat Jean, 1985. Recent changes in mortality in developed countries, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 507–540. Liège, Ordina Édition, 558 p.

Caldwell John C., 1986. The role of mortality decline in theories of social and demographic transition, *in*: United Nations (ed.), *Consequences of mortality trends and differentials*, p. 31–42. New York, UN Department of International and Social Affairs, 191 p. (Population studies no. 95.)

Caldwell John C., 1990. Introductory thoughts in health transitions, *in*: John C. Caldwell, Sally Findley, Pat Caldwell, Gigi Santow, *et al*. (eds.), *What we know about health transition: the cultural, social and behavioural determinants of health*, p. vi–xiii. Canberra, Health Transition Centre, Australian National University, 1–499 p. (Special issue of *Health Transition*, vol. 1[2].)

Clark Colin, 1967. *Population growth and land use*. London, Macmillan, 406 p.

Coale Ansley J. and Hoover Edgar M., 1958. *Population growth and economic development in low income countries*. Princeton, NJ, Princeton University Press, 389 p.

Demeny Paul, 1965. Investment allocation and population growth, *Demography*, vol. 2, p. 203–232.

Depoid Françoise, 1973. La mortalité des grands vieillards, *Population*, vol. 29(4–5), p. 755–792.

Duchêne Josianne and Wunsch Guillaume, 1990. Les tables de mortalité limite: quand la biologie vient au secours du démographe, *in*: Michel Loriaux, Dominique Rémy, and Éric Vilquin (eds.), *Populations âgées et révolution grise. Les hommes et les sociétés face à leurs vieillissements*, p. 321–332. Louvain-la-Neuve, Éditions CIACO, Institut de démographie (UCL), 1118 p. (Chaire Quetelet, 1986.)

Frederiksen Harald, 1960. Malaria control and population pressure in Ceylon, *Public Health Reports*, vol. 75, p. 865–868.

Frenk Julio, Bobadilla José Luis, Stern Claudio, Frejka Tomas, and Lozano Rafael, 1991. Elements for a theory of the health transition, *Health Transition Review*, vol. 1(1), p. 21–38.

Fries James F., 1980. Aging, natural death, and the compression of morbidity, *New England Journal of Medicine*, vol. 303, p. 130–135.

Horiuchi Shiro, 1999. Epidemiological transitions in human history, *in*: Joseph Chamie and Robert L. Cliquet (eds.), *Health and mortality issues of global concern. Proceedings of the symposium Health and mortality, Brussels*, p. 54–71. New York, United Nations and CBGS, 467 p. (Proceedings of the Symposium on Health and Mortality, Brussels, November 19–22, 1997.)

Kirk Dudley, 1944. Population change and the postwar world, *American Sociological Review*, February, vol. 9(1), p. 28–35.

Klarsfeld André and Revah Frédéric, 1999. *Biologie de la mort*. Paris, Odile Jacob, 290 p.

Korenchevsky Vladimir, 1961. *Physiological and pathological ageing*. New York: Hafner, 514 p.

Landry Adolphe, 1934. La révolution démographique. Étude et essai sur les problèmes de la population. Paris, INED, 227 p.

McGinnis Michael J., 1985. The limits of prevention, *Public Health Reports*, vol. 100(3), p. 255–260.

McKeown Thomas, 1965. Medicine and world population, *in*: Mindel C. Sheps and Jeanne Clare Ridley (eds.), *Public health and population change*, p. 25–40. Pittsburgh, University of Pittsburgh, 557 p.

McKeown Thomas, 1976. *The modern rise of population*. London, Edward Arnold, 169 p.

McKeown Thomas and Record R. G., 1962. Reasons for the decline of mortality in England and Wales during the nineteenth century, *Population Studies*, vol. 16(2), p. 94–122.

Medvedev Zhores A., 1975. Aging and longevity, *The Gerontologist*, vol. 1(3), p. 196–201.

Meegama Sriniwasa A., 1967. Malaria eradication and its effects on mortality levels, *Population Studies*, vol. 21, p. 207–237.

Meslé France, 1991. La mortalité dans les pays d'Europe de l'Est, *Population*, vol. 46(3), p. 599–650.

Meslé France, 1995. La mortalité en France: le recul se poursuit, *Population*, vol. 50(3), p. 743–776.

Meslé France and Vallin Jacques, 1988. Les composantes de la mortalité cardio-vasculaire en France depuis 1925: résultats d'une reconstitution historique, *Population*, vol. 43(2), p. 391–425.

Meslé France and Vallin Jacques, 1989. Reconstitution de tables annuelles de mortalité pour la France au XIX[e] siècle, *Population*, vol. 44(6), p. 1121–1158.

Meslé France and Vallin Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in*: UIESP (ed.), *Congrès international de la population*. Montreal 1993, vol. 2, p. 365–382. Liège, UIESP, 493 p.

Meslé France and Vallin Jacques, 1998. L'évolution de la mortalité aux âges élevés en France depuis 1950, *in*: Joël Ankri, Partice Bourdelais, Francis Derriennic, Myriam Khlat, Pierre Mormiche, and Catherine Sermet (eds.), *La santé aux grands âges*, p. 5–48. Paris, INED, 140 p. (Les cahiers de l'INED, no. 147.)

Molineaux Louis, 1985. La lutte contre les maladies parasitaires: le problème du paludisme, notamment en Afrique, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 11–40. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Molineaux Louis, 1985. The impact of parasitic diseases and their control, with special emphasis on malaria in Africa, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 13–44. Liège, Ordina Édition, 558 p.

Mormiche Pierre, 1999. To assess the health status of the population. Between theoretical checks and measurement incertitudes, *in*: ISTAT (ed.), *Towards an integrated system of indicators to assess the health status of the population*, p. 295–321. Rome, ISTAT, 380 p. (Essays no. 4.)

Mosley W. Henry, 1985. Les soins de santé primaires peuvent-ils réduire la mortalité infantile ? Bilan critique de quelques programmes africains et asiatiques, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 101–136. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Mosley W. Henry, 1985. Will primary health care reduce infant and child mortality? A critique of some current strategies, with special reference to Africa and Asia, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 103–138. Liège, Ordina Édition, 558 p.

Mosley Henry W. and Chen Lincoln C., 1984. An analytical framework for the study of child survival in developing countries, *Population and Development Review*, vol. 10, suppl., p. 25–45.

Notestein Frank, 1945. Population, the long view, *in*: T. Schultz (ed.), *Food for the world*, p. 36–57. Chicago, University of Chicago Press, 367 p.

Olshansky S. Jay and Ault A. Brian, 1986. The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases, *The Milbank Quarterly*, vol. 64(3), p. 355–391.

Olshansky S. Jay, Carnes Bruce A., and Cassel Christine, 1990. In search of Mathuselah: estimating the upper limits to human longevity, *Science*, vol. 250, p. 634–640.

Omran Abdel R., 1971. The epidemiologic transition: a theory of the epidemiology of population change, *The Milbank Memorial Fund Quarterly*, vol. 49(4), p. 509–538.

Omran Abdel R., 1982. Epidemiologic transition, *in*: John A. Ross (ed.), *International encyclopedia of population*, p. 172–183. New York, Free Press, 2 vol., 750 p.

Omran Abdel R., 1983. The epidemiologic transition theory: a preliminary update, *Journal of Tropical Pediatrics*, vol. 29, p. 305–316.

Palloni Alberto, 1985. Santé et lutte contre la mortalité en Amérique latine, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 447–473. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Pitkänen Kari J., 1993. *Deprivation and disease. Mortality during the great Finnish famine of the 1860s*. Helsinki, Hakapaino, 176 p.

Pollard John H., 1982. The expectation of life and its relationship to mortality, *Journal of the Institute of Actuaries*, vol. 109(442), part II, p. 225–240.

Preston Samuel H., 1975. The changing relation between mortality and economic development, *Population Studies*, vol. 29(2), p. 231–248.

Robine Jean-Marie, 1993. Allongement de la vie et évolution de l'espérance de vie sans incapacité, *in*: Jacques Vallin (ed.), *L'avenir de l'espérance de vie*, p. 79–95. Paris, INED, PUF, 105 p. (Congrès et colloques no. 12.)

Rogers Richard G. and Hackenberg Robert, 1987. Extending epidemiologic transition theory, *Social Biology*, vol. 34, p. 234–243.

Rosero-Bixby Luis, 1985. Le Costa Rica, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 331–356. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Rosero-Bixby Luis, 1985. The case of Costa Rica, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 341–370. Liège, Ordina Édition, 558 p.

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1996. Health crisis in Russia. I. Recent trends in life expectancy and causes of death from 1970 to 1993. II. Changes in causes of death: a comparison with France and England and Wales (1970 to 1993), *Population, An English Selection*, vol. 8, p. 123–190. (Published in French: *Population*, 1995, vol. 4–5.)

Suarez-Ojeda Nestor and Yunes Joao, 1985. Childhood mortality in the Americas: probable effects of primary health care, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 139–162. Liège, Ordina Édition, 558 p.

Suarez-Ojeda Nestor and Yunes Joao, 1985. La mortalité infantile aux Amériques: effets des soins de santé primaires, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 137–158. Paris, INED, PUF, 542 p. (Travaux et documents, cahier no. 108.)

Tabutin Dominique, 1995. Transitions et théories de mortalité, *in*: Hubert Gérard and Victor Piché (eds.), *Sociologie des populations*, p. 257–288. Montreal, Presse de l'Université de Montréal, 518 p.

United Nations, 1985. World population prospects, estimates and projections as assessed in 1982. New York, United Nations, 521 p.

Vallin Jacques, 1989. Théorie(s) de la baisse de la mortalité et situation africaine, *in*: Gilles Pison, Étienne van de Walle, and Daniel Sala Diakanda (eds.), *Mortalité et Société en Afrique au sud du Sahara*, p. 399–431. Paris, INED, PUF, 446 p. (Travaux et documents, cahier no. 124.)

Vallin Jacques, 1993. Life expectancy: past, present, and future possibilities, *in*: Jean-Marie Robine, Colin Mathers, Margaret Bone, and Isabelle Romieu (eds.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*, p. 63–77. Paris, John Libbey, INSERM, 388 p. (Colloque INSERM, no. 226.)

Vallin Jacques and Chesnais Jean-Claude, 1974. Évolution récente de la mortalité en Europe, dans les pays anglo-saxons et en Union soviétique, 1960–1970, *Population*, vol. 29(4–5), p. 871–898.

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978*. Paris, INED, PUF, 608 p. (Travaux et documents, cahier no. 115 + 7 volumes annexes.)

Vallin Jacques and Meslé France, 1989. A long terme, l'écart d'espérance de vie entre hommes et femmes devrait diminuer, *Population*, vol. 44(6), p. 1244–1251.

Vallin Jacques and Meslé France, 1998. Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925 à 1993, *in*: Gérard Pavillon (ed.), *Enjeux des classifications internationales en santé*, p. 113–156. Paris, Éditions INSERM, 220 p. (Questions en santé publique.)

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX^e et XX^e siècles et projections pour le XXI^e*. Paris, INED, 102 p. + CD-rom. (Données statistiques no. 4–2001.)

Vallin Jacques, Meslé France, Adamets Serguei, and Pyrozhkov Serhii, 2002. A new estimation of Ukrainian losses during the 30's and the 40's crises, *Population Studies*, vol 56(3), p. 249–264.

Vaupel James W., Carey James R., Christensen Kaar, Johnson Thomas E., Yashin Anatoli I., Holm Niels V., Iachine Ivan A., Kannistö Vaïno, Khanzaeli Aziz A., Liedo Pablo, Longo Valter D., Zeng Yi, Manton Kenneth J., and Curtsinger James W., 1998. Biodemographic trajectories of longevity, *Science*, vol. 280, p. 855–860.

Vaupel James W., Manton Kenneth J., and Stallard Eric, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, vol. 16(3), p. 439–454.

Vincent Paul, 1951. La mortalité des vieillards, *Population*, vol. 6(6), p. 181–204.

Walford Roy, 1984. *Maximum life span*. New York, Avon, 230 p.

Wilmoth John R. and Lundström Hans, 1996. Extreme longevity in five countries: presentation of trends with special attention to issues of data quality, *European Journal of Population/Revue Européenne de Démographie*, vol. 12(1), p. 63–93.

S E C T I O N

II

DETERMINANTS OF MIGRATION

Introduction

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Section I of Volume I considered population dynamics and its three components; Section II of Volume I and Section I of Volume II began to explore the determinants of fertility and mortality, respectively. Section II of Volume II now turns to the determinants of migration. Subsequent volumes investigate history and future perspectives of population settlement, the relations between population and society, population policies and theories, and the methods of data collection and complementary tools of analysis.

Within the general system formed by these four volumes, Section II of Volume II deals only with the determinants of migration. Issues concerning sources and analytic techniques were covered in Section I of Volume I. Chapter 13 on longitudinal analysis of population exchanges, Chapter 14 on period analysis, and Chapter 22 on mobility and spatial heterogeneity surveyed the available sources and methods of analysis used to study migration. More specific aspects of sources, concepts, and methods are explored in Section II of Volume IV. Any further elements of analysis or descriptions of sources essential to understanding the determinants are referred to as needed in the various chapters that make up Volume II.

The history of migration is discussed Section I of in Volume III, which includes Chapter 67, written by David Coleman, on the key role of migration in population processes from prehistory to the modern period. Chapters 68 and 69 of Volume III review the internal and external population changes connected with demographic transition and its spread worldwide. Section II of Volume II offers a practical introduction to the study of determinants through a review of the distinguishing characteristics of one of the major migration streams of the past century.

Similarly, the many different consequences of migration, including population genetics, societal impacts (i.e., integration of migrants), economics (i.e., population and development), environmental effects, and the strictly demographic consequences already considered in Volume I in relation to population dynamics, fall within the scope of Section II of Volume III.

It is trivial to say that migration involves the movement of a person from a place of departure to a place of arrival, but the possible types of movement vary widely according to how the investigator defines the places of arrival and departure, the length of absence from the place of departure, the duration of stay in the place of arrival, and the reasons for the move. All of these different characteristics interact with one another. One possible approach to a study of the determinants of migration may be to establish a typology of migration as the basis for a thorough study of the determinants of each type of migration. Such typologies have been posited by many demographers, not least David Coleman, who in Chapter 67 describes a typology for analyzing migration in prehistory or ancient and medieval history. In a different context, Antonio Golini (2000) also developed a typology to account for modern migration. To do so, he argues, means referring to "criteria—distance, recurrence, duration, causes, and legality—that are prevalent in the literature, but generally taken discretely, when they should be combined."

For the first of these criteria, there is no minimum distance for migration, and any change of house, building, neighborhood, municipality, region, country, continent, or any other defining area is a geographic migration that lends itself to analysis, but it will vary in the nature of its determinants and consequences.

Recurrence and duration, two inseparable aspects, may also vary widely. Daily, weekly and seasonal migrations tend to be of short duration, whereas migration stays of longer than a year, several years, or decades are more rarely cyclical, even though they can never be described as permanent during the individual's lifetime.

The type of migration also varies, admittedly widely relative to the preceding criteria, with the cause of the move: labor migration; training; family changes and disunion; the search for housing; a desire or need to live closer to the place of work; education; catastrophes and other acts of nature (e.g., floods, earthquakes); war; repression; military occupation; pressure of political, ideological, or religious events; job requirements (e.g., sales representatives or other mobile occupations, professional or scientific meetings, exploratory missions); tourism; and cultural or religious activities.

Migration is more or less tightly controlled by law—invariably in the case of international migration and frequently so in some countries for some types of internal migration. A distinction may therefore be drawn between legal and illegal migration.

Working from the third criterion, Antonio Golini first distinguishes between three types of territorial mobility: migratory mobility (i.e., any non-involuntary change of place of residence), pseudomigratory mobility (i.e., compelled by natural catastrophes or pressure of political or religious events), and nonmigratory mobility and other territorial moves (i.e., vacationing, occupational, and other moves). He cross-references these three types of mobility with the three other criteria (i.e., geography, time, legality) to obtain the full typology sought.

Sound as its rationale is, this typology is only one classification among others advanced by other investigators who place varying emphasis on different criteria or refer to others more relevant to the aspects they wish to focus on. The concept of migration or mobility varies enormously in different societies and at different times in the long history of humankind and is likely to evolve further in the future. Starting by drawing up a full and final typology of spatial mobility may not be the best way of accounting for the determinants of migration, and it is probably an unachievable goal. In the words of Daniel Courgeau (1988), "the study of spatial mobility involves identifying the changes occurring over time in the relations between an individual or social group, and space."[1] It is a definition that counsels against attempts to develop a universally valid full and final typology of migration, because it does not exist. If it did, it would lock the phenomenon into a rigid framework that could not accommodate the emergence of unforeseeable new kinds of mobility or the disappearance of older ones. The less lofty aim in the following chapters is to discuss the determinants of different types of migration, both spatial (i.e., internal and international migration) and temporal (i.e., from permanent changes of residence to daily commuting), without claiming to give the whole picture.

The propensity to migrate, like the propensity to procreate, depends on a very complex, often interacting set of individual and collective economic, political, social, cultural, ecological, psychological, and other factors that vary in time and space. As with fertility and mortality, a theoretical reference model needs to be constructed as a basis for analyzing and explaining migration and for identifying the singular or conjoint linkages between the migration process and its various factors. Explanatory analysis must be informed by the historical context in which the process takes place and by information about the geographic unit concerned, which encompasses the places of departure and arrival for each migrant.

Because of the complexity of explanatory analysis applied to migration, demographers have often confined themselves to collecting data and developing models of various degrees of sophistication to recreate unreliable or missing information with which to describe migration dynamics (discussed in Section I of Volume I). This is a large area to cover, because economic and political overtones have made migration a popular field of study for other disciplines, especially economics and sociology.

Part I of this section focuses mainly on international migration. Chapter 58 introduces the topic by considering Western migration trends over the past century, and Chapter 59 systematically reviews the main current theories on international migration.

Part II considers the main factors of migration. Migration often originates in a living environment where scarce economic resources compel some individuals to seek better prospects in other lands nearby or far away. From this perspective, migration is arguably a process needed for financial well-being, which may not, however, always be attained (see Chapter 60).

Migration may also be seen as something other than a financially motivated rural-to-urban move. Individuals are not always attracted to towns in search of easier or

[1] "L'étude de la mobilité spatiale consiste à mettre en évidence les changements au cours du temps, qui sont à l'œuvre dans les rapports entre un individu ou un groupe social et l'espace."

better-paid work. They may instead seek better social, cultural, and public health amenities; opportunities for claiming a better place in society; and easier access to basic public services for themselves and their families. These aspirations drive the process of urbanization everywhere in the world, and they are more than ever a major component of spatial mobility in all populations (see Chapter 61).

Whole pages of the history of migration in the 20th century have been written by the victims of medium- or large-scale population flights caused by political crises. Repression, war, ethnic cleansing, and genocide have been behind the forced migration of individuals under threat and seeking refuge or political asylum in nearby or distant countries that were willing to accept them for various periods (see Chapter 62).

Along with what can be broadly described as socioeconomic and sociopolitical factors is the wide range of personal motivations that push migrants toward neighboring countries or distant regions (see Chapter 63).

Part III of Section II looks at two particular types of migration: daily commuting and return migration. These situations vary widely in space and in the interval of time in which the departure and return moves take place. Simple cases are travel between the place of residence and the place of work or education. Such movements may be of very short duration, occurring daily in most cases. Daily commuting is a significant aspect of migration issues and the analysis of spatial interactions (see Chapter 64). On a different scale, return migration after a stay abroad for various periods is a little explored but hugely important field of research because of the total numbers of people concerned (see Chapter 65). Because of their singular characteristics, the two concluding chapters give extensive consideration to issues concerning sources, methods of measurement, and analysis.

References

Courgeau Daniel, 1988. *Méthodes de mesure de la mobilité spatiale: migrations internes, mobilité temporaire, navettes*. Paris, Institut National d'Etudes Demographiques, 301 p.

Golini Antonio, 2000. I movimenti di popolazione nel mondo contemporaneo, *in*: *Atti del Convegno internazionale su "Migrazioni. Scenari per il XXI secolo*," Dossier di ricerca, vol. I, p. 89–160. Rome, Agenzia Romana per la Preparazione del Giubileo, 734 p.

P A R T

I

MAJOR MIGRATION STREAMS AND THEORIES OF INTERNATIONAL MIGRATION

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Accounting for the many facets and different phases of migration dynamics is a daunting challenge. The emigration decision is a supremely subjective one, based on an individual or family's assessment of the comparative advantages and disadvantages of a temporary or permanent move to another country or elsewhere in the country of residence. However, a migration that involves a large number of people over a vast geographic space and long period enters into the common history of a territory, and the determinants that shape it are of a different order. The introduction to Section II of Volume II stressed the importance of the historical and economic contexts in which migration occurs to help identify the factors that determine migration streams and enable them to be set within a theoretical framework of references that help elucidate the phenomenon.

In Chapter 58, Graziella Caselli surveys international migration developments in the 20th century, focusing on the developed countries, where big surges in immigration occurred especially in the first half of the century. She first describes the main migration routes of the late 19th and early 20th centuries, when Europe lost millions of population to the New World, and then analyzes that singular phase of international migrations for which the interwar period was known and the specific characteristics of the migrants compared with those of the departure and arrival countries. Detailed consideration is given to the history of migration in the latter half of the 20th century, when the phenomenon changed to gradually become all embracing in terms of the size of migration streams and their demographic and social composition. From the 1970s in particular, the coexistence of very different demographic groups and stages of economic development, along with the increasingly complex relations between cultural, social, and political systems, wrought radical changes in migration, imprinting the flows of the last 30 years with very different characteristics from those that prevailed in the first half of the 20th century. The chapter concludes with some considerations on the future of international migration and considers how wealthy countries can help immigrants to fit satisfactorily into society in the country of arrival. This chapter does not attempt to give a comprehensive history of mass migration movements. These are covered in some chapters of Volume III, where particular consideration is given to historic migration movements (see Chapter 67) and those that arose with the European demographic transition (see Chapter 68) and its spread across the rest of the world (see Chapter 69). The main aim of Chapter 58 is to introduce the reasons that may prompt mass migration movements.

In Chapter 59, Hania Zlotnik discusses the theories of international migration. She first reviews the main principles of the different theories, frameworks of analysis, and hypotheses from different fields of study and then offers a comparative analysis of their strengths and limitations. Based on the main thrust of her report to the United Nations in 1998, she examines how economic and sociologic approaches, as well as demography and political science, have informed the fundamentals of the various theories.

Reference

United Nations, 1998. *World population monitoring, 1997. International migration and development*. New York, United Nations, 204 p. (United Nations publication sales no. E.98.XIII.4.)

C H A P T E R

58

International Migration in the 20th Century: The Case of the Western Countries

GRAZIELLA CASELLI

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

The economic and social changes that accompanied the industrial revolution were at the roots of another revolution: the demographic transition. The transition from a natural population (i.e., high mortality and high fertility) to a controlled population (i.e., at a low level of both components) was a phase of unprecedented population growth. For hundreds of thousands of years, natural factors had regulated the number of births and deaths and thereby the progression of population numbers. However, from the middle of the 18th century, humans began to change the conditions of mortality without, in the first instance at least, changing those of fertility, thereby provoking an exceptional increase in the population as a result of the growing gap between the two phenomena. The first populations to experience the effects of this demographic trend were those of the northern and western countries of Europe; the southern and eastern countries began the same process a century later, as did the rest of the world after the First World War and particularly after the Second World War.

In the long century that lasted from the beginning of the 1800s until the First World War, Europe's population multiplied by almost 2.5, increasing from 188 to 458 million inhabitants (Livi Bacci, 1998a). An exceptional growth, which did not always go hand in hand with an equally rapid economic and social development and which eventually found an outlet in the great transoceanic migration that lasted about a century, from the middle of the 19th century to the 1930s.

To say that this remarkable population growth was the basis of this great migratory wave is to simplify the explanation of this phenomenon. Reasons for the departure of millions of people from their homelands are more complex. Uneven industrial development was not able to absorb the surplus workforce, which resulted as much from demographic growth as from the departure of part of the population from rural areas, where the increase in productivity made possible by the introduction of the first technologic tools from industry had caused significant unemployment. European cities and early industrial settlements were not able to absorb all the surplus workforce, while on the other side of the ocean, the scarcity of human resources, the extension of available land, and the numerous activities that needed to be developed resulted in a great demand for those looking for work and better living conditions. It is not easy to summarize the reasons for this veritable mass exodus in a few sentences. When emigration, which is a subjective decision taken after a personal and family assessment of its costs and benefits, involves such a large number of people in a collective historical act, the imbalance between resources and population appears in all its complexity (Birindelli, 1989). This is possibly the fundamental reason why some tens of millions of Europeans from the middle of the 19th century to just before the Second World War exiled themselves overseas, thereby lightening the demographic load of the Old Continent and contributing to the population growth of the New World.

During the first decades of the 20th century, the conditions that had led to this great wave of European emigration began to diminish. In this era, northern European countries were already entering a phase of advanced development in which the workforce had declined due to a decrease in the proportion of the population of working age, which itself resulted from the decline in fertility that started a few decades earlier. However, the problem of surplus population and insufficient development remained in the countries of southern and eastern Europe, but their emigrants could also find work in the more developed European countries, given that the demand for foreign labor in the traditional receiving countries had progressively declined. The rules of the labor market had nevertheless reacted very slowly to the drying up of the transoceanic flow of migration; a sudden interruption in migration was caused by the World War I, and after the 1920s, its revival was restrained by the new American policy regarding migration. Ten years later, around the time of the Great Depression in 1929, other northern countries and Latin American countries likewise adopted measures limiting immigration. The time of the great transoceanic migration came to an end quite quickly during the second half of the 1920s.

European migratory flows (not including the forced migration that preceded or followed the Second World War) resumed, with less intensity, in the 1950s. In the meantime, the connotation of the phenomenon and the global context had changed. The new flows directed to the Americas involved mainly family reunification and the systematic reception of war refugees and exiles. In Europe, the economic and political context resulting from the Second World War led the victorious countries (the richest) to look for labor in the poorer countries of the south for reconstruction purposes and then for the development process, which culminated in the economic boom of the 1960s.

From the 1970s, the decline of demographic growth in the European population paved the way (slowly but progressively) to immigration in all the countries of the Old World, a path taken mainly by the inhabitants of the poorest regions of Africa and Asia, who arrive today at the ports from which the Europeans departed a century earlier. All the countries of the world, rich and poor, have had to deal with problems created by migratory movements, whether internal, within their own borders, or external with the arrival of populations from neighboring countries or other continents. The largest of these movements, notably internal migration, was provoked by the technologic revolution and the profound transformations in the global economy that affected most populations of the world, to the benefit of some and to the detriment of others. To these fundamental circumstances, exceptional events have been added and superposed, such as natural disasters (e.g., drought, food shortages), civil wars, and nationalist and ethnic conflicts, which have given a new profile to the movements by upsetting the motives and the traditional structure of different societies, modifying as much their direction as their characteristics.

In recent decades, all internal and external migration has undergone substantial changes; international migration has acquired a global dimension, as much by its new scale as by its demographic and social characteristics (Collison, 1994). The widening of the differences in demographic growth between countries has played a major role in the explosion of the phenomenon and its rapid transformation. Whereas in the highly industrialized countries (notably in Europe), births have progressively fallen to very low levels, rates unimaginable at the end of the 19th century and even in the first half of the 20th century, in the developing countries, the speed of the demographic transition has for several decades resulted in very high growth rates, despite the efforts by international organizations to accelerate the decline in fertility by promoting the adoption of birth control measures.

As in the past, differences in economic growth and the conditions of poverty in which a significant part of the world's population find themselves lead to imbalances in demographic growth and constitute one of the major factors encouraging a growing number of individuals to emigrate from the poor countries to more economically developed countries. Violent internal conflicts add to the difficulties of remaining in regions that offer only minimal survival conditions due to lack of development and compel many people to flee, sometimes to distant lands. Moreover, the recent expansion of modern rapid mass transport have made long distance travel (formerly long and expensive) much more accessible.

These initial considerations suggest that the profound changes that have affected the phenomenon of international migration in the past few decades are caused by worsening of the problems in the poor countries of the planet. However, numerous studies show that these changes are mainly rooted in the development of the economic system of the most developed countries and the balance of international policy that results from it. The impact of the economic recession of the 1970s and 1980s on international migration led to a revision of internal and external economic policies by the rich countries.

The coexistence of different demographic systems and very different levels of economic development, without mentioning the growing complexity of cultural, social, and political relations, have profoundly changed migration and given very different characteristics to the flows of the last 30 years compared with the characteristics of migration in the first half of the 20th century. Because the object of this analysis is the development of international migration in the 20th century, I provide a framework that allows the identification of the different stages of this change. The importance of the great transoceanic migration, which has been the focus of numerous studies, is recalled in its broad outlines, with particular attention to its intensity and characteristics. The analysis then focuses on the past few decades, emphasizing especially the experience of European and North American countries.

TABLE 58–1 Flow of Total and Average Annual Departures from the European Countries and Italy toward Overseas Countries between 1876 and 1915

Period	Total flow[a]		Average annual flow[a]	
	Whole of Europe	Italy[b]	Whole of Europe	Italy[b]
1876–1880	1555	154	311	31
1881–1885	3357	320	671	64
1886–1890	3784	670	757	134
1891–1895	3437	751	68	150
1896–1900	2892	828	578	166
1901–1905	5569	2770	1114	554
1906–1915	13252	5853	1325	585
1876–1915	33846	11346	846	196
1901–1915	18821	8623	1255	574

[a] Flow figures are given in thousands.
[b] Total emigration out of Europe.
From Sonnino and Nobile, 1988; Birindelli, 1989.

I. THE GREAT MIGRATION AT THE TURN OF THE NINETEENTH AND TWENTIETH CENTURIES

Europe, which for 3 centuries, with Iberian and then English imperialism, dominated the rest of the world politically, economically, and demographically, experienced an unprecedented mass emigration in the 19th century. It is estimated that between 1846 and 1932, more than 55 million individuals left the continent in the direction of North America (more than 34 million to the United States and 5 million to Canada) and Latin America (about 7 million to Argentina and Uruguay and 4 million to Brazil). Among the countries that most contributed in populating the Americas are Great Britain and Ireland, with 18 million people; Italy, with 11 million; and Spain and Portugal, with almost 7 million. Significant numbers also departed from Germany (about 5 million) and from the Scandinavian countries (2 million from Sweden and Norway) (Livi Bacci, 1998b).

The phenomenon of the great migration reached its peak during the 40 years straddling the 2 centuries (from 1876 to 1915) when almost 34 million Europeans (including 11 million Italians) crossed the ocean to look for work (Table 58–1). During this period, there was no reduction in the number of departures. In the first 14 years of the 20th century alone, there were about 19 million departures, with a particularly high intensity between 1906 and 1910; the annual average surpassed 1.3 million (of whom 585,000 were Italians). At the same time, more than 9 million people (28% of departures) made the return journey within 40 years (Table 58–2).

The route to America was not the only one for the European populations. While vast hordes were crossing the Atlantic ocean, more than 5 million Europeans crossed the Urals headed for Siberia, and more than 3 million left to inhabit Soviet Central Asia. This emigration that began in Russia after the liberation of the serfs in 1861 ceased only when the fertile lands were totally occupied, and it began again between the end of the 19th and the beginning of the 20th century with the construction of the trans-Siberian and the industrial development of the Asian part of the Russian Empire, which later became the Soviet Union (Livi Bacci, 1998a).

In the years before World War I, the source of the European migratory process changed, from northwest to southeast; whereas previously most emigrants to America (70%) came from countries in the northwest of Europe, from the end of the 19th and the beginning of the 20th century, the Mediterranean, Balkan, and eastern countries of Europe supplied, in the same proportion, the transoceanic migration (Fig. 58–1). A change in the factors of migration and the characteristics of the migrants corresponded to this change in areas of departure, so much so that the flows recorded from the end of the 19th century can be called a "new wave of migration."

Migrants of the previous wave lived mainly in the northern and central countries of Europe and belonged to the intellectual and artistic classes (and were usually very well educated). They were driven by a desire to increase their well-being or by idealism; they were generally accompanied by their families (with both sexes more or less equally represented); and they

TABLE 58–2 Repatriation according to Country of Departure and Country of Return between 1876 and 1915

Country of departure	Country of return[a]						
	Sweden	United Kingdom[b]	Belgium[c]	Netherlands[d]	Italy[c]	Spain	Total
United States	132	3030	9	—	1978	8	5157
Argentina and Brazil	—	—	—	—	1332	214	1546
Australia and New Zealand	—	329	—	—	—	57	386
Other countries	6	1102	6	148	26	996	2285
Total	138	4461	15	148	3336	1275	9373

[a] Flow figures are given in thousands.
[b] Disembarking passengers (nationals and foreigners).
[c] For the period between 1886 and 1915.
[d] Dutch colonies.
From Sonnino and Nobile, 1988.

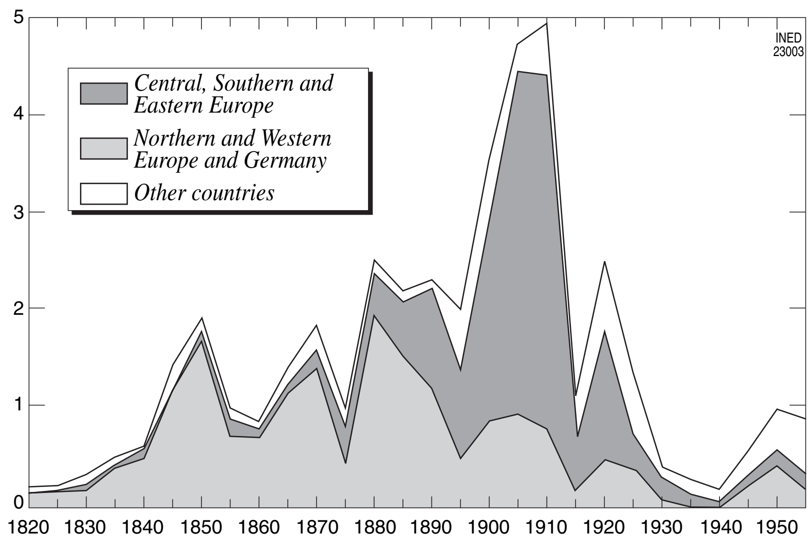


FIGURE 58–1 Old and new migration to the United States between 1820 and 1957 by geographic area of departure for 5-year age groups. (From Reinhard *et al.*, 1971.)

settled in agricultural regions. Later migrants left from the poorest regions of Europe, and they were of humble origins, generally rural, and mainly illiterate; they settled in towns where they looked for work in the nascent industries, and usually being men on their own, they had the intention of returning to their country as soon as they had put enough money aside. Return migration was much higher in this new wave than in the previous wave. In some periods, returns even exceeded the departures. Among the Italians between 1905 and 1915, for 5 million departures to America, 2 million returns (40%) were counted (Assante, 1978). However, the ratio of return journeys over departures increased considerably; between 1907 and 1911, it reached 70% (Foerster 1919, and 1968). According to the *Bollettino dell'Emigrazione* (1908, no. 6, page 4), in December 1907, while 5033 Italians were disembarking at the port of New York, 52,532 embarked to return to Italy. It was difficult to live in a distant country, unable to speak the language and especially separated from their families. They departed alone and remained so. As can be seen in Figure 58–2, the number of females in the great Italian migration was very low, never surpassing 20% of the total number of migrants (generally 22 to 23 women for 100 men).

The lateness with which the Mediterranean and eastern countries of Europe participated in the great migration can be explained by their delayed entry into the demographic transition. At the end of the 19th century, the countries of southern and eastern Europe experienced rapid demographic growth, whereas establishment and development of the industrialization process were delayed. These countries entered into the movement at a time when the development

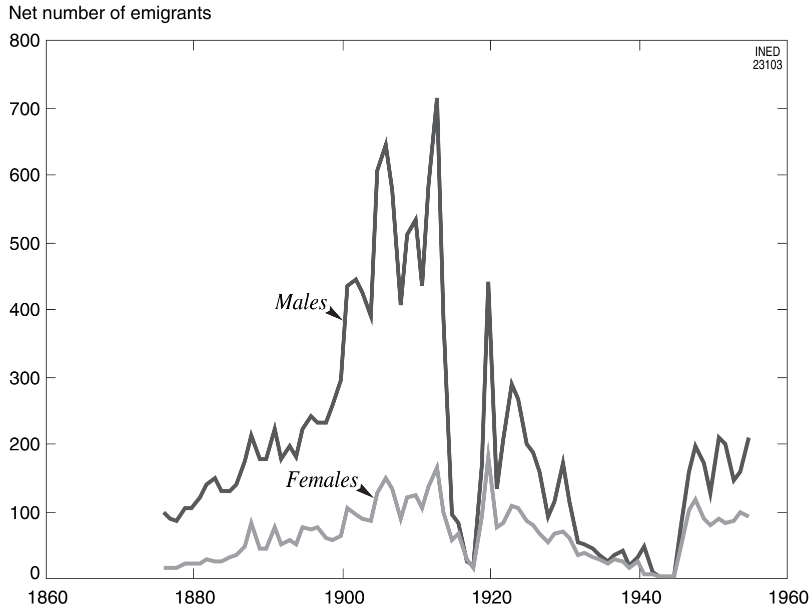


FIGURE 58–2 Italian emigration trends by sex from 1876 to 1955. (Data from Rosoli, 1978.)

TABLE 58–3 Migrants between European Countries from 1876 to 1915, according to Country of Origin and Destination

Country of origin	Country of destination							
	Norway, Denmark	Great Britain	German Empire	Austria–Hungary	France	Switzerland	Other countries	Total[a]
Sweden	119	—	—	—	—	—	47	166
Ireland	—	230	—	—	—	—	—	230
Belgium	1	6	23	1	258	2	53	343
Austria	—	—	—	—	—	—	2275	2275
Italy	—	72	1226	1425	1716	1.340	358	6137
Poland[b]	44	—	2408	—	3	—	—	2455
Russia[c]	—	—	1737	—	—	—	—	1738
Total	164	308	5394	1426	1977	1342	2732	13343

[a] Because of rounded-off numbers, the line totals and column totals may differ from the totals of the corresponding cells. Figures of migrants are given in thousands.
[b] Temporary emigration.
[c] Seasonal movement.
From Sonnino and Nobile, 1988.

process had already reached its peak in the northern and western countries of Europe, for whom the phase of high growth was already over. In these much richer countries, the demand for foreign labor was progressively increasing, to the point that many of them had become, by the turn of the century, the most frequent destination for the poor populations of neighboring countries. Between 1876 and 1915, 6 million Italians emigrated to Austria, Switzerland, and Germany, and 4 million Poles and Russians left for Germany (Table 58–3). As can be seen in Table 58–3, in the 40 years preceding the First World War, while about 34 million Europeans headed for the New World, 13 million others changed countries within Europe. For example, the total number of departures from Italy during this period was close to 15 million.

II. THE NEW MIGRANTS: BETWEEN REGULATION AND PROHIBITION

If the new migratory wave included an important internal component in Europe, it was not only because of the proximity of the more developed countries;

it was also because of the increasing difficulty of entering the American territories, especially the United States, where poor, uneducated immigrants were undesirable. At the beginning, during the first 15 years of the 20th century, this new migration, although undesirable, ended up surpassing the previous one, with 8 of 10 arrivals, because America still needed immigrants, and the new migrants were about the only ones available on the European market.

The Anglo-Saxon and Germanic bourgeoisie that held the economic power and dominated the ethnic and cultural landscape of the United States did not appreciate the new arrivals and saw danger in the introduction of these workers of humble origins, who were unspecialized, poorly educated, and inclined to organize themselves politically and into unions (examples of this had already been seen in Canada and Argentina) (Paris, 1975). Pressure on the political leaders to adopt measures limiting entries was strong and recurrent, and in 1907, the U.S. Senate created a commission (i.e., the Dillingham Commission, named after its chairman) that at the end of 3 years of assiduous work, produced a report suggesting that the new migration had been one of the causes of the economic and social crises of 1893 and 1905 (Paradisi, 1979). Moreover, in the numerous intermediate reports dealing with the issues of immigration, members of the commission insisted on the extremely negative influence that these new migrants would have. Migrants were accused, among other things, of being the cause of increasing overpopulation, public security problems, and decreased viability of urban zones (Tonizzi, 1999).

As early as 1907, a federal law was adopted prohibiting the entry of persons with mental problems or physical deficiencies that could impede on their professional activity and prohibiting unaccompanied children; another law, which foresaw the withdrawal of citizenship from American women married to foreigners, was aimed at limiting Japanese immigration. The work of the Commission and the laws of 1907 toughened the measures already adopted in previous years. The Literacy Act of 1897 (revised and modified to be more restrictive in 1917) stopped the entry of adults of 16 years or older who could only read and write their mother tongue, and a law adopted in 1903 prohibited the entry of anarchists and other subversives into the United States. These initiatives were intended more to establish quality control than to set a true quantitative limit. At the beginning of the 20th century, the United States was experiencing an accelerated phase of industrial development that made immigration necessary. Despite all this, the bourgeoisie and the economic leaders were opposed to this type of development and, in particular, did not share the excessively open attitudes of the democratic presidents on the issue of migration.

Despite these obstacles, between 1901 and 1915, of about 12 million immigrants arriving in the United States, almost 10 million came from the Mediterranean and eastern European countries (Table 58–4). These flows ceased only with the onset of the First World War, but they returned with greater intensity at the end of the war: 0.5 million between 1916 and 1920 and 2.5 million between 1921 and 1925, of which 1.5 million settled permanently in New York.

In 1921, faced with this huge inflow, a quota system was put in place for the first time (i.e., the National Origin Act, better known as the Quota Act), which fixed the maximum annual number of foreign entries

TABLE 58–4 Immigrants to the United States from 1901 to 1915 and Legal Annual Quotas after 1924, according to Country of Origin

Country of origin	Annual number of immigrants to the United States in 1901–1915			Legal annual quotas after 1924	
	Total (thousands)	Annual average (thousands)	%	Number (thousands)	%
German Empire	479	32	4.7	26	15.8
Great Britain and Ireland	1254	84	10.5	84	50.8
Norway and Sweden	560	37	4.0	7	4.5
Austria-Hungary	3035	202	25.4	7	4.0
Italy	2985	199	25.0	6	3.4
Russia[a]	2491	166	20.9	10	6.2
Other countries	1129	75	9.5	25	15.3
Total	11932	796	100.0	165	100.0

[a] Between 1901 and 1915, with Finland included. After 1924, including Poland.
From Sonnino and Nobile, 1988.

at 357,000. The 1924 law (i.e., the Immigration Act) again reduced the annual number of authorized entries to 164,000. However, the effects of the postwar depression and the approach of the great economic crisis of 1929 hindered foreign labor (and not only undesirable foreign labor) in an even more discriminatory way than the laws themselves. The new legislation foresaw quotas by nationality, thereby indicating a clear preference for the previous inflow, especially the one from the United Kingdom. Table 58–4 shows the annual quota granted to Anglophone countries from 1924 as 50.8%, whereas the total entries from these countries between 1901 and 1915 was only 10.5% of the total. The quota granted to Italy at only 11.6% was not in any way realistic, because in previous years, the Italian immigration had supplied 70% of arrivals.

In the years before the First World War, about 12 million Europeans departed for the United States, and another 10 million headed for other overseas destinations (Franzina, 1995): approximately 2 million to Canada, more than 3.5 million to Argentina and Brazil, and more than 1 million to Oceania (Table 58–5). At the time, immigrants to Canada and Australia were mostly Anglophone, and they were the first settlers to inhabit these new territories, seeking improved well-being rather than fleeing from economic necessity or demographic pressure in their home country. The Italians, on the contrary, forced to migrate by poverty, chose to settle in the United States and Latin America, notably in Argentina (Federici, 1979; Rosoli, 1978).

From 1920 to 1930, immigrants to Canada (where the population was a little more than 10 million) exceeded 5 million. At that time, like the United States, the Canadian government was taking hostile measures against immigration from south and east Europe and east Asia, which at the same time as the great crisis of 1929, led to practical closure of borders. Between 1931 and 1935, there were only 72,000 entries, and between 1936 and 1940, the departures exceeded the arrivals.

At the beginning of the 1930s, Brazil and New Zealand established immigration quotas, decreeing laws inspired by those of the United States that clearly discriminated against undesirable populations. The huge masses that had formerly filled the embarkation and disembarkation quays vanished from both sides of the Atlantic. Ellis Island, which was headquarters for receiving immigrants in the port of New York and where the poor third-class passengers had to complete all the necessary formalities to obtain permission to disembark, was almost completely deserted.

At the same time, the West Coast witnessed the departure of millions of Chinese, undesirable immigrants wherever they went. Most of them ended up on the Asian continent itself, few of them finding refuge elsewhere: 5 million people went to the mining regions of Manchuria, almost the same number to Formosa, and about 10 million to Thailand, Indochina and Malaysia, compared with only 300,000 to Brazil, Cuba, Peru, and the United States.

III. FORCED MIGRATIONS FROM THE END OF THE FIRST WORLD WAR TO THE 1950s

The end of the First World War was marked by an exceptional migratory process. In Europe, revolutions, the war, and a series of border changes caused significant transfers of populations from one country to another. About 2 million Russians, Ukrainians, and Byelorussians left the Soviet Union; more than 1.5

TABLE 58–5 Emigration outside Europe from 1901 to 1915, according to Destination

Country of origin	Country of destination[a]						
	Total	United States	Canada	Argentina	Brazil	Australia and New Zealand	Other countries
Norway and Sweden	624	560	48	1	2	13	2
Great Britain and Ireland	3620	1254	1168	22	—	1139	38
German Empire	643	479	39	35	39	27	25
Austria-Hungary	3330	3035	200	56	27	10	2
Italy	4774	2985	119	1099	323	16	235
Russia	2794	2491	97	140	52	8	5
Other European countries	5672	1129	238	1234	748	54	2270
Total	21458	11932	1909	2586	1191	1267	2577
Proportions (%)	100	56	9	12	5	6	12

[a] Figures for emigration are given in thousands.

million Greeks left Turkey; important flows of Magyars left Rumania, Yugoslavia, and Czechoslovakia; and there was a forced exchange of about 2.5 million people between Poland and Germany.

During the 1930s and 1940s, the transoceanic borders remained closed to the most economically disadvantaged European populations, but internal migrations on the continent did not cease. During these 2 decades, France, Switzerland, and Germany were the preferred destinations of 2 million Italians. This is an impressive number, considering that the fascist laws of 1928 and 1931 were intended to limit, if not prohibit, departures from Italy. After agreements between Mussolini and Hitler for the placement of permanent and seasonal labor, approximately 800,000 Italian workers had emigrated to Germany, whereas 250,000 had left Italy to escape the fascist persecutions.

In the years preceding the Second World War, between forced migration, deportations to Siberia (about 5 million), and fugitives from dictatorial regimes, about 15 million Europeans left their countries. This is nothing compared with what happened in the following decade. During the Second World War, 30 million people were deported or exterminated, and in the following years, 35 million others (i.e., East Germans, Baltics, Poles, and Soviets) were transferred out of their countries, whether they had been expelled or were compelled to leave for the Western countries (Paradisi, 1979). Ethnic zoning in Eastern Europe in itself resulted in the displacement of more than 15 million people (Table 58–6).

Europe came out of the conflict impoverished, burdened with millions of refugees, and struck by unemployment. The situation was dramatic in the countries of the Soviet bloc and in the Mediterranean countries, where population problems as acute as those that existed before the war made it even more difficult to recover from the crisis. Emigration again seemed to be the solution. In May 1950, the foreign ministers of the three main Western powers (i.e., Dean Acheson, Ernest Bevin, and Robert Schuman) adopted a resolution in which they declared, “The overpopulation suffered by a number of European states is one of the most important factors in the world’s difficulties and imbalances. A systematic study of the possibilities of emigration could greatly contribute in solving such a problem.” A commission of experts appointed by the Council of Europe in 1950 estimated the surplus population at 5 million and advised that emigration from Europe was the only way to resolve the problem (Paradisi, 1979).

Interest in migration by political leaders favored the reopening the borders of the countries across the Atlantic to European immigrants. Without changing the quota laws, the United States authorized family reunification and lifted all limits on the entry of political exiles. Immigration to the United States therefore increased from 1 million in the 1940–1950 period to 3 million in the 1950–1960 period. For its part, Canada allowed 800,000 European immigrants to enter between 1951 and 1955 (about one-half million entered immediately after the end of the War) by permitting the entry of exiles and technicians capable of integrating the rapid industrialization process started after the war. Even Australia accepted an increased number of immigrants (more than a million between 1947 and 1954), striving to integrate them as soon as they arrived. To this end, the immigrants were chosen according to their origins, age, and physical condition: 48% were British, and the remaining 52% were Polish, German, Italian, Yugoslav, Russian, or from other countries (Reinhard *et al.*, 1968).

TABLE 58–6 Population Movements within Europe from 1945 to 1950, according to Country of Departure and Country of Destination

Country of departure	Country of destination[a]							
	FRG-GDR	Poland	Czechoslovakia	Hungary	USSR	Finland	Italy	Total
Poland	7000	—	—	—	518	—	—	7518
Czechoslovakia	3200	—	—	—	50	—	—	3250
Hungary	225	—	73	—	—	—	—	298
USSR	15000	1496	42	—	—	400	—	3438
Yugoslavia	360	—	—	—	—	—	200	560
Romania[b]	—	—	—	315	—	—	—	315
Total	12285	1496	115	312	568	400	200	15379

[a] Figures for population movements are given in thousands.
[b] Romania, Slovakia, and Yugoslavia.
Data from Tonizzi, 1999.

The period that followed the Second World War was significant by the birth of the new Israeli state, toward which headed hundreds of thousands of Jews who had survived the Nazi camps or who had been exiled all over Europe, as well as in the Muslim countries of North Africa. In December 1955, the inhabitants of Jewish origin numbered about 1.6 million, 88.5% of the total.

At the end of the 1950s and at the beginning of the 1960s, European emigration slowed down, and a new chapter in international migration began.

IV. THE GLOBALIZATION PROCESS AND INTERNATIONAL MIGRATION

The expansion of the global economy and the development of communication methods have contributed to changing the mobility conditions of populations between different regions of the world. From the 1950s, the nature of migration changed progressively, acquiring, like the flow of goods and capital, a global character, as much from the point of view of volume as from that of space and structure. After 1945 to 1950, during which some countries (including France and Great Britain) had to resort to immigration for their postwar reconstruction, several phases in the evolution of international migration can be identified over the course of the last 50 years (Salt, 1989). The first, which occurred approximately from 1950 to 1973, coincides with the period when the main economic strategies of large transfers of capital were decided by the richest countries of Europe and North America (and by Argentina, Venezuela and Australia). It was toward these countries that hordes of workers from the least-developed countries of the south and west coasts of Mediterranean Europe headed (Fassmann and Münz, 1994; Castles and Miller, 1993; King, 1993). At the same time, France and Great Britain opened their borders to immigrants from their former colonies, and in America, significant south-to-north migration was developing (Fig. 58–3). Access to the wealth of oil-producing Arab countries created in certain Gulf states an attraction for North African and Asian populations, encouraging the development of a new immigration destination.

The 1973 oil crisis radically changed the picture. It initiated a period of recession that led the developed countries to modify their economic strategies. To overcome the crisis, they were obliged to restructure their production systems and modify their working methods, to such an extent that the global economy was completely transformed. This caused a profound change in migration policies and consequently in the nature, volume, and direction of the flows (Fig. 58–4). In this way, a new phase of internal and international migration began, which saw some southern European countries, traditionally exporters of emigrants, become receiving countries. From the 1970s onward, these effects accentuated a process already underway in the poorest countries of the world, with millions of people leaving the poor rural areas to settle in towns, opening the way to a migratory chain, which by increasing over time, resulted in a decisive change in the intensity and direction of international migration flows. The arrival of large masses of population in towns still devoid of economic, social, and sanitary structures caused a dramatic crisis, even at the level of social values, forcing

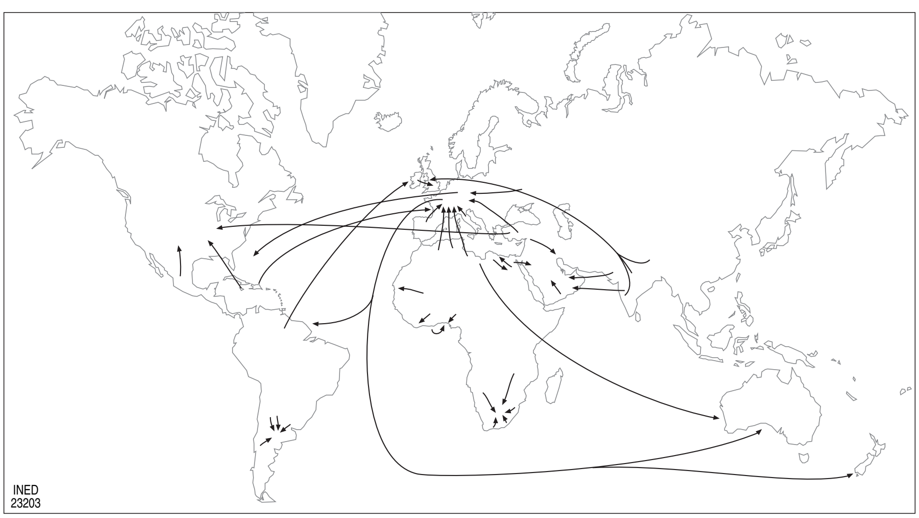


FIGURE 58–3 Direction of global migratory flows from 1945 to 1973. The dimensions of the *arrows* do not indicate the intensity of the movements. (From Castles and Miller, 1993.)

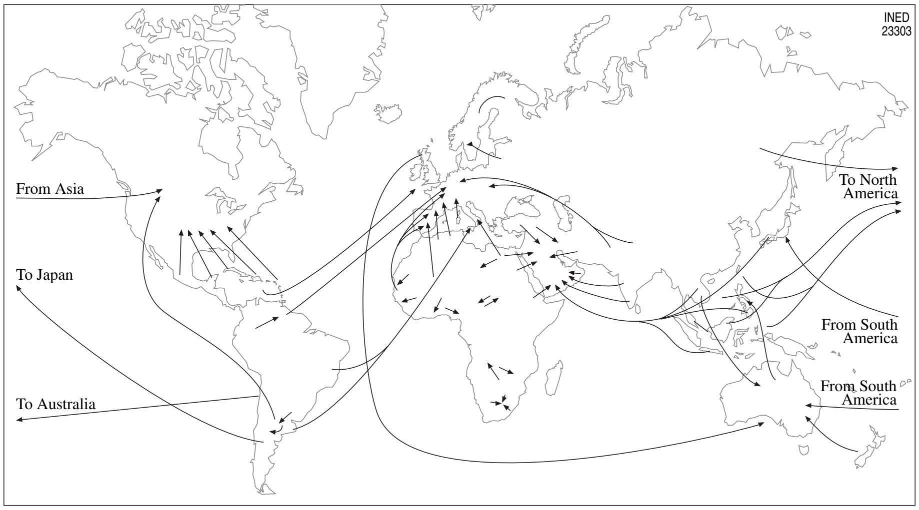


FIGURE 58–4 Direction of global migratory flows since 1973. The dimensions of the *arrows* do not indicate the intensity of the movements. (From Castles and Miller, 1993.)

a growing number of people to look for a better life in countries distant but better developed economically (whether traditional immigration countries or countries considered up to that time as countries of emigration, but sometimes of easier access).

From the second half of the 1980s and particularly at the beginning of the 1990s, international migration underwent new changes. While maintaining the same direction as in the preceding years, migratory flows became more associated with the political and economic globalization process that began at the end of the Cold War. The political and economic events of 1989 to 1996 that overturned the old political order and its economic relations opened the way for new migratory flows (Gould and Findlay, 1994). In effect, the political collapse of the Soviet Union and the creation of new independent states that significantly interrupted the internal trade exchanges in the Soviet bloc and the resulting economic crisis affecting all the countries in the region propelled the populations to look for work in the richest countries or the nearest countries from an ethnic and geographic point of view (Kupiszewski, 1994).

This crisis also affected nonmember countries (e.g., Yugoslavia, Somalia, Eritrea) that had privileged business relations with the Soviet bloc. In some of these countries, the wish to emigrate was increased by internal ethnic, religious, or political conflicts that obliged many minorities to leave their countries to seek political asylum in the Western countries. Often, a significant number of refugees, sometimes difficult to distinguish from economic migrants, were added to the normal flow of labor migrants (the volume of which is sometimes difficult to estimate). The case in recent years of migrants from Albania and the former Yugoslavia is emblematic.

Generally, these problems are also those of migrants from several countries in the Third World, where political and social conflicts in the context of a demographic explosion exacerbate the disastrous effects of the collapse of an economy that never really got off the ground. Today's events suggest that a new stage in migration history is beginning. The gap between rich and poor countries and between north and south is widening. While Eastern Europe's economy is experiencing difficulty emulating that of the rest of Europe, the economic barrier that separates the United States and Mexico and the one across the Mediterranean between Europe and North Africa continue to create enormous migratory pressure on the world's richest regions, sometimes contributing to the conversion of traditional emigration zones into immigration zones, as is the case in southern Italy. The old social and economic problems are exacerbated, and new ones appear. The emergence of new economic rivalries sometimes favors the adoption of protectionist measures, justified ethically by the necessity for self-protection.

Nevertheless, the migratory flows resulting from the globalization process are minimal compared with those produced by regional conditions. If it is true that the economy is becoming more and more international, it is no less true that a geographic polarization of investments is occurring, by virtue of which migration between neighboring countries and countries that have a shared history or culture is much more widespread. The migratory processes derive part of their momentum from political and economic treaties established between countries in the same geographic zone. This was one of the functions of the European Union, just as for other economic treaties such as the North American Free Trade Agreement (NAFTA), which was negotiated in the 1990s among the United States, Canada, and Mexico (Werner, 1994).

Among the new characteristics of international migration is its feminization, not only as a result of the policies developed in the last decades in favor of family reunification, but also because of the evolution in the social roles of both sexes and the *status* of women (Zlotnik, 1995). Formerly, the number of female immigrants was low (see Fig. 58–1), and they were mostly women following their husbands, but today, the decision to emigrate is their own, and they migrate to find work in countries where female labor is needed or to obtain permission to enter as refugees. In Italy, for example, asylum seekers from the former Yugoslavia are mainly women.

TABLE 58–7 Estimate of Net Migration between 1950 and 1989 in Europe and North America

Regions	1950–1959[a]	1960–1969	1970–1979	1980–1989	1950–1989
Western Europe	–545	1150	2995	4275	7875
Center[b]	3805	4520	2815	3470	14610
North[c]	–1025	–195	–25	220	–1025
South[d]	–3325	–3175	205	585	–5710
Canada and United States	4885	4855	9445	5790	24975

[a] All figures for net migration are given in thousands.
[b] Austria, Belgium, France, Federal Germany, Liechtenstein, Luxembourg, Monaco, Netherlands, and Switzerland.
[c] Iceland, Denmark, Finland, Ireland, Norway, and United Kingdom.
[d] Andorra, Gibraltar, Greece, Italy, Malta, Portugal, Spain, and the former Yugoslavia.
From Macura, 1994.

V. ACCELERATION OF MIGRATORY MOVEMENTS IN EUROPE AND NORTH AMERICA

The availability of data is the preliminary condition for any analysis or consideration regarding the political and social actions that affect migration, but collecting data on the volume and characteristics of the phenomenon causes serious problems. Even in the countries where a migrant registration system is in place, migrants from one zone to another do not always declare their transfer; the same is true for those going back to their country of origin, who do not always declare their return. The provisions in this area vary from one country to another, and this often makes comparative analyses impossible. The main problem is that the flows toward receiving countries often occur illegally, so that the registered migration is only part of the real migration. The literature on this aspect is abundant (Salt, 1998; CMS/IUSSP, 1987), and it does not seem helpful to review the regulations of each country nor to refer to all the definitions of migration here (see Chapter 13). We must content ourselves with an evaluation of the migratory movements based on existing information, national and international, concerning recorded migration and the indirect estimates of illegal migration and entries by requests for asylum.

Before going into detailed analysis, it may be useful to define the background of population movements since the Second World War for the geographic zones discussed here. From 1950 to 1989, Western Europe had a migratory net balance of approximately 8 million people, and North America had almost 25 million (Table 58–7). Almost two-thirds of this balance was acquired after 1960. In both cases, the speed is obvious, even though they occur differently: in North America, after doubling in the 1970s compared with the 1960s, the balance was reduced by one-half in the 1980s compared with the 1970s, whereas among the European countries as a whole, the balance has been doubling roughly every decade since the 1960s.

The European situation is, however, far from homogeneous. We know that the southern countries had a negative net migration total until the 1970s as a result of more than 3 million departures per decade. In Italy, for example, more than 5 million people departed over a period of 20 years from 1950 to 1970, and more than 3 million returned home; the final net migration was 2.5 million (Rosoli, 1978). The migratory balance became positive in southern Europe in the 2 decades that followed.

Western Europe had a positive migratory balance of more than 14.5 million people between 1950 and 1989. The Federal Republic of Germany and France were the main beneficiaries, with net immigration rates, particularly for the Federal Republic of Germany, of 4% annually on average, while natural growth was only 2% (Macura, 1994).

At the same time in North America, while an increase in the population had always been guaranteed by natural growth, the contribution from migration was 4% in Canada and 3% in the United States, supplying one-fourth of the population growth of both countries.

The most recent data, even if incomplete or fragmentary and not always comparable to preceding data, show a new growth in migratory movements for the first years of the 1990s. In Canada between 1990 and 1996, 231,000 people entered on average each year, compared with 138,000 between 1985 and 1989. In the United States each year, an average of more than 1 million foreigners legally entered the country, compared with 606,000 between 1985 and 1989 (UN/ECE, 1994). European countries were not left out; between 1990 and 1996, it is estimated that more than 2 million net annual immigration entries occurred in Western Europe, compared with 500,000 per annum in the 5 preceding years. The average annual estimate for Germany is about 900,000 between 1990 and 1996, compared with 332,000 per annum between 1985 and 1989. In the same period, net immigration for Italy was estimated at about 140,000 per year, compared with 66,000 per annum in the 5 preceding years.

Migration evolved differently in Eastern Europe (i.e., the former Soviet Union and the central and eastern European countries). After the Second World War, migration developed differently from that of the other European countries (Table 58–8). The Cold War period strengthened the Soviet regimes and contributed to maintaining a positive migratory balance in favor of the USSR until the end of the 1960s, despite the departures of members of certain specific ethnic or religious groups. In the 2 decades that followed, the entries were not sufficient to counterbalance the departures. Almost 700,000 people left the Soviet Union between 1950 and 1989, notably 370,000 Jews between 1973 and 1979, and a significant number of people followed during Gorbachev's presidency between 1986 and 1987 (Münz, 1995). The other countries in the region belonging to the Soviet bloc had a permanently negative balance during the whole of the Cold War period, on the order of 6 million during the first 2 decades (1950 to 1969) and 3.2 million in the following 2 decades (1970 to 1989), of which a significant part crossed from East to West Germany.

A veritable exodus accompanied the fall of Communism and the economic crisis that affected all the countries in the Soviet influenced zone. In only 4 years (1990 to 1993) the whole of Eastern Europe lost more than 3.5 million inhabitants through emigration, of which 1.4 million left from the former Soviet Union and 2.3 from the rest of the region. In the next 4 years (1994 to 1997), Russia alone lost almost 1.2 million people (Salt, 1998).

On the threshold of the 21st century, there are two contrasting Europes: all the Western countries that receive immigrants and all the Eastern countries that are experiencing a growing exodus of their population, who head mainly for the rich countries of Western Europe.

VI. IMMIGRANTS IN EUROPEAN COUNTRIES AT THE END OF THE TWENTIETH CENTURY

If the scale of the two components of the migratory balance (i.e., immigration and emigration) is often unknown, we do know that it depends mainly on three types of flow that are difficult to separate: migrants seeking work, political asylum seekers, and family reconciliation.

In recent decades, the issue of refugees has been added to the problem of the continual increase in

TABLE 58–8 Evolution of Migratory Balances in Europe, from 1950 to 1993

Regions	1950–1959[a]	1960–1969	1970–1979	1980–1989	1990–1993	1950–1993
Eastern Europe	–3960	–1820	–1430	–2710	–3660	–13610
Center-East[b]	–4000	–1920	–1090	–2280	–2310	–11600
USSR/CIS[c]	40	100	–370	–430	–1350	–2010
Scandinavian countries[d]	–50	70	200	240	220	680
Southern Europe[e]	–2910	–3080	620	1620	670	–3080
Northern Europe[f]	4,250	5,090	2,520	2,470	3,860	18,190
Total	–2670	260	1,880	1,620	1,090	2,180

[a] All figures for migratory balances are given in thousands.
[b] Former socialist countries, including Albania and the former Yugoslavia and, until 1991, the GDR.
[c] Including the Baltic countries.
[d] Denmark, Norway, Sweden, and Finland.
[e] Greece, Italy, Portugal, and Spain.
[f] Other European countries and, since 1991, reunified Germany.
From Macura, 1994.

numbers of workers from the Third World. For the receiving countries, it is essential to know the volume of these inflows to construct legislative measures that are adapted to the new characteristics of the phenomenon. To obtain the required statistics, it is necessary to regulate the different flows while taking several factors into consideration, such as the needs of the national labor market, the integration of the immigrants, the ability to cope with the costs engendered by increasing demands for service from the foreign population, and the means to prevent immigration from becoming a catalyst for social conflicts and confrontations (Golini and Bonifazi, 1987). Despite the countless registration difficulties, official bodies and individual researchers on a national and international level have produced numerous and detailed estimates of migration by type of flow, destination, and migrants' characteristics.

Before examining estimates of flows, a few words should be said about the volume of foreign presence in the territory, as seen through official available data that, although only partially representative of real migration, are an interesting reference point on the evolution of the phenomenon. These data do not include children born in the receiving country to foreign parents, naturalized refugees, citizens of former colonies who have obtained citizenship of the receiving country, nor illegal immigrants, whose numbers are relatively high in the more recent receiving countries of Mediterranean Europe. Table 58–9 shows that the ratio of the foreign population to the native population has practically doubled every 20 years since the end of the Second World War in the whole of Western Europe, from 1.3% to 2.3% between 1950 and 1970 and from 2.3% to 4.5% between 1970 and 1990, with more sustained growth rates in recent years. Today, more than 19 million foreigners live in these countries, compared with 5 million in 1950 and 11 million in 1970.

In 1998, the western European countries with the highest proportion of foreigners were Luxembourg and Liechtenstein (34.9% and 34.3%, respectively), even though the numbers are very low (148,000 and 11,000 people, respectively), especially comparison with Germany (more than 7 million) or France (more than 3.5 million in 1990). Germany holds the European record for the total number of foreigners reached between 1960 and 1970, with the entry of refugees from

TABLE 58–9 Foreign Populations in the European Countries

Country	1950		1970–1971		1982		1990		1998	
	Number[a]	%	Number	%	Number	%	Number	%	Number	%
Germany	568	1.1	2,976	4.9	4,667	7.6	5,338	8.2	7,366	9.0
France	1,765	4.1	2,621	5.3	3,660	6.8	3,534	6.4	3,263[b]	5.6[b]
Great Britain	NA	NA	NA	NA	2,137	3.9	1,904	3.4	212[c]	3.6[c]
Switzerland	285	6.1	1,080	17.2	926	14.7	1,127	16.7	1375	19.4
Belgium	368	4.3	696	7.2	886	9.0	903	9.1	903	8.9
Italy	47	0.1	NA	NA	312	0.5	580	1.1	992	1.7
Netherlands	104	1.1	255	2.0	547	3.9	692	4.6	678	4.3
Austria	323	4.7	212	2.8	303	4.0	482	6.2	734	9.1
Spain	93	0.3	148	0.4	183	0.5	407	1.0	610	1.5
Sweden	124	1.8	411	5.1	406	4.9	484	5.6	522	5.9
Denmark	NA	NA	NA	NA	102	2.0	161	3.1	250	4.7
Norway	16	0.5	76	2.0	91	2.2	143	3.4	158	3.6
Luxembourg	29	9.9	63	18.4	96	26.4	109	28.0	148	34.9
Portugal	21	0.3	32	0.4	64	0.6	108	1.0	175	1.8
Ireland	NA	NA	137	4.6	232	6.7	80	2.3	111	3.0
Greece	31	0.4	15	0.2	60	0.7	173	1.7	162	1.5
Finland	11	0.3	6	0.1	13	0.3	26	0.5	81	1.6
Liechtenstein	3	19.6	7	36.0	9	36.1	10	38.5	11	34.3
Total	5,100[d]	1.3[d]	10,900[d]	2.3[d]	14,700	3.1	18,600	4.5	19930	4.6

NA, not available.
[a] All figures for foreign populations are given in thousands.
[b] According to the 1999 census.
[c] Data from 1997.
[d] Including missing data obtained by interpolation.
Data from Frassmann and Münz, 1994; Ferruzza *et al.*, 1997; Natale and Strozza, 1997; Council of Europe, 1999.

East Germany. This position was reinforced at the beginning of the 1990s with the arrival of migrants from the Soviet bloc. From 1970 to 1998, the number of foreign residents rose from 3 million to more than 7 million. In France and the United Kingdom, the number of foreigners is lower because it does not include the arrivals from former colonies that came in large numbers to their parent state in the first decades after the Second World War, at the end of the colonial era of the 20th century.

Closer reading of Table 58–9 shows the variations in the trends of the phenomenon. Decelerations and reversals in trend appear clearly during the 1970s. Numerous developed countries reacted to the economic recession that followed the 1973 oil crisis by adopting measures to restrict the entry of immigrants. However, in countries such as France, the decrease in the number of foreigners was above all the result of the increase in naturalizations. In Switzerland and Sweden, a decrease in foreign presence was observed. In Switzerland, the proportion of foreigners fell from 17.2% between 1970 and 1971 to 14.7% in 1982; in Sweden, it fell from 5.1% to 4.9%. In France and Great Britain, the same reversal in trend occurred in the following decade; between 1980 and 1990, the proportion of foreigners fell from 6.8% to 6.4% in France and from 3.9% to 3.4% in Great Britain. This shift, in France at least, was a result of the introduction of policies promoting family reunification that allowed entries, even during a period of economic crisis, and beyond the response to immigration needs (Fig. 58–5). These policies were beneficial from a demographic point of view; by increasing the number of women and children, the age and sex structure of the immigrant population became more homogeneous with that of the native population.

From the 1970s onward, the southern countries of Europe, traditional emigration zones, became immigration zones. In Italy particularly, the combination of the effects of the border closures of northern European countries and rapid economic growth resulted in the return of many emigrants (i.e., returns quickly exceeded departures) and the beginning of immigration from the poor Mediterranean countries, sub-Saharan Africa, and Asia and, from the second half of the 1980s onward, from the countries of the former Soviet bloc.

Tables 58–10 and 58–11 give an estimate of immigrant entries, and political asylum seekers in some European countries between 1980 and 1996. Despite the paucity of information on illegal migration and the diversity of data collection methods used in different countries, which limit comparison, these estimates give an idea of geographic differences resulting as much from each country's legislation as from the political and economic events that affected the whole continent. In the 1980 to 1985 period, in almost every country with traditionally high migration, a decrease in entries was observed, as mentioned previously in the cases of France and Great Britain, followed by a resumption of previous levels with the increase in requests for political asylum. At the end of the 1980s and at the beginning of the 1990s, for France and Sweden, the increase in migration seems entirely attributable to the entry of asylum seekers (Fig. 58–6).

For certain countries, this demographic profile is not surprising, considering that between 1980 and

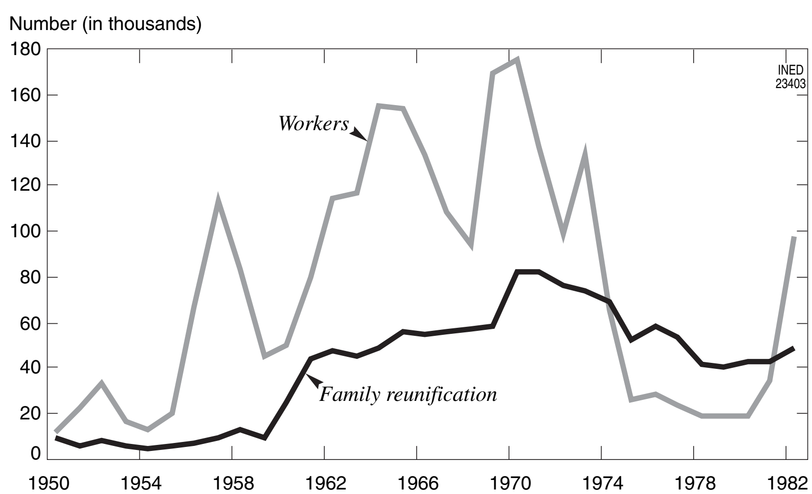


FIGURE 58–5 Annual trend in the number of immigrants in France from 1950 to 1982 according to the two main motives for entry.

TABLE 58–10 Flow of Immigrants in Some European Countries between 1980 and 1996

Year	Belgium[a]	Denmark	Finland	France[b]	Germany[c]	Ireland	Italy[d]	Luxembourg	Netherlands	Norway	United Kingdom	Spain	Sweden	Switzerland
1980	46.8	14.5	1.9	59.4	632.3	—	88.3	7.4	79.8	11.8	107.0	3.0	34.4	70.5
1981	41.3	14.5	2.3	75.0	502.0	—	91.5	6.9	50.4	13.1	93.0	11.3	27.4	80.3
1982	36.2	15.3	2.4	144.4	322.4	—	100.1	6.4	40.9	14.0	104.0	2.8	25.1	74.7
1983	34.3	15.0	2.8	64.2	276.4	—	98.3	6.2	36.4	13.1	108.0	3.6	22.3	58.3
1984	37.2	17.9	2.7	51.4	333.3	—	86.9	6.0	37.3	12.8	106.0	4.4	26.1	58.6
1985	37.5	24.6	2.6	43.4	400.0	—	82.2	6.6	46.2	15.0	122.0	6.2	27.9	59.4
1986	39.3	26.6	2.7	38.3	479.5	—	75.7	7.4	52.8	16.8	130.0	4.3	34.0	66.8
1987	40.1	23.8	2.8	39.0	473.3	17.2	104.5	7.2	60.9	23.8	113.0	5.3	37.1	71.5
1988	38.2	22.2	3.2	44.0	648.6	19.2	85.8	8.2	58.3	23.2	127.0	9.7	44.5	76.1
1989	43.5	24.4	4.2	53.2	770.8	26.7	81.2	8.4	65.4	18.5	146.0	14.4	58.9	80.4
1990	50.5	26.2	6.5	63.1	842.4	33.3	96.7	9.3	81.3	15.7	161.0	13.7	53.2	101.4
1991	54.1	29.2	13.2	65.3	920.5	33.3	70.9	10.0	84.3	16.1	150.0	10.6	43.9	109.8
1992	55.1	29.1	10.4	87.7	1208.0	40.9	72.3	10.7	83.0	17.2	116.4	18.2	39.5	112.1
1993	53.0	28.2	10.9	116.0	989.8	35.0	51.1	10.1	87.6	22.3	120.0	15.4	54.8	104.0
1994	56.0	28.9	7.6	82.8	774.0	31.5	52.3	10.1	68.4	17.9	133.0	–18.6	74.8	91.7
1995	53.1	45.9	7.3	88.0	792.7	31.2	68.2	10.3	67.0	16.5	154.0	–19.5	36.1	87.9
1996	51.9	31.5	7.5	74.0	777.5	21.5	—	10.0	77.0	17.2	160.0	–16.7	29.3	74.3

[a] All figures for flow of immigrants are given in thousands and do not include asylum seekers.
[b] Immigrants considered as permanent according to the definition of the 1971 Immigration Act.
[c] Until 1990, the data relate only to Federal Republic of Germany, and afterwards to Reunified Germany.
[d] Data taken from entries and removals from the register.
Data from UN/ECE, 1996; Salt, 1999; ISTAT, 1999.

TABLE 58–11 Entry of Foreigners followed by a Request for Asylum in some European Countries between 1980 and 1997

Year	Austria[a]	Belgium	Denmark	Finland	France[b]	Germany	Greece	Italy	Norway	Netherlands	Portugal	United Kingdom[b]	Spain[c]	Sweden	Switzerland
1980	9.3	2.7	0.2	—	18.8	107.8	—	—	0.1	1.3	1.6	9.9	—	—	6.1
1981	34.6	2.4	0.3	—	19.8	49.4	—	—	0.1	0.8	0.6	2.9	—	—	5.2
1982	6.3	3.1	0.3	—	22.5	37.2	—	—	0.1	1.2	0.4	4.2	—	—	4.2
1983	5.9	2.9	0.3	—	22.3	19.7	0.5	3.1	0.2	2.0	0.6	4.3	1.4	4.0	7.9
1984	7.2	3.7	4.3	—	21.6	35.3	0.8	4.6	0.3	2.6	0.2	2.2	1.1	12.0	7.4
1985	6.7	5.3	8.7	—	28.8	73.8	1.4	5.4	0.8	5.6	0.1	6.2	2.3	14.5	9.7
1986	8.6	7.6	9.3	0.1	26.2	99.7	4.3	6.5	2.7	5.9	0.1	5.7	2.8	14.6	8.5
1987	11.4	6.0	2.7	0.1	27.6	57.4	6.3	11.0	8.6	13.5	0.2	5.9	3.7	18.1	10.9
1988	15.8	4.5	4.7	0.1	34.3	103.1	9.3	1.4	6.6	7.5	0.3	5.7	4.5	19.6	16.7
1989	21.9	8.1	4.6	0.2	61.4	121.3	6.5	2.3	4.4	13.9	0.1	16.8	4.1	30.0	24.4
1990	22.8	13.0	5.3	2.7	54.8	193.1	4.1	4.7	4.0	21.2	0.1	38.2	8.6	29.4	35.8
1991	27.3	15.4	4.6	2.1	47.4	256.1	2.7	31.7	4.6	21.6	0.2	73.4	8.1	27.4	41.6
1992	16.2	17.3	13.9	3.6	28.9	438.2	2.0	2.6	5.2	20.3	0.6	32.3	11.7	84.0	18.0
1993	4.7	26.9	14.4	2.0	27.6	322.6	0.8	1.6	12.9	35.4	2.1	28.0	12.6	37.6	24.7
1994	5.1	14.3	6.7	0.8	26.0	127.2	1.3	1.8	3.4	52.6	0.8	42.2	12.0	18.6	16.1
1995	5.9	11.4	5.1	0.8	20.2	127.9	1.4	1.7	1.5	29.3	0.5	55.0	5.7	9.0	17.0
1996	7.0	12.4	5.9	0.7	17.2	116.4	1.6	0.6	1.8	22.9	0.3	37.0	4.7	5.8	18.0
1997[d]	6.7	11.5	5.1	1.0	20.0	104.3	—	1.7	2.3	34.4	0.4	32.5	5.0	9.6	23.9

[a] All figures for entry of foreigners are given in thousands and do not include asylum seekers and constitute provisional data.
[b] Not including minors.
[c] Including minors.
[d] From January to September 1997 for Luxembourg and Spain.
Data from OECD, 1992, 1995, and 1998.

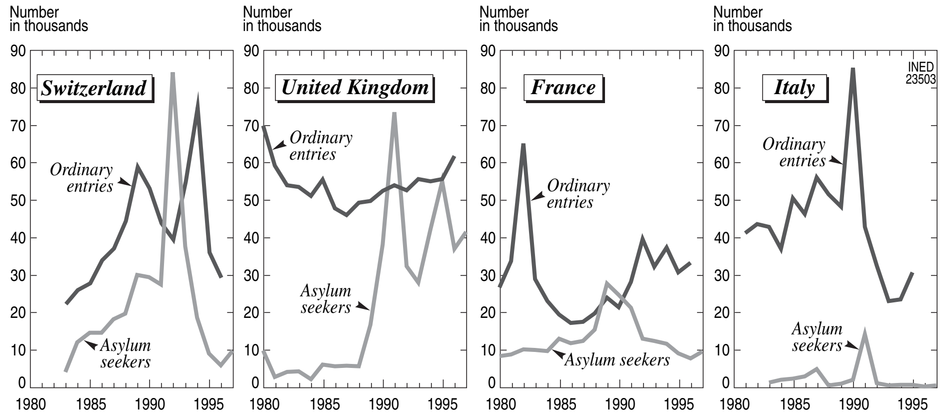


FIGURE 58–6 The trend of ordinary flows (not including asylum seekers and seasonal or border workers) and entries for asylum seekers in some European countries, with numbers given in thousands. (Data from ISTAT, 1999; Salt, 1998; OECD, 1992, 1995 1998.)

1993, more than 6 million people left Eastern Europe and the Soviet Union (net balance), of whom 3.6 million left between 1991 and 1993 alone, and that a very high proportion of them headed for Western Europe. Between 1985 and 1992, for Germany alone, the flow of labor migrants from the East was about 1.5 million, as well as 1 million asylum seekers (Fassmann and Münz, 1994). Table 58–12 completes this general view with a distribution of immigrants in 1992 by country of origin. In the first place, it confirms the previous statement regarding Germany, where, in that year immigrants from the former Communist countries made up more than one-half the total immigration figures, but the table also shows that it is different for the other countries that receive mostly nationals from Western European countries. Only Portugal, France, and Italy (for this country, see Natale and Strozza, 1997) are exceptions to this rule. In France, 44% of immigrants in 1992 came from Africa; in Italy, 53% came from African and Asian countries; and in Portugal, 52% came from America. Asylum seekers cannot be considered as permanent immigrants because most of them stay only for a short period in the receiving country because refugee status is granted to few of them.

The situation is different (at least concerning inflows from the countries of the former Communist bloc, Africa, and Asia) from what is seen from the data that do not include illegal immigration. The proportion of illegal immigration varies significantly from one country to another. In southern Europe, notably in Italy, illegal immigration is much higher than elsewhere, as numerous factors (e.g., border geography, proximity to countries of emigration, legislative tardiness, lack of adequate policies, weakness of control measures) have contributed in facilitating their entry. Numerous estimates confirm, for example, that in southern European countries, illegal flows from the poor Mediterranean and sub-Saharan countries are far higher than official flows. According to a Council of Europe estimate, in 1995, there were approximately 1.5 million illegal immigrants in Italy, Spain, Greece, and Portugal from non-European countries, and more than 0.5 million of these immigrants were from North Africa (Schoorl *et al.*, 1997). The same source also provides estimates of the trend in African migratory flows into several European countries, distinguishing between the two components, legal and illegal. In 1985, the African population in European countries, calculated using data from census data was about 2.3 million, and the estimate obtained by adding illegal immigrants was as high as 2.7 million. Only 9 years later, the gap between these two figures had more than doubled. In Europe in 1994, there were more than 3 million registered immigrants, but by adding illegal immigrants, the estimate was 4 million; the largest gap between registered immigrants (288,000) and the total number including illegal immigrants (935,000) was observed in Italy (Schoorl *et al.*, 1997).

In Italy, the issue of the number of illegal immigrants has caused lively debates because of the inability to agree on an estimate of the number of illegal immigrants. The complexity of the Italian administrative system makes it difficult, if not impossible, to correctly estimate this component of migration; the subdivisions of "regularized," "nonregularized," and "illegal" immigrants confuse rather than clarify the issue. Until the end of 1999, foreigners residing in Italy could be in order regarding their residence permit but not registered with the *anagrafe* (as is stipulated by

TABLE 58–12 Proportional Distribution of Immigrants in Some European Countries in 1992, according to Country of Origin

Country of origin	Receiving country										
	Belgium	Denmark	France[a]	Germany	Greece	Ireland	Luxembourg	Netherlands	Portugal	United Kingdom	Spain
Western Europe	59.9	65.3	22.9	28.4	67.5	88.7	75.3	45.6	12.3	59.1	64.6
Turkey	4.0	2.7	8.3	5.4	0.2	—	0.1	8.0	0.0	0.4	—
Former Yugoslavia	2.2	1.1	0.9	25.5	0.9	—	10.0	4.6	0.3	0.9	—
Eastern Europe[b]	1.8	3.9	3.0	26.7	12.7	—	2.8	3.1	29.1	1.8	—
America	7.3	4.8	6.7	2.2	4.0	4.1	4.7	11.6	51.7	8.3	—
Africa	14.9	4.1	43.8	4.6	6.2	—	1.8	14.2	2.7	4.1	—
Asia	7.3	13.8	13.0	5.5	5.2	—	3.7	11.4	0.2	18.0	—
Other	3.0	4.3	1.4	1.7	3.6	—	1.6	1.5	3.7	7.4	—
Total	100.0	100.0	100.0	100.0	100.0	—	100.0	100.0	100.0	100.0	—
Total immigrants[c]	72.4	43.0	116.6	1645.8	32.1	40.8	10.7	103.3	13.7	84.9	38.9

[a] The low proportion of Westerners in the flow of arrivals in France in 1992 results from the fact that they were only partially included in the statistics (Tribalat, 1993).
[b] Including the former Soviet Union, but excluding the former Yugoslavia.
[c] Including refugees. Figures of total immigrants are given in thousands.
From OECD, 1995.

law), or they could be registered with the *anagrafe* (and even registered on the list of job seekers) without having a residence permit and therefore considered nonregularized or illegal. According to the new legislation, foreigners cannot register with the *anagrafe* unless they have a residence permit, but the *anagrafe* services, which are supposed to check the validity of each permit every year, rarely make these checks, at least not in the large cities. The application of legislation also varies from one region to another. Fortunately, in recent years, the problems and incoherencies introduced by legal and administrative complications have not prevented researchers from producing several estimates that appear to tally on several important points (Strozza, 1994). According to the first estimate, the number of immigrants in Italy between 1997 and 1998 should be about 1.2 or 1.4 million, more than double the number estimated using data from the 1991 Census (ISMU, 1995; ISTAT, 1994; CENSIS, 1993a; Natale, 1990) but quite close to the number of foreigners registered with the *anagrafe* in 1998 (about 992,000 according to ISTAT, 1999) and close to the number of residence permits issued in the same year (a little more than a million) (Caritas, 1999).

In Italy, more than 80% of the immigrants identified live in the central and northern regions (51% in the northern regions alone). More specific estimates show that in the *Mezzogiorno* region, immigration is lower and has different characteristics, so much as to suggest a hypothesis of temporary migrations. The southern regions may constitute an easily accessible first stage for immigrants intending to move further (Ferruzza *et al*., 1997). The statistics for residence (residency) permits issued in 1997 confirm the trend of progressive growth in the foreign population in the north and the center of the country, with proportions of more than 50% (Table 58–13).

According to a recent estimate, more than 30% of the permits were issued in Milan and Rome (Bonifazi and Strozza, 2003; Bonifazi, 1998); 65% of the 649,000 residence permits at the end of 1991 were issued for the purpose of seeking employment, a trend at odds with that of most of the traditional European immigration countries (Fig. 58–7).

The characteristics of migration have changed in recent years in many European countries as a result of measures taken to limit the entry of labor migrants. In Italy, because of the incoherencies and delayed adaptation, the effects are not yet evident, but in the countries where policy measures have been in place for a certain length of time, the impact on the flows is significant. In France, the long-term trend has been to limit new entries, with the exception of family reunification (see Figure 58–7B). This example was partially emulated elsewhere, which is partly responsible for one of the most significant changes in the structure of migrant populations: its feminization. In all the northern European countries, women currently make up about 45% to 46% of the total immigrant population, even though the phenomenon varies quite significantly according to the country of origin. Although it is true that the emigration of women depends on the sociocultural context in which they live, including events and circumstances at the time (e.g., war, food crisis, political repression), emigration reflects even more so the situation in the receiving country, as much because of policies favoring family reunification as the opportunities offered by the employment market. Figure 58–8 shows, for example, the heterogeneity of female presence according to country of origin (for five nationalities) and according to the receiving country (for five countries plus Europe).

The number of women from Mauritius is greater than the number of men (values more than 1000) in

TABLE 58–13 Foreign Population (Resident and Nonresident) in Italy in 1991 and Percentage of Valid Residence Permits Issued in 1991, 1994, and 1997

Geographic distribution	Foreigners in 1991 census		Residence permits issued in					
			1991		1994		1997	
	Number[a]	%	Number	%	Number	%	Number	%
North–West	149.6	28.0	181.4	27.9	198.1	29.2	299.5	30.4
North–East	105.4	19.7	127.6	19.8	147.1	21.7	203.4	20.6
Center	169.6	31.7	220.9	34.0	224.5	33.1	311.7	31.6
Mezzogiorno	110.4	20.6	119.0	18.3	108.1	15.9	171.5	17.4
Total	535.0	100.0	648.9	100.0	677.8	100.0	986.0	100.0

[a] All population figures are given in thousands.
Data from Ferruzza *et al*., 1997; ISTAT, 1997.

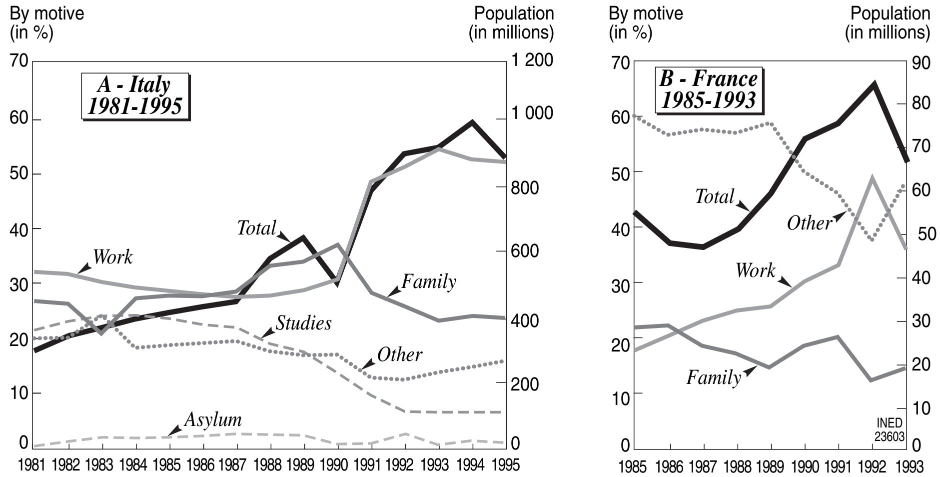


FIGURE 58–7 Distribution of foreign populations according to motive of entry and total foreign population. **A:** Italy, 1981 to 1991. **B:** France, 1985 to 1993. For Italy, the residence permit data are from the Ministry of the Interior, and the values for 1995 are from the August 31, 1994, inventory; from 1989, the work category includes those registered at registry offices. For Italy, the distribution concerns the movement of the "stock" of the foreign population, estimated from valid permits, and for France, it concerns the net "flow" evolution of foreign entries, estimated from permits issued annually. All values are expressed in thousands. (From Schoorl *et al*., 1997).

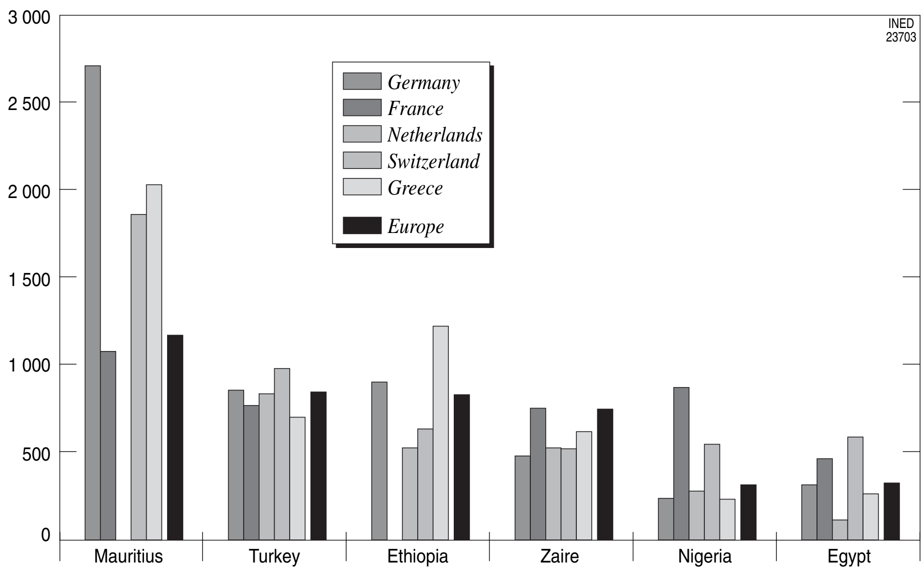


FIGURE 58–8 Sex ratio for each foreign population residing in different European countries in 1990 (number of women per 1000 men). The data are only for legal residents. (From Schoorl *et al*., 1997.)

each receiving country, and it is almost three times greater in Germany. This may indicate that they have autonomy from men regarding the migration project or that there are employment opportunities for women in some sectors in the receiving country. The emigration of Egyptian women is significantly lower than that of Egyptian men; the decision is generally made within the family and often includes women giving up their wish to emigrate.

VII. RECENT IMMIGRATION TO CANADA AND THE UNITED STATES

To estimate the number of foreigners in Canada and the United States, we can refer to the most recent censuses. In Canada, the number of immigrants was more than 4 million in the 1990 census, which is 16% of the total population. In 10 years, it had grown by approximately 400,000, but its proportional increase in the total population was only 0.7% (from 15.4% to 16.1%). The trend has been more distinct in the United States, where according to the most recent censuses, the number of immigrants increased from a little more than 14 million in 1980 to almost 20 million in 1990, which is an increase from 6.2% of the total population to 7.9%. This increase is even more surprising when compared with the 9 million immigrants recorded in the 1970 census; the immigrant population has doubled in a period of only 20 years (Table 58–14).

Detailed analysis of the annual flows reveals the migratory differences and similarities of both countries. In the United States in the 1970s, 450,000 entries were registered annually on average, close to values of the preceding years, and in the 1980s (Table 58–15, last row), the flow of entries was higher, with about 650,000 per year (Castles and Miller, 1993). In Canada, the 1970s was a period of higher immigration, with the entries, which in preceding years were about 100,000 annually, doubling before falling in the first half of the 1980s and stabilizing for several years at a lower level than that of previous years (see Table 58–15). These differences in trend are associated with different migratory histories, reflecting different political and administrative decisions (Papademetriou, 1994).

It is only in the second half of the 1980s that a real convergence between the two countries is observed, with resumption in the increase of migratory flows at the beginning of the 1990s, exceeding 200,000 per year in Canada and 1 million in the United States. This convergence is the consequence of policy measures adopted by both countries, such as those in favor of family reunification, the regularization of illegal immigrants, and the entry of certain categories of refugees. At least for the United States, a revival in the economy also contributed to a wider opening of the borders.

Measures for the regularization of illegal immigrants and the authorization of family reunification allowed more detailed information to be collected on the degree of the migratory movements and particularly on the composition of the foreign population.

TABLE 58–14 Immigrants Recorded in Various Censuses

Censuses	United States		Canada	
	Number[a]	%	Number	%
1970	9,619.3	4.7	—	—
1980 (to 1981)	14,079.9	6.2	3,908.0	15.4
1990 (to 1991)	19,767.3	7.9	4,342.9	16.1

[a] All population figures are given in thousands and percent of the total population.

From OCDE, 1995.

Figure 58–9 illustrates the effect of these measures on the increase in the number of immigrants, and Table 58–15 shows the role it played in the distribution of immigrants according to country of origin.

In Canada from 1980 to 1991, the flows from Africa and the Middle East increased from 6% to 18% of the total, whereas in recent years, they have decreased to 16%, with almost 30,000 entries per year. At the same time, the proportion of South Americans increased from 9% to 16%, only to decrease to 8% in more recent years. Despite a constant increase in the flows, the proportion of immigrants from Europe decreased between 1980 and 1996 (from 29% to 20%). The distribution of the flows also changed in the United States (particularly in the years after the regularizations of 1986 and 1990). From 1980 to 1991, the proportion of Mexicans increased from 11% to 51%, whereas that of Europeans, despite the increase in immigration, fell from 14% to 7%. The proportions of Europeans, Asians, and Latin Americans, which at the end of the 1970s were 18%, 44%, and 35%, respectively, changed to 11%, 38%, and 47% after the regularizations of 1986 (i.e., Immigration Reform and Control Act) and to 7%, 22%, and 67% after the regularizations of 1990 (i.e., Immigration Act).

The U.S. Congress did not have to wait long for the results expected from the new legislation of 1990. With the regularizations, the 1 million immigrants observed in 1989 grew to more than 2 million in 1991. The succession of legislative measures favoring immigration should have improved on the previous situation and regularized or eliminated illegal entries. This did not happen. On the contrary, it is estimated that the number of illegal immigrants is currently about 3.5 million (of which 80% to 90% are Mexicans) and that despite border control, 200,000 illegal immigrants enter annually. In the absence of new legislative measures, the number of legal immigrants has decreased, falling more or less to the levels observed in the 1980s (see Table 58–15). With regard to asylum seekers, at the end of the Second World War the United States favored the entry of numerous European immigrants, especially those from the Communist bloc countries. From 1946 to 1980, it is calculated that about 2 million refugees entered the United States to settle there permanently. A comparison with data from recent years is practically impossible, because new legislation in 1980 radically changed the definition of a refugee. However, it can be stated that entries by asylum seekers have made up for about one-sixth of the total flow each year from 1980 onward (see Fig. 58–9).

In 1990, the United States and Canada adopted migration policies setting a maximum number of immigrants for 1995 based on a significantly higher quota than that of preceding years. Canada basing

TABLE 58–15 Immigrants to the United States and Canada from 1980 to 1996, according to Geographic Region of Origin

Region of origin	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
Canada[a]																	
Africa and Middle East	9.1	10.3	9.9	7.6	8.6	9.0	12.3	19.8	22.3	31.0	38.2	41.6	41.6	36.5	29.4	32.8	36.1
Asia and Pacific	69.3	45.7	38.5	34.2	38.0	34.3	35.3	57.9	70.0	76.5	89.3	97.6	120.9	130.8	128.2	112.7	124.1
America (except United States)	13.6	15.8	17.2	15.7	13.8	15.5	21.6	28.9	22.4	25.5	28.4	36.9	37.9	33.9	21.4	20.4	18.4
United States	9.9	10.6	9.4	7.4	6.9	6.7	7.3	8.0	6.5	6.9	6.1	6.6	7.5	8.0	6.2	5.2	5.8
Europe	41.2	46.3	46.2	24.3	20.9	18.9	22.7	37.6	40.7	52.1	51.7	48.1	44.9	46.6	38.6	41.2	39.7
Total	143.1	128.7	121.1	89.2	88.2	84.4	99.2	152.2	161.9	192.0	213.7	230.8	252.8	255.8	223.8	212.3	224.1
United States[a,b]																	
Europe	72.1	66.7	69.2	58.9	64.1	63.0	62.6	61.2	64.8	82.9	112.4	135.2	145.4	158.3	160.9	128.2	147.6
Asia	236.1	264.3	313.3	277.7	256.3	264.7	268.4	257.7	264.5	312.1	338.6	358.5	357.0	358.0	292.6	267.9	307.8
Africa	14.0	15.0	14.3	15.1	15.5	17.1	17.5	17.7	18.9	25.2	35.9	36.2	27.1	27.8	26.7	42.5	52.9
Oceania	4.0	4.2	3.8	3.5	3.8	4.1	3.6	4.0	3.8	4.4	6.2	6.2	5.2	4.9	4.6	4.7	5.3
Canada	13.6	11.2	10.8	11.4	10.8	11.4	11.1	11.9	11.8	12.2	16.8	17.2	15.2	17.2	16.1	12.9	15.8
Mexico	56.7	101.3	56.1	59.1	57.6	61.1	66.5	72.4	95.0	405.2	679.1	946.2	213.8	126.6	111.4	89.9	163.6
Caribbean	73.3	73.3	67.4	73.3	74.3	83.3	101.7	102.9	112.4	88.9	115.1	264.8	—	—	—	—	—
Central America	21.0	24.5	23.6	24.6	24.1	26.3	28.3	29.3	30.7	101.0	146.2	79.2	—	—	—	—	—
South America	39.7	35.9	35.4	36.1	37.5	39.1	42.1	44.4	41.0	58.9	85.8	182.7	55.3	53.9	47.4	45.7	61.8
Others[c]	0.2	0.2	0.2	0.1				0.1	0.1	0.1	0.2	0.2	155.0	157.6	144.7	128.7	161.1
Total	530.6	596.7	594.1	599.9	543.9	570.1	601.8	601.6	643.0	1090.9	1536.3	2026.4	974.0	904.3	804.4	720.5	915.9

[a] All immigrant population figures are given in thousands.
[b] The high number of immigrants to the United States after 1990 is due to regularizations.
[c] From 1992, figures include Caribbean and Central America.
Data from OCDE, 1992,1995, and 1998.

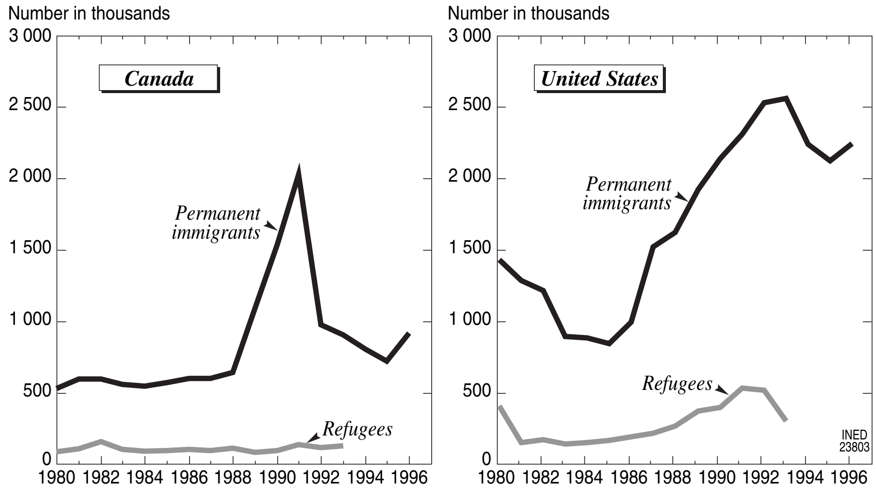


FIGURE 58–9 Evolution in the number of permanent immigrants (except refugees) and refugees in North America, in absolute figures. (From OECD, 1992, 1995, 1998.)

its quota on migration observed in 1990 fixed the maximum number at 250,000 immigrants per year, with the stipulation that labor migrants should not make up more than 50% of this number. Meanwhile, the United States set the average number of entries for the 1992 to 1994 period at 714,000, and the maximum number at 738,000 for 1995. In this same year, the proportion of labor migrants was restricted to 16% to 17% of the total in order that family reunification could constitute 70% of the total. It was possible to verify *a posteriori* whether the legal entries surpassed the limits set. In 1993, about 256,000 entries were recorded in Canada, whereas in the United States, 975,000 were recorded in 1992 and 904,000 in 1993. In the latter case, it appears that the number of entries respected the quota in 1995 but not in the following year, when the number of immigrants increased to 916,000.

VIII. INTEGRATION: THE END OF THE JOURNEY

The American experience calls into question the certitude of a guaranteed control of movement through the programming of entries. The historical reality and the current social, political, economic, and demographic situation show that the rich countries will continue to be the target for a growing number of emigrants. Sufficient economic growth in the Third World countries cannot be counted on to bridge the gap between them and the rich countries. Even at the level of international cooperation and despite efforts made in recent years, the innumerable interventions have not produced tangible results. Current economic agreements that allow the free flow of goods and capital are not sufficient to combat the demographic pressure in the least-developed countries, even if they can have an indirect effect in limiting the free flow of illegal workers. Moreover, if natural population growth cannot ensure a sufficient population of working-age people for development needs, there will be no choice but to resort to foreign labor in the future in many rich countries, notably in Europe. Numerous European countries that are today declaring their desire to close their borders may tomorrow be obliged to continue their economic development through immigration.

The objective of the governments regarding immigration on the other side of the Atlantic is different from that of the European countries. The United States is more concerned with blocking the entry of goods and capital than that of immigrants, contrary to European countries that, although pursuing different policies, converge around a basic position aimed at controlling the quantity and quality of immigration flows (Bonifazi and Gesano, 1994). The significant migratory pressure of African populations and the inexorable flood of people from the countries of the former Communist bloc have contributed to reinforcing the wish to restrict the entry of new immigrants. There are, however, different attitudes according to the type of flow and particularly about the necessity of at least favoring family reunification for foreigners already residing legally in the territory and the admissibility of asylum seekers.

Beyond their differences regarding migration policy, all the European governments are in agreement about the fact that conditions must be created for the complete economic, political, and social integration of immigrants who have been residing in the country for a long time (Salt, 1998; Tapinos, 1994). This is merely an agreement in principle, and the issue of the strategies for achieving this objective remains very controversial between different countries and within each of them. Total integration depends on equal opportunities for each individual within each society, without any cultural, religious, or racial discrimination. The process leading to integration is nevertheless extremely complex, and as Corrado Bonifazi and Giuseppe Gesano (1994) stated,

> Different ways of evolving the reference models exist on the basis of which the relation between the immigrant and the host society is organized and structured. It is a fundamental policy goal of the migratory process, but also a major problem which remains despite numerous efforts, as shown by the fact that the different approaches selected up to now in receiving countries, in Europe or elsewhere, have not fully achieved their objective, generally leave immigrants facing significant integration problems.[1]

In the failure of the integration models developed in the different countries, the main objective of legislation seems to be to favor the integration of legal immigrants, taking care to preserve their native culture and traditions in harmony with those of the receiving country (Bonifazi, 1998; Cagiano de Azevedo *et al.*, 1993, 1994). Among the fundamental points, the legislation is based (this is the case in Italy since 1986) on the recognition that immigrants have a right to be treated in the same manner, to have the same working conditions, to have the same access to social and health services, and to have the same schooling and accommodation as the native population (Coleman, 1993, 1994).

Even if measures are adopted to guarantee migrants at least satisfactory integration, if not complete immersion, we know that they vary greatly from one country to another, as much in their definition as in their application. Difficulties in intervention do exist, even in those areas where integration policies would seem to have broken all the barriers. Examples of discomfort are numerous. As in the past, immigrants are generally employed in the toughest employment sectors and usually have the most insecure jobs (Entzinger, 1990), without mentioning circumstances (as is underlined in the last OECD [1998] report on international migration) that resulted in unemployment rates increasing more rapidly for immigrants than for the native population in all the Western countries until 1996.

If employment facilitates the immigrant's integration into the host society, access to decent housing is an equally important factor. We know, however, that in all the countries of the world, housing for most immigrants is often very precarious and far from meeting the standards of a modern society. Unfortunately, this problem is not easy to resolve in countries where a large proportion of the native population is facing the same housing problems as the immigrants (CENSIS, 1993b).

Education is a fundamental element of integration. The first measures concern the immigrants' learning of the receiving country's language and setting up of special programs for their children. In all the receiving countries, the level of education of second-generation children is inferior to that of the local population. In countries that have not been receiving immigrants for a long period, the right to education for immigrants' children is rarely guaranteed, and even when it is, mediocrity is often the rule rather than the exception (Coleman, 1994). It may be thought that this poor education is the result of linguistic difficulties and the poor cultural context in which the children live. However, it depends greatly on factors such as the teachers' lack of preparation and the lack of multicultural education programs.

Recognition of citizenship (or of nationality) possibly constitutes the most fundamental stage in the integration process, because it is the conclusion of the migration process and is a major element in the acquisition of political rights, without which the immigrant's access to a full participation in the civic life of the receiving country is not possible. In Italy, for example, acquiring citizenship is a particularly complicated problem because the procedures in force are particularly restrictive and dissuasive, as much because of the duration of permanent residence required as the cumbersome nature of the administrative process, which culminates with a decree by the President of the Republic, on the suggestion of the Ministry of the Interior, after recommendation by the Council of State. Certain countries, even without naturalization, aim at encouraging the political integration of immigrants by giving them voting rights in local elections; this already appears in the provisions of the Maastricht Treaty for nationals of countries in the European Union, but it would be timely to offer this possibility to other nationals, at least for immigrants who have been resident in the receiving country for a long time, to allow them to exercise the democratic principle of participating in collective decisions and to protect their own interests.

The ideal conclusion of the migration process for recent generations of immigrants is still far away. The present climate of economic crisis and unemployment

[1] "Ha diversi esiti possibili al mutare dei modelli di riferimento in base ai quali viene organizzato e si struttura il rapporto tra immigrati e società d'arrivo. Un nodo politico fondamentale del processo immigratorio, ma anche un problema ancora aperto e dalle molteplici soluzioni, come dimostra il fatto che delle diverse strade scelte in passato dai paesi d'immigrazione europei ed extraeuropei, nessuna ha realizzato pienamente i suoi obiettivi, mentre generali rimangono le difficoltà di inserimento per gli immigrati."

make its achievement even more difficult, even improbable, as the immigrants' problems coincide with those of the native population. The immigrant can be considered by the most disadvantaged citizens as being a potential rival, and the actions of public authorities on his behalf can result in dangerous xenophobic and racist feelings that lead to aggressive and violent behavior and that are in contradiction with the fundamental principles of democracy. In this case, we lose sight of the basic problem: the search for a solution to control the flows without harming anyone. This is especially true for recent receiving countries (especially Italy), which must face the traditional problems of immigration and those resulting from the unavoidable globalization process of migratory movements. Two-thirds of humanity have access to less than one-third of global wealth, and this proportion of the planet's population is expected to increase rapidly for several decades before achieving its demographic transition, whereas the rich countries are experiencing an enormous reduction in births.

In the years to come, the developed world will not easily escape migratory pressure from the poor countries. Attempts to solve this through border closure policies will only aggravate the problem of illegal immigration. From now on, it appears evident that nothing will be done to improve cooperation between the north and the south, because the trend of the employment market in the developed countries does not provide any hope for mass recruitment of foreign workers. It is equally certain that numerous factors that could limit migratory flows are associated with the integration and globalization processes of the economy, a process capable of encouraging the movement of capital from the rich countries to the poor countries, where the labor force is based, and increasing international business (Malacic, 1996; Tapinos, 1996). Although such a prospect is no doubt desirable, it is not feasible in the long or short term as long as all the countries of the world are not working together toward the same goal.

Acknowledgment

This chapter was translated from the French version by Accenta Ltd. The chapter draws on an article published by the *Enciclopedia Italiana* (Caselli, 2001).

References

Assante Franca (eds.), 1978. *Il movimento migratorio italiano dall'unità nazionale ai giorni nostri*. Geneva, Librairie Droz, vol. 2, p. 369 + 439.

Birindelli Ana Maria, 1989. Le migrazioni con l'estero. Chiusura di un ciclo e avvio di una nuova fase, *in*: Eugenio Sonnino (ed.), *Demografia e società in Italia*, p. 189–224. Rome, Editori Riuniti, 311 p.

Bonifazi Corrado, 1998. *L'immigrazione straniera in Italia*. Bologna, Il Mulino, 291 p.

Bonifazi Corrado and Gesano Giuseppe, 1994. L'immigrazione straniera tra regolarizzazione dei flussi e politiche di accoglimento, *in*: Antonio Golini (ed.), *Tendenze demografiche e politiche per la popolazione. Terzo rapporto IRP (Istituto Ricerche sulla Popolazione)*, p. 259–289. Bologna, Il Mulino, 376 p.

Bonifazi Corrado and Strozza Salvatore, 2003. Le migrazioni internazionali in Europa dagli anni cinquanta ai nostri giorni, *in*: Nicola Acocella and Eugenio Sonnino (eds.), *Movimenti di persone e movimenti di capitali in Europa*, Rome, Il Mulino.

Cagiano de Azevedo Raimundo, Bacchetta Paola, Di Santo Paola (eds.), 1993. *International migrations as a factor of security and cooperation in Europe and in the Mediterranean region. European seminar proceedings.* Rome, Università degli studi di Roma "La Sapienza," 190 p.

Cagiano de Azevedo Raimondo, *et al*. (eds.), 1994. *Immigrants integration policies in seven European countries*. Rome, Università degli studi di Roma "La Sapienza," 236 p.

Caritas Diocesana, 1999. *Immigrazione: dossier statistico.* Rome, Anterem, 352 p.

Caselli Graziella, 1998. Migrazioni internazionali, *in: Enciclopedia del Novecento*. Suppl. II, p. 252–264. Rome, Istituto della Enciclopedia Italiana, 875 p.

Caselli Graziella, 2001. Le migrazioni internazionali, in *Eredità del Novecento*, vol. 2, p. 622–640. Rome, Istituto della Enciclopedia Italiana, p. 470–1091.

Castles Stephen and Miller Mark J., 1993, 1998. *The age of migration. International population movements in the modern world*. New York, Guilford Press, 336 p.

CENSIS, 1993a. *Italy today: social picture and trends, 1992*, Rome, Franco Angeli, 231 p.

CENSIS, 1993b. *I nuovi scenari dell'immigrazione: tra competizione et solidarità*, Rome, Centro Studi Investimenti Sociale, 76 p.

Coleman David A., 1994. Le monde est-il en mouvement? La migration internationale en 1992, *in: European population conference: proceedings*, March 23–26, 1993, vol. I, p. 307–401. Geneva, New York, United Nations, 447 p.

Coleman David A., 1994. International migrants in Europe: adjustment and integration processes and policies, *in*: Miroslav Macura and David Coleman (eds.), *International migration: regional processes and responses*, p. 41–76. New York, United Nations, 200 p.

Collinson Sarah, 1993. *Europe and international migration*. London, Pinter Publisher for Royal Institute of International Affairs, 189 p. (Italian translation: *Le migrazioni internazionali e l'Europa. Un profilo storico comparato*. Bologna, Il Mulino, 1994, 269 p.)

CMS/IUSSP (Center for Migration Studies/International Union for the Scientific Study of Population), 1987. Measuring international migration: theory and practice, *International Migration Review*, vol. 21(8), p. XII, 619.

Entzinger Han B., 1990. L'emergenza delle politiche di integrazione per gli immigrati in Europa, *in*: Albert Bastenier *et al*. (eds.), *Italia, Europa e nuove immigrazioni*, p. 179–200. Turin, Edizioni della Fondazione Giovanni Agnelli, 260 p.

Faron Olivier and George Pierre, 1999. Les migrations européennes de la grande guerre à nos jours, *in*: Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe*, vol. III. *Les temps incertains 1914–98*, p. 323–358. Paris, Librairie Arthème Fayard, 792 p.

Federici Nora, 1979. Migration and regional imbalances, *in*: Ronald E. Krane (ed.), *International labor migration in Europe*, p. 147–155. New York, Praeger Publisher, 250 p.

Ferruzza Angela, Baldazzi Barbara and Corrado Helen-Cristin, 1997. L'inserimento lavorativo degli stranieri presenti in Italia:

l'analisi di alcuni dati ufficiali, *in*: Piero Giorgi and Salvatore Strozza (eds.), *Studi di popolazione. Temi di ricerca nuova*, p. 437–462. Rome, Dipartimento di Scienze Demografiche, Università di Roma "La Sapienza," 727 p.

Förster Robert Franz, 1968 (originally published in 1919). *The Italian emigration of our times*. New York, Russell & Russell, 256 p.

Franzina Emilio, 1995. *Gli italiani al nuovo mondo: l'emigrazione italiana in America, 1492–1942*. Milan, Mondadori, 644 p.

Fassmann Heinz (ed.), 1994. *European migration in the late twentieth century. Historical patterns, actual trends and social implications*. Aldershot, Edward Elgar, 274 p.

Golini Antonio, Bonifazi Corrado, 1987. *Tendenze demografiche e migrazioni internazionali nell'area occidentale*. Rome, Consiglio Nazionale delle Ricerche, Istituto Ricerche sulla Popolazione (IRP), 91 p.

Gould William T.S. and Findlay Alan M. (eds.), 1994. *Population migration and the changing world order*. New York, Chichester, Wiley, 293 p.

ISMU (Fondazione Cariplo per le Iniziative e lo Studio sulla Multietnicità), 1995. *Primo rapporto sulle migrazioni 1995*. Milan, Angeli, 333 p.

ISTAT (Sistema statistico nazionale, Istituto nazionale di statistica), 1994. *Rapporto annuale 1993*, vol. 2. Rome, ISTAT, 365 + 246 p.

ISTAT (Sistema statistico nazionale, Istituto nazionale di statistica), 1997 and 1998. *La presenza straniera in Italia: caratteristiche demografiche*, vol. 2. Rome, ISTAT, 226 + 132 p.

King Russell (ed.), 1993. *Mass migrations in Europe. The legacy and the future*. London, Belhaven, 334 p.

Kupiszewski Marek, 1994. *Migration from Eastern Europe to European community: current trends and future development*. Leeds, University of Leeds, 40 p. (School of Geography Working Paper no.94/4.)

Livi Bacci Massimo, 1998a. *La popolazione nella storia d'Europa*. Rome, Bari, Laterza, 300 p.

Livi Bacci Massimo, 1998b. *Storia minima della popolazione mondiale*. Bologna, Il Mulino, 300 p.

Livi Bacci Massimo (ed.), 1972. *The demographic and social pattern of emigration from the South European countries*. Florence, Dipartimento statistico-matematico dell'Università di Firenze, 393 p.

Macura Miroslav, 1994. Overview, *in*: Miroslav Macura and David A. Coleman (eds.), *International migration: regional processes and responses*, p. 1–20. New York, United Nations, 200 p.

Macura Miroslav and Coleman David A. (eds.), 1994. *International migration: regional processes and responses*. New York, United Nations, 200 p.

Malacic Janez, 1997. Migrations internationales: les solutions de rechange, *in*: Council of Europe (ed.), *Mediterranean conference on population, migration and development. Conférence mediterranéenne sur la population, les migrations and le développement*, Palma de Mallorca, October 15–17, 1996, p. 213–226. Strasbourg, Council of Europe, 473 p.

Münz Rainer, 1995. Where did they all come from? Typology and geography of European mass migration in the twentieth century, *in*: EAPS (ed.), *Evolution or revolution in European population*, European population conference, Milan, September 4–8, 1995, vol. I, p. 95–154. Milan, Angeli, 266 p.

Natale Marcello, 1990. L'emigrazione straniera in Italia: consistenza, caratteristiche e prospettive, *Polis*, vol. 4(1), p. 5–40.

Natale Marcelloand Strozza Salvatore, 1997. *Gli immigrati stranieri in Italia. Quanti sono, chi sono, come vivono*? Bari, Cacucci, 272 p.

OECD (Organisation for Economic Cooperation and Development), 1992, 1995, and 1998. *Trends in international migration: continuous reporting system on migration*. Paris, OECD, 157 + 243 + 272 p.

Papademetriou Demetrios G., 1994. International migration in North America: issues, policies, implications, *in*: Miroslav Macura and David A. Coleman (eds.), *International migration: regional processes and responses*, p. 77–107. New York, United Nations, 200 p.

Paradisi Bruno, 1979. Migrazioni umane, *in*: IEI (ed.), *Enciclopedia del novecento*, vol. IV, p. 287–304. Rome, Istituto della Enciclopedia Italiana, 1060 p.

Paris Robert, 1975. L'Italia fuori d'Italia, *in*: Ruggiero Romano and Corrado Vivanti (eds.), *Storia d'Italia, tomo 4: "Dall'Unità a oggi"*, p. 509–619. Turin, Einaudi, 820 p.

Reinhard Marcel R., Armengaud André, and Dupaquier Jacques, 1968. *Histoire générale de la population mondiale*. Paris, Montchrestien, 708 p. (Italian translation: Storia della popolazione mondiale, Bari, Laterza, 1971, 1040 p.).

Rosoli Gianfausto (ed.), 1978. *Un secolo di emigrazione italiana, 1876–1976*. Rome, Centro Studi Emigrazione, 385 p.

Salt John, 1989. A comparative overview of international trends and types, 1950–80, *International Migration Review*, vol. 23(3), p. 431–455.

Salt John, 1998. International migration in the UNECE region: patterns, trends, policies, *in*: *European regional conference on population*, Budapest, December 7–9, 1998, p. 61. Geneva, Université de Genève-Laboratoire de démographie économique and sociale.

Schoorl Jeannette J., 1997. Migration from African and Eastern Mediterranean countries to Western Europe, *in*: *Mediterranean Conference on Population, Migration and Development/Conférence mediterranéenne sur la population, les migrations and le développement, Palma de Majorque, 15–17 October 1996*, p. 133–223. Strasbourg, Council of Europe, 473 p.

Sonnino Eugenio and Nobile Anuntiata, 1988. Questione demografica e grandi migrazioni, *in*: *La Storia*, vol. VI, p. 315–355. Milan, UTET, 775 p.

Sori Ercole, 1979. *L'emigrazione italiana dall'unità alla seconda guerra mondiale*. Bologna, Il Mulino, 512 p.

Strozza Salvatore, 1994. La presenza straniera in Italia: un esame critico delle possibilità di rilevazione e di stima, *in*: Luciano Ciucci and Filomena Racioppi (eds.), *Studi di popolazione: nuovi approcci per la descrizione e l'interpretazione*. Convegno dei giovani studiosi dei problemi di popolazione, Roma, 2–4 giugno 1993, p. 185–214. Rome, Dipartimento di Scienze Demografiche, Università di Roma "La Sapienza," 742 p.

Tapinos Georges Photios, 1997. International migration and development, *Population Bulletin of the United Nations*, no. 36, p. 1–19.

Tapinos Georges Photios, 1997. Development, cooperation and international migration: The European Union and the Maghreb, *in*: *Mediterranean Conference on Population, Migration and Development/Conférence mediterranéenne sur la population, les migrations and le développement*, Palma de Majorque, October 15–17, 1996, p. 301–345. Strasbourg, Council of Europe, 473 p.

Tonizzi M. Elisabetta, 1999. *Le grandi correnti migratorie del Novecento*. Turin, Paravia Scriptorium, 176 p.

Tribalat Michèle and Simon Patrick, 1993. Chronique de l'immigration, *Population*, vol. 48(1), p. 125–182.

UNECE (United Nations Economic Commission for Europe), 1996. *International Migration Bulletin*, no. 8, p. 1–12.

Werner Heinz, 1994. Regional economic integration and migration: the European case, *Annals of the American Academy of Political and Social Science*, vol. 534, p. 147–164.

Widgren Jonas, 1987. International migration: new challenges to Europe, *Migration News*, vol. 36(2), p. 3–35.

Zlotnik Hania, 1995. The South-to-North migration of women, *International Migration Review*, vol. 29(1), p. 229–254.

CHAPTER

59

Theories of International Migration

HANIA ZLOTNIK

Population Division, United Nations, New York, United States

Attempts to establish a theory of international migration have a long history. In the late 19th century, E. G. Ravenstein (1885, 1889) formulated the *laws of migration* on the basis of empirical analysis of migration in England and established the first theoretical framework for the analysis of the phenomenon. Although some of the laws formulated by Ravenstein do not have the universality that the term *law* suggests, they nevertheless encapsulate the principles that were to guide most of migration research during the 20th century. Ravenstein stated that "the major causes of migration are economic" and that "the major direction of migration is from the agricultural areas to the centres of industry and commerce," foreshadowing the attention that 20th century researchers would accord to the study of rural-urban migration and of the economic aspects of migration in general. Today, at the beginning of the 21st century, there is still no comprehensive theory of migration, but considerable progress has been made in devising explanatory frameworks that are useful in understanding the determinants of migration and that permit the derivation of testable hypotheses. Given the multifaceted nature of migration, it is not surprising to find that several disciplines have focused on its study and that each has posited some theoretical explanation about the nature, genesis, and development of specific types of migration. This chapter reviews the main tenets of the different theories, frameworks, or hypotheses that have been proposed within the various disciplines and then offers a comparative analysis of their strengths and limitations. The chapter draws on material prepared for the 1997 World Population Monitoring of the United Nations (United Nations, 1996).

I. THE ECONOMIC APPROACH

From the economic perspective, international migration is a mechanism to redistribute labor. Consequently, international migrants are equated with workers and existing economic theories are aimed at explaining *labor migration*, meaning the international movement of economically active individuals. Perhaps the oldest theory seeking to explain the occurrence of labor migration is that based on the thinking of Adam Smith and other economists of the period who suggested that the migration of workers is caused by differences in the supply of and demand for labor in different locations. Adam Smith observed that the laws in England that hindered the poor from moving from one locality to another, as well as those that prevented people from changing trades, served only to maintain inequality between places. He therefore argued in favor of dismantling barriers to the free circulation of labor to permit the natural tendency of workers to move from low-wage to high-wage areas to express itself and consequently to facilitate the economic progress of the areas of origin and those of destination and the progress of the migrants themselves. Adam Smith advocated the promotion of the completely unfettered international movement of capital, goods, and labor so that market forces could ensure maximum economic expansion and a reduction of poverty.

1. The Neoclassic Theory of Migration

Modern scholars (Todaro, 1976; Harris and Todaro, 1970; Sjaastad, 1962) have modified the classic theory, developing the *neoclassic theory of migration*, which posits that countries with a large endowment of labor relative to capital will tend to have a low equilibrium market wage, whereas countries with a limited labor endowment relative to capital will tend to have high market wages. The resulting wage difference entices workers from low-wage countries to move to high-wage countries. As a result of that movement, the supply of labor decreases and wages rise in the countries of origin, and the supply of labor increases and wages fall in the countries of destination, leading eventually to a new equilibrium at which wage differences reflect only the costs of migration. At that point, international migration ceases.

The macroeconomic formulation presented has a microeconomic counterpart that focuses on individual migrants acting as rational actors and deciding to migrate on the basis of cost-benefit calculations (Todaro, 1976; Harris and Todaro, 1970; Sjaastad, 1962). People are seen as choosing to move to wherever they can be most productive given their skills. However, to attain higher productivity, they must invest in their migration, which involves tangible and intangible costs (e.g., the costs of travel, of job search, of adaptation to a new environment). When applying this theory to actual case studies, the benefits of migration are measured in terms of net returns at the individual level discounted over a certain time horizon. Net returns are estimated as the difference between expected earnings in the country of destination and expected earnings in the community of origin. Expected destination earnings are the earnings at destination estimated on the basis of an individual's skills multiplied by the probability of that individual getting a job at destination. Similarly, expected earnings in the community of origin are the observed earnings multiplied by the probability of employment at origin. Net returns are summed over the time horizon discounted by a factor that reflects the greater utility of money earned in the present than in the future, and the sum is taken as an estimate of the benefits of migration (Massey *et al*., 1993; Todaro, 1976; Harris and Todaro, 1970; Sjaastad, 1964). The subtraction of estimated costs leads to a measure of the expected gains from migration. This approach to the quantification of the expected gains from migration implies that its occurrence depends on wage differences between countries and on differences in unemployment levels between the country of origin and that of destination. Because expected earnings depend on an individual's characteristics (including human capital), factors leading to the selectivity of migration are explicitly taken into account. Any factor that reduces the costs of migration tends to increase its likelihood of occurring. In practice, the problem of measuring costs is far from straightforward, because it is not easy to attach a value to preferences, personality traits, family constraints, or perhaps more importantly, to the obstacles faced by prospective migrants when trying to secure entry to countries characterized by restrictive admission policies.

The neoclassic theory of international migration has well-known limitations, including its failure to take into account the international political and economic environment, as well as the effects of state-level economic and political decisions that influence individual decisions regarding migration (Papademetriou and Martin, 1991). The macroeconomic model assumes that there is homogeneity of skills between areas of origin and destination (i.e., that labor is completely interchangeable) and that full employment is maintained in both areas during the migration process. Because those assumptions are generally not fulfilled in practice, the effects of labor migration tend not to conform to those predicted by neoclassic theory (Martin, 1991). The microeconomic model has generally not been put to a rigorous test in explaining international migration (Massey *et al*., 1994). However, there is ample evidence that wage differences play a significant role in determining the size of migration flows in many settings, so that it is widely accepted that the existence of such differences is a necessary, albeit not a sufficient, condition for the migration of labor to occur.

2. The New Economics of Migration

Recognition of the limitations of neoclassic theory has led to the proposition of alternative theories. The most recent, known as the *new economics of migration*, focuses on the microeconnomic level, but instead of assuming that migration decisions are made by individuals acting largely on their own, it assumes that people act collectively, typically within households or families, to maximize expected income and to minimize risks and loosen constraints associated with a variety of market failures that are particularly common in developing countries, where crop insurance markets, futures markets, or unemployment insurance do not exist or are beyond the reach of most people (Stark, 1991). According to this approach, households attempt to minimize risks to their economic well-being by diversifying the allocation of family labor. From this perspective, sending certain family members to work in another country where wages and employment conditions are largely independent of local economic conditions is a form of insurance against the deterioration of the latter. International migration and the remittances it generates can also allow households to obtain the capital needed to increase the productivity of assets in the community of origin. Such a function of migration is considered to be particularly important in developing countries where capital markets are weak, many people do not have access to banking services, and families lacking adequate collateral find it virtually impossible to borrow at reasonable interest rates.

The approach to the new economics of migration also questions the assumption that income has a constant effect on utility for different actors across socioeconomic settings. Instead, it argues that households send family members to work abroad to improve income in absolute terms and to increase income relative to other households in a reference group (Stark, 1991; Stark and Taylor, 1989). If in a community the income of affluent households increases, whereas that of poor households remains unchanged, the relative deprivation of the latter increases, and their incentive to participate in international migration rises as well, even if no change in expected wages takes place.

The theories derived from the new economics of international migration have important implications for the interrelations between international migration and development. First, they imply that migration can occur even in the absence of wage differentials between areas of origin and destination, because migration may be fueled by the desires of households in the place of origin to diversify risks. Second, because there are strong incentives for households to engage in international migration and in local economic activities, an increase in the return to the latter may heighten the attractiveness of migration as a means of overcoming capital and risk constraints on investment in local activities. Successful economic development within areas of origin need not reduce pressures for international migration (Massey *et al*., 1993). Third, international migration will not necessarily stop once wage differentials between countries of origin and destination have been eliminated, because as long as other markets within countries of origin are nonexistent, imperfect, or inaccessible, incentives for migration may continue. Fourth, governments of countries of origin can influence migration through labor-market policies and through those that shape insurance and capital markets. Government policies and economic changes that modify the income distribution will change the relative deprivation of some households and will affect the incentives to migrate. In particular, when poor households in areas of origin do not share equitably in the income gains of other households, their propensity to send some family members abroad will increase.

Evidence supporting the premises and the implications of the new economics of international migration has been growing steadily. However, most of it relates to cases of long-standing migration processes linking rural areas in developing countries to developed countries (Massey *et al*., 1994). The migration involved is usually temporary and precludes the relocation of complete households that, were it to occur, would largely reduce to a minimum the economic nexus between migrants and the community of origin. Typically, the studies available focus on the experience of just a handful of rural communities, which are not selected to be representative of the whole population, and in all cases, large wage differentials between areas of origin and destination continue to exist. Nevertheless, the evidence supports several of the conclusions predicted by the theory, including that remittances increase productive investment in the areas of origin (e.g., greater use of machinery, land, and hired labor) and promote the acquisition of income-producing assets (e.g., livestock, equipment, education). As a result, remittances can raise household income over time by more than the value of the remittances themselves (Taylor, 1992).

3. Family Migration and the Selectivity of Migration

Although not formally a part of the new economics of migration, extensions of the neoclassic migration model have been proposed to take into account that a large part of migration consists of the movement of complete families, not just of workers. Mincer (1978) considered the effects of earnings differentials across space on family migration. Such an approach is more realistic because, in many contexts, one family member may anticipate gains in potential earnings while another expects losses in the place of destination. To the extent that a family includes more than one person that is economically active, the possibility of conflicting goals exists. Assuming that a family consists of at least two married adults, with or without dependent children, Mincer (1978) shows that migration can create "tied movers" or "tied stayers" and that the difference between the individual's and the couple's optimal strategy depends on the degree of correlation in the gains from migration of the husband and wife. Only when there is a perfect correlation do the optimal strategies of the individual and the married couple coincide.

Borjas and Bronars (1989) use Mincer's approach to analyze the selectivity of migration to the United

States. Based on a model involving only individual migrants, Borjas (1987) shows that the extent of income inequality between the places of origin and destination influences the type of person that migrates. If income inequality is greater at the place of destination, people with higher than average skills have an incentive to migrate because they can earn a higher relative wage in the area of destination, other things being equal; conversely, if income inequality is greater in the place of origin, such persons have less incentive to migrate than those with lower skills. However, when married couples are considered, the selectivity effect weakens because some low-skilled migrants who would not have migrated on their own to a place with a high level of income inequality may do so if they are married to skilled persons. Such tied migrants with low skills are not likely to fare as well in the labor market of the area of destination and therefore may reduce the average wage for migrants as a whole, as Borjas and Bronars (1989) show in the case of married-couple migration to the United States.

4. The Dual Labor Market Theory

The neoclassic economic theory of migration and the new economics of migration assume that migration is mainly the result of rational choices made by individuals or families. In contrast, other economically based theories take a macroeconomic perspective and argue that international migration is mainly the product of international forces that transcend individual choice and set constraints on it. The *dual labor market theory* posits that international migration results from a permanent demand for foreign labor that is inherent to the economic structure of developed countries (Piore, 1979). Several factors contribute to create such structural demand for foreign workers. First, wages not only reflect conditions of supply and demand; they also confer status and prestige. Consequently, to preserve an established occupational hierarchy, wages cannot respond freely to changes in the supply of workers. Under those conditions, employers seeking to attract unskilled workers for jobs at the bottom of the hierarchy cannot simply raise wages, because by so doing, they would change the socially defined relationships between status and remuneration: An increase of wages at the bottom would put pressure to increase wages at other levels of the occupational hierarchy. To avoid such pressures in times of labor scarcity, employers have strong incentives to seek a cheaper solution: the importation of foreign workers willing to work for low wages.

A second factor against the use of local workers to fill in jobs at the bottom of the occupational hierarchy relates to the motivational problems associated with jobs that provide few avenues for upward mobility. Employers need workers who see bottom-level jobs simply as a means to earn an income and who disregard status implications or the lack of upward-mobility prospects. Foreign workers from low-income countries, especially those having a tentative migration status (as legal temporary workers or as undocumented migrants), are usually willing to satisfy that need because even a low wage in a developed country can be several times greater than the average wage in a developing country.

The third factor contributing to the existence of an unmet structural demand for unskilled labor in advanced industrial societies is the development of a dual economy characterized by the coexistence of a capital-intensive primary sector and a labor-intensive secondary sector and by a segmented labor market. Workers in the primary sector usually hold stable, skilled jobs that allow them to accumulate firm-specific human capital. Workers in this sector tend to be unionized and to work under contracts that require employers to bear a substantial share of the costs of any idle periods (through severance pay and unemployment benefits). They are expensive to let go and become in that sense more similar to capital. In contrast, workers in the secondary or low-productivity sector generally hold unstable, unskilled jobs and may be laid off at any time with little or no cost to the employer. They become the means of adjustment during cyclical downturns, when they can be forced to bear the costs of their own unemployment. Because native workers have no incentive to take the unstable jobs characterizing the secondary sector, shortfalls in labor supply in that sector are common and are often filled by having recourse to foreign labor. That is most likely to happen in societies where other sources of labor for the secondary or low-productivity sector—women and teenagers—are in short supply because of social constraints or better occupational opportunities in the case of women or because of declining cohort sizes and increasing years of schooling in the case of teenagers.

Although studies aimed at validating the existence of segmented labor markets and at identifying the primary and secondary sectors have been common, few have addressed the linkages between the existence of a segmented labor market and international migration. If, as the theory suggests, migrant workers were mainly recruited into the secondary labor market, we would expect to observe lower returns to education, skills, and work experience among migrants than among nonmigrants. Evidence supporting that expectation has been found with respect to Mexican migrants in the United States (Massey *et al*., 1994). Studies of other migrant groups in the United States (e.g., Cubans, Chinese) have shown, however, that urban labor markets are segmented, but particularly in cities with large migrant populations, they appear to be divided into three sectors rather than two: a primary sector, a secondary sector, and an ethnic enclave. The existence of ethnic enclaves facilitates the incorporation of migrants to the labor force, and although the jobs they initially get in the enclave pay low wages and lack prestige, the enclave provides greater opportunities for advancement than the secondary sector (Portes and Bach, 1985).

According to Piore (1979), foreign labor recruitment is the principal means through which unsatisfied demand in the secondary sector of the labor market is expressed. This assertion is validated mostly by the experience of the labor-importing countries of Europe during the late 1950s and 1960s, when the hiring of foreign workers was often done through official recruitment programs. Recruitment, whether through official channels or private ones, was also instrumental in initiating the flows of Mexican and Filipino workers to the United States and of workers from some Eastern and Southeastern Asian countries to Japan and some newly industrializing economies. However, despite the undeniable importance of recruitment in starting and building the momentum of international labor migration flows, it is not necessarily the case that, as Piore (1979) postulates, such migration is largely demand driven. In statistical analyses of the determinants of migration flows that include variables relative to both the country of origin and that of destination, those relative to the country of origin appear to have greater explanatory power (Massey *et al*., 1994). However, the analyses considered were not meant to test Piore's assertion and consequently do not test the effect of labor demand in the country of destination.

5. The World Systems Theory

A fourth approach to explain the causes of international migration is based on the thoughts of Marx regarding the functioning of international capitalism (Papademetriou and Martin, 1991; Wallerstein, 1974). Known as the *historical-structural model* or the *world systems theory*, this approach posits that the origins of migration can often be traced to the institutional and sectoral imbalances brought about by a state's incorporation into the world capitalist system. The world's labor is seen as divided into three geographically distinct zones: core, semi-periphery, and periphery. The nature of political and economic interdependence between the three zones, as well as the direction and nature of capital and commodity flows, structures the patterns of labor movements between the zones (Portes, 1995; Sassen, 1988; Petras, 1981). In core countries, the demand for migrant labor is not uniform because it usually reflects the relative scarcity of persons prepared to work for low wages. Demand for cheap labor arises in sectors, such as agriculture, that cannot rely on productivity increases to maintain profit levels. It is also linked to the overall shift of the economies of core countries from a manufacturing to a service base, coupled with the downgrading of jobs in the manufacturing sector, especially in high-technology industries (Grasmuck and Pessar, 1991). In peripheral countries, the commercialization of agricultural production generally leads to land consolidation, the substitution of cash crops for staples, the intensive use of modern inputs to produce high crop yields, mechanization and the consequent reduction of labor demand. Such changes result in the displacement of labor from the land and an increase in rural-urban migration, which contributes to producing a surplus labor force in urban areas. Typically, the growing urban labor force cannot be fully accommodated by the creation of new urban employment and high rates of unemployment and underemployment ensue, being accompanied by a growing number of persons in informal and low-paying service-sector occupations. Many families become marginalized, even during periods of rapid economic growth, and income inequality increases. The coexistence of such developments with the unmet demand for low-cost labor in core countries results in the international migration of workers from the periphery, a migration that is facilitated by the cultural and ideologic bonds between countries in the core and certain countries in the periphery and by the transportation and communication links between them, both of which are the by-products of the capitalist penetration of the periphery by core countries.

According to world system theory, migration is more likely to occur between past colonial powers and their former colonies and is facilitated by the cultural, linguistic, administrative, transportation, and communication links existing between the two. Within core countries, the management of the world economy is concentrated within a relatively small number of urban centers where banking, finance, professional services, and high-tech infrastructure coalesce. Given the strong demand for services from unskilled workers and the decline of heavy industrial production in those cities coupled with their highly developed transportation and communication links with peripheral countries, they are likely to attract a large proportion of international migrants (Sassen, 1991).

II. THE SOCIOLOGIC APPROACH

Although economic factors are almost always important, if not dominant, in migration decisions, the economic theories of international migration have often been viewed as too narrow because they cannot account for movements triggered by other considerations, including the need to flee life-threatening situations, the desire to move to a better climate, or the intention to search for a spouse in a better marriage market. Recognition of the relevance of a broader range of factors in explaining migration decisions and the fact that migration is not always voluntary underlie the approach proposed by Lee (1966), according to whom migration is caused by positive factors in possible areas of destination and by negative factors in the place of origin or current residence. Both areas of origin and destination are characterized by sets of positive factors, or forces of attraction or retention (i.e., pull factors), and negative factors, or forces of repulsion (i.e., push factors). The greater the perceived difference in the net forces of attraction (i.e., positive minus negative factors) in places of origin and destination, the more likely migration is to occur. Lee observed that migrants who respond primarily to pull factors in the place of destination tend to be positively selected, whether by age, education, skills, or motivation, whereas those responding primarily to push factors in the area of origin tend to be negatively selected. The former group is more likely to migrate voluntarily than the latter. An implication of this theory is that emigrants from a given origin are likely to be characterized by a bimodal distribution according to certain characteristics (e.g., educational attainment, skill levels). Such a bimodal distribution has been observed in many empirical studies.

Lee's approach built on the work of Stouffer (1940, 1960), who suggested that migration was a function of the relative perceived attractiveness of areas of origin and destination and that migration flows were limited by the existence of intervening opportunities and obstacles, both of which increased with the distance separating origin and destination. Despite the declining costs of travel, distance continues to be an important factor reducing migration propensities or, as the economic models would put it, increasing the tangible and the intangible costs of migration.

As in the case of economically based theories of international migration, those based on sociologic considerations have evolved toward recognizing the important role played by the family as a decision-making unit or as an institution providing the needed support for migration to take place. Migration often involves family groups, not just workers, as posited by economic theory. Family units may move simultaneously or in stages, with one family member relocating first and others following. The migration policies of countries of destination condition whether family migration can take place and the manner in which it can occur. Evidence showing that migrants often rely on kin already established in the country of destination to obtain accommodation, find a job, or secure financial and other types of support during an initial period of adaptation implies that kinship ties between potential migrants in the country of origin and residents of the country of destination can reduce the costs of migration, lower the risks involved, and increase the returns of migration. The concept of a migration network, which encompasses all interpersonal ties that connect migrants, former migrants, and nonmigrants in areas of origin and destination, provides a generalization of the kinship group because the ties considered include those based on kinship and those resulting from friendship or shared community origin (Massey *et al*., 1993; Gurak and Caces, 1992). Network connections can be considered as a form of social capital that people can draw on to gain information, material, or psychological support to facilitate migration and the adaptation process. Key elements of a migration network are the ties between migrants already established in the area of destination and persons remaining in the area of origin. Because of the assistance provided by the former to the latter, the costs and risks of migration decrease and the probability of further migration increases. Migrant networks contribute therefore to maintain the migration momentum even after the factors responsible for initiating the flow have lost their relevance. As the costs and risks of migration decline, the flow becomes less selective in socioeconomic terms, and migrants become more representative of the sending community as a whole. Because of the efficient operation of migration networks, governments may have difficulty in controlling the size of migration flows, particularly when their migration policies reinforce the operation of networks by facilitating family reunification or the admission of new migrants on the basis of kinship ties with previous migrants.

III. THE MOBILITY TRANSITION

At a time when demographers were positing that the path to modernization brought with it specific patterns of demographic change (i.e., the so-called demographic transition from high to low mortality and then from high to low fertility), Zelinsky observed that modernization generally was associated with specific changes in the patterns of mobility. He posited the hypothesis of the mobility transition (Zelinsky, 1971), which is a model associating patterns of variation of different types of spatial mobility with the phase in the path to modernization in which a society finds itself in. One valuable aspect of Zelinsky's approach is that it considers explicitly the existence of different types of migration: international migration; migration to the frontier areas of a country (frontierward); rural-urban migration; urban-urban and intra-urban migration; and a whole set of short-term types of mobility that he called *circulation*. According to Zelinsky, the five types of mobility identified vary in relative importance over time but remain correlated with the stage of the demographic transition as suggested by the diagrams presented in Figure 59–1.

Zelinsky also considered the likely impact of better transportation on mobility. As transportation improves, people can stay in their places of origin and commute to work, whereas during phases in which transportation was poor, they would have had to move. As modernization advances, some forms of migration are "absorbed" by circulation (Fig. 59–1, diagram F). As electronic communication improves, more people may be able to work at home, avoiding the need to commute altogether, as is suggested by diagram G in Figure 59–1.

Skeldon (1997) observes that Zelinsky's mobility transition has several weaknesses, two of which stem

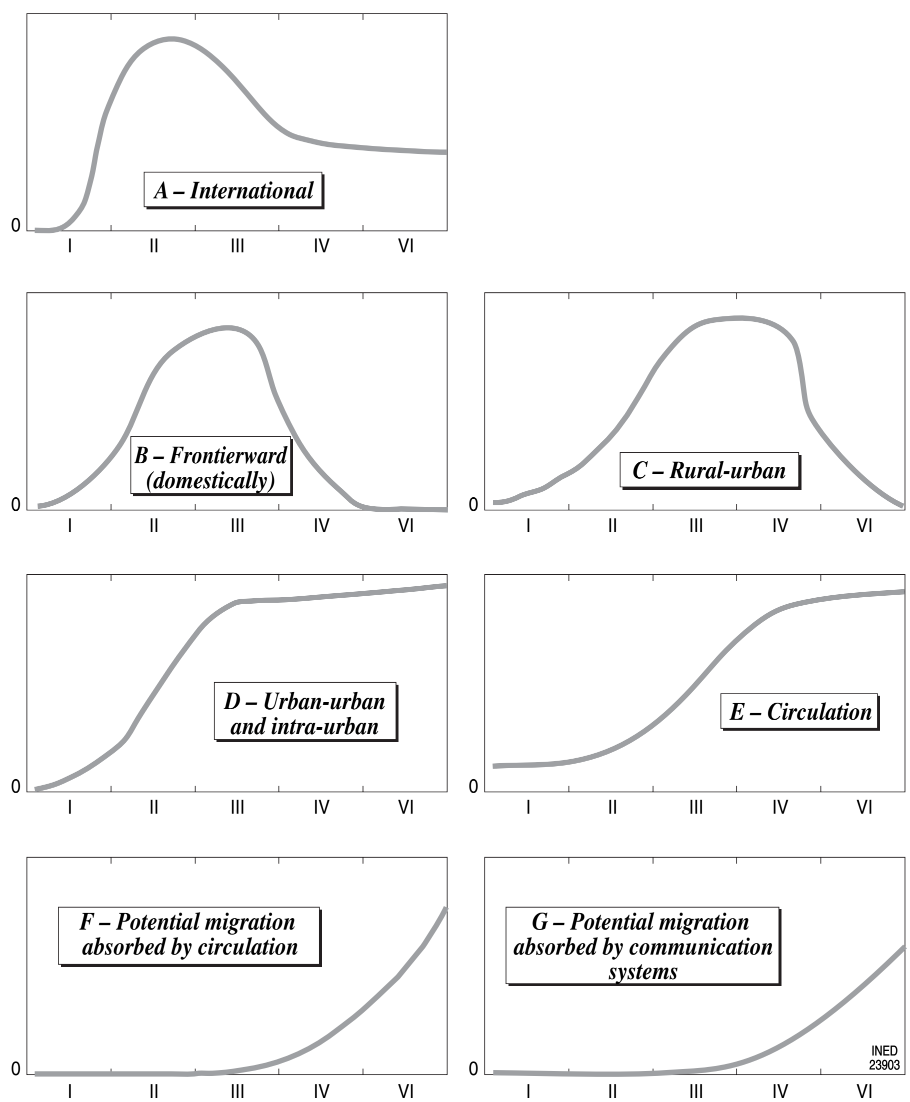


FIGURE 59–1 Variation of different types of migration in the course of the mobility transition. (From Zelinski, 1971.)

from problems in the basic assumptions made. Zelinsky's model assumes that mobility is almost nonexistent or rare in premodern societies, an assumption that is not validated by the evidence. The model assumes that the modernization process is unilinear—that societies move steadily toward ever-higher levels of prosperity. To be fair, that assumption is also at the basis of the theory of the demographic transition. However, today it is better known that there is no single path through the transition from high to low fertility and mortality. It is therefore unlikely that migration, which is an even more varied phenomenon, will change in similar ways in all development contexts.

Despite its drawbacks, Zelinsky's idea of the existence of specific migration patterns related to the stage of development or the stage in the demographic transition, which is reminiscent of the attempt by Ravenstein to establish laws of migration, has had sufficient appeal so as to be echoed in other attempts to provide explanations for the changes observed in international migration. For instance, in the 1990s, the concept of the *migration transition* was evoked when trying to identify the factors that had brought about the transformation of labor-surplus economies of emigration into labor-deficit economies of immigration among the newly industrializing countries of Eastern and Southeastern Asia (Abella, 1994; Fields, 1994). However, as Skeldon (1997) amply documents, despite the attractive simplicity of the model, it seldom fits well with actual experience. In the case of the Asian countries, inflows and outflows of migrants coexist; even the cases of Hong Kong or Malaysia, there has been no lengthy period since 1960 in which net migration has remained negative, so it is hard to argue that they have passed from being countries of emigration to being countries of immigration.

IV. THE CONTRIBUTION OF POLITICAL SCIENCE

As Zolberg (1981) observed, theories that model migration as the response of individuals to unevenly distributed opportunities make no significant distinction between internal and international migration, according secondary importance to factors such as barriers restricting exit or entry, deliberate recruitment efforts, or forced departures. Such theories cannot explain why most of the world's population, which should rationally opt to move to another country, does not migrate internationally. To explain the low international mobility of people, it is necessary to take account of the organization of the contemporary world into mutually exclusive and legally sovereign states, each containing some approximation of a single society, whose social boundaries coincide, by and large, with the frontiers delineated by international law as those of the state. In this context, international migration can be seen as an interaction between states, whereby there is a transfer of jurisdiction so that international migrants cease being members of one society and become instead members of another. From this perspective, international migration constitutes an exception from the prevailing norm of current social organization, which holds that society is a territorially based, self-reproducing cultural and social system whose human population is assumed, tacitly or explicitly, to renew itself endogenously over an indefinite period. The exceptional character of international migration is further buttressed by the fact that, whereas international instruments enshrine the right to leave any country including one's own, there is a universal and unambiguous consensus on the very opposite principle, that every state has the right to restrict the entry of foreigners, a right that is essential to maintain the integrity of society.

The analysis of the factors giving rise to international migration must be approached through a framework that takes the anomalous aspect of international migration as a point of departure rather than merely as an incidental feature. The world can be conceived as consisting, on the one hand, of individuals seeking to maximize their welfare by exercising a variety of choices, including that of transferring from one political jurisdiction to another and, on the other, of mutually exclusive societies, acting as organized states to maximize collective goals by controlling the exit and entry of individuals and their ultimate membership in society through nationality and naturalization laws. International migration is related to a fundamental tension between the interests of individuals and those of societies. However, the emphasis on tension does not preclude the possibility that individual and collective interest may coincide under particular circumstances, as with respect to the freedom to stay in, leave, or return to one's own country. The key point is to approach the study of the determinants of international migration in terms that consider to what extent individual and societal interests coincide or are antagonistic to one another, as well as to consider the interests of the state of destination and those of the state of origin and the societies each represents.

With respect to labor migration, this approach focuses on the conflicting interests of receiving societies (to maximize the labor supply and to protect cultural integrity), which lead to a dilemma with a limited number of possible solutions, the most common of which is to confine foreign workers strictly to their economic role by reinforcing barriers against their integration in society. Such barriers may include the physical isolation of migrants by concentrating them in special areas, legal restrictions on their length of residence or on the employment they may hold, constraints on family reunification, restrictions on their access to social services, barriers to their naturalization, and other legal devices that imply the erection of a societal boundary within the territorial confines of the receiving society, a boundary that is meant to offset the consequences of physical presence. In some countries, tolerance of unauthorized or illegal migration can be seen as another way of maintaining a separation between international migrants and the host society, given that the very status of those migrants ensures that they will work but not become incorporated (Zolberg, 1981).

With respect to forced migration, although refugees constitute a pervasive presence in the contemporary world, the processes that produce refugees have received little attention from theorists of international migration. Refugees are often the product of the nation-building process, whereby states try to conform to the nation-state model, according to which the state represents a homogeneous society with a uniform culture. Societies that are culturally or ethnically heterogeneous face the problem of increasing their homogeneity to conform to that model. Cultural homogeneity becomes an indivisible public good whose attainment may require the transformation of individuals through, for instance, religious conversion or linguistic assimilation. When transformation is impracticable or does not succeed, the state may expel or segregate the individuals involved (Zolberg, 1981). Because such actions by the state are tantamount to persecution, if the targeted individuals manage to leave their country, they normally become refugees.

State building entails regime construction, which stresses political uniformity or ideologic conformity (Zolberg, 1981). Because ideologic orientations are usually founded on objective interests related to position in the social structure, those evincing heterogeneity or lack of conformity are often strata or classes, or at least segments of them. Sometimes, the groups targeted are constituted by the descendants of migrants originally brought in by the colonizing power to serve as commercial intermediaries. When a political strategy of uniformity and an economic strategy of autarchy are simultaneously espoused by emerging states, migration policies preventing the exit of most citizens are likely to be adopted simultaneously with measures that force members of certain groups to leave (Zolberg *et al.*, 1989; Dowty, 1987). Although on the surface such policies appear as contradictory, expulsions of selected groups and prohibitions against the exit of the remainder of the population are complementary techniques of control. A comprehensive model of the causes of international migration should allow for their coexistence.

V. THE SYSTEMS APPROACH

Each of the approaches previously described presents a partial explanation of the causes and likely dynamics of international migration. In an attempt to combine the different approaches and recognizing that there is considerable interdependence between the migration experiences of groups of countries of origin or between that of groups of countries of destination, an approach focusing on migration systems has been proposed. Although the idea of using a systems approach to study migration can be traced back to the work of Mabogunje in the early 1970s (Mabogunje, 1970), its implications for the study of international migration began to be explored only in the late 1980s, and its formalization as a framework for analysis is still in the formative stages. The most comprehensive specification of what constitutes the systems approach resulted from the work of the Union for the Scientific Study of Population (IUSSP) Committee on International Migration, which published the results of its activities in 1992 (Kritz *et al.*, 1992). However, the IUSSP Committee's work marked only a first step in promoting the use of the systems approach in the study of international migration and left considerable space for more precise specification of models, hypotheses, and related measurement and estimation issues that must be addressed for the systems approach to fulfill its potential as a useful analytic tool.

The systems approach is based on a number of premises. The first is that migration creates a unified space, encompassing both the place of origin and that of destination. To understand migration and its causes and impacts, consideration of this unified space is necessary. The second is that migration is but one of the processes linking areas of origin and destination and is closely associated with those processes, often of long historical gestation. Consequently, most migration flows occur between areas that already have some historical, cultural, political, or economic linkages. The third is that the processes linking areas of origin and areas of destination modify, over time, the conditions in both areas, giving rise to feedback mechanisms likely to transform the initial processes. Migration will most probably change in magnitude and characteristics because its impact on the area of destination and the linkages it implies with the areas of origin varies over time. The fourth is that the state plays a crucial role in determining international migration flows through explicit policies to shape migration or by fostering economic, strategic, or political linkages with other states that may indirectly lead to migration. The fifth is that it is necessary to identify the mechanisms through which the macroeconomic-level forces that influence migration flows are translated into the determinants of migration at the level of the actual, individual decision-makers. One of the mechanisms identified is the operation of migrant networks, which are constituted by migrants, their relatives, friends, and acquaintances, whose social and kinship relationships act as links at the microeconomic level between the communities of origin and those of destination. Migrant networks help channel resources and information between communities, facilitating the further migration of individuals. Other mechanisms of interest include the operation of labor recruiters, multinational corporations, training and educational institutions, the military, and others engaged in fostering international flows of people. Consequently, the systems approach is based on "the conceptualization of a migration system as a network of countries linked by migration interactions whose dynamics are largely shaped by the functioning of a variety of networks linking migration actors at different levels of aggregation" (Kritz *et al*., 1992, p. 15).

Central to the use of the systems approach is the identification migration systems. A group of countries may be considered as generating a system if they satisfy the following conditions:

1. They have a similar position relative to migration; they can be identified as attraction poles for migrants or as sources of migrants.
2. If the countries concerned are mainly countries of destination, the migration flows converging to them should show considerable overlap in terms of the countries of origin involved.
3. The countries concerned should have the same level of development, a high degree of cultural affinity, and similar institutional structures.
4. They should also show some degree of coherence in the policies that they use or have used to control migration, and they should be linked by strong economic or political ties.

Given that migration creates a unified space encompassing the places of origin and those of destination, it is necessary to focus on a system as a whole to understand migration dynamics. Because international migration is generally just one of the processes linking areas of origin and those of destination and is closely associated with other processes, it is important to take into account the impact that the other processes may have in shaping migration. To do so, use of a combination of macroeconomic and microeconomic variables is necessary. Similarly, macroeconomic-level considerations are imperative to take account of the role that states play in initiating and molding migration flows. The appropriate methodology needs to be developed to ensure that the effects of feedback mechanisms are an explicit part of any model used.

Until now, the use of the systems approach to analyze international migration has mostly remained at a descriptive level. General observations have been made about the similarity of linkages among the countries generating the system (i.e., core countries) and among countries of origin, but the relevance of such linkages has not been formally or quantitatively tested. A major obstacle in advancing further the use of the systems approach is the lack of comparable and comprehensive data on international migration. Even for systems containing only a few core countries in the same part of the world, such data are generally not available. Exploring the relative strength of possible macroeconomic-level factors determining migration flows also demands that comparable indicators on a variety of socioeconomic issues be available for most countries of the world. Data limitations are likely to constrain the types of models that can be examined, although attempts at the examination of a partial system have been carried out (Kritz, 1995, 1996).

VI. COMPARISON OF EXISTING THEORIES OF MIGRATION

Table 59–1 compares the theories of migration previously discussed. A key point that the table makes clear is that most of the theories available relate to both internal and international migration without, in general, making any explicit differentiation between the two types of flows (the only exception is Zelinsky's mobility transition, which deals explicitly with internal and international migration separately). All the economic theories that focus on migration at the microeconomic level make no distinction between migrants moving within the borders of their own countries and those moving internationally, although the models proposed by the neoclassic theory and the economics of family migration can make some allowance for the difference between internal and international migration through higher wage differentials or through the increased costs usually involved in international migration. In the case of the new economics of migration, the emphasis the theory puts on the institutional constraints that prompt people to move is based on the

TABLE 59–1 Comparison of the Theories of Migration

Theory type	Theory	Main proponent	Type of migration		Level of analysis	Unit of analysis	Type of constraints explained	Part of the process
Economic	Neoclassic (1)	Harris-Todaro	Internal or international	Labor	Macro	Community or country	None	Initiation, maintenance, and end
	Neoclassic (2)	Harris-Todaro	Internal or international	Labor	Micro	Individual	Costs of migration and probabiliy of employment	Initiation
	New economics of migration	Stark-Taylor	Internal or international	Labor	Micro	Family	Lack of adequate financial or insurance institutions	Initiation and maintenance
	Economics of family migration	Mincer	Internal or international	Labor	Micro	Family/couple	Migration to the same destination	Initiation
	Segmented labor market	Piore	International	Labor	Macro	Community or country	Structural constraints at destination reduce socal mobility	Initiation and maintenance
	World systems	Wallerstein	International	Labor	Macro	Country	Constraints imposed by position in the capitalist system	Initiation and maintenance
Sociologic	Push/pull	Lee	Internal or international	All types	Micro	Individual	Some constraints may be interpreted as push factors	Initiation
	Networks	Various authors	Internal or international	All types	Micro >> macro	Individual/ family	None	Maintenance
Transition	Mobility transition	Zelinsky	Internal and international	All types, excluding forced	Macro	Countries	None	Process
Political	Political science	Zolberg	International	All types, including forced	Macro >> micro	State in relation to individual	Those imposed by States of origin and destination	Process
Systems	Systems approach	Kritz *et al.*	International	All types	Macro >> micro	State, community, and individual	Those imposed by states of origin and destination, by the structure of production, and by institutions	Process including feedback mechanisms

premise that areas of origin lack the financial institutions typical of the advanced market-economies so that the theory implicitly restricts its explanatory power to flows of migrants from developing and relatively backward communities to more advanced ones. This is somewhat of a simplification, because the theory's premise about the importance of risk diversification in determining the decision to migrate can be applied to a wide range of circumstances, but there is no denying that the most likely scenario in which the theory may be relevant is in cases where migration originates in developing countries or countries with economies in transition where people cannot normally rely on local institutions to diversify risk, obtain insurance, or get access to capital.

Table 59–1 shows that all theories that focused exclusively on international migration have devised explanatory frameworks at the macroeconomic level and have emphasized the constraints that migrants face in terms of the global capitalist system (Wallerstein) or the economic system of the community of destination (Piore) or the state (Zolberg). When the state is considered explicitly, it is recognized that the interplay between the state of origin and that of destination is relevant in determining the constraints actually faced by a migrant. The political science and the systems approach emphasize that in understanding migration at least three players need to be considered simultaneously: the state of origin, the state of destination, and the migrant. No other theory makes explicit mention of the state, although constraints emanating from the relation between migrant and state can be accommodated by the push-pull approach proposed by Lee.

As expected, all theories using an economic approach focus their explanatory power on the dynamics of labor migration (i.e., the movement of economically active individuals). The normal premise of such theories is to consider that workers are unattached individuals that can make decisions based entirely on the workings of the labor market (e.g., Harris-Todaro, 1970) or whose decisions are conditioned by the workings of the labor market (e.g., Piore, 1979). In such a context, the issue addressed by Mincer adds a welcome dose of realism to the theory and illustrates the complexity of incorporating a decision-making in which the interests of several individuals are at stake. Mincer's approach was developed to analyze the constraints faced by migrants moving within advanced market-economy countries. Only in that context did it seem necessary to attach equal importance to the economic goals of husband and wife. In general, by being gender blind, existing theories of migration tend to disregard the specificity of female migration and do not necessarily reflect the realities of the opportunities and constraints that women face when they engage in international migration.

In a similar vein, although there are several theories that do not explicitly delimit their universe of interest to specific types of migrants, they do not necessarily accommodate with equal ease the dynamics of voluntary migration and those characterizing forced migration. The political science approach is the only one providing a specific framework for the analysis of forced movements, although the push/pull approach has been used to assess the factors leading to such movements and both systems and networks theory can be used to understand how forced migration arises or may be maintained.

Massey and colleagues (1993, 1994) have argued that it is important to distinguish between theories that provide explanations for the initiation of migration and those that provide reasons for the maintenance or continuation of migration flows once they have started. Most of the available theories fail to state explicitly whether their focus of concern is the genesis of migration or its maintenance over time, but according to my interpretation, many of them address the processes triggering migration and those sustaining it after it has started. The new economics of migration, for instance, suggests that migration starts as a risk minimization strategy and to palliate the negative effects of the lack of local financial institutions or the limitations of existing ones. However, because migration is unlikely by itself to transform local institutions or to eliminate existing risks, it generally needs to be maintained in order for those left to keep on benefiting from its palliative effects. The appropriate word is not *maintained* but rather *expanded*, because as the theory itself expounds, as migrants become successful in improving the well-being of those that they have left, others will want to emulate that success by migrating. The development of migrant networks is the logical next step to ensure that the required expansion happens in the most efficient way possible. In this sense, the networks theory that was presented earlier is really part of the complex process that is part and parcel of migration. It is that holistic approach that is embodied by the systems approach.

Another important distinction that can be made between theories is that some refer explicitly to migration as a process whose characteristics change through time or, as Zelinsky (1971) put it, according to the different stages in which society finds itself in. Zolberg's framework (Zolberg, 1981) for the generation of forced migration also posits that the likelihood of such migration depends on the stage that the state of origin has reached in the nation-building process. The sense of the existence of some sort of evolution or life cycle of migration streams influences the idea that flows start, are maintained, and end. The key difference between considering a single stage in the "life" of a flow and addressing likely changes through time or in relation to the evolution of other processes, as the systems approach suggests, is that the second strategy is likely to provide a richer and more comprehensive understanding of the feedback mechanisms through which migration interacts with other factors. That is the essence of the dynamic framework that the systems approach attempts to provide.

The theories available suggest that it is generally more straightforward to address the mechanisms shaping migration at the macroeconomic level and their interactions than to stipulate the modes of articulation between the macroeconomic and the microeconomic levels. Only three approaches attempt to describe such modes of articulation, and in all cases, the description of the process involved remains too vague to be a useful tool for furthering understanding. The systems approach posits that social networks are one of the mechanisms for the translation of macroeconomic effects to the microeconomic level, but in the systems and the networks approaches, the networks themselves are described as a set of transnational relations between individuals that are not easily amenable to analysis through standard social science instruments. To be more precise, existing instruments have been successful in documenting the existence of networks and in elucidating some of the functions that networks perform, but they have not gone much beyond providing descriptive information on the issue. Theories need to go beyond mere description.

One particularly important aspect of the micro-macroeconomic articulation is the interaction between the state and the individual. The political science approach emphasizes the importance of such interactions, and the systems approach suggests that they become operationalized through institutions. The types of institutions and their mode of action depend on the nature of the state. In the case of a powerful state that is responsive to public demands, it is likely that institutions can work in partnership with the state, participating in the process of policy formulation in ways open to them. However, if the state is weak or authoritarian, institutions need to set parallel mechanisms of operation to supply the functions that the state has failed to undertake or to combat those that the state has put in place. The important role that institutions may play in influencing international migration has been underscored by Massey and colleagues (1993), who have gone so far as to claim that the operation of institutions produces yet another approach to the understanding of migration. In contrast, the systems approach suggests that institutions provide another mechanism for translating macroeconomic processes to the microeconomic level.

This overview suggests that, although considerable progress has been made in elucidating the mechanisms that give rise and maintain migration flows and in establishing their interrelations, there is a need for a more holistic approach to the issue, especially because of the importance of interactions among the main actors in the international migration process (i.e., the state of origin, the state of destination, and the migrant) and with a series of "intermediaries" that mediate the interactions between the main actors and are themselves players in the process and therefore have an impact in shaping flows. The systems approach tries to provide a framework for the analysis of the interactions involved but has not been fully successful in integrating the various elements that need to be taken into account. Consequently, its analytic usefulness remains largely untested. Other, more partial theories about specific aspects of international migration have been more successful in providing useful insights about the functioning of the process but their limited perspective reduces the generality of their findings. There are still major lacunae in the realm of theory, especially in regard to the integration of the role of the state or the analysis of forced migration. Without expanding existing theories further, we may discover that most our laws of migration have questionable validity.

References

Abella Manolo, 1994. Introduction. Turning points in labour migration, *Asia and Pacific Migration Journal*, vol. 3(1), p. 1–6.

Böhning W. Roger, 1984. *Studies in International Labour Migration*. London, Macmillan, 318 p.

Borjas George J., 1987. Self-selection and the earnings of immigrants, *American Economic Review*, (Nashville, TN), vol. 77(4), p. 531–553.

Borjas George J. and Bronars Steven G., 1989. *Immigration and the family*. Unpublished document.

Dowty Alan, 1987. *Closed borders*. New Haven, CT, Yale University Press, 270 p.

Fields Gary S., 1994. The migration transition in Asia, *Asian and Pacific Migration Journal*, vol. 3(1), p. 7–30.

Grasmuck Sherri and Pessar Patricia R., 1991. *Between two islands: Dominican international migration*. Berkeley and Los Angeles, University of California Press, 280 p.

Gurak Douglas T. and Caces Fe, 1992. Migration networks and the shaping of migration systems, *in*: Mary M. Kritz, Lin Lean Lim, and Hania Zlotnik (eds.), *International migration systems: a global approach*, p. 150–176. Oxford, UK, Clarendon Press, 354 p.

Harris John R. and Todaro Michael P., 1970. Migration, unemployment, and development: a two sector model, *American Economic Review* (Nashville, TN), vol. 60, p. 126–142.

KRITZ Mary M., 1995. *Population growth and international migration: is there a link?* New York, The International Center for Migration, Ethnicity and Citizenship, New School for Social Research, 50 p. (Migration Policy in Global Perspective Series, occasional paper no. 1.)

KRITZ Mary M., 1996. *U.S. investment, population growth and GNP as determinants of emigration to the USA*. Unpublished document, 33 p. (Paper presented at the Annual Meeting of the American Sociological Association, New York City, New York, August 1996.)

KRITZ Mary M., LIM Lin Lean, and ZLOTNIK Hania (eds.), 1992. *International migration systems: a global approach*. Oxford, UK, Clarendon Press, 354 p.

LEE Everett S., 1966. A theory of migration, *Demography*, vol. 3(1), p. 45–47.

MABOGUNJE Akin, 1970. Systems approach to a theory of rural-urban migration, *Geographical Analysis*, vol. 2(1), p. 1–18.

MARTIN Philip L., 1991. *The unfinished story: Turkish labour migration to Western Europe*. Geneva, International Labour Office, 50 p.

MASSEY Douglas S., 1994. International migration theory: the North American case, *Population and Development Review* (New York), vol. 20(4), p. 699–751.

MASSEY Douglas S., ARANGO Joaquín, HUGO Graeme, KOUAOUCI Ali, PELLEGRINO Adella, and TAYLOR J. Edward, 1993. Theories of international migration: a review and appraisal, *Population and Development Review* (New York), vol. 19(3), p. 431–466.

MINCER Jacob, 1978. Family migration decisions, *Journal of Political Economy* (Chicago, IL), vol. 86(5), p. 749–773.

PAPADEMETRIOU Demetrios G. and MARTIN Philip L.D., 1991. *The unsettled relationship: labor migration and economic development*. New York, Greenwood Press, 336 p.

PETRAS Elizabeth M., 1981. The global labor market in the modern world economy, *in*: Mary M. Kritz, Charles B. Keely, and Silvano M. Tomasi (eds.), *Global trends in migration: theory and research on international population movements*, p. 44–63. Staten Island, NY, Center for Migration Studies, 465 p.

PIORE Michael J., 1979. *Birds of passage: migrant labour in industrial societies*. Cambridge, UK, Cambridge University Press, 229 p.

PORTES Alejandro, 1995. Economic sociology and the sociology of immigration: a conceptual overview, *in*: Alejandro Portes (ed.), *The economic sociology of immigration: essays on networks, ethnicity, and entrepreneurship*, p. 1–41. New York, Russell Sage Foundation, 328 p.

PORTES Alejandro and BACH Robert L., 1985. *Latin journey: Cuban and Mexican immigrants in the United States*. Berkeley, Los Angeles, University of California Press, 432 p.

RAVENSTEIN Edward G., 1885. The laws of migration, *Journal of the Statistical Society*, vol. 48, p. 167–227.

RAVENSTEIN Edward G., 1889. The laws of migration, *Journal of the Statistical Society*, vol. 52, p. 241–301.

RICHMOND Anthony, 1995. The environment and refugees: theoretical and policy issues, *Population Bulletin of the United Nations* (New York), no. 39, p. 1–17.

SASSEN Saskia, 1988. *The mobility of labor and capital*. New York, Cambridge University Press, 236 p.

SASSEN Saskia, 1991. *The global city: New York, London, Tokyo*. Princeton, Princeton University Press, 412 p.

SJAASTAD Larry A., 1962. The costs and returns of human migration, *Journal of Political Economy* (Chicago, IL), vol.. 705, p. 80–93.

SKELDON Ronald, 1997. *Migration and development: a global perspective*. Harlow, UK, Addison Wesley Longman, 256 p.

STARK Oded, 1991. *The migration of labor*. Cambridge, MA, Basil Blackwell, 406 p.

STARK Oded and TAYLOR J. Edward, 1989. Relative deprivation and international migration, *Demography* (Washington, DC), vol. 26(1), p. 1–14.

STOUFFER Samuel A., 1940. Intervening opportunities: a theory relating mobility and distance, *American Sociological Review*, vol. 5, p.845–867.

STOUFFER Samuel A., 1960. Intervening opportunities and competing migrants, *Journal of Regional Studies*, vol. 2(1), p. 1–26.

TAYLOR J. Edward, 1992. Remittances and inequality reconsidered: direct, indirect, and intertemporal effects, *Journal of Policy Modeling*, vol. 14, p. 187–208.

TODARO Michael P., 1976. *Internal migration in developing countries*. Geneva, International Labour Office, 106 p.

UNITED NATIONS, 1998. *World population monitoring, 1997. International migration and development*. New York, United Nations, 204 p. (United Nations publication sales no. E.98.XIII.4.)

WALLERSTEIN Immanuel, 1974. *The modern world system, capitalist agriculture and the origins of the European world economy in the sixteenth century*. New York, Academic Press, 419 p.

ZELINSKY Wilbur, 1971. The hypothesis of the mobility transition, *Geographical Review*, vol. 61(2), p. 219–249.

ZOLBERG Aristide R., 1981. International migrations in political perspective, *in*: Mary M. Kritz, Charles B. Keely, and Silvano M. Tomasi (eds.), *Global trends in migration: theory and research on international population movements*, p. 3–27. Staten Island, NY, Center for Migration Studies, 465 p.

ZOLBERG Aristide R., SUHRKE Astri, and AGUAYO Sergio, 1989. *Escape from violence: conflict and the refugee crisis in the developing world*. New York, Oxford, Oxford University Press, 400 p.

PART

II

THE PRINCIPAL DETERMINANTS OF MIGRATION

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

The review of theory in Chapter 59 provides the basis for identifying the main determinants of migration dynamics. Historical population growth inequalities that reflect the economic growth and poverty differentials affecting most of the world's population are still among the main reasons why a growing number of individuals move from poor countries to economically more developed ones. Radical changes in economic systems and the balance of international political power between developed countries prompted an overhaul of internal and external economic policies, which caused a backlash of economic crises in some developing countries that seemed poised in the 1970s and 1980s to make an early exit from poverty.

Economics was and remains a key variable in international migration dynamics. In Chapter 60, Marc Termote sets out to clarify the fundamental importance of these variables and to provide a general theoretical framework for analysis of the different kinds of migration that are found and to isolate the main methodologic implications for each.

Although many parallels are readily found in the theoretical concepts between this and the previous chapter, both were included despite the inevitable overlap, because they set the theoretical foundations of economic migration in very different contexts. Chapter 59 dealt with the theories of international migration only on a comparative basis of economic versus more sociologically or culturally oriented theories. Chapter 60 considers only economic migration, but it is confined neither to the theoretical aspects nor international migration alone, dealing also with internal migration.

The introduction to this volume referred to the importance of urbanization as a factor of spatial mobility. This process may over the centuries have been responsible for more population shifts than all international and inter-regional migration movements together. Denise Pumain addresses this idea in Chapter 61, working from a suggested definition of rural and urban areas to concepts of counting the urban population in developed and Third World countries, followed by an analysis of some aspects of the phenomenon that depart from economic theories of migration that are likely to explain away urban development as a simple aggregation of individual migration decisions.

Political migration is a special case within the general situation of international migration, but almost all of the host countries acknowledge it a legitimacy often denied to other forms of migration. During most periods in the history of migration, repression, war, genocide, and a wide range of political factors have resulted in forced movements of millions of people. In Chapter 62, Luc Legoux examines the roots of such movements, discusses the scale and reasons for their development worldwide, and assesses future prospects. He defines and clarifies the distinction between asylum seekers and refugees, and he shows how the official categorization as an asylum seeker or refugee can be influenced by migration policies. Tightening of regulations for economic migration prompts some potential migrants to seek entry as asylum claimants, and restrictive entry policies result in asylum being refused to claimants genuinely at risk in their countries of origin.

Analyzing the origins of migration behaviors requires consideration of the contexts in which individuals live, including the demographic and social characteristics, and consideration of all possible interactions among the different variables in the equation. Daniel Courgeau and Éva Lelièvre, who co-authored Chapter 63, analyze the determinants of migration through this prism, and they expand the field of study, moving from the macro approach at the aggregate level to the micro approach at the individual level. By combining these two approaches, they produce a theory of migration behaviors that brings into sharper focus the roles of the different levels of aggregation.

C H A P T E R

60

The Economic Determinants of Migration

MARC TERMOTE

National Institute of Scientific Research, University of Quebec, and Department of Demography, University of Montreal, Canada

The aim of this chapter is threefold: to present the fundamental significance of migration within the workings of the economic system, to construct a general theoretical framework that will make it possible to analyze the contribution of various migration theories and models already in existence, and to establish the main methodologic implications of this second analysis.

In economic theory, migration is a condition on which the system's optimal performance depends. This is demonstrated by two theoretical approaches, each representing one side of a single mirror: trade theory, which studies optimal trade flows between given production locations, and localization theory, which studies the optimal localization of production activities. Although migration appears to be a necessary process in finding economic equilibrium within a given space (whether defined in terms of regions or countries), it does not follow that this equilibrium is necessarily reached. A succinct presentation of the workings of a spatially defined labor market helps to demonstrate this.

This spatial version of the labor market model is enhanced to take into account the set of economic factors determining migratory movements. Rather than enumerating the various migration models and theories, which would only be incomplete and fastidious, it seems preferable to construct a general framework that makes it possible to identify each model's contribution and clarify its meaning.[1] Instead of a critical review of the various existing models, this chapter offers a theoretical grid of analysis of these models and theories. The methodologic implications of these approaches are discussed.

I. MIGRATION AS AN EQUILIBRIUM PROCESS

"Although the resource of emigration seems to be excluded from such perfect societies as the advocates of equality generally contemplate, yet in that imperfect state of improvement, which alone can rationally be expected, it may fairly enter into our consideration", wrote Malthus 2 centuries ago (1803, Book III, Chapter IV) in his famous *Essay on the Principle of Population*.

It is not really because our societies are in an imperfect state of improvement that migration needs to be introduced into the demographer's or the economist's language, but because migration represents a necessary process in spatial economic equilibrium. For too many years, economists, like most other researchers in

[1] This generalization encompasses countries said to be underdeveloped as well as developed countries. Very few models have been elaborated specifically for underdeveloped countries; Todaro's model (1980) is the most notable exception. In most cases, a model designed for developed countries has been applied to a context of underdevelopment.

the social sciences, have considered migration between countries or within a single country to be the result of a malfunction of the economic social system. The migrant has been seen as an outcast, as the reflection of some unbalance; on the contrary, migration constitutes, according to economic theory, a necessary condition to the optimal functioning of the economic system.

There are two ways of attaining spatial economic equilibrium: by the localization of production factors or by trading goods (material and immaterial) produced by these factors. Economists have mainly studied the second process. By developing trade theory, they tried to bring to light the optimal flows of goods and services between given localizations of production factors, but they rarely studied the question of the optimal localization of production factors. In other words, they postulated that an economic spatial optimum would automatically be attained simply by trading goods, without it becoming necessary to delocalize production factors, that is without having to envisage the migration of these factors.

Trade theory itself demonstrates this postulate's invalidity and therefore, the need for the migration of production factors. For the simple trading of goods to cause the equalization of the remuneration of production factors and thereby eliminate the need for any factor migration, the Hecksher-Ohlin theorem conditions, as made explicit by Samuelson, must be filled (Samuelson, 1949; Ohlin, 1935; Heckscher, 1919). This entails (1) that demand functions be homogeneous and identical in all places (i.e., identical tastes and consumer preferences, whatever the place of residence); (2) that mobility (e.g., by sector, by profession) within each spatial unit be perfect and without any cost attached; (3) that perfect competition and full employment reign everywhere; (4) that, for any given good, the production function be linear and homogeneous (with no economies of scale) and identical in all places (i.e., production conditions should be the same everywhere); (5) that the marginal productivity of factors be always positive and decreasing; (6) that there be no "excessive" specialization anywhere; (7) that the trading of goods be completely free; and (8) that the transportation costs of goods (in the broadest sense) be nil.

If all these hypotheses are simultaneously respected, the localization of production factors is of little importance, because the simple act of trading of goods and services will make it possible to reach spatial economic equilibrium and the spatial equilibration of factor remuneration. Migration is of no use in such a world because there are no remuneration disparities. Unfortunately, all of these hypotheses are unrealistic. The eighth hypothesis even leads to a sort of tautology. If we suppose the absence of any transportation costs, of any spatial friction, we negate the existence of space. In that sort of nonspatial world, the question of the delocalization of factors (i.e., the migration of labor and capital) becomes superfluous.

Rather than studying spatial economic equilibrium through the trading of goods between given locations (as in trade theory), it is possible to touch on the question by determining these locations themselves by explicitly introducing space and transportation costs. This is precisely the aim of localization theory, which provides a second demonstration of the need for migration as a process for reaching spatial economic equilibrium using a Löschian framework.

August Lösch[2] begins by posing a series of hypotheses concerning spatial uniformity. Production factors, population and natural resources, are evenly distributed across space; the transport conditions are always the same (i.e., the cost of transporting one unit of goods over one unit of distance is the same whatever the place of origin, and this rate does not depend on the overall distance); and the production and consumption functions are the same at every point.

The key feature is that, despite these hypotheses concerning spatial uniformity, according to Lösch's theory, a spatial concentration of production in the form of a hierarchical pyramid of *central places* is reached by taking transportation costs into consideration. This way of thinking leads to a contradiction. The population in its role as consumer is distributed evenly over space, but this same population in its role as producer of goods is concentrated in a hierarchy of central places. The only way of resolving this contradiction is to introduce migration, but by breaking the spatial uniformity of consumption, this reinforces the movement toward the concentration of production.

If instead trade theory proves the need for production factors to migrate, a final question remains: Can the equalization of the remuneration of the labor factor not be obtained by exchanging goods along with a capital transfer? The mere migration of capital (i.e., without labor migration) is not enough for at least three reasons. First, several types of capital exist. Doubtless, *liquid capital*, or financial capital, is very mobile, but like all production factors, it flows toward places where remuneration is highest. This optimal localization is seldom a place with low wages (e.g., even if the interest rate is sometimes high, the real return is often quite low, possibly even negative, due

[2] An English translation of Lösch's book (*Die räumliche Ordnung der Wirtschaft*, 1940) was published in 1954 (*The Economics of Location*, New Haven, CT, Yale University Press, 520 p.)

to uncertainty, devaluation risks), and it can vary quickly and in such a way that capital will seldom be reinvested to become a true production factor. *Immobilized capital* is made up of equipment (it is then difficult to move, and when it is moved toward a poor destination, its use often requires the parallel displacement of qualified workers) or human capital. In both cases, labor migration becomes necessary.

Human capital (e.g., scientific and technologic, progress of knowledge, know-how) leads us to the second reason, which takes into account the dynamic character of the economic process. This type of production factor is increasingly important, and capital transfers in their traditional form cannot replace them, except possibly in the long term. A third reason is related to the existence of production factors that are immobile or those that can be moved only at great expense. More precisely, the link between labor and land or natural resources must be considered. Capital migration cannot replace low reserves of natural resources, except in a few specific cases.

The fact that labor migration is a necessary process in the optimal running and equilibrium of the spatial economic system does not, however, mean that it necessarily leads to equilibrium. Contrary to the beliefs of neoclassic thinkers, economic theory provides no *a priori* solution concerning the effect of economic migration.

Most classic 19th century theoreticians thought migration represented an economic equilibrium process in so far as the subjacent fall in labor supply that it implied for the place of emigration caused an increase in this factor's remuneration. "All possible help should be accorded these poor unfortunates who . . . sacrifice themselves for their brothers by distancing themselves from them," affirmed Sismondi in 1817 in his *New Principles of Political Economy, or Wealth in its Relationship with Population*.[3] For John Stuart Mill (*Principles of Political Economy*, 1848), migration is a means of fighting the tendency of diminishing returns of the land, it is a remedy for low wage rates and for the decline in the profit rate ("the more . . . we send away, the more we shall possess and be able to retain at home").

Not everyone shares this opinion. Malthus remains consistent when, in his laws of population, he insists on the inadequacy of emigration as an answer to overpopulation. Even if emigration caused the wage rate to rise above subsistence level among the population who had stayed in the country, the subsequent natural population growth would soon neutralize the beneficial effects of emigration. Marx adopts this idea by stressing that the agricultural revolution and the natural growth rate of the non-emigrating population can neutralize the positive effects of emigration, as he points out was the case in Ireland, where the production of a *relative overpopulation* more than compensated the *absolute depopulation*.

Classic investigators were essentially concerned with the consequences of migration on the economy experiencing the emigration, and they thought that the advantage for the place of immigration (which was often a colonial country) was somehow self-evident. However, to study the role of migration as a "equilibrium engine," the effects of the place of emigration and immigration need to be studied simultaneously. To systematize this analysis, we can use a spatial model of the labor market.

Let us consider a place (region or country, depending on whether it is an internal or international migration) named *A*, where labor supply by workers is low and very sensitive to wage levels (because there are few workers and there is no labor reserve) and where labor demand is high and equally sensitive to wage levels (because there are quite a few production units and they are all highly capitalized). In this case, the meeting point between the labor supply curve (which has a positive slope, because the higher the wages, the greater the number of workers offering their production capacities) and the labor demand curve (with a negative slope) fixes a high wage level. Consider a second place, *B*, which is densely populated but has relatively few production units and where the meeting point of the supply and the demand of labor leads to a much lower level of wages. If the difference is significantly higher than the cost of migrating between these two localities, workers will emigrate from *B* to *A*.

According to neoclassic theory, this migration will lead to decreased wages in *A*, because the arrival of immigrant labor implies a rise in labor supply; in other words, the curve representing labor supply will cross the labor demand curve at a much lower point. In *B*, the decrease in labor supply resulting from this emigration will cause an increase in wages. Overall, the level of wages will rise in the poorest place and fall in the richest place. Migration from *B* to *A* will continue for as long as the wage gap is higher than the cost of migrating. Migration will have made the equalization of wages possible; it will have fueled an equilibrium process within the interregional and international economic system.

Many theories and models have been inspired by and continue to be based on this framework, and despite 200 years of international migration (some

[3] Nouveaux principes d'économie politique, ou de la richesse des nations dans ses rapports avec la population.

massive) and almost constant rural depopulation, what is observed is the persistence of or increase in wage differences between rich and poor regions or countries. Many elements help to explain the classic model's lack of relevance as far as migration is concerned (the introduction of technologic progress with the aim of producing the needed dynamization of the previously mentioned theory, is not the least of these elements). However, one fundamental weakness invalidates the classic framework: the nonintegration of the delocalization of consumption resulting from migration.

In the neoclassic model, migration is only seen as a means of delocalizing labor, and its effect on wages reflects only to the displacement of the labor supply. However, the migrant is both a producer and a consumer. He is a producer because he is a consumer. The migrant who moves from a poor locality to a rich one contributes to an increase in labor supply and to the rise in the demand for goods and services and therefore to a further increase in labor demand on the part of firms producing these goods and services. In the emigration site, the fall in labor supply is accompanied by a fall in the demand for goods and services and by a fall in labor demand.

Because of this double delocalization of production capacity and consumption capacity, to determine the economic consequences of migration on the emigration and immigration sites, it is not enough to state the production functions (i.e., technical production conditions) in each of the locations. It is also necessary to consider the consumption functions (without forgetting that these functions are not usually the same for migrants and nonmigrants) and to establish the relationship between these consumption functions and labor demand that will make it possible to fill consumption demand.

According to the framework described, there will be an increase in labor supply in the place of immigration, and there will be an increase in labor demand after an increase in consumption linked to immigration. It is impossible to know beforehand whether this supplementary labor demand will be greater than or less than the supplementary labor supply resulting from the immigration. It all depends on the immigrants' consumption function (more precisely, on their propensity to consume local goods and services), the relationship between the demand for goods and labor demand, and on the level and elasticity of labor supply and demand compared with wage levels. The same can be said of emigration. In other words, it is impossible to determine the effects of migration on the places of emigration and immigration without knowing the consumption and production functions, as well as the functions for labor supply and demand.

Given these conditions, migration can be a disequilibrium process for the international and interregional economic system, as well as an equilibrium process. It can reduce spatial wage disparities, but it can also increase them. The theory and the model now appear undetermined.

This theoretical conclusion is accompanied by an important methodologic implication. Migration as an answer to a spatial disparity can aggravate the condition as easily as it can reduce it. An analysis of the causes of migration runs the risk of being biased if it does not simultaneously include a study of its consequences. In other words, migration is both a dependent and an independent variable, and an analyst should be wary of models based on a single equation. Only a system of simultaneous equations makes it possible to obtain theoretically well-founded results.

This interdependence needs to be studied within a spatially disaggregated system, making it possible to identify the consequences of migration, wherever they may be. This is particularly important in the case of international migration. If we study the effect of international immigration at the global national level, we run the risk of underestimating it, because international immigration is most often concentrated in a few metropolitan regions. Almost all studies concerning the economic consequences of international immigration indicate a negligible effect or none at all, at least for the receiving country. This finding reflects the fact that the studies have been carried out within a spatial framework that is far too wide and in which the consequences are diluted. The causes and consequences of migration need to be analyzed at the level of the spatial units, where the phenomenon manifests itself.

II. IN SEARCH OF A GENERAL THEORETICAL FRAMEWORK

The conclusion that migration does not necessarily lead to spatial equilibrium was based on a simplified model in which only worker migrations exist and in which the only determinant of these migrations is wage levels and their only consequence is to modify wages. This framework needs to be extended.

In addition to wage levels in each location, the model must consider the probability of finding a job (which will make it possible to take advantage of these wages) or the unemployment rate. However, it is not enough to compare the present level of wages (i.e., income thesis), weighted by the probability of receiving a salary (i.e., job vacancy thesis). The migrant must also bear in mind the likelihood of being able to count on this salary in the future; he will estimate a discounted value of this weighted salary, taking into account the anticipated evolution of wages and unemployment. This approach characterizes Todaro-type models, developed to analyze internal migrations (rural-urban) in Third World countries; curiously, it has seldom been adopted for use on industrialized countries.

The future value of wages is discounted because migration is seen as an investment whose gain should be maximized. The length of time during which this investment can be made profitable is important. This naturally leads to the introduction of age as a determinant of migration. An elderly worker will not have much time to take advantage of his migratory investment compared with a young worker, who will have his entire working life to profit from migration or to correct the effects of a bad migratory calculation by a new migration.

This first addition to the model concerns only worker migrations that are exclusively motivated by wage levels (weighted against unemployment levels). To continue enriching the model, factors linked to the individual seen as a consumer need to be added. When the price of goods and services varies significantly over distance, wage rates need to be deflated to take into account disparities in the cost of living, established in real wage terms rather than in monetary wage terms.

In the same vein, local, regional, and national (in the case of international migrations) tax levels can also be pertinent. It is, however, important to tread carefully because wage levels may already incorporate this type of factor, in the same way that they may already take price levels into account.

For the same worker-consumer factor, it makes sense to consider the availability and the cost of some special goods and services, such as *land*, *housing*, and *public services*. Similarly, immaterial goods, such as *environmental and climatic conditions*, or even certain negative goods, such as *crime rates and pollution*, may be taken into account by some migrants. It is necessary to check that wage levels do not already reflect the level of housing costs, the availability of public services, or even the presence of negative goods.

Taking into account these nontraded goods leads to a new theoretical approach. The initial assumption in the neoclassic model (and therefore for all generalizations in this model) is that migration is a response to disequilibrium. Toward the end of the 1970s, when the evaluation of the environmental costs of large industrial projects became a research priority, a key question was whether a theoretical, but empirically verifiable, link existed between the value of compensation offered to residents negatively affected by such projects and the amplitude of migrations that might result from these negative effects. If compensation is insufficient, migration will ensue. However, under the terms of traditional microeconomic theory, measuring compensation requires the hypothesis of initial equilibrium (before the shock caused by the project) and a subsequent equilibrium, after the agents have adapted their behavior to the modifications caused by the project.

Based on this new approach, spatial wage disparities are presumed to be equilibrium disparities (i.e., the difference between wages is explained entirely by migration costs) or possibly compensatory disparities for the presence of nonmarket goods (i.e., *location-specific amenities*), whether they are positive goods, such as a pleasant climate, or negative goods, such as pollution. This sort of spatial wage disparity should not cause migratory movements. Migration should happen only when equilibrium is broken to correct the effects of a previous bad migratory calculation by making a new migration.

Location-specific amenities cause migration. Variations in demand can come from a change in income or a modification of the relative prices of certain market goods that can neutralize the effects of nonmarket goods (e.g., a fall in the cost of air conditioning units in regions with hot, humid climates). The only way for an individual or a household to satisfy a change in the demand for such goods is to migrate. The intensity of migratory relations between two localities does not, however, depend on the absolute difference between the exogenous variables being studied, but on changes in the absolute value of exogenous variables that produce a modification in the demand for nonmarket goods.

The demand for nonmarket goods that makes certain locations particularly attractive is closely linked to individuals' personal characteristics and the stage reached by them in the life cycle. The availability of microdata becomes an important condition for carrying out empirical studies based on this approach. Because this type of data remains uncommon, the use of this new theoretical approach is not yet widespread despite promising initial results.

Because migration has a cost, there will always be spatial disparities between wages, even in equilibrium situations. However, this cost of migration covers more than the cost of the simple physical displacement of people. Other than the *monetary cost* associated with displacement (of the migrant and his belongings), migration also has a *temporal cost* (the preparation of a migration takes time; for example, the person's housing needs to be sold and new housing found at the immigration point) and a *psychological cost* (i.e., linked to the loss of a support network of family and friends or difficulties in adjusting to the new environment, to cultural, linguistic, or even religious differences that need to be faced).

An often-neglected cost is that linked to the process of collecting information concerning utilities or disutilities offered by various points in space. This *information cost* depends on individuals' locations (e.g., a migrant candidate living at a spatial system's center of gravity will have access to better quality information at a lesser cost than a person living on the outskirts). Regardless of the initial localization, there will usually be a considerable quantitative and qualitative inequality between the information concerning the point of departure and the information relative to the destination point. The use of this asymmetric information is a relatively new and useful way of understanding a system of migratory flows.

Because information is expensive and takes time to collect, the migratory answer to a modification in the socioeconomic context will become noticeable only after a certain *time lag*. The migratory reaction of individuals is characterized by a strong dose of inertia. (Adam Smith [1776] remarked that "a man is, of all sorts of luggage, the most difficult to be transported"). It becomes necessary to enrich the model by introducing *explanatory variables* for migration integrating a time lag, with migration observed at *t* being, for instance, a function of the wage rate and the unemployment rate observed at *t-n*. Because the lag (value of *n*) is seldom longer than 1 or 2 years (several models obtain highly satisfactory results using a lag expressed in terms of a few months), this type of development requires chronologic series based on population registers or administrative files, which very few countries can provide.

After a certain number of people have emigrated, information flows from the place of immigration to these migrants' place of origin become richer, more trustworthy, and less expensive. A chain reaction of migrations takes place. One method sometimes used to study this "multiplier effect of parents and friends" (Nelson, 1959) consists of making the migration observed in *t* a function of the migration observed in *t-n*. We should be wary of this approach because we run the risk of uncovering only the temporal stability of migratory flows by using this type of model specification. We find that the model is more descriptive than explanatory.

Having defined a first series of generalizations linked to the migrant in his production role, a second series of generalizations linked to the migrant in his consumer role, and a third linked to the generalization of the spatial system to a system in equilibrium rather than in disequilibrium, this fourth series of generalizations, which refers to the various cost dimensions of migration, turns out to be particularly rich. However, faced with the multiplicity of the different costs that have an impact and the difficulty in obtaining adequate statistical data, we can try to simplify the problem by using a single variable capable of summarizing to some extent all those linked to migration costs.

The various migration costs are related to *distance*. It is understandable that a considerable investment has been made to try to uncover the mathematical function that best expresses the relationship between the intensity of migratory movement and distance. In this regard, we cannot ignore the considerable success obtained by the *gravitational approach*. Inspired by Newtonian physics, this type of model takes as its starting point the idea that the number of migrants moving between two places is a direct function of the mass (i.e., population) present in each of these places and is inversely proportional to the distance separating them.

The key problem resides in the definition of these masses. When talking about the migratory relationship between two places of population rather than the relationship between two celestial bodies, this mass needs to be specified in economic terms and not in physical terms. It becomes necessary to weight the numbers using the level of income, the unemployment rate, and other factors. By doing this, we are led back to the classic framework, as presented previously, in such a way that we could question the theoretical significance of this type of approach (Le Bras, 1990).

The reference to the law of gravity allows us to introduce a fifth type of generalization. Newtonian physics does not study the relationship between two specific masses, but rather a *system* of masses. Until now, we have thought only about the migratory relationship between two places (i.e., regions or countries). However, a candidate for migration does not make his migratory decision based solely on one possible destination. He studies a system of locations and chooses the one that will most easily provide employment opportunities or a higher income (i.e., Zipf's law of least effort [1949], which directly inspired several migration models). The key feature is not so much the distance separating two places, *i* and *j*, but the set of opportunities offered between these two places. The more opportunities that are lost in *i*, the more candidates for emigration among the inhabitants of this place and the more migrants there will be from *i* to *j*. A candidate emigrating from *i* is not alone in trying to improve his life by leaving *i*; he finds himself competing with other migration candidates (including some in a better position because they are closer to *j*), who are also looking for employment or higher wages. The combination of these various ideas led Stouffer to his celebrated model of *intermediate opportunities and competing migrants*.

The problem with this model resides in the difficulty of measuring the number of opportunities lost and offered. The simplest solution (and it is also the one adopted by empirical applications) consists in approximating the number of lost opportunities by the number of emigrants and the number of opportunities offered by the number of immigrants. After all, if the number of emigrants from a given place is high, it is because the number of jobs lost (or jobs at low wages) is high, and if there are many immigrants in this place, it is because the number of employment opportunities there is high. By proceeding in this manner, we are reduced to explaining migration by migration.

This type of circular reasoning is common in migratory analysis. It can be found in a number of theories on rural depopulation that aim to explain rural-urban migration by the attractiveness of cities, although that fact is precisely what needs to be explained. It is true that Stouffer's model has the advantage of being easy to apply while giving impressive empirical results. The coefficient of determination is often close to 0.9, which is rare in social science. Moreover, the model makes it possible to estimate origin-destination flows based on the total emigration and immigration flows for each place. However, we are far from an explanatory analysis.

Moreover, in this model of intervening opportunities and competing migrants, the spatial system only functions on a local basis (for each origin-destination pair). The model does not cover all flows in the entire system simultaneously. It does not take into account the *interdependence between various* regional labor *markets*. To cross into this new area of generalization, we need to use matrix formulation. This provides an immediate advantage over the statistical formulation implicit in the models discussed previously (i.e., models in which the migratory variable depends on a series of independent variables) because it forces the user to respect a certain number of accounting identities. The most obvious identity is the requirement that the sum of all regional migratory balances in a closed system must be nil (i.e., migration is a zero-sum game; what one place loses, another gains).

This *zero-sum constraint* implies that in a statistical analysis of interregional migratory flows, the sum of each variable's coefficients, as well as those of the regression constants, should be equal to zero when the various equations are all added up (one equation per region). This condition is necessary for the sum of the estimated values for the dependent variable (migratory flows from *i* to *j*) to be nil for the overall regional system. Few models respect this constraint, but neglecting it can lead to serious errors of econometric inference. This is all the more astonishing because the procedures for resolving the equations have existed for decades.

We can also study the entire space by constructing a *composite indicator of attractive and repulsive forces* that effects migration from *i* to *j*; this is a measurement in which the value of one variable in each place is weighted by the distance between this place and each of the other places. This weighted value is sometimes summed for all the points of origin (i.e., the global attractive force of *j*, or the relationship between *j* and the rest of the spatial system) and sometimes for all the destinations (i.e., the repulsive force of *i*, or at least the relationship between *i* and the rest of the system). This approach, suggested by Alonso (1978), adopts an idea developed by the gravitational school, the concept of potential, and is a part of the *family of spatial interaction models* (Wilson, 1970) that incorporates systemic variables of the same nature.

We still need to examine a sixth area of generalization concerning the basic framework. Until now, we have implicitly supposed that all migrations were labor factor migrations and that the population concerned was homogeneous.

The introduction of the cost of migration into the model (and thereby the approach by which migration is a sort of investment in space) naturally leads to the rejection of the homogeneity hypothesis, because the migrant's age (and therefore the length of time over which the migrant will be able to write off the cost of his migration and gain a profit from his investment by receiving higher wages) is a fundamental determinant in the migratory calculation. Likewise, the level of education determines information costs for utilities and disutilities offered by each place in the spatial system. Not every person migrating for economic reasons necessarily participates in the labor market, as is the case for members of a household accompanying a migrant worker or for migrations of retired people.

These considerations lead us to take account of the *life cycle*. This concept is not new, but its application to the study of migrations is relatively recent, although Hägerstrand (1963) did use the *life line* concept at the beginning of the 1960s to study migrations in Sweden. Although the different stages of the life cycle (i.e., marriage, divorce, finishing education, labor market entrance, giving birth, education and departure of children, and retirement from activity) are crucial to migratory decision, their importance was ignored for many years. The first models explicitly incorporating life cycle stages were formulated at the end of the 1970s. Three reasons explain this neglect.

The first reason is related to the fact that the necessary microdata have been collected for only a decade or two and that the collection of microdata implies costly surveys. We should, in this regard, underline the pioneering work carried out in France in 1981 by Daniel Courgeau and his team in their survey of "family, professional, and migratory biography" (Courgeau, 1985).

The second reason is methodologic. The biographic analysis raises the problems of the interaction between phenomena and the homogeneity of the population that the traditional demographic analysis tended to brush aside and to which are added the problems inherent in all retrospective surveys (e.g., incorrect or incomplete data) and the search for appropriate econometric techniques for microdata (e.g., probit or logit analyses).

A third reason is of a theoretical nature. The introduction of life cycle stages transfers the decision-making unit for migration from the individual to the household. The study of the economic causes and consequences of migration have traditionally been formulated in a context of maximization of individual utilities (based on the usual vision of microeconomic theory). The change from the individual to the household considerably complicates the theoretical analysis and the resulting models. The same type of problems can be found when household forecasts are attempted rather than population forecasts.

The links existing between household members imply negative personal external factors that are generally internalized by the household and that normally tend to slow down migration. "Tied people" (Mincer, 1978) include those for whom the gains resulting from migration are dominated (in absolute terms) by gains (or losses) obtained by the other members of the household. Mincer showed that by increasing these links, the increased participation of women on the labor market leads to a fall in the propensity to migrate, while increasing the household's (matrimonial) instability. This instability encourages migration and female labor market participation.

The complexity that results from using the household rather than the individual as the basic unit can be seen in the case of interregional and international migrations and in the context of intraregional or urban mobility (metropolitan). The simultaneous determination of the optimal location for the household's (single) residence and the optimal location for each household member's place of work constitutes a considerable analytic challenge.

III. METHODOLOGIC IMPLICATIONS

We can identify a few methodologic lessons from the preceding theoretical analysis. A first conclusion concerns the way in which to define the migratory problem. Too often, we tend to analyze the causes and consequences of migration separately. However, what determines migration is affected by migration. Using a system of simultaneous equations is the only way of grasping the set of interdependencies between migration and socioeconomic variables. This conclusion is particularly valid when migration is measured over a relatively long period, as is the case for countries that do not have a register for recording migratory events.

When the migratory analysis is directed at a closed system, the sum of regional migratory balances is necessarily zero. It is important to introduce a constraint aimed at maintaining this accounting identity in the specification of any explanatory model of internal migration. Few models respect this requirement.

The study of the various enhancements needed in the basic model (obtained from the spatialization of the labor market) clearly shows that the number of socioeconomic factors that should be taken into account and the complexity of the links between them are so overwhelming that no model could possibly incorporate them all at the same time. In this regard, the specification of the migratory variable is at least as important as the choice of independent variables. The effects directly related to the measurement of migration often are neglected. Too much attention is paid to the specification of the independent variables, when the latter specification should be a function of the way in which the phenomenon to be explained is measured.

The classic example of this carelessness is the use of net migration. Many studies (including some recent ones) try to identify the socioeconomic determinants of regional migratory balances. This begs the question of what is the rationality behind the migratory calculation of a net migrant! What would be said of a theory and a model that tried to uncover the determinants of the balance of births and deaths? The use of the concept of net migration runs the risk of leading to rather dubious interpretations.[4] The only sensible analysis of migratory balances must be aimed at the macroregional level; by doing so, a region's (or country's) demographic-economic dynamism (or the lack of dynamism) is studied rather than individual migratory behavior.

The set of problems and the model need to be specified in different terms, depending on whether

[4] On this point, the article by Andrei Rogers (1990) significantly entitled *Requiem for the Net Migrant* is required reading.

migration is expressed in absolute numbers or rates. To obtain impressive results in terms of coefficients of determination, the absolute number of migrants should be chosen as the dependent variable, and the population of origin (and destination, if origin-destination migratory flows are being studied) will therefore be introduced as an explanatory variable. This said, all we have done is introduce a size effect (mass, as we would say in the gravitational approach), and we have barely moved beyond the descriptive level.

If origin-destination flows are used as the dependent variable, the contents of the "costs of migration" variable will be completely different compared with a model that is defined in terms of total emigration and immigration flows; the way in which the overall spatial system is taken into account will also be different. Likewise, the fact that migration is measured in terms of events (as provided by registers and administrative files) or in terms of people having experienced certain events (surviving migrants obtained in a census) is significant. In the explanatory analysis of migration, we tend to neglect the fact that migration is a renewable phenomenon. The existence of multiple migrations and return migrations, however, significantly complicates the interpretation of results based on data mixing various periods of migration together.

If no longitudinal microdata are available (obtained from retrospective surveys), the analysis in terms of life cycle cannot easily be developed (even if several authors have tried to proceed using period data). However, retrospective surveys that produce data covering the set of crucial stages in the life cycle are rare, not to mention seldom sufficiently trustworthy or statistically significant. The sample size is an obstacle often underestimated. To pass this obstacle and ensure the sample's homogeneity with regard to exogenous events that can affect the behavior of the sample's members, it is probably inevitable to limit the sample to one single cohort (as in the case of the previously mentioned French survey).

With the exception of the first conclusion, which states that the causes of migration can be studied only at the same time as their consequences, the various methodologic implications derived relate mainly to the specification of the dependent migratory variable. This does not occur by chance. Before drawing up a set of problems, before constructing a theory and specifying a model, it is first necessary to define what we intend to analyze. In the study of the causes and consequences of migration, the means used to measure migration and those used to explain it often are inadequate. The slow advancement made in the demographic analysis of migration is related to the difficulties and weaknesses associated with economic analysis in this field.

Acknowledgments

This chapter was translated from the French version by Jane Muir Dickinson.

References

Alonso William, 1978. A theory of movements, *in:* Niles M. Hansen (ed.), *Human settlements: international perspective on structure, change, and public policy*, p. 197–211. Cambridge, MA, Ballinger, 297 p.

Chalmers James A. and Greenwood Michael J., 1985. The regional labor market adjustment process: determinants of changes in rates of labor force participation, unemployment, and migration. *Annals of Regional Science*, vol. 19(1), p. 1–17.

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life-cycle: a French survey. *European Sociological Review*, vol. 1(2), p. 139–162.

Evans Alan W., 1990. The assumption of equilibrium in the analysis of migration and interregional differences. *Journal of Regional Science*, vol. 30(4), p. 515–531.

Graves Philip E., 1979. A life-cycle empirical analysis of migration and climate, by race. *Journal of Urban Economics*, vol. 6, p. 135–147.

Graves Philip E., 1983. Migration with a composite amenity: the role of rents. *Journal of Regional Science*, vol. 23(4), p. 541–546.

Graves Philip E. and Linneman Peter D., 1979. Household migration: theoretical and empirical results. *Journal of Urban Economics*, vol. 6, p. 383–404.

Greenwood Michael J., 1975. Simultaneity bias in migration models: an empirical examination. *Demography*, vol. 12(3), p. 519–536.

Greenwood Michael J., 1975. A simultaneous-equations model of urban growth and migration. *Journal of the American Statistical Association*, vol. 70, p. 797–810.

Greenwood Michael J., 1985. Human migration: theory, models, and empirical studies. *Journal of Regional Science*, vol. 25(4), p. 521–544.

Greenwood Michael J. and Hunt Gary L., 1984. Migration and interregional employment redistribution in the U.S. *American Economic Review*, vol. 74(5), p. 957–969.

Greenwood Michael J. and Hunt Gary L., 1984. Econometrically accounting for identities and restrictions in models of interregional migration. *Regional Science and Urban Economics*, vol. 14, p. 113–128.

Greenwood Michael J., Hunt Gary L., Rickman Dan S., and Treyz George I., 1991. Migration, regional equilibrium, and the estimation of compensating differentials. *American Economic Review*, vol. 81(5), p. 1382–1390.

Hägerstrand Torsten, 1963. Geographic measurements of migration. Swedish data, *in:* Jean Sutter (ed.), *Entretiens de Monaco en sciences humaines. Les déplacements humains*, p. 61–83. Paris, Hachette, 239 p.

Harrigan Frank J. and McGregor Peter, 1993. Equilibrium and disequilibrium perspectives on regional labor migration. *Journal of Regional Science*, vol. 33(1), p. 49–67.

Heckscher Eli F., 1919 (1949). The effect of foreign trade on the distribution of income, *Ekonomisk Tidskrift*, p. 497–517. Swedish translated into English: *in:* Howard S. Ellis and Lloyd A. Meltzer (eds.), *Readings in the theory of international trade*, p. 272–300, Homewood, IL, Richard D. Irwin, 637 p.

Knapp Thomas A. and Graves Philip E., 1989. On the role of amenities in models of migration and regional development. *Journal of Regional Science*, vol. 29(1), p. 71–87.

Krumm Ronald J., 1983. Regional labor markets and the household migration decision. *Journal of Regional Science*, vol. 23(3), p. 361–376.

Le Bras Hervé, 1990. Population et migrations, *in:* Xavier Greffe, Jacques Maresse, and Jean-Louis Reiffers (eds.), *Encyclopédie économique*, p. 1233–1273. Paris, Economica, 2188 p.

Malthus Thomas R., 1803 (1960). *An essay on the principle of population, or a view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*, 2nd ed. New York, The Modern Library, 1960, 352 p.

Mincer Jacob, 1978. Family migration decisions. *Journal of Political Economy*, vol. 86(5), p. 749–773.

Molho Ian, 1986. Theories of migration: a review. *Scottish Journal of Political Economy*, vol. 33(4), p. 396–419.

Mueller Charles F., 1982. *The Economics of Labor Migration: A Behavioral Analysis*. New York, Academic Press, 199 p.

Muth Richard F., 1971. Migration: chicken or egg? *Southern Economic Journal*, vol. 37, p. 295–306.

Nakagome Masaki, 1986. The spatial labour market and spatial competition. *Regional Studies*, vol. 20(4), p. 307–312.

Nakagome Masaki, 1989. Urban employment and the spatial structure of labor markets: an examination of the "Todaro paradox" in a spatial context, *Journal of Regional Science*, vol. 29(2), p. 161–170.

Nelson Phillip, 1959. Migration, real income and information, *Journal of Regional Science*, vol. 1(3), p. 43–73.

Ohlin Bertil, 1935. *Interregional and international trade theory*. Cambridge, MA, Harvard University Press, 617 p.

Polachek Solomon W. and Horvath Francis W., 1977. A life cycle approach to migration: analysis of the perspicacious peregrinator, *in*: Ronald Ehrenberg (ed.), *Research in labor economics*, vol. 1, p. 103–149. Greenwich, CT, JAI Press, 376 p.

Porell Frank W., 1982. Intermetropolitan migration and quality of life. *Journal of Regional Science*, vol. 22(2), p. 137–158.

Rogers Andrei, 1990. Requiem for the net migrant. *Geographical Analysis*, vol. 22(4), p. 283–300.

Samuelson Paul Anthony, 1949. International factor-price equalization once again, *Economic Journal*, p. 181–197.

Sandefur Gary D. and Scott Wilbur J., 1981. A dynamic analysis of migration: an assessment of the effects of age, family, and career variables. *Demography*, vol. 18(3), p. 355–368.

Sandell Steven H., 1977. Women and the economics of family migration. *Review of Economics and Statistics*, vol. 59, p. 406–414.

Schachter Joseph and Althaus Paul G., 1989. An equilibrium model of gross migration. *Journal of Regional Science*, vol. 29(2), p. 143–159.

Sismondi Claude-Charles de, 1817 (1971). *Nouveaux principes d'économie politique, ou de la richesse des nations dans ses rapports avec la population*. Paris, Delaunay, 2 vol. (Re-ed. 1971: Paris, Calman-Levy, 379 p.)

Sjaastad Lawrence, 1972. The costs and returns of human migration. *Journal of Political Economy*, vol. 70, p. 80–93.

Smith Adam, 1776 (1965). *An inquiry into the nature and causes of the wealth of nations*. London, George Routledge, 780 p. (Re-ed. 1937, 1965, New York, The Modern Library; 976 p.)

Stark Oded, 1991. *The migration of labor*. Cambridge (Mass.), Basil Blackwell, 406 p.

Stouffer Samuel A., 1960. Intervening opportunities and competing migrants. *Journal of Regional Science*, vol. 2(1), p. 1–26.

Stuart Mill John, 1848 (1996). *Principles of political economy with some of their applications to social philosophy*. (Re-ed. 1996, London, Routledge and Kegan Paul, 1166 p.)

Tabuchi Takatoshi, 1985. Time-series modeling of gross migration and dynamic equilibrium. *Journal of Regional Science*, vol. 25(1), p. 65–83.

Termote Marc, 1993. Causes et conséquences économiques de la migration internationale—théorie et réalité (Causes and consequences of international migration—theory and reality). *Études internationales*, vol. 24(1), p. 51–61.

Todaro Michael P., 1980. Internal migration in developing countries: a survey, *in:* Richard A. Easterlin (ed.), *Population and economic change in developing countries*, p. 361–402. Chicago, University of Chicago Press, 581 p.

Topel Robert H., 1986. Local labor markets, *Journal of Political Economy*, vol. 94, p. S111–S143.

Wilson Alan G., 1970. *Entropy in urban and regional modelling*. London, Pion, 166 p.

Zipf George Kingsley, 1949. *Human behaviour and the principle of least effort*. Cambridge, MA, Addison-Wesley Press, 573 p. (Re-ed. 1965, photomechanic reproduction of the 1949 edition, New York, Hafner Publishig company, 573 p.)

CHAPTER

61

The Urbanization Process

DENISE PUMAIN

Institut national d'études démographiques (INED), Paris, France

Urbanization is an important factor in the spatial mobility of populations. It is likely that urbanization has displaced many more persons than the interregional and intercontinental migrations that have gradually shaped the world's population settlements. Urbanization is a universal process that has modified and is still modifying the spatial distribution of population settlements in all parts of the world, and this process seems to be irreversible.

Even though cities began to appear several thousand years ago, their predominance, to the extent that they have become the most common form of human settlement, is a phenomenon that is only 2 centuries old. In developed countries, urban habitat has dominated for 50 years, and urban population now represents 80% of the population. According to estimates, in the year 2000, more than one-half of the world's population was living in cities. Historical and comparative analyses of urbanization processes reveal regular patterns, making it possible to predict in part future trends and their impact on population movements.

The urbanization process can be globally defined as the transformation of a rural type of settlement made up of villages—small settlements of no more than a few hundred or a few thousand inhabitants, located not far from each other, and spread over the territory and rather homogenous in terms of size—into an urban type of settlement, which is concentrated, organized on a hierarchical basis, and made up of cities—dense population clusters of very unequal size, from 1000 to 100,000 times larger than a village. The shift from one type of settlement to the other corresponds to the transformation of agrarian economies to industrial and tertiary economies. This change occurred relatively quickly on the broad scale of human history, although with variable intensities and durations, depending on the region of the planet.

The urbanization process led to a specific type of migration called *rural exodus*, whereby villages were drained of their populations by the cities. The urban system also encompasses more ordinary forms of migration, such as migrations between cities, a phenomenon that maintains differential growth rates and modifies, sometimes rather selectively, the population composition of a city, and migrations toward urban peripheries, which do not correspond to a return to the countryside but to spreading urbanized zones over the surrounding rural environment.

I. CITIES AND URBANIZATION

To measure and analyze urban migrations, clear definitions are needed for *city* and *countryside* in terms of rural space and urban space. Practically and theoretically, it is a difficult question.

1. Interpreting Urban Phenomena

A definition of the city must be based on a theoretical interpretation of the urbanization process. Cities, however, are complex social and geographic entities, and it is difficult to confine them to a global definition. Urbanization is a multifaceted phenomenon that cannot be reduced to what the theoretical apparatus of a single discipline has to offer. Urban phenomena can be considered from anthropologic, cultural, economic, geographic, political, legal, religious, and social viewpoints, and depending on the choice, the definition of the city can vary considerably.

From a demographic point of view, the city is a large and extremely dense population cluster, permanently grouped together on a limited area. The city also is defined morphologically by the permanence and continuous nature of its constructions and by the presence of urban planning rules underlying their organization. Such a conception implies that within the limited space it occupies, the city is unable to produce the total amount of food needed to feed its inhabitants. Their survival depends on the exchange of the goods and services they produce for the necessary agricultural products. The result is a functional, social, and economic definition of the city as a center of nonagricultural activity and innovation through the development of a more or less complex social division of labor; this enables it to support itself and grow because of a diversified economic basis.

Exchanges between city and countryside are always unequal, because a much higher value is attributed to manufactured products than to food products (Camagni, 1993; Braudel, 1979). The advantage held by cities is linked to their legal status as centers of power. Whether religious or political, this power always goes along with economic and territorial privileges, and city authorities exert this power over the population of neighboring localities. This specialized city occupies central functions for the surrounding region; it is a center through which all exchanges must pass. Unlike villages, which develop the resources of their immediate environment, cities develop their importance within a network, the scope of which depends on the size of the city and its type of specialization (Reymond, 1981). In this respect, a city is never isolated but always belongs to a network of cities; it is a node in a settlement system that changes according to coherent trends and in conjunction with the other elements of the system (Pumain, 1992).

Through the urbanization process, cities have become prevalent as a form of human settlement. This historical process has been interpreted in as many ways as there are definitions of cities. Several researchers have underscored the importance of symbols and of the sacred in the emergence of social forms organized on the basis of centralized power and exceeding in size all other forms of hunter-gatherer or sedentary agricultural societies (Racine, 1993; Weatley, 1971); others have highlighted the links between the development of cities and the political organization of the territories they control (Duby, 1985). Urbanization also can be explained by the change in the nature of economic activity. According to Paul Bairoch (1985), only the accumulation of agricultural reserves and of a surplus that can be sold can explain the almost simultaneous and independent emergence of cities in various regions of the world approximately 2000 to 3000 years after the Neolithic revolution. At the same time, the powerful urban growth of the 19th century cannot be separated from industrialization, even though a closer look at the phenomenon demonstrates that the links between urbanization and industrialization are quite complex. Henri Pirenne (1925) underscored the importance of trade and the possibilities for urban development offered by its expansion. In the current context of globalization, particularly the globalization of trade and the rapid development of international exchanges, this factor must again be taken into consideration because it may help to determine the nature of future urbanization processes.

Over the centuries, the same terms, *city* and *urban population*, have been used all over the planet to designate realities that are qualitatively and quantitatively very different, regardless of whether the point of view is social, cultural, environmental, political, or economic. What is there in common between a Greek *polis* and the cities of the Silk Road, between African mud towns and Manhattan skyscrapers, between medieval cities corseted behind their walls and sprawling Los Angeles, or between a small provincial town and a huge buzzing metropolis? Urban populations are by definition heterogenous, because the urban environment defines itself through the invention of new activities and new ways of life, in contrast to those of the rural world. The use of identical terms to designate such diverse forms has one advantage—it enables us to address the issue of the universal nature of urban phenomena and the uniqueness and coherence of their historical development. It also shows how important it is to develop relevant measuring tools for the purpose of comparison.

2. Defining the Urban Population

In practice, modern states have set up statistical instruments to distinguish between rural and urban populations or classify localities as towns or villages. Comparisons between one country and another are very difficult because official definitions are based on different criteria, which do not necessarily reflect the specificity of the local forms of urbanization but do reflect the complexity of urban phenomena. The *Demographic Yearbook* of the United Nations (UN) refers to these official definitions of urban populations, whose noncomparability was evaluated in two recent publications concerning countries of the European

Union (Pumain *et al.*, 1991) and concerning all countries of the world (Moriconi-Ebrard, 1993). The criteria adopted by the national institutes of statistics refer to the numerous concepts of city:

1. The *administrative status* of the city is derived from the legal definition of the city, often quite ancient. Localities are declared urban by political or administrative authorities. This criterion is applied, for instance, in Germany, in the United Kingdom, in Egypt, and in Tunisia.
2. An arbitrarily determined *population threshold* (e.g., 10,000 inhabitants in Spain or Italy) defines urban localities. Minimum thresholds vary considerably throughout the world, from 250 in Denmark to 50,000 in Japan, even though these differences may not reflect a systematic variation in the distribution of the population. In France, the threshold of 2000 inhabitants has not been modified since 1856. In other countries, the density threshold is used (e.g., in the Philippines, localities are considered urban if the population exceeds 1000 per square kilometer).
3. *Socioeconomic criteria* generally refer to a maximum threshold of the proportion of population in agriculture is retained, but this proportion can vary from 10% to 50%, depending on the country. The presence of typically urban infrastructures may sometimes be a determining factor, such as urban planning networks (in Bangladesh) or various service establishments (in Czechoslovakia).
4. Many *mixed criteria* are applied, such as size and urban planning as in Panama, size and legal status in India, or size and importance of nonagricultural activities in Zaire and in Zambia.

These criteria are generally applied within the framework of administrative subdivisions of limited size (e.g., the *communes* in France or local units in Europe). These vary considerably from one country to another. For example, in China, an urban *commune* always includes the population of a large rural territory, which often comprises several towns, and only partially corresponds to the reality of the population's spatial distribution. Defining the limits of urban entities is an essential issue, particularly for the detailed analysis and interpretation of migratory movements.

II. DEFINING THE LIMITS OF URBAN ENTITIES

As long as urban constructions were contained behind walls and as long as difficult traveling and transport conditions prevented towns from becoming too large, the town as a notion could be assimilated in the municipal territory of a *commune* or locality. When the town spread beyond the walls or the legal borders of the urban territory, these borders were periodically readjusted to correspond to the continuity of the built-up area, because urban growth remained slow.

With the industrial revolution and the urban explosion of the 19th century, the suburbs of cities extended, engulfing neighboring localities. Only after World War II did statistical institutes begin to consider the problem of the increasing difference between the city taken as a legal, political, and municipal entity and its sociologic, economic, and geographic reality. In several countries, as in Japan, Sweden, or West Germany (where the number of municipalities fell from 24,000 to 8500 between 1968 and 1978), some towns merged, or more often, new statistical entities that did not correspond to administrative limits were created.

The notion of an urban entity made up of several *communes* is based on the morphologic criterion of the continuity of buildings, which is used to define coherent urban entities. There are elementary urban forms regrouping a certain amount of population within a continuous built-up area as observed on maps or photographs (and sometimes on satellite images). According to the 1978 UN recommendation, an urban agglomeration is defined as a group of neighboring constructions separated by a space not exceeding 200 meters (500 meters in Latin America, where settlements are not as dense). Public areas, areas used for industrial or commercial purposes, as well as rivers with bridges crossing over them, are not taken into account to calculate this distance.

A population grouping that is continuously spread over several administrative units forms a *multicommunal agglomeration*. These agglomerations are usually adjusted according to communal limits, so that they coincide with a whole number of administrative entities. In France, the notion of multicommunal agglomeration appeared around 1954 (Marpsat, 1986). Communes with a population cluster of at least 2000 inhabitants are called *isolated towns*. If the agglomeration covers several communes, the multicommunal urban agglomeration is defined by identifying all the built-up zones regrouping at least 2000 inhabitants. Among the communes linked together, those in which the population cluster represents at least one-half of the total population of the commune are multicommunal urban agglomerations.

With minor variations, the concept of urban agglomeration—in the sense that the city is defined as a morphologic entity perceived in terms of the coherence of the urban fabric—is used by about 50 countries in the world. To the extent that the administrative networks of different countries can vary considerably in size and regularity, urban agglomerations defined in such terms are more comparable than cities defined as municipalities. In countries where settlements are very ancient, as in Europe, most urban activities, functions, and jobs are concentrated in the agglomeration; the latter can be considered a relevant framework for measuring the importance and impact of a city from a demographic or economic point view and for establishing national or international comparisons.

In many countries, the concept of urban agglomeration does not exist. The large-scale comparative studies of world urbanization conducted by Tertius Chandler and Gerald Fox (1974) or by Kingsley Davis (1969, 1972) were faced with the problem of the diversity of sources, as concerns the definition of a city. For this reason, the historical databases established for Europe between 1500 and 1800, on the basis of a single concept of urban agglomeration, by Jan de Vries (1984) are extremely useful, as are those of Paul Bairoch and colleagues (1988) for Europe from 800 to 1850 and Geopolis, the world database developed by François Moriconi-Ebrard (1994), in which the concept of morphologic agglomeration represents the first harmonized definition, applicable to the population of all cities of more than 10,000 inhabitants in the world from 1950 to 1990 (updated to 2000 on the web site www.geo.univ-avignon.fr).

From Cities to Urban Population Areas

According to the previously explained concepts, the city and the fact of living in it is what defines the urban population. With the spread of urban lifestyles to the rest of the territory, it is more and more the urban population itself, defined by its occupations and movements within an area of daily contacts, that determines the limits of the zones considered urban. The less compact urban forms observed in countries of more recent settlement, such as the United States, have encouraged some to suggest a more extensive definition of urban spaces, particularly around large cities. The use of cars and the preference for single-family houses and for a more rural type of environment have promoted the development of extensive residential suburbs, while city centers remain very dense employment areas with *central business districts* (Dupuy, 1995). This distribution has reinforced the dependence between the center and the peripheral zones of the city, with people massively commuting between home in the periphery and work in the center. The expression *daily urban systems* has been used to describe these configurations. Beginning in the 1950s, this model spread to many other countries, with the development of *banlieues* or immediate suburbs (i.e., entirely urbanized peripheral zones), followed by the creation of more distant suburban or peri-urban zones (i.e., population settlements in rural localities situated in the periphery of cities). Unlike classic suburbs, these peri-urban zones still have agricultural activities and non–built-up areas; sometimes called *rurban*, these forms of settlement reflect complex social transformations.

The Standard Metropolitan Statistical Area (SMSA) was created by the United States Census Bureau in 1949. An SMSA is centered on a city (or two contiguous cities) with a population of at least 50,000. It includes the entire surrounding county and the neighboring counties that do not have a large working population of farmers and where a certain proportion of residents commute to the city center for work. Several countries have used this definition to extend the concept of city to that of *employment area* or *functional urban area*. The aim is to include in the definition of an urban entity the peripheral zones polarized by a center that they depend on for employment and population growth. The center is defined as a composite entity made up of the central municipality and its immediate suburbs, and the urban area includes an urban unit or central city and the peripheral municipalities that are directly linked to it.

In Europe, four countries have officially defined functional urban areas: Belgium, France, Luxembourg, and the Netherlands. In Belgium, the area has a center called the *operational agglomeration*, and in Luxembourg, the center is an *urban agglomeration* of at least 50,000 inhabitants. In the Netherlands, the central nucleus of the agglomerates is made up of two contiguous municipalities with a population of at least 100,000. In France, the central nucleus of the *industrial and urban settlement zones* (ZPIU, as defined from 1962 to 1990) can be a multicommunal urban unit or an isolated town or a rural commune with an industrial activity (more than 100 jobs grouped in businesses employing at least 20 persons). To these centers are added the neighboring municipalities that have preferential commuting links with this center and that have characteristics that are considered specifically urban. The minimum population level determining an urban area varies greatly from one country to the next. An urban area must have a population of at least 100,000 in Luxembourg and in the Netherlands, 80,000 in Belgium, and only 2000 in France. Since 1990, a new definition of *urban areas* has replaced the ZPIU, which had become too extensive. These areas are situated around centers with at least 5000 jobs and regrouping communes, from which more than 40% of the residents commute to the center or any other commune dependent on the latter.

Employment areas or functional urban areas are more relevant as entities than morphologic agglomerations, and they are extremely useful for the study of current urbanization processes, including changes in land use, the extension of urban property, and the decentralization of residences or of certain types of occupation that have moved to peripheral zones not necessarily linked to the main agglomeration through continuously built-up surfaces. However, given the size of these areas and the difficulty of determining their limits with simple criteria (because there can be discontinuous parts or enclaves) and given their uneven status in terms of development, these entities are difficult to use in terms of their global weight in national and international comparisons.

The different types of statistical description of spatial forms can be grouped into four categories covering the huge variety of urban population conglomerations in different countries:

1. *Urban localities* are defined by administrative limits or by legal status.

2. *Urban agglomerations* or *urban units* group together urbanized centers made up of continuous built-up areas, covering a portion of an administrative unit or grouping together several units.

3. *Functional urban regions*, which include a central city and its sphere of influence, often are defined on the basis of commuting patterns.

4. *Polynuclear urban regions* or *conurbations* can represent a continuous built-up expanse but contain several polarizing centers. They often represent the combination of several agglomerations or urban regions that are initially distinct but that later merge as they spread in space.

III. THE URBAN TRANSITION

During the centuries before the industrial revolution, some cities were born while others disappeared. Their appearance and function underwent many transformations. The size of urban populations fluctuated, sometimes to a considerable extent, as for example during the great plague of the 14th century in Europe, which killed one-half the population of cities. However, throughout that period, the urbanization rate—the proportion of urban dwellers among the total population—remained relatively stable. Even though it may have varied from one period and one region to the next, on average, the rate remained about 10%, within a range that only exceptionally reached 30% or even 40%, for example in 17th century Netherlands (Bairoch, 1985).

The industrial revolution caused extensive changes in the organization of economic activities, lifestyles, and social structures, and it completely modified the organization of settlements in space. During a major transitional period, which in developed countries lasted about 2 centuries, the settlement system completely changed, and only about 20 years ago was a new balance apparently reached. The relationship between city and countryside has been totally reversed. Even though agricultural activities remain dominant in terms of land use and landscape, the population and its living conditions, activities, and lifestyles have become to a large extent urban, including in most communes still considered rural.

This major transition has been called *urban transition* by some researchers (Zelinski, 1971) by analogy with the demographic transition (see Chapters 68 and 69), which occurred at different times in the history of different countries of the world. Several aspects of the historical urbanization process can be linked to the way innovations are spread in space. A first hint is given by the specific shape of the curve describing the trend in the proportion of urban population within the total population, which takes on everywhere the shape of a logistic curve, similar to the curve of the proportion of population that adopts a given innovation (Hägerstrand, 1953). This simple indicator is called the *urbanization rate*. It measures the proportion of the population that has chosen to live in the city; it describes the relative weight of cities in the settlement pattern. This rate underwent specific variations during the period of urban transition. The settlement systems of the pre-industrial period showed a relatively low rate of urbanization; they remained quite stable, with a slow pace of progress and occasional small fluctuations. The urban growth of the 16th and 17th centuries in Europe, which led to a significant increase in the size of the urban population and in the number of cities, went along with a very slight increase in the relative share of the population living in towns of more than 5000 inhabitants. This proportion grew from 10% to 11.5% around 1500 to only 12% to 13% in 1700 (Bairoch, 1985; de Vries, 1984). The transitional phase began at a time when the urbanization rate began to increase faster than during previous periods and especially when this increase persisted steadily over several successive periods; thus began an irreversible process which in the space of a century led to an urban population of 80%. The end of the urban transition was marked by the stabilization of the urbanization rate, which meant that the industrial settlement system had reached a certain level of development.

This stability corresponds to a new balance in the spatial distribution of the population. However, this balance is a dynamic one, because each component of the settlement system continues to change. Even if some cities grow while others stagnate or decline, globally the ratio between city dwellers and country dwellers remains stable. A simple criterion such as the urbanization rate measures only one aspect of urban growth; it is a global measure that describes the organization of the system as a whole. The spatial extension of cities, for example, is another aspect of urban growth whose consequences are felt mostly on a different scale, that of local planning. The latter process is ongoing, and it has even gained momentum in recent years.

The relationships between migration and urbanization are complex. Although it is true that the overpopulation of the countryside can generate migration flows toward the city, overall, on that scale, it cannot be said that migrations are the cause of growing urbanization. The surplus of farming production, the emergence of new types of activity, and the intensification of exchanges are the decisive factors in the skyrocketing growth of urban populations, and they have an influence on country-city migrations. Conversely, on a local scale, residential migrations are the cause of recent transformations in the urbanization process in regard to its morphologic aspects, even if these are due to improved living standards and the generalization of automobiles, which must be taken into account but are not sufficient to explain the urban sprawl.

Globally, the urban transition appears to be a universal phenomenon in world history. However, a detailed study of the history of the urbanization process shows that there are many variants to this phenomenon, because its characteristics, pace, and explanatory factors seem to differ from one continent to the other and especially between developed countries and Third World countries (Fig. 61–1). Overall and over time, a general relationship is observed between urbanization and economic development, which is also reflected in technical progress, industrialization, and the social division of labor. The openness of the country to exchanges also seems to be a necessary condition, as shown by the decline of urbanization between 1500 and 1850 in China despite high demographic growth. At that time, the country was undergoing a period of total isolation, whereas during the 15th century, its situation had been comparable to that of Europe in terms of development level and urbanization rate (Bairoch, 1985).

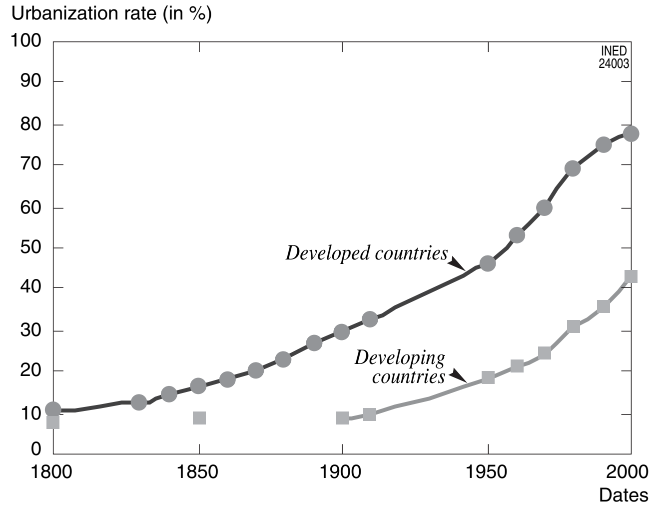


FIGURE 61–1 Urbanization rate since 1800 in developed and developing countries. (Data from Moriconi-Ebrard, 1994; Bairoch, 1985.)

1. The Urban Transition in Developed Countries

Toward the end of the 18th century, the urbanization rate in Europe began to increase and rose from about 20% to 80% in about 150 years. At that time, the average annual rate of growth of the urban population was close to 2% annually, about twice as high as that of the total population. During the 19th century, the increase in agricultural productivity drove unemployed farm hands to the city, and these migrations were responsible for one-half to two-thirds of the urban growth. At the same time, the population was beginning to increase at a very high rate due to the demographic transition, and this phenomenon also fueled urban growth. Nonetheless, urban dwellers had lower fertility and higher mortality rates than rural inhabitants (prompting the phrase *man-eating cities*); this mainly resulted from higher population density, resulting in a higher risk of epidemics, which at the time were difficult to control.

The urban transition first appeared in Great Britain, which at that time was not the most urbanized country, but it was the country where agricultural productivity had increased most and the first to have experienced the industrial revolution. In France, where the demographic transition preceded the urban transition, urbanization began later and at a slower pace, with an average rate of growth of the urban population of about 1% per year during the entire 19th century. This explains why in France the major part of the population was still rural until 1950, whereas in Great Britain, this had ceased to be the case at the turn of the century. Overall, the progress of urbanization followed that of the industrial revolution; it began first in northwestern Europe, was very rapid in the United States and in other "new" countries, and was somewhat delayed in some countries of southern Europe, even though these were the most urbanized at the start of the process.

After World War II, thanks to strong economic growth and to the baby boom, urbanization processes accelerated, especially in countries where they had been delayed. Migrations from the countryside to cities also fueled urban growth in the 1950s and 1960s; they later stabilized and even slightly declined in the 1980s. In these countries, the countryside—whose population had reached its maximum level in the last third of the 19th century but had since then consistently been declining—lost a considerable part of its population. In the space of 1 century, between 1866 and 1975, the French countryside lost one-half of its population, more than 12 million persons, mainly due to emigration. However, certain rural localities situated in the periphery of large cities increased their population thanks to urban expansion and the spread of the urban population. The suburbanization process, which began in the 1950s in the United States, spread to Northern European countries in the 1960s and 1970s, and this suburban spread continued throughout Europe during the 1980s and 1990s.

2. The Urban Explosion in Developing Countries

In 1950, the urban transition had not yet begun in developing countries; the urbanization rate was then only 15% on average. The urban population explosion that then took place was far more intense than anything ever experienced in industrialized countries. Between 1950 and 1990, the urban population increased each year by 4.5%, a rate almost 2.5 times higher than that observed in developed countries during the same period and twice as high as that witnessed in developed countries in the 19th century during the most intense moments of the urban transition. The pace of increase of the urbanization rate in developing countries is now twice as high as that of the developed countries in the last century. Instead of lasting about 150 years, it is likely that the urban transition will last only about 50 years in developing countries (see Fig. 61–1).

This unprecedented urban growth does not have the same causes. Migrations from country to city only account for one-half of this growth. The fertility rate of urban populations remain identical to that of rural populations, and the mortality rate is lower because health care facilities are more numerous. As Paul Bairoch (1985) said, "The city has ceased to be an impediment to its own growth." Instead of declining as in Europe during the urbanization process, the rural population continues to increase, and demographic pressure is strengthening.

There is a controversy concerning the factors that determined the urbanization process in developing countries. If we compare them with industrial countries during the past century, it seems that in the former, industrialization and agricultural productivity did not progress at the same pace as the urbanization rate. The causes of urban expansion are rooted in the very high rate of demographic growth, unprecedented in nature and linked to the rapid pace of the demographic transition in the Third World (see Chapter 69). The wide gap between urban and rural incomes, with the former twice to three times as high as the latter (compared with differences of 15% to 30% in 19th century Europe), can also be ascribed to the development of international relations and of links with the developed world, which generate dual economic systems. This wide gap, along with a broader reach of information and education, may explain the strong impact of the "urban mirage" (Bairoch, 1985).

Foreseeable population movements will be massive. For example, in China, during the first 2 decades of the 21st century, about 300 to 400 million persons will be compelled to leave the countryside to find employment in the city. The several million persons who have already moved to the city to find work represent an unstable population group, whose status remains unclear. Whatever its origins and given its momentum and very rapid pace, the urban transition in the Third World is different from that of industrial countries, or at least it represents a specific variation of this process. This new form is marked by a much higher concentration of the urban population in very large cities, much higher than that observed for the same urbanization rate in developed countries. It is not surprising that the distribution of large cities in the world has been completely reversed. In 1900, only 1 of 16 cities with a population of at least 1 million was in a developing country; by 1950, the figure had risen to 32 (37%) of 86 and to 174 (58%) of 298 in 1990 (Moriconi-Ebrard, 1993).

IV. THE DYNAMICS OF SETTLEMENT SYSTEMS

A settlement system on a given territory is defined by the inhabited areas (e.g., villages, towns, cities) and by communication networks (i.e., transport and information networks) that connect these areas of settlement. The notion of territory is understood here in its administrative or political sense, as a portion of the earth's surface that has been appropriated by a group.

This implies a certain degree of homogeneity of socioeconomic conditions and global control over the possibilities of circulation within that territory. The networks are created by the population of the different settlements and make possible numerous forms of interaction: exchanges of material and monetary goods (i.e., trade), exchanges of persons (i.e., migrations) or information. Throughout history, the system's environment, made for the most part of the mining and agricultural resources of the given territory, became less and less dependent on its natural characteristics and increasingly built-up as the industrial and urban society developed and took over the environment.

Settlement systems throughout the world have very powerful characteristics, which are relatively invariant depending on period or place. These systems are strongly hierarchical structures, reflected in the difference in size of the elements of the system, from a population of a few dozen living in small hamlets to over 10 million in large, modern metropolises. The general model that accounts for this organization is a reverse geometric progression of the number of cities according to their size (Zipf, 1949) or a log-normal distribution of the size of the settlement units (Pumain, 1982; Gibrat, 1931). Historical studies (de Vries, 1984; Carroll, 1982) and archaeologic studies (Fletscher, 1986), as well as very recent international comparisons (Guérin-Pace, 1993; Moriconi-Ebrard, 1993), all show the universal nature of this model in space and time.

This hierarchical structure is marked by regular patterns in the distance separating the settlement units; the larger the cities, the longer the distance between them. The theory of central places (Christaller, 1933; Reynaud, 1841, quoted by Robic, 1982) explains this configuration by way of the centrality principle, according to which the production of services situated at the same level and designed for a population dispersed in the surrounding region (or zone of influence) is regrouped in a center that attracts the client population. The centers are organized hierarchically because there are several different levels of services, defined by their spatial scope (i.e., the distance the consumer is willing to cover to obtain the service) and by thresholds of demand that determine the creation of a given service (i.e., the volume of clientele necessary to ensure that the production of the service remains profitable). The zones of influence of the different centers overlap, because the higher-level centers usually offer all the services that can be found at the lower levels. The consequence of this competition between centers is a regular pattern of distance between cities, and the hierarchy of service levels is reflected in the smaller number and longer distance between centers as we go up in the hierarchy.

Another important characteristic is the very high level of interdependence in the development of elements belonging to the same settlement system. Changes in morphologic configuration (e.g., city planning), in activity types (e.g., industrialization, tertiarization), in the social composition (e.g., development of salaried work, increase in the level of qualification), and in the number of inhabitants (e.g., rural exodus, urban growth) very seldom appear as isolated or sporadic phenomena happening in a single place. On the contrary, detailed analyses of the changes occurring in all the cities of a given territory show that the transformations pertaining to a given period were taking place simultaneously and in a very homogeneous way (Sanders, 1992; Pumain and Saint-Julien, 1978). Because these changes occur with practically the same intensity everywhere, the inequalities and differences that exist at a given moment between cities persist for periods. Over a given period, cities that belong to the same system shift in the socioeconomic space, and the settlement system adapts to this shift without modifying its structure.

The only way to modify in a deep and lasting way the structure of these city systems, as well as the urban population's ability to survive by adapting to the endless innovations that are part of urbanization dynamics, is through the selective adoption of innovations (e.g., when they are based on the development of sporadic mining fields, such as cities built around coal and iron mines during the industrial revolution, or towns built on the Riviera with the development of tourism) or, on the longer term, through major changes in political and economic organization (e.g., those that gradually displaced the centers of world economies, according to Fernand Braudel's concept).

If we observe in each city the growth of urban populations, encouraged during the urbanization process by these systemic dynamics, we distinguish two major types of development. In countries where settlements are ancient, they can be for a large part described by Robert Gibrat's (1931) stochastic model, which explains the log-normal distribution of city sizes through a random distribution of their growth rates in space and time. To this overall process, it must be added that over the long term, large cities tend to grow more than smaller ones, and this tends to reinforce contrasts in the hierarchy of city sizes within the system. This tendency can be explained by the fact that innovations spread according to a hierarchical pattern whereby overall development first takes place in large cities, which benefit more from the initial advantage. However, it is also explained by the historical tendency to reduce duration of travel with increasing speed of communication, which hampers the development of smaller towns. The urban network of these countries continues to develop along the same patterns; contrasts between city sizes are increasing, and hierarchy is simplified at the lower levels (Fig. 61–2).

In countries of more recent settlement, urban growth takes on other forms. In a country being colonized, there is a negative correlation between urban growth and the size of cities because of the continuous emergence of new towns on frontiers and the phenomenon of mushroom towns. This type of development continues until the territory is saturated. It seems that distances between cities are on average longer and much less regular in newer countries settled at times when means of communication were more rapid than they were at the time settlements were created in older countries. The average distance between two cities of more than 10,000 inhabitants in the European Union is of about 13 km; it is 17 km in India, but it exceeds 60 km in North America (Moriconi-Ebrard, 1994). In

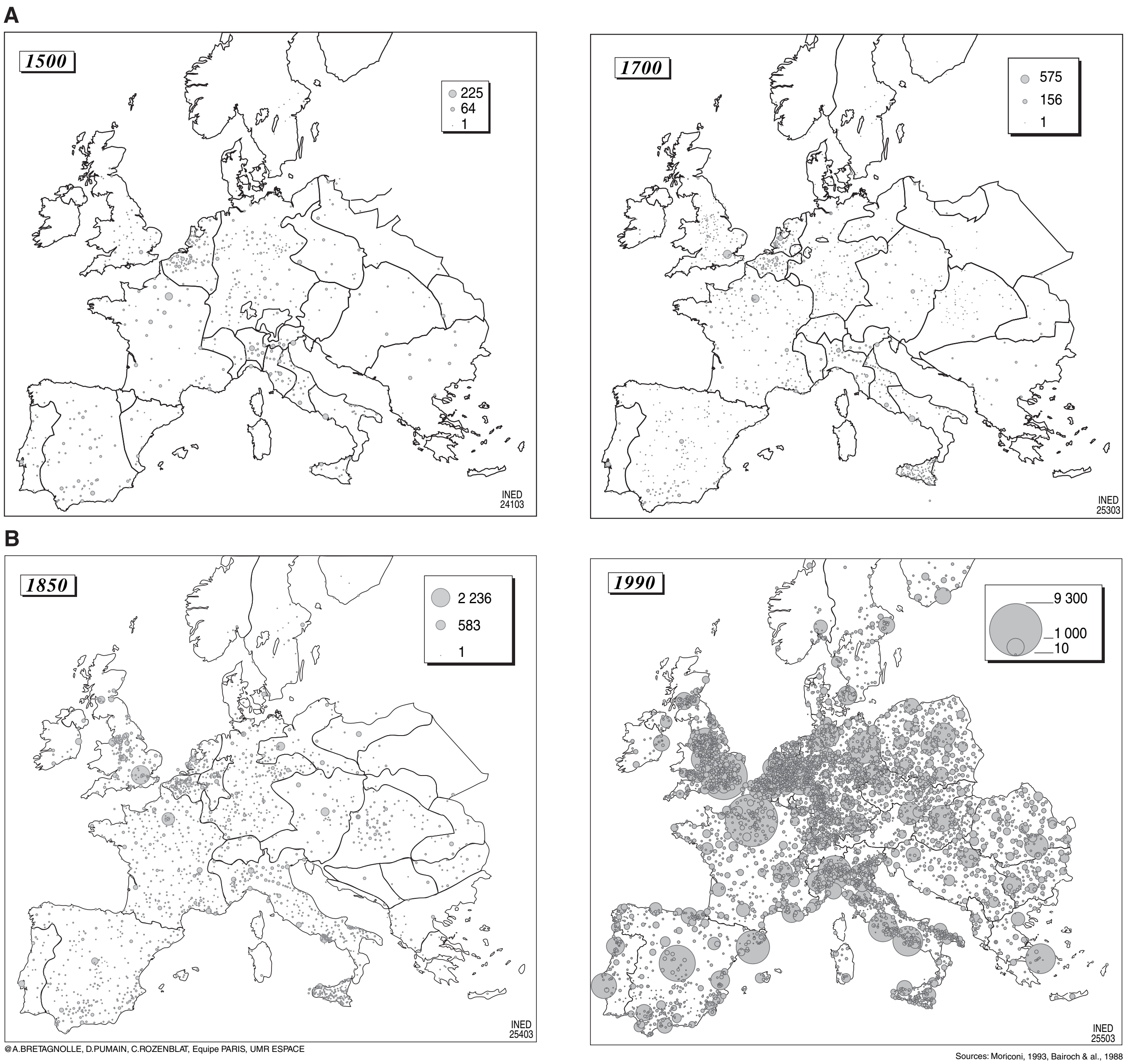


FIGURE 61–2 **A:** Population growth in the European city system in 1500 and 1700 (figures given in thousands). **B:** Population growth in the European city system in 1850 and 1990 (figures given in thousands).

America, the volume and density of the population is much lower, but big cities are much larger than European big cities, and the total number of urban units is much smaller. In Europe, the differences in distance patterns correspond to different types of settlement: more dense in countries along the Rhine, less dense and more contrasted in the western and northern periphery, and more regular in the eastern marshes, which were colonized later (Fig. 61–3).

Given the strong regularity of such patterns, researchers studying and analyzing urban populations must always consider two geographic scopes of investigation and interpretation: that of individual characteristics and of a person's position within the set of relations linking him or her to the rest of the city's inhabitants, with the city being a relatively autonomous subentity, and that of the variables that describe this conglomeration (i.e., the system formed by these persons), depending on its relative position (mostly represented by size, specialization, location, image, and other factors) in the system of cities. These two levels of organization have an undeniable impact on individual characteristics and movements, in the same way as a city cannot emerge as a coherent entity without going through a long history of formation of interaction networks. For this reason, newer approaches to explaining cities and city systems are based on evolutionary theories (Pumain, 1997).

V. MIGRATIONS AND URBANIZATION

Inequalities of urban growth more often result from the variable intensity of the city's power to attract migration than differences in the natural growth of the nonmigrating population. Although according to economic theory the explanatory factor of urban development would be the individual decision to migrate, it seems that migrations are very much influenced by the structural characteristics of the urban system in which they occur and that the latter's own dynamics highly influence individual movements.

1. Centripetal Migration: The Rural Exodus

In all countries where the urbanization process took place, it led to significant population movements.

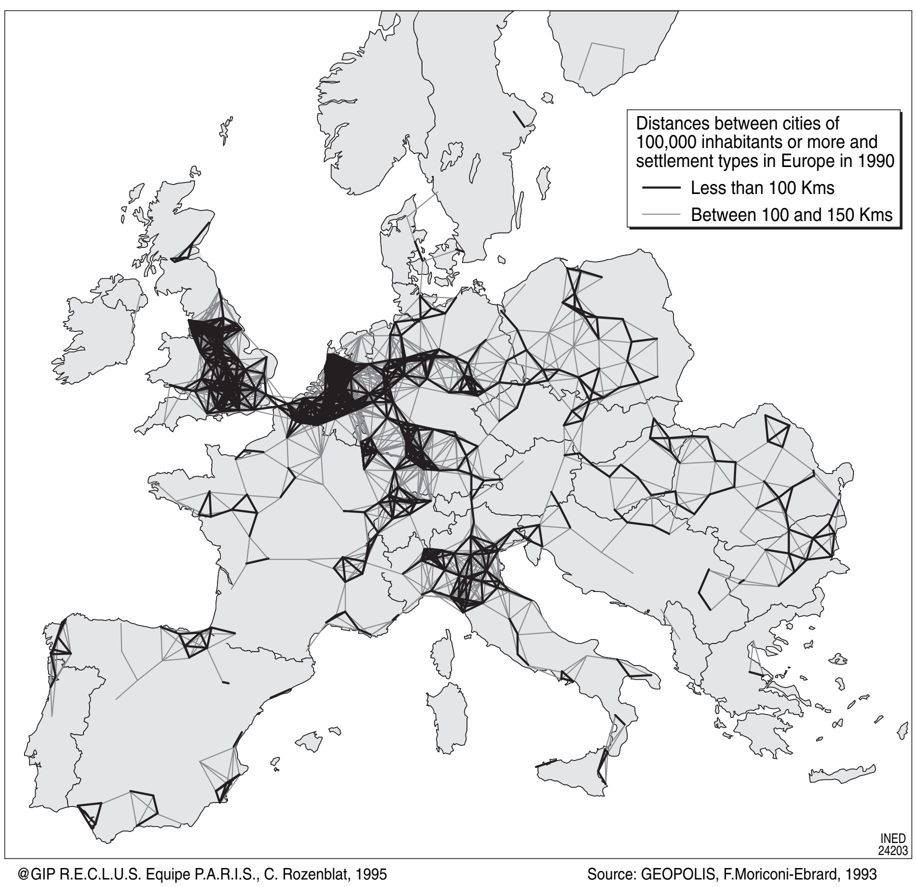


FIGURE 61–3 Distances between cities of 100,000 inhabitants or more and settlement types in Europe in 1990.

Although Edward Ravenstein's *laws of migration* (1885) established that all migrations occurring between different parts of a territory were made up of a flow and a counterflow that were more or less equivalent in volume, the rural exodus has been persistently asymmetric, because the flow from country to city has long exceeded the flow going the other way. These flows follow the gravitational model; if we measure the flows between two areas, we find that they are roughly proportional to the product of the populations located in the departure and arrival areas and inversely proportional to the distance separating the two zones. For a long time, the gradient of the decrease in the intensity of migrations from areas around cities remained very sensitive, reducing the relative volume of migrations coming from farther away than several tens of kilometers (Courgeau, 1970). The area around the city from which population moved to the city grew gradually larger, but today still, the more intense migratory exchanges, such as around French regional capitals (i.e., cities of 250,000 to 1 million inhabitants), do not extend beyond the first circle of *départements* surrounding the city.

The population migrating into cities mainly comes from the city's immediate surroundings. Many studies have shown that migrations from country to city most often occur in several stages, either during the life of a single individual or over successive generations, and that these stages generally correspond to a move upward in the urban hierarchy. Small and medium-sized cities often play the role of an intermediary stopover stage in the journey from country to city (Courgeau, 1983).

The rural exodus was also a selective process. Those who left the countryside first were those with the weakest ties to the land, because artisans and shopkeepers migrated before farmers. Results of studies underscore persistent differences in the individual characteristics of those who move to the city and those who remain in the country (Courgeau, 1987).

2. Migrations between Cities

During the urbanization process, because of the distribution of migratory processes into separate stages and to the positive net migration of larger cities compared with smaller towns, migrations between cities played a part in the concentration of population in the larger cities and in the development of urban hierarchies. Although this pattern was first considered valid for migrations in general, it was later limited to the working population when the urbanization process reached a phase of saturation. This selective process based on urban hierarchy remains nonetheless very significant.

When the rural exodus ended, migrations between cities increasingly contributed to the differential growth of the urban population. Although the cities' rates of natural increase are becoming more and more homogenous, their net migrations remain very uneven (although their overall contribution to urban growth is declining), and their impact is selective. This impact is more significant from a qualitative than from a quantitative viewpoint. Populations migrate differently depending on age. Young adults, students, or persons looking for their first jobs prefer big and dynamic cities, whereas older adults prefer medium-sized cities, and senior citizens prefer small towns. However, the social or economic composition of migratory flows seems to be determined to a higher extent by preexisting urban specializations. Workers are attracted to working-class towns but also leave them, and the composition of the population arriving in and leaving tertiary cities is just about similar to that of the resident population (Pumain and Saint-Julien, 1989). The role of migrations in urban dynamics is rather difficult to determine. Their impact on the rejuvenation of urban populations and the concentration of the young, working population in big cities is important, but it is less significant as concerns social or economic modifications in cities, because these factors seem to be determined mainly by endogenous factors (Baccaïni and Pumain, 1998). At the most we may observe, as a recurrent phenomenon, a selective emigration of senior-level professionals trained in the universities of large cities, who then move back to smaller towns.

3. Centrifugal Movement: The Urban Sprawl

The distribution of urban populations in urban space, usually observed in the area of residence, follows general processes such as competition for the occupation of the most accessible areas, and these often are the most central areas. This distribution also depends on technologic factors (i.e., speed and transport systems) and on social, legal, and cultural issues. Even though population densities vary considerably, from several hundred to more than 100,000 inhabitants per square kilometer, they usually follow rather systematic patterns. The density of cities varies, as does the general population density of the territory as a whole; on average, it is lower in the United States than in Europe, and it is higher in Asia. Within a single territory, the larger the city, the higher the average density. Within a single city, density decreases very strongly when moving from the center to the periphery, following an exponential or power function according to the distance from the center. Historically, urban densities first increased during an initial phase lasting from the Middle Ages, when the walls surrounding cities restricted their spatial extension, until the invention of urban streetcars. From then on, urban peripheries began to extend in space, and the density of suburbs increased faster than that of the center, reducing density gradients. In parallel, competition for land use on the part of economic activities, top-level services in particular, has created a central crater in big cities where residential density is lower (although the daytime density of the working population is very high).

The growing use of cars also generated specific population movements and urban forms that continue to influence migration.. Automobiles, along with single-family housing and very nonrestrictive land use standards, have led to the emergence of specific urban forms on the American suburbia model. However, extreme forms can be found in other new countries. In Australia, for example, Sydney's population of 3 million shares an area the size of greater New York City, which has a population of 24 million. From the 1960s on, the suburbia phenomenon began gaining momentum in the United States and spread to Europe in a slightly different form, called *urbanization of the countryside* and *peri-urbanization*. It first appeared in 1970 in Scandinavia and Great Britain, and it later occurred in France, reaching southern Europe in the 1980s. This phenomenon gave rise to pessimistic interpretations. Some came to believe that the end of cities, a process they called *counter-urbanization* (Champion, 1989; Berry 1976), was near (Chombart de Lauwe, 1982). However, the consequence of the urban sprawl, which led to a loss of population in city-centers and the migration of households toward rural peripheries, was not so much the revitalization of rural zones as the extension of the urban influence in terms of employment and services. These residential migrations went along with the development of commuting practices, because rather than take advantage of increased speed to reduce time spent in transport, urban populations preferred to increase the distance covered and gain access to more spacious and cheaper housing in a more natural and less urbanized environment.

This trend will probably continue, and it seems to have begun in the big cities of the Third World. On the part of urban populations, rather than being a sign of defiance faced with the generalization of the urban way of life, it appears to be a manifestation of the wish to take advantage of both the nearness of facilities and services made possible by concentration and of improved living spaces and a natural environment. This aspiration is not new and explains in part the fragmented, fractal aspect of urban settlement configurations (Frankhauser, 1993). Does this mean that cities that are more adapted to automobile traffic and that now have the highest living standards represent a universal urban model toward which all cities tend to converge? Let us not forget that cities are complex entities, the result of long-standing historical processes, and that the various processes that have influenced urban forms explain contemporary diversity and can guarantee the sustainability of cities.

CONCLUSIONS

The study of the long history of the urbanization process enables us, with a reasonable degree of certainty, to make a few predictions concerning the future of cities. Urbanization belongs to the category of processes involving the spread of innovations, and in that sense, it is possible to determine, especially in developing countries, which stage has been reached—all the while taking into account the country's specific historical characteristics—to determine how to go about facing the later phases of urban growth. The distribution of urban population growth among cities and the modifications of urban hierarchies are relatively foreseeable in the medium term if we apply what we know about the dynamics of city systems.

It is more difficult to foresee the future of this process after the populations have been more or less completely urbanized. Some believe that urban forms are linked to available technologies and to the economic activities they make possible; in their view, it is possible that new information technologies may undo previous population concentrations. The information society may return to more dispersed forms of settlement, and rural regions may be repopulated. Others insist on the strength of the links tying current urban systems to their past history. The networking of cities, which is a powerful phenomenon in Europe (Cattan *et al*., 1994) and which goes along with the growing internationalization of exchanges, corresponds to the relative concentration and the selection by large cities of the most skilled population and the most modern activities. These trends have been observed in the past, occurring at the beginning of every large cycle of innovations. Over the centuries, the city system has constantly reacted by adapting its structure to successive qualitative and quantitative changes in human activities, and it may very well adapt again to the innovations brought about by the *information society*.

Acknowledgment

This chapter was translated from the French version by Zoé Andreyev.

References

Ascher François, 1995. *Metapolis ou l'avenir des villes*. Paris, Odile Jacob, 347 p.

Baccaïni Brigitte and Pumain Denise, 1998. Les migrations dans le système des villes françaises de 1982 à 1990, *Population*, vol. 53(5), p. 947–978.

Bairoch Paul, 1985. *De Jéricho à Mexico, villes et économie dans l'histoire*. Paris, Gallimard, 708 p. (Arcades.)

Bairoch Paul, Batou Jean and Chèvre Pierre, 1988. *La population des villes européennes de 800 à 1850*. Geneva, Droz, 336 p.

Berry Brian J.L., 1964. Cities as systems within systems of cities, *Papers of the Regional Science Association*, vol. 13, p. 147–163.

Berry Brian J.L. (ed.), 1976. *Urbanisation and counter-urbanisation*. Beverley Hills, CA, Sage, 334 p.

Braudel Fernand, 1979. *Civilisation matérielle, économie et capitalisme (XV–XVIII siécles)*. Paris, A. Colin, three vol., 544 + 600 + 606 p. (I. Les structures du quotidien, II. Les jeux de l'échange, III. Le temps du monde.)

Camagni Roberto, 1993. *Economia urbana*. Roma, La Nuova Italia Scientifica, 431 p.

Cattan Nadine, Pumain Denise, Rozenblat Céline, and Saint-Julien Thérèse, 1994. *Le système des villes européennes*. Paris, Anthropos, 208 p.

Champion Anthony G., 1989. *Counterurbanisation*. London, Edward Arnold, 266 p.

Chandler Tertius I. and Fox Gerald, 1974. *3000 years of urban growth*. New York, Academic Press, 431 p.

Chombart de Lauwe Paul-Henri, 1982. *La fin des villes: mythe ou réalité*. Paris, Calman-Levy, 249 p.

Christaller Walter, 1933. *Die zentralen Orte in Süddeutschland*. Jena, Gustav Fischer, 331 p.

Courgeau Daniel, 1970. *Les champs migratoires en France*. Paris, INED, 158 p.

Courgeau Daniel, 1983. *Trois siècles de mobilité spatiale en France*. Paris, UNESCO, 71 p. (Rapports et documents de Sciences sociales, no. 51.)

Courgeau Daniel, 1987. Constitution de la famille et urbanisation, *Population*, vol. 42(1), p. 57–82.

Davis Kingsley, 1969, 1972. *World urbanization, 1950–70*. Beverley, University of California, vol. 1, 1969, 318 p., vol. 2, 1972, 319 p. (Population monograph series, nos. 4 and 9.)

Driant Jean-Claude (ed.), 1992. *Habitat et villes: l'avenir en jeu*. Paris, L'Harmattan, 162 p.

Duby George (ed.), 1980–1985. *Histoire de la France urbaine*. Paris, Seuil, vol. 1, 1980, 607 p., vol. 2, 1981, 656 p., vol. 3, 1982, 655 p., vol. 4, 1983, 667 p., vol. 5, 1985, 672 p.

Dupuy Gabriel, 1991. *L'urbanisme des réseaux. Théories et méthodes*. Paris, A. Colin, 216 p.

Dupuy Gabriel, 1995. *Les territoires de l'automobile*. Paris, Anthropos, 216 p. (Villes.)

Fletscher Robert, 1986. Settlement in archaeology: world-wide comparison, *World Archaeology*, vol. 18(1), p. 59–83.

Frankhauser Pierre, 1993. *La fractalité des structures urbaines*. Paris, Anthropos, 291 p. (Villes.)

George Pierre, 1959. *Questions de géographie de la population*. Paris, INED, PUF, 231 p. (Travaux et documents, cahier no. 34.)

Gibrat Robert, 1931. *Les inégalités économiques: applications aux inégalités de richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc. d'une loi nouvelle, la loi de l'effet proportionnel*. Paris, Librairie du recueil Sirey, 296 p.

Gottmann Jean, 1961. *Megalopolis, the urbanized northeastern seaboard of the United States*. Cambridge, MA, MIT Press, 810 p.

Guérin-Pace France, 1993. *Deux siècles de croissance urbaine*. Paris, Anthropos, 205 p.

Hägrestrand Torsten, 1953. *Innovation diffusion as a spatial process*. University of Chicago Press, Chicago and London, 334 p.

Hall Peter, Gracey Harry, Drewett Roy, and Thomas Ray, 1973. *The containment of Urban England*. London, Allen & Unwin, vol. 1. Urban and metropolitan growth processes; or Megalopolis denied, 648 p., vol. 2. The planning system: objectives, operations, impacts, 464 p.

Marpsat Maryse, 1986. Les agglomérations urbaines multicommunales: évolution des définitions et de leur mise en oeuvre, *Courrier des Statistiques*, vol. 39, p. 28–32.

Moriconi-Ebrard François, 1993. *L'urbanisation du monde depuis 1950*. Paris, Anthropos, 372 p. (Villes.)

Moriconi-Ebrard François, 1994. *Geopolis, pour comparer les villes du monde*. Paris, Anthropos, 246 p. (Villes.)

Pirenne Henri, 1925. *Les villes et les institutions urbaines*. Paris, Brussels, Alcan, 2 vol., 431 + 298 p.

Pumain Denise, 1982. *La dynamique des villes*. Paris, Economica, 231 p. (Géographie économique.)

Pumain Denise, 1992. Le peuplement, *in*: Antoine Bailly, Robert Ferras, and Denise Pumain (eds), *Encyclopédie de géographie*, p. 439–462. Paris, Economica, 1132 p.

Pumain Denise, 1993. L'espace, le temps et la matérialité des villes, *in*: Bernard Lepetit and Denise Pumain (eds.), *Temporalités urbaines*, p. 135–157. Paris, Anthropos, 319 p.

Pumain Denise, 1997. Pour une théorie évolutive des villes. *L'Espace Géographique*, no. 2, p. 119–134.

Pumain Denise and Saint-Julien Thérèse, 1978. *Les dimensions du changement urbain*. Paris, CNRS, 202 p.

Pumain Denise and Saint-Julien Thérèse, 1989. Migrations et changement urbain. *Revue d'Economie Régionale et Urbaine*, vol. 3, p. 509–530.

Pumain Denise, Saint-Julien Thérèse, Cattan Nadine, and Rozenblat Céline, 1991. *Le concept statistique de la ville en Europe*. Luxembourg, Statistical Office of the European Communities, 88 p.

Racine Jean-Bernard, 1993. *La ville entre Dieu et les hommes*. Paris, Anthropos, 354 p.

Ravenstein Edward, 1885. The laws of migration. *Journal of the Royal Statistical Association*, vol. 48, p. 167–235.

Reymond Henri, 1981. Pour une problématique de la géographie humaine, *in*: Hildebert Isnard, Jean-Bernard Racine, and Henri Reymond (eds.), *Problématiques de la géographie*. Paris, PUF, 263 p.

Reynaud Jean, 1841. Villes, *in*: *Encyclopédie nouvelle*, vol. VIII, p. 670–687. Paris, Gosselin.

Robic Marie-Claire, 1982. Cent ans avant Christaller . . . une théorie des lieux centraux, *L'Espace Géographique*, vol. 4, p. 5–12.

Roncayolo Marcel, 1985. Nouveau cycle ou fin de l'urbanisation? *in*: Marcel Roncayolo (ed.), *La ville aujourd'hui. Croissance urbaine et crise du citadin*, p. 13–131. Paris, Le Seuil, 671 p. (Histoire de la France urbaine, sous la dir. de George Duby, vol. V.)

Sanders Lena, 1992. *Système de villes et synergétique*. Paris, Anthropos, 274 p.

United Nations, 1978. *Recommendations for the 1980 censuses of population and housing in the ECE region*. New York, United Nations.

Vries Jan de, 1984. *European urbanisation*, 1500–1800. London, Methuen, 398 p.

Weatley Paul, 1971. *The pivot of the four quarters: a preliminary enquiry into the origins and character of the ancient Chinese city*. Edinburgh, Edinburgh University Press, 603 p.

Zelinski Wilbur, 1971. The hypothesis of the mobility transition. *Geographical Review*, vol. 61(2), p. 219–249.

Zipf George Kingsley, 1949. *Human behaviour and the principle of least effort*. Cambridge, MA, Addison-Wesley, 573 p. (Re-ed. 1965 of the 1949 edition: New York, Hafner Publishing, 573 p.)

CHAPTER

62

Political Pressure

LUC LEGOUX

Institut de démographie de l'université de Paris (IDUP), Paris, France

Whether it is repression, war, ethnic cleansing, or genocide, the political reasons for the forced movement of population seem innumerable. On the contrary, possible destinations are far fewer; they are limited to neighboring countries whose hospitality is often forced and to developed countries that are becoming increasingly hostile toward immigration. However, political immigration has a separate status in international immigration because most countries acknowledge its legitimacy over other types of migration. Consequently, if political violence determines the pressure to leave, the magnitude and destination of political migration depends on the receiving country's translation of this into a formal acknowledgment of the right to migrate. There is a wide gap between the theoretical right to look for political asylum in democratic countries and the effective possibility of actually doing it. Today, the major obstacle facing asylum seekers is no longer the defunct Iron Wall but Western policies of limiting immigration. Judging from the difficulties facing asylum seekers in Europe, these policies are effective. However, should this be called effectiveness when it is associated with a growth in tolerated illegality? Legal measures allow for the legitimate refusal of asylum, but they do not authorize the expulsion of asylum seekers to their countries if such countries are in the middle of violent conflict.

I. THE ROOTS

Violence has always been at the origin of forced movements of population. Twenty-five centuries ago, Aeschylus' oldest play, *The Suppliants*, depicted a perfectly contemporary request for asylum. The 50 daughters of Danaos requested and obtained asylum from the people of Argos to escape forced marriage. Today, recognition of refugee status for this type of human rights abuse is not always the case in Europe, although Canada recognizes gender persecution. Escape from wars and religious persecution has existed in all of human history, but until the 16th century, asylum was only for common law offenders. Sacred sanctuaries of antiquity and, later, Christian churches were asylums placed under the protection of the gods, where criminals could obtain temporary protection from the private vengeance of victims. The function of this asylum was to correct the weaknesses of the state justice system, and in reducing private violence, it indirectly reinforced the monopoly of legitimate violence detained by the lords. Because the Prince's interests took precedence over those of society, the inviolability of sanctuaries was often transgressed in political cases (Bettati, 1985). From the 16th century, the development of monarchies and civil justice questioned the basis of asylum for common law delinquents. Protection from the gods was no longer necessary for maintenance of order, and because it cast a shadow on the absolute power of the king, it had to disappear. In the opposite direction, political asylum found its own roots in the relativity of crimes of opinion, as the states, becoming increasingly powerful and protective of their territorial prerogatives, were no longer all condemning the same convictions at the same time. The exile of about 200,000 Huguenots in the 17th century and their generous welcome in France's neighboring states marked the beginning of the modern tradition of asylum in Europe.

When subjected to reason of the state, the political asylum of foreigners was very controversial, and the legitimacy of the protection given to the politically persecuted established itself very slowly. The French Revolution was a new stage that exalted resistance to oppression and placed political asylum in the constitution of 1793: "The French people shall provide political asylum to foreigners banished from their fatherland for the cause of freedom. It shall refuse asylum to tyrants" (article 120). This proclaimed right to asylum remained purely virtual because France at that time was no more a land of asylum than its neighbors, and the construction of national identities was progressively imposing a control of international migration (Noiriel, 1991).

II. THE INTERNATIONAL SYSTEM FOR MANAGING REFUGEE MIGRATION

The need for an international system for managing refugee migration appeared forcefully after the First World War. In a few years, Europe had had to cope with the resettlement of war refugees, the influx of about 1 million refugees of the Russian Revolution, and the consequences of the Armenian massacres in Turkey. Under the auspices of the League of Nations, a series of ad hoc international agreements were signed to regulate the status of the various groups of refugees. In 1921, the Nansen passport,[1] ancestor to the travel document of current refugees, was created to provide Russian refugees with an identity document to enable them to travel. The provision of this passport was then extended to other groups of refugees.

Before the Second World War, the mass arrival of Spanish and German refugees fleeing the Franco and Nazi regimes increased the necessity for a concerted management of these populations. The League of Nations consequently set up the Intergovernmental Committee for Refugees (ICR) in 1938, which adopted a very general definition for refugees and consequently for political immigration. After the Second World War, the United Nations Organization succeeded the League of Nations and created the International Organization for Refugees (IOR), which managed the resettlement of more than a million persons in a few years.

[1] Taken from the name of the President of the Norwegian delegation to the League of Nations. Founder of this travel document, he was an Arctic explorer and a renowned activist for refugees.

In 1951, the United Nations High Commission for Refugees (HCR) replaced the IOR, and during the same year, the Geneva Convention on the status of refugees was drawn up during the Cold War (Table 62–1). The jurisdiction of this convention, initially limited to events that took place before 1951 in Europe,[2] was extended to the whole world without restriction by the additional New York protocol of 1967.[3] This convention, ratified by more than 120 nations, has become the fundamental text for modern asylum. It defines a refuge as a person who "owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it."[4] It also defines the rights of refugees in their host countries and specifically stipulates that they must at least enjoy the rights of non-refugee foreigners. It also protects refugees against expulsion or transfer to a country where they may be threatened.

The large number of civil wars in Africa and the ambiguity of the Geneva Convention concerning persecution related to this type of situation[5] led the Organization for African Unity (OAU) to write its own convention in 1969, which explicitly includes in the definition of a refugee, victims of all types of persecutions, irrespective of the circumstances. Latin America did the same with the Cartagena Declaration in 1984. Since then, political migration is managed by a dual system; in the Third World, the legal definition of a political refugee corresponds to the classic sociologic definition that considers all persons fleeing persecution,[6] whereas in developed countries, legal recognition of refugees giving them residency rights are

[2] From the beginning, the signatory countries had the possibility of extending the effects of their membership to events in Europe or elsewhere.

[3] Often called the Bellagio Agreement, it was named after the city where it was written.

[4] See HCR 1979, for the text of this convention accompanied by its interpretation.

[5] According to certain restrictive interpretations of the Geneva Convention, it does not apply to any person who is not individually persecuted by his or her own country, which effectively excludes civil war, whose victims are not individually persecuted.

[6] We should not see in this the evidence of more generosity; because refugees were already present in the Third World, recognizing them did not lead to more migration while enabling them to benefit from international aid if necessary.

TABLE 62–1 Major Texts Cited on the Protection of Refugees

Texts	Date	Supporting institution	Territories concerned	Subject
Geneva Convention on Refugees	1951	United Nations	Initially Europe, then the world since 1967	Definition of the refugee and protection against expulsion at the border
New York Protocol	1967	United Nations	World	Extension of the Geneva Convention to the rest of world
Organization for African Unity Convention	1969	Organization for African Unity	Africa	Explicit recognition of generalized violence in the definition of a refugee
Cartagena Declaration	1984	Government of Colombia	Latin America	Explicit recognition of generalized violence in the definition of a refugee

subjected to constraints of migration policies, which are very restrictive in Europe.

III. GROWTH IN THE NUMBER OF REFUGEES IN THE WORLD

At the creation of the HCR in 1951, the problem of refugees was perceived as temporary[7]; it would disappear with the integration of all exiles in the first host or resettlement countries. In the 1960s, the question of refugees still existed but changed location. If movement toward Western countries was low, new waves generated by conflicts associated with the Cold War begin to appear in Third World countries. It is the strategic importance of an international management of these movements (HCR, 1995) that made possible a lifting of the limitations in time and space initially prescribed by the Geneva Convention.[8] A few years later, unending growth in the number of refugees began. First from Latin America and then Asia (e.g., 1 million Boat People from South East Asia, 5 million Afghans), Africa (i.e., from Mozambique to Liberia through Somalia and Rwanda), and Europe (4 million ex-Yugoslavs), all the continents successively contributed their share of refugees. Despite the wish of the international authorities to encourage a return at the end of every conflict, the number of refugees and other persons under the authority of the HCR rose from 2.4 million in 1975 to 27.4 million in 1995, a quasi-regular increase of 25 million in 20 years. Since then, new conflicts have succeeded the ones that are ending, but the number of persons under HCR authority has been on the decline, falling below the figure of 20 million on January 1, 2002 (Fig. 62–1).

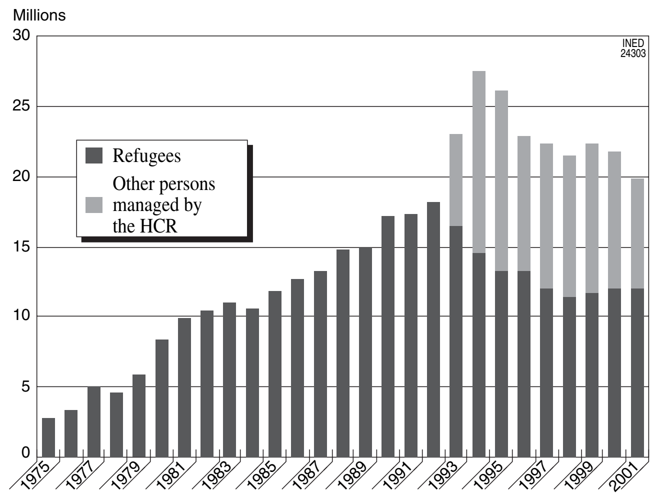


FIGURE 62–1 Trends in the number of persons managed by the United Nations High Commission for Refugees.

These figures do not include the 3 million Palestinian refugees registered by the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA). This office was created in 1949 to cater originally for 750,000 Palestinian refugees, and one-half century later, this problem remains unresolved, and the number of refugees has increased because of the continuation of the conflict (Israeli-Arab war of 1967) and especially because of natural demographic growth.

The estimate of the number of persons displaced by conflict within their own country and not under the control of the HCR, which by nature is even less precise than that of refugees, is greater than 20 million persons. Currently, political migration may be between 40 and 50 million persons in the world.

[7] The initial mandate of the HCR, drawn up to be very temporary for three years, has since been regularly renewed for 5 year periods.

[8] France ratified the New York Protocol in 1971, after speculating for a long time whether there was a risk of refugees from Africa arriving.

IV. DATA QUALITY

Between the political implications of the number of refugees, the materially difficult conditions of data collection, and the absence of standards in definitions, the available figures on refugees can hardly be better than gross estimates.

The political implications take many forms. Certain countries do not like to recognize the fact that their nationals are fleeing persecution. Others find it useful to attribute a large number of departures to their opponents. For others, misinformation on the number of refugees is an integral part of their political strategy, as was the case in the debate concerning the number of Hutu refugees in ex-Zaire and what became of them during the Kabila offensive.[9] In the host countries, diplomatic considerations sometimes bar access to information on the real figures. To admit receiving refugees from a country with which one maintains very good economic and political relations may be seen as embarrassing. There are also economic implications for the host countries, because most often, the amount of international assistance given to refugees depends on the media coverage of events accompanied by the most shocking images and the most impressive figures possible.

In the zone of origin, conditions do not often allow a precise count of the refugee population. When the situation is unstable and camps move frequently, estimates become even more approximate. The use of indirect methods becomes more common. For example, Médecins sans Frontières (Doctors without Borders) estimated the population of camps in the region of the African Great Lakes by calculating their surface with the aid of the global positioning system (GPS)[10] and then multiplying this figure by a coefficient determined through a survey of enumeration areas. Identification by aerial and satellite photography is also used.

Even in host countries with a strong statistical tradition such as in Western Europe, statistics are not reliable. Even if the validity of data provided by the national authorities is not in doubt, enormous progress remains to be made in the standardization of estimates. For example, until 2002, French figures were based on the number of files of asylum requests, which did not include child minors, whereas German figures often included all members of the family. France still makes available only figures concerning statutory refugees as strictly defined by the Geneva Convention, whereas the German figures include all refugees, whether or not they have obtained the status of refugee according to the Geneva Convention.[11] Another example shows that integration within the host population does not make political refugees disappear from statistics at the same pace in all countries. France more easily naturalizes refugees than Germany. If in the past few years, the European Union has been trying to standardize statistics on asylum seekers (i.e., entries within the common territory); it is totally uninterested in the exits and makes no effort to measure the number of statutory refugees present in the territory every January 1. The European Union does not publish global statistics on this population. It is improbable that this relative ignorance of the total number of refugees protected by the Geneva Convention in Europe is fortuitous, and such a strategy is unfortunate. In France, statistics on statutory refugees, confidential for a time, show that their number has been on the decrease in recent years.

The number of refugees in the world depends equally on the notion employed; the population covered by the HCR includes conventional refugees (recognized within the Geneva Convention, the OAU Convention, the Cartagena Declaration, or HCR Statutes) and various other categories of persons whose definition changes regularly. On January 1, 1995, the 27.4 million persons depending on the HCR consisted of 14.5 million conventional refugees, 3.9 million recently repatriated refugees, 3.5 million other persons, and 5.4 million internally displaced refugees. On January 1, 2002, there were 12 million refugees (but these now included the various precarious or humanitarian sub-statuses created in Western Europe), including almost 1 million asylum seekers, 500,000 repatriated refugees, 5 million internally displaced persons, 240,000 internally displaced repatriated refugees, and 1 million "others." Even independently of these changes in definition, a time-series study of the HCR statistics is very problematic. For example, the distribution between conventional refugees and "others" in the German figures is stunning. At the end of 1999, 2000, and 2001, the number of conventional refugees was, respectively, 252,000, none, and 518,000; the category "others" was, respectively, 723,500, 906,000, and 385,000. The total, however, seems more coherent with, respectively, 975,500, 906,000, and 903,000.

[9] In this case, Western countries supporting rival parties are equally responsible for the misinformation.

[10] Equipment initially designed for aerial and sea navigation which provides a precise position in space based on calculation according to several satellites.

[11] The way of identifying nonstatutory refugees sometimes varies. Between January 1, 1995, and January 1, 1996, the number of refugees in the German statistics compiled by the HCR declined by more than one-half after a change in definition.

The category "conventional refugees" in the strict sense is far from homogeneous. The 14.5 million refugees estimated in 1995 were based on two very different notions. In the Third World, they were defined as victims of persecution of all origin, whereas in Europe, the restrictive interpretation of the Geneva Convention excluded the victims of general violence.[12] If this restrictive definition were to be applied all other the world, the number of refugees in the strict sense of the word would be much less.

V. THE CAUSES OF GROWTH

The increase in the number of refugees in the world since the mid-1970s goes hand in hand with the evolution in the nature of conflicts. A very high level of political instability and a certain "inward withdrawal" of Western Europe marked the second half of the 1970s (Rufin, 1996). Within a few years we witnessed the fall of Saigon; the arrival of new Marxist regimes in Angola, Mozambique, Guinea Bissau, Nicaragua, and Ethiopia; the entry of Vietnam into Cambodia; the invasion of Afghanistan by the USSR; the Lebanese war; the crisis in the Western Sahara; the Tamil insurrection; the Iranian revolution; and the invasion of East Timor by Indonesia. All these upheavals were accompanied by interminable guerrilla warfare whose first victims were civilians. The major characteristic of these so-called low-intensity conflicts is that direct combat between armies is rare, but civilians and economic infrastructures are systematically chosen as targets, thereby leading to an increasing number of refugee movements.

The collapse of the USSR at the beginning of the 1990s reopened a new period of great instability. As observed by Jean-Christophe Rufin (1996), "Since 1990, the possibility of a radical definition of territorial demarcations is no longer taboo, thereby opening a new pathway for the action of armed movements." The disintegration of ex-Yugoslavia is a perfect example of a redefinition that resulted in almost 4 million refugees and displaced persons in Europe. In Africa, the end of the Cold War opened up new political and economic perspectives[13] to guerrillas abandoned by their Western or Communist tutors, and the disintegration of certain countries such as Somalia or Liberia is almost complete. In Rwanda, the Tutsi genocide by the Hutus, followed by the military victory of the Tutsi guerrillas, led to the displacement of 4 million persons, of whom 2.2 million were Hutu refugees outside the country.

Major refugee movements are brought about by civil war. Even if persecutions and individual infringements of human rights are extremely high in the world, victims of collective violence are responsible for most of these movements.

VI. THE DESTINATION OF THE FLOW OF REFUGEES

Refugee movements in the world are extremely local. According to refugee statistics by country of asylum and country of origin published by the HCR, more than 95% of refugees in Africa and Asia found refuge in a neighboring country. Even in Europe, reputed for receiving refugees from all continents, more than one-third of refugees have been European since the wars in ex-Yugoslavia and the disintegration of USSR. These figures concern *stricto sensu* refugees or people recognized as such by a country or by the HCR.[14] It excludes essentially all local movements of "persons displaced within their own country," who are almost as many.

Beyond the neighboring countries and on a worldwide scale, Western Europe and North America (i.e., United States and Canada) are essentially the only destination for Third World refugees.[15] They receive a little more than 1.5 million non-European refugees, who make up about 13% of the total number of refugees and roughly two-thirds of the total in Europe. Migration of refugees is growing equally in all of Eastern Europe. Most of it is purely a local response to the collapse of frontiers in the former Eastern Bloc, but there are also arrivals of refugees from the Third World. For the latter, Eastern European countries are generally only one step toward Western European countries, whose direct access has been made very difficult by restrictions on immigration.

If the destination of "neighboring country" can be immediately understood, the logic governing the destination of "Western country" is much more complex, and an analyst does not systematically perceive the nationalities that have the highest number of refugees in the flows toward developed countries.

[12] Because Europe could not alienate itself from exiles present in its own territory, most of the victims of the conflict in ex-Yugoslavia are recorded in the category "others" of the HCR.

[13] For many leaders of armed movements, profits from illegal goods allowed by a war situation are at least as high a motivation as ideology.

[14] In Europe, Germany includes in its statistics protected war refugees who are not protected by the Geneva Convention.

[15] The other very developed democracies accept only a marginal number of refugees (e.g., 3200 in Japan; 55,000 in Australia; 5300 in New Zealand).

For example, if Afghanistan, Iraq, and Somalia, which were, respectively, the first, third, and sixth country of origin of world refugees at the end of 2001, were the main nationalities of refugees in Western countries; Burundi, Sudan, and Angola, which were, respectively, the second, fourth, and fifth, do not figure. We could also compare the 122,000 refugees from Sri Lanka in the world at the end of 2001, who were mostly present in Western countries, with the 111,000 refugees from Bhutan present at the same time but almost nonexistent in the same countries.

Explanatory factors are multiple and interrelated. First is the very nature of disorder that generates refugees. If most of the movements are purely local, this is because they are of victims of collective violence and that fleeing from massacre leaves no choice of destination. Absolute famine sometimes created by predator guerrillas themselves does not leave a choice either. These major collective movements of flight correspond to the logic of minimum displacement as the destination is generally fixed by the geography of the zones of conflict and refuge is found behind the first international boundary. It is this same logic of minimum displacement that is equally at the origin of the high number of "displaced persons," those who find refuge in another region of their own country and live in the same precariousness as refugees who succeed in reaching a neighboring country. In emergencies, these refugees or displaced persons, even if they wanted to, have absolutely no means of reaching Western countries. The major movements of refugees given the widest media coverage contribute little to the migratory pressure in Europe, except when these movements last a long time.

The destination of individual departures is more complex. Emergency flight in the wake of individual persecution, whether by a dictatorial state or armed groups, is common, but contrary to common opinion, all individual departures are not totally improvised. On the contrary, when the consequences of conflict in social organization is at the source of exile or when the state is deliquescent and the resulting arbitrariness, corruption, and generalized violence make the situation dangerous and life impossible, people start to leave. Not everybody leaves at the same time; the departure date depends on the level of each person's fears, the level of saturation, and especially the level of available resources. These financial and human resources, especially knowledge of migration networks or previous ties with a foreign country, determine to a large extent the chosen destination. Depending on available means, the possibility of choice for refugees will be limited to neighboring countries (i.e., individual movements in the region of origin) or will extend to rich countries. The itinerary to developed countries is most often indirect because of difficulties in the country of origin, particularly the lack of freedom to leave the country if it is run by a dictatorship, and because of restrictions on immigration in the developed countries themselves.

Preexisting organized migratory networks, or those created in recent globalization exchanges[16] (Angoustures, 1996), play a major role in the destination of movements toward the West. In this case, the notion of political asylum and economic immigration slowly blend to become one among asylum seekers.

VII. ASYLUM SEEKERS

The term *asylum seeker* was coined in the 1970s to differentiate refugees who had obtained their status from the mass of persons awaiting the decision that was becoming more and more negative. This expression, which had previously not been used, became necessary when statistically the asylum seekers stopped being potential refugees to become potential rejected asylum seekers.

If certain justified applications for asylum based on indisputable individual persecution have for a long time been rejected because of restrictive jurisprudence that is progressively being abandoned in European countries (France was the last to abandon this at the end of 2002), there are also asylum applicants whose motivations are purely economical. However, these requests are fewer than Western countries would like to make people believe, and in the many cases in which the economic component is strong, it is by no means unique. China is not a model of democracy; the most elementary human rights are often contravened, torture has been institutionalized, and the death penalty is common. One-fourth of the requests for asylum from Asians in 1996 in France came from Chinese. However, it would seem that political opposition members cannot come to France because a major part of the applications come from the Province of Zhejiang, especially the city of Wenzhou, and that the only persecution expressed is that of the one-child policy. According to asylum seekers from Wenzhou, they come to France to improve their conditions of life and escape the control on number of births, especially when their first child is a girl. This predominance of applicants from the city of Wenzhou has historical

[16] The control of commercial migration networks is sometimes in the hands of armed opposition movements, such as the Kurdish groups that control the bus companies and the Tamil groups that manage the networks of travel agencies.

roots going as far back as the First World War. France invited about 140,000 Chinese, almost all repatriated at the end of the war, with the exception of a little group among whom were people from Wenzhou. The latter then initiated a network of migration that prospered between the two wars, was interrupted during the Second World War, and was stopped during the Mao regime, but which took off again with the opening of China since 1979 (Porquet, 1997). France now considers that economic motivations are confirmed and that French society can no longer accept limitless Chinese immigration. The French policy is that limitation of births is not persecution, and consequently, France cannot but systematically reject all Chinese asylum requests. However, major world conferences on population have all affirmed that individuals have "the fundamental right to freely decide in all responsibility and without constraint the number and spacing of their children" (Mexican Declaration on Population and Development). It is true that this declaration aims at promoting family planning and birth limitation programs, but does a fundamental right have to be relative to the country in question? According to the perfect structure of all tragedies, the asylum seekers whose requests have been rejected by OFPRA and OFPRA[17] itself have the law on their side.

There are more examples for which economic factors coexist with political factors, and they show the importance of migratory networks. In effect, the existence of migratory networks has become more and more inevitable with the beginning of very restrictive migratory policies in Europe that practically refuse any legal arrival of asylum seekers.

VIII. ASYLUM POLICY

Speaking of asylum policy in the singular is no doubt excessive, because European countries have not harmonized their methods. However, the dominant attitude, which is clearly repressive, has led to a certain uniformity in practice.

The policy of restrictive asylum is based on two pillars. The first consists in refusing asylum seekers so as not to have to refuse them asylum (Marie, 1996). The second one is based on an interpretation of the Geneva Convention that disconnects the status of refugees from that of the need for protection. In democratic countries, asylum seekers are not refused as such; they instead suffer from measures of general immigration control whose first aim is to prevent them from reaching Western countries and then to transfer them to another country if they arrive.

1. Retention

Retaining asylum seekers at the borders of developed countries is the objective of two complementary measures: imposition of visas on all nationals of potential asylum-seeking countries and fines on transporters of persons not traveling with obligatory travel documents (i.e., visas). The aim is to retain these illegal immigrants at their source by delegating the power of the police to the transport firms (Julien-Laferrière, 1996). These measures have the advantage of replacing the refusal of admission at the borders and the costly and difficult procedures of expulsion with that of refusing embarkation at the country of origin. The effectiveness of these measures lies in the level of the fines imposed and on the list of countries subjected to visas. Currently, 126 nationalities are subjected to visas by all the countries who signed the Schengen Convention.[18]

France provides an example of how general migratory measures affect political immigration. The combined measures of visas and fines on transporters (the two measures are effective only when combined) distort the spirit of the Geneva Convention (Legoux, 1995). A person persecuted in his or her country can expect to encounter difficulties in obtaining all the necessary travel documents, and article 31 of the Geneva Convention stipulates that irregular entry or stay should not bar refugees. In France, article 10 of the 1952 law on the right to asylum (2002 version)[19] confirms this general disposition: "admission may not be refused for the sole reason that the foreigner does not have the documents or visas mentioned." However, article 20b of the ordinance of 1945 on the entry and stay of a foreigner in France[20] indicates that the fine on transporters is not inflicted in two cases only: if the foreigner "who asks for asylum has been admitted into French territory or if the application is not obviously

[17] The French Office for the Protection of Refugees and Stateless Persons (OFPRA) is the organization responsible for recognizing the status of refugee.

[18] France adds 21 other states to a common list of Schengen nations, and the fines for transporters is 10,000 francs (about $1500) per irregular passenger transported. According to a parliamentary report on underground immigration, called the Philibert-Sauvaigo report, the fines inflicted on Air France alone in 1995 amounted to 5 million francs ($750,000), to which was added 6 million francs ($900,000) of charges for the return and the escort of illegal immigrants.

[19] This article borrows extensively from the former article 31b of the 1945 ordinance in its 1993 version.

[20] The ordinance of November 2, 1945, modified version, on conditions of entry and residence of foreigners in France (2002 version).

unjustified" or "if the transport firm establishes that the necessary documents were presented to it at embarkation or if the presented documents do not have any obvious irregularity." Because the agents of the transport firm cannot determine that the border police will recognize the obviously unjustified nature of the asylum request nor the quality of the documents presented, they tend to refuse any passenger without a visa, even if she or he declares wanting to seek asylum on arrival in France. Formally, the absence of a visa is not a condition for refusing asylum requests, but in practice, an asylum seeker cannot use regular transport firms and is forced to enter France illegally. These measures were very effective for the first few years they were in effect, and the number of requests fell from 61,422 in 1989 to 17,405 in 1996. Since then, the networks of irregular entries have adapted themselves, and the applications increased to 47,291 in 2001.

2. Expulsion

When asylum seekers manage to reach a country that signed the Schengen Convention, expelling them is the objective of re-admission agreements. Because of the policy of visas, direct itineraries between the asylum seekers' country of origin and Western countries have become rare and have been replaced by complex itineraries ending with a land entry through the borders of Schengen countries. These countries "have thereafter instrumentalized these itineraries and developed concepts of 'third-party host' or 'secure third-party country' " (Créach, 1997). According to the London resolution of 1992,[21] third-party host countries are those in which the life or liberty of asylum seekers is not threatened and where they have already obtained some protection or *have had the possibility to ask for it*. As soon as it is possible to send an asylum seeker back to such a country, the examination of their request may be refused. Consequently, any country in which the asylum seeker has stayed, even if it is in transit for a couple of hours, may become a destination to return him or her to, if it is considered safe and if it accepts to readmit the asylum seeker. The readmission agreements are precisely aimed at obtaining the agreement of these countries and are sometimes accompanied by financial compensation. The agreement between Germany and Poland provided 120 million German marks (60 million dollars) to the police of Poland, and another between Germany and the Czech Republic provided 60 million German marks (30 million dollars) to the Czech security forces involved with guarding the borders.

The generalization of readmission agreements[22] gives rise to the possibility of true "expulsion chains" resulting from cascading readmissions from third-party country to third-party country, pushing asylum seekers to the periphery of Western countries. The ultimate objective of the readmission agreement is to constitute "a network of third-party countries promoted to the grade of advanced border-guards of the Schengen plan of action. . . ." The major consequence of this "advance protection" is to remove all meaning from the principle of "non-refoulement" or nonexpulsion of the Geneva Convention by replacing it with the principle of nonentry (Marie, 1996).

3. Interpretation of the Geneva Convention

After asylum seekers have reached a Western European country, applied for asylum, and are not the object of a readmission agreement, the final obstacle is the analysis of their request by the competent authorities. The level of rejection is very high, about 90% of requests in most countries in the Schengen area.[23] This rejection is based on an extensive use of the notion of "obviously unjustified request" and on the systematic doubt concerning persecutions not proved beyond reasonable doubt. The notion of "obviously unjustified" makes it possible for most countries to use procedures of rapid and expeditious judgment that often dispense with providing temporary residence rights to the seeker. The proof of persecution is not formally demanded, but the benefit of the doubt is never to the advantage of the asylum seeker. Illegal entries due to visa refusals and the fear of being expelled oblige asylum seekers to lie about their itinerary, which increases the doubt.

In total opposition to the HCR understanding of the Geneva Convention, France and Germany, followed by other European countries, wanted to reduce the number of refugees by excluding all nonstate persecutions from the Geneva Convention. Persons recognized as being in danger in their countries but whose request for asylum had been rejected could nevertheless, in certain cases, obtain the right to a more or less protective residence for humanitarian reasons. This is

[21] "Resolution on a harmonized approach of issues relative to third party host countries" was adopted in London on December 1, 1992, by the Ministers of Member States of the European Committee in Charge of Immigration.

[22] The asylum section of the Schengen agreement is already a readmission agreement between all the member states.

[23] The proportion of recognition displaced by France is higher (up to 20% in 1996), but close to one-half of the recognition considered relates to purely official requests of children of refugees already present in France and approaching 18 years old.

called subsidiary protection, humanitarian status, or B status. This restrictive interpretation of the Geneva Convention, which contributed significantly to the development of subsidiary protection in the past 10 years, has been abandoned, but the sub-status designations remain.

IX. NEGATIVE CONSEQUENCES

Effectively prohibiting the regular arrival of asylum seekers provokes fraudulent commercialization of assistance to cross the borders of developed countries, which itself has serious negative effects. The first consequence is the selection by money; the cost of introduction into Western Europe by Mafia-type professionals is between 20,000 French francs ($3000) for Turks to 100,000 francs ($15,000) for Chinese from Zhejiang; the trip between Shanghai and the United States can cost as much as $50,000 (Porquet, 1997; Marie, 1996). Even if a large part of this amount is payable on arrival after having found an illegal job, the victims of persecution forced to hastily leave their country are not in a position to employ these indispensable networks to cross the borders of developed countries. Another penalizing effect for political asylum seekers is that the authorities consider the use of immigration networks as an indicator of economic motivation for the asylum request. The sale to candidates of false papers and stereotyped asylum requests along with the trip contributes to this disqualification of request, especially because the necessity of dissimulating the conditions of their arrival pushes them to tell unconvincing tales. The traffic of persons is often accompanied by that of legal and illegal products, and this association highly discredits political immigration in public opinion. The cost of access to the asylum request provokes indebtedness, which makes return immigration impossible and imposes an underground life for asylum seekers whose applications have been rejected.

The measures designed to hold back the incoming flow of asylum seekers create the conditions for rejecting numerous requests, thereby creating a vicious circle in which a lack of respect for the right to asylum is self-justified.

X. INFORMAL ASYLUM

Before immigration restrictions were implemented in Europe in the mid-1970s, numerous political exiles found refuge in the Western democracies under the cover of the status of economic refugee. Consequently, the political immigration of Greeks and Portuguese hardly used the status of refugee, and many Republican Spaniards who arrived in France in 1939 after the Franco regime took power did not ask for refugee status until 10 or 20 years later. At that time, economic immigration was a possible alternative to political immigration, and it offered the advantage of not breaking ties with the country of origin. In France, even after 1974, many refugees from Chile did not ask for asylum until many years after their arrival, when the status became indispensable to avoid being expelled. When immigration for work was stopped, these incoming flows of political exiles very logically transformed themselves into those of asylum seekers. Currently, it is political asylum that is less accepted, and political exiles use the latest avenues possible: temporary protection provided by the duration of the examination of the asylum application and underground immigration.[24]

Confronted with massive rejections of their asylum requests, these exiles have obviously developed alternative strategies using the request for asylum as a means of temporarily regularizing their situation. Multiple applications in many countries (i.e., asylum shopping) have been a strong justification for the asylum section of the Schengen plan of action and then of the Dublin Convention, but this plan of action does not seem to be dissuasive. In France, the example of Romanians is very instructive about the use of asylum procedures to obtain protection normally refused by the host country. After the fall of Ceausescu, France decreed that Romania had become a democratic country and undertook to remove asylum status from all previously recognized Romanian refugees; under these conditions, new applications for asylum have been systematically rejected. In 1996, there were more than 4000 Romanian applications, amounting to 23% of the total applications from all nationalities. Apparently, the geographic proximity reduces the efficiency of procedures for holding back the flow. According to many studies,[25] it seems that these applicants, often Gypsies, have assimilated the fact that their applications would be rejected, but they have continued to use the asylum procedure in a sort of circular migration that provided them the financial resources necessary to fight against the social and economic discrimination

[24] In certain cases, it may be preferable to speak of illegal immigration. In France, the public struggle of the Sans-Papiers (those without papers) has shown that they are not illegal aliens in the strict sense of the word.

[25] Studies were directed by Olivier Brachet, an economist and the director of Forum-réfugiés (Refugee Forum, formerly Comité rhodanien d'accueil des réfugiés et de défense du droit d'asile), which is based in Lyon.

of which they were victims in Romania. Since 2002, Romanian nationals are exempted from the visa obligation to enter France, and the requests have greatly diminished.

Underground existence is the most precarious form of protection, but it is the ultimate refuge for many seekers whose applications have been rejected and who are awaiting a hypothetical regularization of their papers. Depending on the country, the presence of illegal aliens is more or less accepted. In France or in Italy, this seems to be a quasi-normal way of managing persons (Brachet, 1997), and this informal extralegal asylum contributes to the debate around conventional asylum.

XI. TRENDS

Even though it takes place in the context of democratic countries, the subjection of asylum policy to migration policy leads to negative practices that are shaking the whole edifice of the right to asylum. The retention of refugees outside developed countries, the legal controls that promote the nonrecognition of the status of refugee, and the multiplication of substatus designations and informal asylum show that the great principles of the protection of victims of political persecutions, at one time recognized for its universal nature by the Geneva Convention, are not respected in European democracies. For many years, the fear of a decline in the protection of human rights in the case of renegotiation of what is sometimes called *The Geneva Regime* (Zolberg, 1992) has led defenders of asylum rights to affirm the unsurpassable character of the 1951 Convention. However, the Geneva Regime is being progressively eaten from the inside by its very founders, and it is now time to explore new solutions.

1. Temporary Protection

For many years, provision has been made for resorting to temporary protection. A former director of international protection at the HCR already mentioned this in 1992 as a possible solution for Europe (Moussalli, 1992). In abandoning the right to permanent residence for statutory refugees, temporary protection makes a partial disassociation between asylum policy and migration policy possible. During the Kosovo crisis, most countries in the European Union set-up temporary *de jure* or *de facto* protection. France created a special plan of action for people from Kosovo, offering them temporary residence rights and the right to work if they did not ask for conventional asylum. Because this crisis had exposed evidence of differences in the treatment of temporary refugees in the European nations, the European Council in 2001 adopted a directive on temporary protection meant for massive incoming flows of displaced persons. It is an exceptional mechanism that offers a 1-year renewable protection and the right to work and to social benefits. It can be set in motion by the Union's Council of Ministers. The principle of double accord, that of the persons concerned and that of the host country, preserves the sovereignty of states in this European procedure.

Subsidiary protections are also a form of temporary protection. In France, territorial asylum was created in 1998 because of the issue of Algerians persecuted by groups of armed Muslim militia who could not be recognized as refugees because of jurisprudence excluding nonstate persecution from the protection of the Geneva Convention. At the end of 2002, France gave up this jurisprudence but retained territorial asylum, which was meant to be renamed subsidiary protection and administered by OFPRA[26] in a new reform of asylum planned for 2003. This temporary asylum is explicitly subjected to the interests of the host country in the very text of the law:

> In the conditions compatible with the country's interests, territorial asylum may be accorded to a foreigner by the Minister of Interior after consultation with the Minister of Foreign Affairs, if the foreigner establishes that his or her life or liberty is threatened in his or her country or that she or he is exposed to treatment contrary to article 3 of the European Convention of the protection of human rights and fundamental freedom. The decisions of the Minister do not have to be motivated.[27]

Statistics show a rapid growth in territorial asylum: 6984 applications in 1999; 11,810 in 2000; and 31,000 in 2001, according to figures cited in the asylum reform project (but only 15,700 in the preliminary statistics of the Interior Ministry). The number of residence permits delivered is very low: 292 in 1999, 351 in 2000, and approximately the same in 2001.

The major problem in using temporary protection is the duration of the temporary period. In the case of quick repatriation, obliging refugees to return to their countries if all security conditions are met. However, when conditions at the origin of exile cease after many years, certain refugees who are well integrated in the host society may see possible expulsion as a new form of persecution based on their nationality. The principle of temporary protection therefore separates asylum from the hazards of the migration policy only very

[26] Communication of the Minister of Foreign Affairs during the Council of Ministers meeting of September 25, 2002.

[27] The law on asylum rights of July 25, 1952, was modified by law no. 98-369 of May 11, 1998.

imperfectly, especially if the motivation for exile is perceived as resulting from a long-term situation. For example, persecutions based on sex are associated with cultural considerations that are reputed to evolve only very slowly, and it is unlikely that protection given for such a reason can be considered really temporary.

2. Defense of the Right to Remain in One's Own Country

A statement from the HCR perfectly summarizes the evolution in United Nations doctrine on the problem of refugees:

> During the years of the Cold War, the determination of Eastern countries to prevent the departure of their citizens led the international community to concentrate most of its attention on the right to leave one's home country. . . . Those who succeeded were transformed into heroes, an attitude which reflected and conditioned the exile-oriented approach as a response to the problem of refugees. . . . But since the collapse of the Soviet Bloc, the wheel has turned such that it is more a question today of the right to remain in one's own country and that of not being displaced (HCR, 1995).

The fine-tuning of the human rights issue is glaring. From an instrument in the East-West conflict, acceptance of refugees has become a liability that henceforth must be borne by the countries of origin. This total inversion of perspectives explains the current interest in the prevention of refugee movement and consequently the prevention of conflict.[28] At the same time, we discover the negative effect of resettlement programs or the prolonged assistance of refugee populations that have sometimes been an obstacle to the search for solutions in their countries of origin.[29]

The HCR firmly states that actions in the country of origin must not be transformed into arguments to justify the closure of borders to refugees nor used as a means of locking them in their countries under the pretext that international humanitarian organizations have created zones of security within these countries, because these zones are more often virtual than real, as has been shown by the massacres in the humanitarian enclaves of Bosnia.

The new paradigm of the right to remain in one's country cannot be conceived without political development. Respect for all individual freedom is necessary for the exercise of the right to remain. It cannot be conceived without economic development, because economic exploitation is at the origin of many conflicts, and periods of economic regression are the most conflicting ones because of the increase in competition for the control of resources (HCR, 1993). This paradigm therefore revisits the old problem of development as an alternative to purely economic migration and encounters the same limitations. At the first stage, a certain freedom regained from the emergence of a new way of life may increase migration instead of reducing it. At a second stage, development as a solution to migration implies a very strong reduction in inequalities between countries, whereas the process of globalization tends to amplify them.

While recognizing the declared wish of the United Nations to attack the causes of political migration, it is impossible not to recognize its chronic impotence, whether in Sarajevo, Grozny, or the catastrophic extermination of Rwandan refugees in the forest of Zaire. The right to remain in securely within one's own country is still an elusive objective like that of a better world permanently rid of war. Before that great day arrives, political pressure will remain a strong determinant of international migration.

[28] This interest is not new; in 1985, the UNO affirmed "the causes of refugee movements constitute an essential aspect of the solution and international efforts must also be geared toward the elimination of these causes. We also should be more interested in the causes and the prevention of refugee movements" (Michel Moussalli 1992).

[29] "Currently, it is believed that the international community's response to refugees in the 70s and 80s is perhaps not foreign to the level of the problem in the 90s" (HCR 1995).

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Angoustures Aline and Pascal Valérie, 1996. Diasporas et financement des conflits, *in*: Francois Jean and Jean-Christohpe Rufin (eds.), *Économie des guerres civiles*. Paris, Hachette, 593 p.

Bettati Mario, 1985. *L'asile politique en question*. Paris, PUF, 205 p.

Brachet Olivier, 1997. L'impossible organigramme de l'asile en France, *Revue Européenne des Migrations Internationales*, vol. 13(1), p. 7–36.

Marie Claude-Valentin, 1996. L'Union Européenne face aux déplacements de populations. Logiques d'État et droits des personnes, *Revue Européenne des Migrations Internationales*, vol. 12(2), p. 169–209.

Créach Xavier, 1997. La notion de 'pays tiers sûr' ou l'instrumentalisation des itinéraires par les États d'accueil, *Recherches & Asile*, vol. 1(2), p. 23–31.

Eschyle, 1964. *Théâtre complet*. Paris, GF-Flammarion, 250 p.

HCR (United Nations High Commission for Refugees), 1979. *Guide des procédures et critères à appliquer pour déterminer le statut de réfugié*. Geneva, HCR, 107 p.

HCR (United Nations High Commission for Refugees), 1993. *Les réfugiés dans le monde. L'enjeu de la protection*. Paris, La découverte, 191 p.

HCR (United Nations High Commission for Refugees), 1995. *Les réfugiés dans le monde. En quête de solution*. Paris, La découverte, 263 p.

Julien-Laferrière François, 1996. Les accords européens, la raison des États et la maîtrise des flux migratoires, *Hommes & Libertés*, no. 89–90, *Hommes & Migrations*, no. 1198–1199, p. 49–54.

Legoux Luc, 1995. *La crise de l'asile politique en France*. Paris, CEPED, 344 p. (Coll. Les Études du CEPED, no. 8.)

Moussali Michel, 1992. La contribution du HCR pour la protection des réfugiés, *in*: OFPRA (ed.), *Les réfugiés en France et en Europe: quarante ans d'application de la convention de Genève, 1952–1992*, p. 33–44. Paris, OFPRA, 637 p. (Actes du colloque, 11–13 juin 1992.)

Noiriel Gérard, 1991. *La tyrannie du national. Le droit d'asile en Europe*. Paris, Calmann-Lévy, 355 p.

Philibert Jean-Pierre and Sauvaigo Suzanne, 1996. *Immigration clandestine et séjour irrégulier d'étrangers en France*, vol. 1. Paris, Assemblée Nationale, 2 vol., 232 + 489 p.

Porquet Jean-Luc, 1997. *Les clandestins*. Paris, Flammarion, 403 p.

Rufin Jean-Christophe, 1996. Le temps du monde rebelle, *in*: Jean-Marc Balencie, Arnaud de La Grange (eds.), *Mondes rebelles*. I. *Amériques, Afrique*, p. I–XXVI. Paris, Michalon, 564 p.

Zolberg Aristide R., 1992. Le régime de Genève dans la nouvelle donne internationale, *in*: OFPRA (ed.), *Les réfugiés en France et en Europe: quarante ans d'application de la convention de Genève, 1952*, p. 19–21. Paris, OFPRA, 637 p. (Actes du colloque, 11–13 juin 1992.)

C H A P T E R

63

Individual and Social Motivations for Migration

DANIEL COURGEAU AND EVA LELIÈVRE

Institut national d'études démographiques (INED), Paris, France

Any study of the motivations for migration behavior must also examine those that cause individuals to stay in the same place. Moreover, there is no reason to suppose that these motivations are the same for people who move short distances (e.g., intracommunal migration) as for people who move longer distances (e.g., interregional migration) or for nonmigrants, who do not change residential location. In this chapter, therefore, we examine both nonmigrants and migrants according to the type of *mobility* that they experience.

As a general rule, the term *mobility* is used here to denote all the moves undertaken by the individuals in the group being studied. Depending on the type of study to be conducted and the data available, it is possible to consider some moves as *migrations* while excluding others from analysis.[1] However, the individual lives in a space where administrative boundaries (e.g., communes, departments, and regions in France) often play a limited role in mobility within the country while possibly playing a more important role in international migration.

An analysis of the motivations for migration behavior also needs to consider them in interaction with all the other elements in the life of the individual. The reasons that "push" individuals to move, like those that cause them to stay in their group of origin or current dwelling, can only be understood when related to their past experience and present circumstances in the domains of family, work, and so on. The attraction of a higher income powerfully favors a move for a young adult living alone, whereas it has far less impact for an older individual with a wife who works and children, because of the risk of losing the wife's income and a reluctance to uproot the children from their school. In these conditions, motivations cannot be spoken of in absolute terms, but always with reference to the context of individuals in domains such as family life and career.

It is also necessary to introduce the information available to the potential migrant about the advantages and disadvantages of the choices between different destinations. Individuals are more likely to choose a destination region about which they have ample information through an extensive network of contacts than one about which their information is limited and obtained merely though the press. However, an individual who is well integrated in a neighborhood or town or city where most of his friends and relatives live has little incentive to move to another place.

The definition and measurement of motivations can vary greatly depending on the level of aggregation at which we are working. When survey data are being used, the motivations for migration can be introduced directly, if the survey subjects have been questioned about them. It is also possible to examine the dependencies between residential relocations and the individual's other life-history processes. The latter

[1] The definition of migration depends on two criteria—distance and period of residence—which are determined in the context of each individual study (see Chapter 22).

approach, less subjective than the former, is better suited for identifying the nature of the reasons for moving. Rather than working at the individual (micro) level, an aggregated (macro) level can be adopted when using census data, for example. Stronger assumptions then have to be made about the migration flows observed and about the various characteristics measured for each zone. A model of individual migration behavior underlies such an analysis (Puig, 1981). Validation of these underlying models calls for a more detailed analysis of the linkages that exist between micro-level and macro-level models.

I. THE COMPLEXITY OF MIGRATION BEHAVIOR

Before considering the individual reasons that affect mobility decisions, we need to take another look at the complexity of the analysis of geographic moves and the influence on them of other events. This analysis, i.e. an examination of the endogenous factors that motivate moves, only becomes meaningful if we envisage the interactions between multiple processes. For although a move can be motivated by a family event, another event affecting the family can be a cause of nonmigration or be the consequence of a move.

There are two elements to consider: mobility just as much as immobility and the implications of mobility for the demographic processes occurring in the other life domains of individuals. The latter aspect is not the subject of this chapter and is therefore not discussed further. Nonetheless, we felt it important to stress the need to envisage reciprocal dependencies in the analysis of individual behavior (Lelièvre, 1992).

1. Another Look at the Effect of Age

Migration rates by age based on data from the French 1999 Census exhibit a highly characteristic profile (Fig. 63–1). These are annual rates for residential mobility (movers) and for interregional migration.[2] This is a relatively universal profile, very close to that observed in other national contexts. Up to age 17, individuals' mobility parallels that of their parents, i.e. that of economically active adults between the ages of 30 and 47 years. Between ages 18 and 30, a sharp increase in mobility is observed, which is attributable to young people entering the labor market, leaving the parental home, moving into first union—in short, the mobility corresponding to the transition to adulthood. A small rise is observed around retirement (between ages 60 and 70), followed by an increase in mobility in old age, especially for women, corresponding to moving into care accommodation or to being looked after by one's children. The differences between the sexes reflect the age differences at which family and occupational transitions are made by individuals of opposite sex.

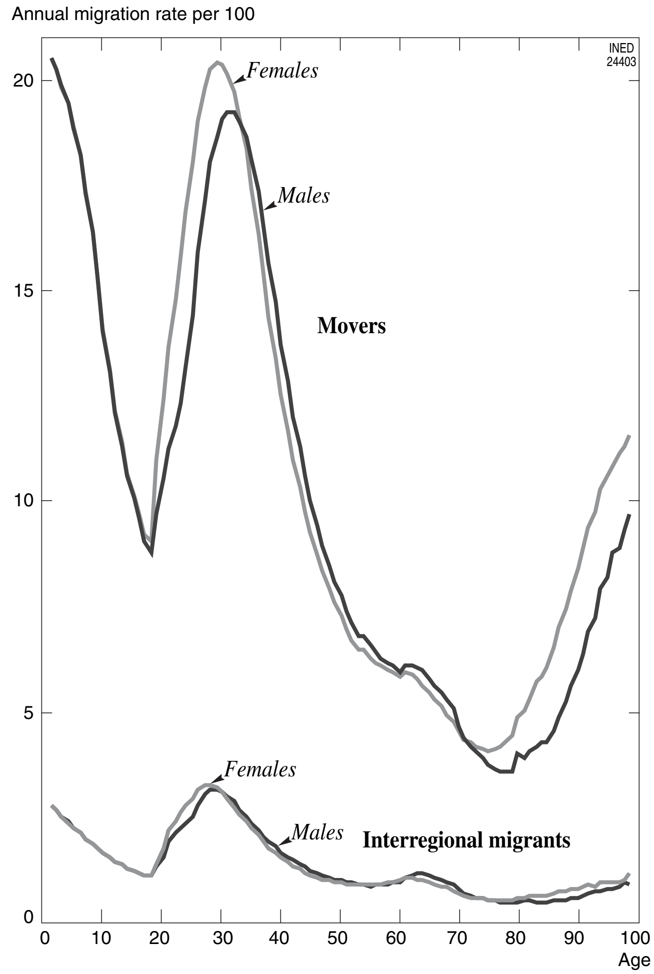


FIGURE 63–1 Residential mobility and interregional migration by age and sex (1999 French census). These estimates differ from those given by Baccaïni (2001) because of the inclusion of changes made between 1982 and 1990 and between 1990 and 1999 in the parameters used to convert inter-census data to annual data (Courgeau, 1973, 1979). These new estimates were produced using results from the French Labor Force Surveys for 1990 to 1999 and from the Jeunes et carrières (i.e., Young People and Careers) survey (INSEE, 1997) by Franck L'Hospital (2001). They are also consistent with data from France's electricity supply company (EDF) (Courgeau *et al.*, 1999, 2000).

The curves for residential relocation and interregional migration have broadly the same form, with just a difference in the multiplying factor. This confirms the choice made in the introduction of considering mobility as comprising the set of all spatial moves, with the

[2] In the case of France, this is a division into 22 regions. The direction of the flows is not discussed here (see Baccaïni, 2001).

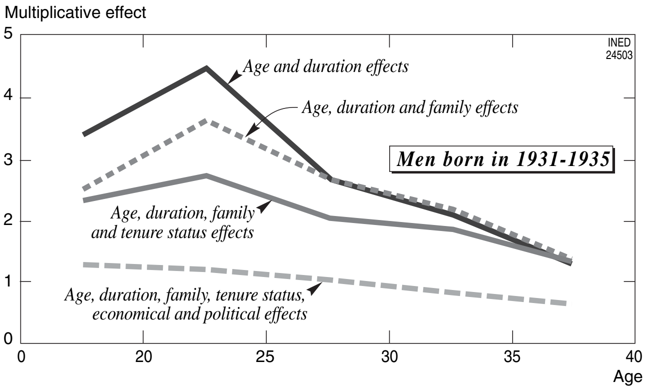


FIGURE 63–2 Multiplicative effect of age on male mobility by variables in the model used in the French Triple Biography survey of family, employment, and migration histories.

option then of considering only some of these as migratory moves. We shall see that the motivations involved are the same, some having greater impact for long-distance migration, others for short-distance moves.

In addition to a discussion of this classic distribution, a description based on analysis of individual-level data is now possible. Figure 63–2 shows the distribution (for the male population) of the effects of age, as estimated by a Gompertz type regression, modelling age at last move and duration of stay (Courgeau, 1985). A series of models are estimated, into which are introduced an increasing number of variables controlling the stages reached by the individual in his or her various life domains.

The age effect, controlling for the duration of stay, is very similar (disregarding the scales) to the curves in the previous diagram: a peak of mobility observed around age 22.5 gives way to stability. This time, however, the successive models make it possible to separate clearly what corresponds just to an age effect from what is the result of the variables characterizing the stages in the individual trajectories.

By first introducing the variables controlling for the stages of the family life-course trajectory (essentially marriage, divorce, and successive births), the effect of age alone is seen to be reduced between ages 15 and 28. Introducing the move into owner-occupation brings a further reduction in the age effect, this time across all the ages considered. In the full model, which introduces variables characterizing labor force status (military service, economically active, unemployment), type of job held, and macroeconomic conditions (e.g., period of wartime, crisis of 1931 or of 1974), the age effect disappears completely.

Age is clearly a proxy variable that captures, combines, and synthesizes the effects of the other processes at work in individual trajectories. It is therefore pertinent to try to identify the specific influence on mobility. Empirical distributions suggest that these effects occur at particular points in the trajectory and modify mobility over the lifetime of individuals. This leads us to favor, when the data are available, a longitudinal event history approach (see Chapter 23), which is the only satisfactory method for studying these changes.

2. Measurement of Mobility over the Individual Life-Course Trajectory

Event history data are extremely rich and allow mobility to be studied in a variety of ways that produce results that could be thought contradictory. It is therefore necessary to review these approaches whose implicit assumptions have a powerful influence on the results.

First, event history data files enable the changes in the migration behavior of individuals to be observed in relation to changes in, for example, marital status and work histories. For the case of marriage or divorce, both of which correspond to times of high residential mobility, it can be shown that these events also initiate new and highly specific periods of mobility. After the initial residential changes, marriage has a powerful stabilizing effect on individual mobility; divorce, by contrast, marks the start of a period of residential instability (Bonvalet and Lelièvre, 1991).

Second, at the time of the survey a summary statement of mobility (in the form of the number of dwellings) of the individuals questioned is obtained. This can be broken down according to the available characteristics, such as marital status at the time of the survey, family size, and occupational status at the time of the survey.

Because the key events in family life (e.g., entry into union, birth of children) and in occupational life (e.g., first job, changes) are very likely to be accompanied by a residential relocation, we can safely predict a higher number of dwellings for those with the most status changes in their family and occupational lives. The presentation of such a migration history provides the basis for valuable comparisons (Lelièvre, 1990a, 1990b; Bonvalet and Lelièvre, 1989).

One way of using these categories defined on *a posteriori* criteria is to differentiate, for example, the parents of three children or the permanently single at any point in the life course. This technique, although potentially highly instructive in a later phase of the analysis, produces results that must be used with caution, and for this reason, it is not recommended for an initial approach. It involves differentiating before family formation has even begun the future parents of three children from those who will have only one child

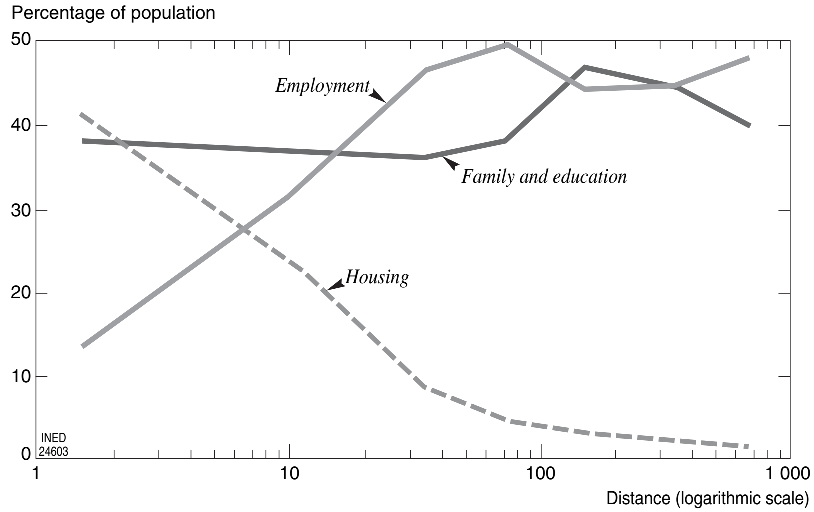


FIGURE 63–3 Reasons for migrations relative to distance (in kilometers) as found in the French Triple Biography survey of family, employment, and migration histories.

or the permanently single from those who will later form a union. It can be used, however, to test for possible selection processes. In the study of migration to France's principal metropolitan areas,[3] it could be shown that the fertility regime of migrant women was different from that of nonmigrants outside the metropolitan areas even before they moved to the highly urbanized regions. In other words, a definite selection process was at work. In the case of migration away from the metropolitan areas, no selection occurred because no distinction could be made between the fertility of future movers and nonmovers in the metropolitan areas (Courgeau, 1987; Courgeau and Lelièvre, 1989, 1992, 2001).

Although summary statements should not be ignored, they give an overly aggregate view of the different facets of individual life course trajectories. An average value for the number of moves per status at the time of the survey cannot take account of a fall followed by a rise in the propensity of individuals to move over their lifetime. Such summaries can only be established for groups exposed to the risk of moving over periods of the same length. An average number of dwellings for individuals in a 10-year age range lumps together in a way that varies with age group people who are in a phase of intense mobility with others who are in a period of stability. This biases the summaries.

The procedure whereby selective behavior is tested by examining the differential behavior of future categories must remain an empirical exercise to be used with caution.

[3] The regions of Paris, Lyon, and Marseille (see Courgeau, 1987; Courgeau and Lelièvre, 1989, 1992, 2001).

When longitudinal data are available, an analysis by duration of stay (if the data permit) can alone reveal the changes in mobility behavior over the lifetime of individuals.

3. The Motivations for Mobility

Determining the relative importance of the motivations quoted by individuals to explain their moves is a particularly difficult exercise. In part this is because these moves are very likely to be distributed differently depending on cultural context and historical period. Family-related reasons appear to dominate, at least for residential mobility, and work-related reasons dominate for longer-distance migrations. In the American context, Peter Rossi (1980) finds "the major function of mobility to be the process by which families adjust their housing to the housing needs that are generated by the shifts in family composition" (Rossi, 1980, p. 35). For France, Brigitte Baccaïni (1991) has shown how the reasons invoked to explain or justify a residential relocation vary with migration distance (Fig. 63–3). Over short distances (i.e., intracommunal moves), housing factors (e.g., entry to owner-occupation, home improvement, or enlargement) are the most important, whereas employment factors are negligible. The latter rapidly become more important with migration distance and form the majority of migrations at medium distances (i.e., intraregional moves), whereas family-related reasons remain at the same level up to a distance of 100 km. For the longer distance interregional migrations, housing factors account for a very small minority, and family and employment factors are on the same level. All of these motivations have a role, regardless of migration distance, but their importance varies with this distance.

Although the stages of the family life-course trajectory have a direct impact on the propensity of individuals to migrate, these stages occur within a context in which motivations are evaluated and acted on differently by individuals. Gary Sandefur (1985) has shown that the importance of motives for migrating depends strongly on the stage individuals are at in their life-course trajectory. Also, these reasons are reconstructed, often *a posteriori* as is the case in a retrospective survey, and only have the value attributed to them by the individuals required to justify or explain a move in the interview.

When event history data are available, a less subjective way to determine the nature of the motives for migration is to test the dependencies between mobility and other processes. The moves considered can be diverse in kind, ranging from simple residential relocations or even daily home-to-work journeys, up to interregional migration, depending on the level of precision of the data collection.

Despite the pertinence of event history methods, their application has limits that first need to be identified. Although a study of the interactions between marriage and migration proves particularly suited to this approach, we might want to carry out such a study between divorce and migration, and this is less straightforward. In the former case, selecting an origin that is common to both processes[4] presents no problem; for example, we could examine first marriage and the move into an independent dwelling. In the case of divorce, it appears virtually impossible to give an unequivocal definition of the specific move (which will be the same for all the individuals) from among the migration events to associate with the divorce, without making explicit reference to the breakdown of the couple. However, if we choose the move that precedes or follows the divorce, a study of interaction becomes meaningless.

Another difficulty, less trivial than it appears at first sight, stems from the structural transformations affecting relations between individuals. Although living as a couple still concerns the majority of individuals, the growth of cohabitation makes it harder to identify the entry into union. In the same way, well-defined events like divorce do not necessarily correspond to the point in time when the partners actually separated.

To illustrate these various points we now give a brief overview of the results obtained using event history data.

[4] This is one of the conditions for implementing the analysis (Courgeau and Lelièvre, 1989, 1992, 2001).

4. Some Significant Results

The following results are derived mainly from the analysis of data from the French Triple Biography survey of family, employment, and migration histories (also known as the 3B survey), conducted by the Institut National d'Études Démographiques (INED) in 1981. This highly innovative survey took the form of a retrospective collection of longitudinal event history information on individuals born between 1911 and 1935. The sample was nationally representative and contained 4602 individuals (Courgeau, 1985). These data allow us to examine the residential trajectories of these cohorts using the approaches described previously. We begin with the case of the different effects of marriage on individual mobility, which we believe is very explicit.

a. Marriage, a Factor of Mobility and of Greater Residential Stability

The survey data capture the mobility linked to this key event in the family trajectory of individuals. For the cohorts born between 1911 and 1935, marriage was associated with a move to a new residential location (Table 63–1). Marital status at the time of entry into a first independent dwelling explains the differences observed by sex. Forty-seven percent of men moved into an independent dwelling before they married, whereas women in these cohorts, who moved out of the parental home on average a year earlier than the men, more often moved at the time of their marriage (51% of women versus 39% of the men in the French 3B survey). Women are clearly more likely to change place of residence at the time of their marriage, and this move corresponds more often for them than for their partner with leaving the parental home.

Residential mobility is closely related to family events. Marriage usually results in a move by at least one of the partners, whereas divorce works in the opposite direction, resulting in the departure of at least one of the former partners. The results in Table 63–2 confirm these tendencies. Individuals who were single at the time of the survey were the least mobile, with

TABLE 63–1 Proportion of Individuals Changing Residence when They Married as Reported in the French Triple Biography Survey

Group	Cohort 1911–1925	Cohort 1926–1935	Total
Men	44.9	50.8	47.8
Women	64.8	66.4	65.5
All	55.6	58.7	57.4

TABLE 63–2 Average Number of Dwellings at Age 45 Years by Marital Status at the Time of the French Triple Biography Survey

Marital status	Average number of dwellings (standard error, 95% level)	Number
Single	2.13 (± 0.34)	189
Married	3.15 (± 0.12)	1385
Remarried	3.87 (± 0.48)	86
Divorced or separated	0.80 (± 0.13)	164

TABLE 63–3 Average Number of Dwellings Occupied for More Than 1 Year according to Marital Status at Each Age and at the Time of the French Triple Biography Survey

Marital status by age and at time of survey	Average number of dwellings (standard error, 95% level)	Number of people surveyed
25 Years		
Single–single	0.80 (± 0.13)	396
Single–married	0.90 (± 0.10)	1017
Married–married	1.55 (± 0.06)	1959
30 Years		
Single–single	1.19 (± 0.17)	396
Single–married	1.34 (± 0.20)	367
Married–married	2.06 (± 0.06)	2609
35 Years		
Single–single	1.52 (± 0.19)	396
Single–married	1.42 (± 0.30)	134
Married—married	2.51 (± 0.06)	2842

one in three of this group having never moved out of the parental home (many were farmers); these were usually individuals who never married and had very limited migration histories, and whose behavior contrasted sharply with that of single people who were not yet married. Next came married people, and the highest mobility was observed for the divorced and remarried. However, these results in no way take account of changes in mobility in the course of the trajectory.

As part of the attempt to measure the effect of marriage, Table 63–3 presents a description by current marital status and that observed at the time of the survey to capture the mobility of single people and the impact of marriage. At the three ages considered (25, 30, and 35 years), individuals who are single have practically the same mobility whatever their subsequent marital outcome. For married individuals, the average number of dwellings occupied is markedly higher at each age, revealing an effect of marriage on residential mobility.

This set of three tables is characteristic of the descriptive results commonly presented. Table 63–1 measures mobility at the time of marriage, Table 63–2 gives a summary comparable to that we would obtain with period data, and Table 63–3 is an attempt to capture mobility across the life-course, which turns out to be not very convincing.

Event history analysis is alone able to take into account the changes in the migration behavior of individuals, and its application reveals the second characteristic of the longer-term influence of marriage, the stabilizing effect (Courgeau, 1985), which is particularly clear, corresponding to a large slowdown that reduces mobility by one-third (Fig. 63–2). We see that an average value for the number of residential locations cannot reflect a rise followed by a fall in the migration propensity of individuals over their lifetime. Information derived only from this type of result is therefore completely inadequate for capturing the influence of an event on the mobility of individuals, and an event history analysis of such data has become indispensable. Implementing such an analysis is straightforward using widely available software.

b. Individual Mobility and the Arrival of Children

An event history analysis of the interactions between mobility and parenthood reveals in a few precise instances an effect associated with the birth of children. The increase in mobility that would be expected after successive births is observed only for women who married before age 22 (Courgeau, 1985). It can be assumed that these couples have had to adjust the size of their housing *a posteriori* to that of their family. This contrasts with couples who married later and who probably already had a dwelling more adapted to their desired future family size, which would explain the absence of any significant increase in mobility after the births. For the couples who married later, a solution employed by these cohorts to have more dwelling space was the move to home ownership, whose stabilizing effect has been combined with that of marriage.

In this case, when a summary is prepared showing the average number of dwellings occupied according to family size reached at the time of the survey, we find that childless couples have on the whole moved less than the others, but the differences observed thereafter

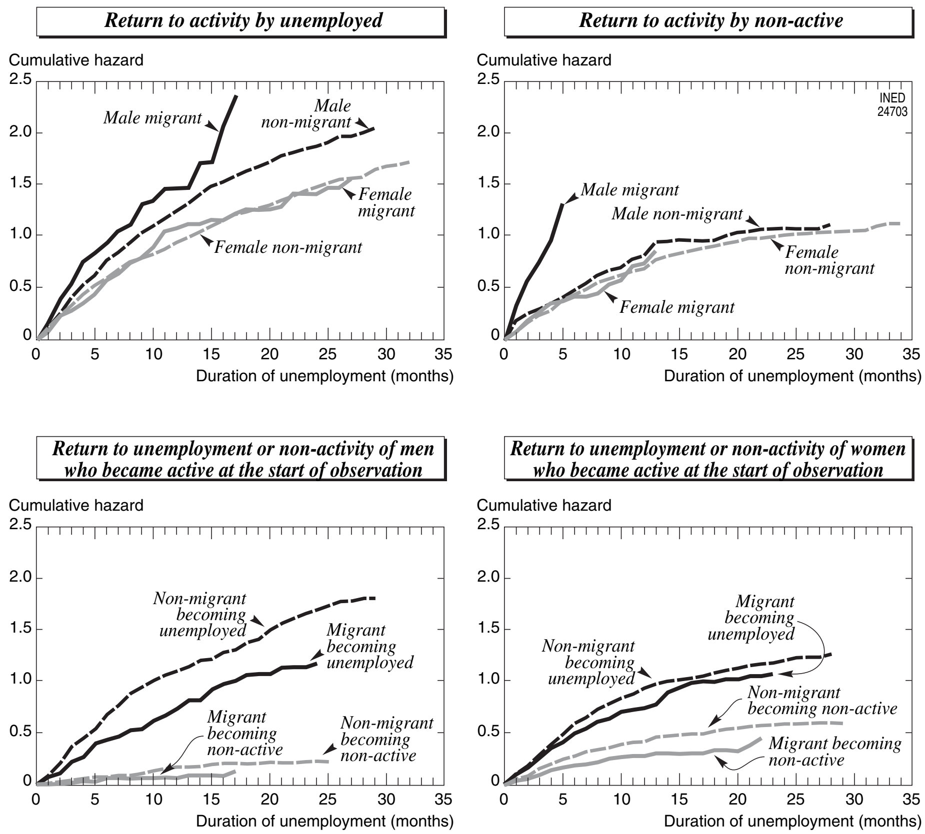


FIGURE 63–4 Mobility and economic activity in France between 1991 and 1995 as found in the Labor Force survey (INSEE). (From Courgeau and Meron, 1995, p. 22.)

are not statistically significant. In this information, we capture the description of the mobility of very specific categories and not the effect of births over the lifetime of individuals.

A similar analysis of the interactions between intraregional or interregional mobility and the birth of children was also conducted for Norway using data from the Population Register (Baccaïni and Courgeau, 1996). For the cohorts born in 1948 and 1958, the birth of a first child leads to increased short-distance mobility in the year of this birth and the following year. After this 2-year period, this mobility is reduced for interregional migrations or for higher-order births. In this country, the adjustment of housing size, when it is necessary, occurs after the first birth, and it is not observed for subsequent births.

c. Mobility and Economic Activity

Spatial mobility also has complex links with employment. In France in the early 1990s, it was observed that migration by a couple was more likely to lead to unemployment of one partner, usually the woman, the greater the distance moved (Courgeau and Meron, 1995; Figure 63–4). Twenty-seven percent of economically active women who were employed and who changed region became unemployed, compared with only 4% of those who did not move. For economically active men, these rates are 8% and 2%, respectively, which shows that they are also subject to this effect but to a lesser degree. It might be thought that because these migrants are cut off from some of their existing networks for integrating the labor market, they would have more difficulty finding work. However, this is not what is observed. Male migrants more often find work than nonmigrants, and no difference in behavior is found for female migrants. Moves are not made to a region picked at random, but to a region where the individual already has well-established relational networks.

Unemployed and economically inactive persons are more likely to find work when they are migrants than nonmigrants. Among nonmigrants, 41% of unemployed persons and 14% of the economically inactive find employment in the course of the year, and these percentages rise to 57% and 47%, respectively, among movers. Turning next to examine the stability of the work found by these individuals, we see that the return to unemployment or nonactive status is much slower for men and women who moved than for the nonmigrants. We can conclude that persons who have moved to find a job are better placed to keep it than those who took what was available in their existing place of residence. The cost of migrating must be evaluated in relation to the more stable nature of the work obtained after moving. The small numbers of individuals who take this course are clear evidence of the difficulty associated with such a reorientation.

5. Choice of Destination: The Role of the Personality of the Individuals and the Information Available to Them

Humans do not live in a homogeneous space about which they have all the relevant information necessary to decide about a possible migration. They inhabit a heterogeneous space about which they possesses information that is imperfect and that changes over time. The same applies to the means for acquiring this information through relational networks, through the press, or by any other means of diffusion.

One way to get a clearer view of this information is to capture it using surveys that identify the mental maps of individuals living in various locations, according to their age (Gould, 1975). Research on this subject has shown the construction of mental maps during an individual's childhood through adulthood. These maps reflect a sharply contrasted knowledge of the zones studied, with a maximum of information around the place where the individual lives and with relative maxima around the most densely populated centers. This information can then be modeled as a function of the distance of the individual and the populations from the different places in the country. Variables other than age can usefully be introduced; educational level, occupation or nonactivity, and the rural or urban milieu in which the individual lives are all characteristics with the potential to modify these mental maps. It is easy to imagine the complexity of the information that must be collected to facilitate understanding of these various maps.

A second method is to model this information in a more *macro* approach to individual behavior. Let us assume that the individual in a given region has perfect information about the other regions to which he can migrate. The probability for him of choosing region j is equal to the attractiveness specific to this region, divided by the summed attractiveness of all regions. Such attractiveness can be measured as a function of the various characteristics of region j: its unemployment rate, average income levels, the proportion of the population working in the agricultural sector, and so on. However, we must take account of the fact that the choice of the region of destination is not independent of the region of origin. The information the individual may have about the different destinations is a function of the proximity of the places, his or her occupation, income, and other factors. This means that we can write the probability of choosing destination j, knowing that the individual lives at i, in the following form:

$$\text{Prob. (choose } j/\text{living at } i) = \frac{p_j q_{ij}}{\sum_{k=1}^{r} p_k q_{ik}} \qquad (1)$$

In Equation 1, p_j is the attraction specific to region j independently of the origin, and q_{ij} is a measure of the information that pushes an individual present at i to move to j (Ginsberg, 1972). Depending on the characteristics introduced, this formulation produces various models of migration behavior in general use. If we measure the specific attractiveness of a region by its population, P_j, and the information by a function of the distance separating regions i and j, d_{ij}, we can write the migration flow between these regions:

$$M_{ij} = \frac{P_i P_j f(d_{ij})}{\sum_{k=1}^{r} P_k f(d_{ik})} = k P_i P_j f(d_{ij}) \qquad (2)$$

We obtain a model belonging to the class of gravity models, based on a behavioral theory. These models can introduce various other measures of the information between these zones and many other characteristics of both origin and destination zones: distance in terms of intermediate jobs, cost of living, income, leisure facilities, climate, and density of population in both origin and destination zones (Cadwallader, 1992; Rogers, 1967). These models can be estimated using linear regression methods when the logarithms of the two sides of the equations are taken. It is necessary to verify the conditions of application of such a model (Poulain, 1981).

This approach assumes that individual behavior derives from choices that depend on perception of aggregate characteristics. An individual will prefer the regions whose living conditions he or she knows best relative to other regions, while at the same time comparing it with living conditions in his or her region of residence. The logic and estimation of the model can then deal with flows and characteristics measured at the *macro* level, under the assumption that individual behavior based on rational choices is not simultaneously influenced by personal characteristics. However, we saw earlier that this assumption was not verified.

In the next section, we need to synthesize these two approaches.

Having presented the complexity of migration phenomena and the various approaches employed for their analysis, we now turn our attention to the more innovative lines of inquiry that can offer deeper insights into mobility.

II. TOWARD A BROADER UNDERSTANDING OF MIGRATION BEHAVIOR

The need to understand the decision-making mechanisms that lie behind the migration decisions and locational choices of individuals has led many researchers to look for a framework that allows individual strategies to be situated in their spatial context. Two main currents of research are presented here. These are currently undergoing major methodologic developments and are contributing, from very different perspectives, to the integration of the study of migration into a broader context.

In the study of urbanization processes, Véronique Dupont and Françoise Dureau (1988) directed their attention not at the abstract entity of the city, but at the actors of the processes of urbanization. For their analysis of urban dynamics, they found it simplistic to focus uniquely on the population living in the urban space delimited by its built environment and necessary to go beyond the usual practice of linking individuals to a single place of residence.

Another line of research allows the inclusion of more complex parameters derived from widely available data to analyze the interactions between the results obtained at different levels of aggregation. In particular, we need to try to identify the interaction between the individual level and the aggregate level. These two approaches are examined in order.

1. The Life Space and Its Modification

An initial means of going beyond the study of simple individual motivations is by situating the actors of mobility in their life space, defined as "that part of space in which an individual carries out all his or her activities . . . not only the places of transit and residence, but also all the places with which he or she interacts" (Courgeau, 1980). This concept issues from the conjunction of work in cognitive psychology and sociology, and it was first applied in demography by Daniel Courgeau (1972) for exploring and mapping the networks of relations between persons. The objective in this research was to use these networks to study individual mobility. Its basic postulate is as follows: migration by individuals or by households is not just the result of economic and social conditions but also depends heavily on the ties that individuals form, the advice they receive and the information to which they have access. In this perspective, a survey was conducted in a rural setting (followed by another in an urban setting). In this way were traced the first steps for operationalizing the life-space concept in a quantitative experiment. When the network of relations is considered, the exact location of its members gives its spatial distribution at a given point of time. Any modification over time in this set of places constitutes a change in the life space, causing it to undergo a contraction, an extension, a shift, or a complete redefinition.

Definitions of the life space are varied, ranging from a space structured by a network of relations to a space defined by a set of places corresponding to functions rather than to persons. In the generally available statistics, individuals are identified at a single place, that of their residence, and changes in this place over time define their migrations. In the context of a quantitative analysis, which applies a narrower definition of the life space, we can assume that it comprises the place of residence and place of work of an individual at a point in time, for example.

A study carried out in India (Dupont and Lelièvre, 1993) operationalized a specific definition of the life space of the migrant. Taking as their starting point the possible separation of place of work and place of residence, the researchers decided to incorporate the plurality of places of residence and work for any given individual. This allowed them take into account the fact that a migrant of rural origin can combine a job in the town with doing agricultural work and can consider his village of origin as a more important place of residence than his dwelling in the town, with the village remaining the main locus of emotional, familial, and social investment. In the study in question, "la dynamique d'une ville moyenne en Inde" (i.e., the dynamics of a medium-sized town in India), the objective was to "apprehend the town beyond the space of human concentration, such as to relate it to the rural dynamics that are its constituent parts." India presents the paradox of having a low rate of urbanization (25.7% of the population was urban in the 1991 census) even though it has the world's second largest urban population (218 million in 1991) after China. Using this approach also allows for the possible reversibility of migration flows. The preference accorded at any given point in time to one of the constituent poles of the life space can be modified, causing the movement to change direction. This approach to the actors of the urbanization processes and their life space led us to invert the classic view of the town as a pole possessing its zone of influence and to propose a new perspective, particularly revealing in the case of the small and medium-sized towns in India: The town appears more as an outpost of the village, a locus of investment and of diversification of the activities of country dwellers. The concept possesses considerable analytical power.

In the Population, Espace de Vie, et Environnement (Population, Life Space, and Environment) survey conducted by INED in 1991, Philippe Collomb and France Guérin-Pace collected data intended to identify the representations and practices of French people regarding the environment, for which purpose they introduced the life-space concept. Applied in a concrete approach to the spatial practices of the respondents, this concept enabled the authors to gain insight into the universe of individual representations and to get access to the larger notion of collective environment (Guérin-Pace, 1994). The life space the researchers observed with this survey is more accurately the space that is frequented and moved through, and around which individual existence is constructed and which structures daily life (Di Meo, 1991), which is also referred to as the *daily-life environment* (i.e., *spazio utilizzato*) (Barsotti and Bottai, 1994). A large part of the survey is devoted to the journeys made on foot in the vicinity of the dwelling, with a distinction being made between the familiar territory where the respondent goes with pleasure and the more functional and utilitarian territory where he or she goes from necessity. Information is also collected on the reasons why places are frequented, those that determine the attraction of some places or the repulsion or lack of interest for other zones.

The data from the Biographies et Entourage (Event Histories and Contact Circle) survey conducted in France between 2000 and 2001 opened up new and particularly promising perspectives for the study of the individual's life space and its evolution over the lifetime (Lelièvre *et al.*, 2002). The survey catalogues the mosaic of places that makes up the respondent's geographic universe, each place corresponding either to a *function* (i.e., place of residence and work, regularly visited places: school boarding facilities, barracks, weekend cottage, place of holiday) or to a *bond* maintained with the contact circle (i.e., places of residence of parents, siblings, uncles or cousins). The location of the kinship group at the time of the survey and at different stages in the respondent's lifetime, allows us to study the spatial dynamics and strategies at the level of the contact circle and to re-examine the analysis of diverse forms of mobility in interaction. Description can embrace commuting journeys (e.g., residence or work, family home or hall of residence) and international migration (e.g., place of origin or place of destination), and it can shift from residential mobility to space experienced and structured by networks of relations or space defined by personal mobility. By rethinking the temporal dimension of the processes and by placing individual strategies in their spatial context, we can avoid the single criterion of distance and the view of individuals as having a single reference point, while achieving a more balanced view of the economic determinants of migration decisions and envisaging mobility in its entirety rather than migration events in isolation.

We have gone from a narrow definition of migration as a change of residential location to a more general definition of migration as a change of life space. The former definition, elaborated on the basis of classic demographic statistics, is replaced by a definition that is more general and multiform in its quantitative operationalizations. Both of these definitions include the physical space in which the movements occur, which makes them familiar to geographers (Frémont, 1974). They differ from other definitions that are more tied to the conceptual and emotional space of individuals (Mangalam, 1968; Eisenstadt, 1953). This concept offers countless possibilities, because its flexibility of application means it can be adapted to the diverse problems associated with the range of types of mobility and immobility present in the field of migration.

2. Individual and Aggregate Approaches: The Need to Integrate the Different Levels Simultaneously

We saw earlier how various individual characteristics can influence migration behavior; this is possible with event history analysis. We also saw how different characteristics of the origin and destination areas, measured at the aggregate level, can influence the percentages of migrants; this is a task for which regression analysis is well adapted. The next step is to explore the relations that exist between the results produced by these two analyses.

In 1950, William Robinson showed that a correlation measured at an aggregate level is not necessarily the same as a correlation measured at the individual level. This introduced the concept of the ecologic fallacy, which occurs when we want to measure an individual correlation using aggregate data, the most often available. Almost 30 years later, Glen Firebaugh (1978) related the problems of aggregation to the theory of group effects and confirmed that an aggregate-level variable often measures a different characteristic than that measured by the same variable defined at the individual level. In 1992, Michael Von Korff and colleagues showed that analyses that relate aggregate-level characteristics to outcome variables measured at the aggregate or individual level do not inherently produce ecologic bias. The ecologic bias is the result of improper interpretation of the correct relationship measured at the aggregate level.

A broader perspective then becomes possible with a multilevel analysis, which acknowledges the importance of individual and aggregate characteristics in determining migration behavior. We need to examine in more detail the theoretical and practical associations that exist between these different levels and to demonstrate the value of this type of multilevel approach (Courgeau and Baccaïni, 1997, 1998; Courgeau, 1994, 1995, 2003).

Without going into the detail of the demonstrations, we have verified that for the study of migration behavior, the effect of aggregate characteristics was virtually independent of that of individual characteristics in the cases of French and Norwegian interregional migration. It follows that it is entirely legitimate to introduce individual-level and aggregate-level characteristics simultaneously to get a more comprehensive explanation of migration behavior.

The effect of some of these characteristics, depending on whether they are treated as aggregate or individual, can lead to apparent paradoxes that have to be explained through a more detailed examination of the situations. Let us take as an example the case of French women working in farming, whose migration probability is much lower than that of the other occupational groups. However, the fact of living in a region where women working in farming are more numerous increases the probability of moving out of that region (Courgeau, 1994, 1995).

This apparent paradox is explained by subdividing the population at risk into two distinct groups: women in farming and women not in farming. Regression is then used to estimate the logarithm of the probability of leaving the region as a function of the proportion of women in farming in the various regions (Fig. 63–5).

Women in farming have a consistently lower migration probability, regardless of the proportion they represent in the region. This confirms the negative parameter obtained at the individual level (–0.333). At the same time, we see that the migration probability for all women, whether in or not in farming, increases as the proportion of women in farming increases. The result is a positive value (+3.785) for the parameter associated with this aggregate variable. This illustrates the danger of inferring assumptions about individual behavior from results obtained at the aggregate level. A high proportion of women in farming is associated with a higher migration probability for all categories of the population, partly because of the scarcity of nonagricultural employment in these regions. This does not mean that women working in farming have a higher probability of migrating than others; at the individual level, the exact opposite is observed. This result holds regardless of the region of origin.

This suggests that to analyze migration behavior correctly, we need to introduce simultaneously the

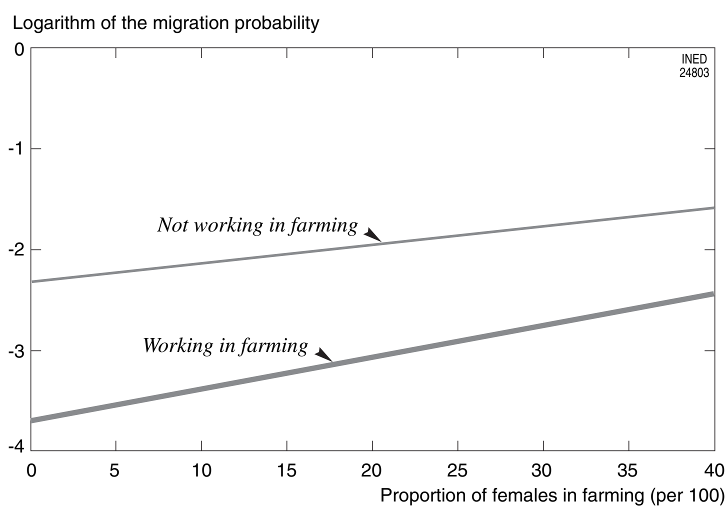


FIGURE 63–5 Logarithm of the migration probability for women working in farming and for other categories as a function of the proportion of women in farming in each region as found in the French Triple Biography survey of family, employment, and migration histories. (From Courgeau, 1994, p. 22.)

characteristics defined at different levels of aggregation: at the individual level and at the level of larger geographic units (e.g., regions, towns). At the same time, more far-reaching inquiry is essential into the significance of the different levels considered to identify the most relevant levels of aggregation for inclusion in the analysis. Does a choice have to be made between spatial divisions, such as that by communes, departments, or regions, or should they all be treated as relevant for inclusion in the analysis? Only further research in this field can provide the answer to such questions.

CONCLUSIONS

Our study of the motivations for migration has led us to broaden the scope of analysis by moving from an approach at the *micro* level to one at the *macro* level and then by attempting a synthesis of these two approaches.

The individual-level approach is a valuable tool for determining the motivations associated with the family and work domains of potential migrants, either through direct questioning about the reasons for migrating or by means of an event history analysis of migrations in interaction with a range of family- and work-related events. We have found this second approach preferable to the first, which in the case of retrospective surveys risks producing reasons reconstructed *a posteriori*. Event history analysis makes possible a very detailed analysis of these motivations, which are reconstituted from the succession of life stages of the respondents, such as marital migration, migration associated with divorce or widowhood, migration related to the birth of children, and work-related migration.

The aggregate-level approach is valuable for examining the motivations associated with the position of the individual in a space that is both social and geographic. The social space of origin is formed by the milieu in which the individual lives, which is apprehended through various aggregate characteristics of the place of origin. From this origin, there are various social spaces of destination that are more or less well known to potential migrants as a function of the physical or social distance that separates them. The attraction of these different spaces can then be modeled by introducing their various characteristics captured more or less precisely according to the distance between origin and destination zones. This approximate modeling results in aggregate migration models, which explain the flows in relation to the characteristics of the origin and destination areas and to a measurement of the information exchanged between these areas.

These approaches then must be generalized to give a more detailed view of migration in the form of change in life space and to attempt a synthesis linking the micro and macro levels. The first generalization involves conducting surveys with which to track the individual's contact circle (*entourage*), a concept whose complexity and value for social science we have already demonstrated (Bonvalet and Lelièvre, 1995; Lelièvre *et al*., 1997 and 1998). The second generalization leads to formulating models in which the characteristics measured at different levels of aggregation are introduced simultaneously. Such multilevel models can be used to test the independence between the effects of the characteristics measured at different levels. They show that in some cases the effect of the individual characteristic can run counter to that of the aggregate characteristic, although because of this independence, there is no paradox.

Much remains to be done to develop a theory that explains the role of the different levels of aggregation in migration behavior. The effect of individual characteristics can be taken to measure migration propensity according to personal situation while allowing for individual freedom of choice. Social constraints operate at a more aggregate level and influence the migration flows in a different way from the characteristics of the individuals who are going to migrate. The question then must be asked whether some levels of aggregation are more relevant than others for an understanding of migration behavior. This constitutes an entire field of research, both theoretical and empirical, which has only begun to be explored and for which many more studies will be needed.

Acknowledgment

This chapter was translated from the French version by Godfrey I. Rogers.

References

Baccaini Brigitte, 1991. *Mobilité, distances de migration et cycle de vie.* Paris, Université de Paris I, UFR de Géographie, 520 p. (Doctoral thesis.)

Baccaini Brigitte, 2001. Les migrations internes en France de 1990 à 1999: l'appel de l'Ouest, *Economie et Statistique*, 344(4), p. 39–79.

Baccaini Brigitte and Courgeau Daniel, 1996. The spatial mobility of two generations of young adults in Norway, *International Journal of Population Geography*, vol. 2(4), p. 333–359.

Barsotti Odo and Bottai Marco, 1994. La ricerca "Spazio Utilizzato." Obiettivi e primi risultati, *in*: Marco Bottai and Odo Barsotti (eds.), *Lo spazio e la sua utilizzazione*, p. 16–60. Milan, F. Angeli, 219 p.

Bonvalet Catherine and Lelièvre Éva, 1989. La mobilité résidentielle en France et à Paris depuis 1945. I. Bilan résidentiel d'une génération, *Population*, vol. 44(3), p. 531–560.

Bonvalet Catherine and Lelièvre Éva, 1991. Nuptialité et mobilité *in*: Thérèse Hibert and Louis Roussel (eds.), *La nuptialité: évolution récente en France et dans les pays développés*, p. 237–253. Paris, INED, PUF, 280 p. (Col. Congrès et Colloques, no. 7.)

Bonvalet Catherine and Leliévre Eva, 1995. Du concept de ménage à celui d'entourage: Une redéfinition de l'espace familial, spécial Une nouvelle morphologie sociale. *Sociologie et Sociétés*, vol. 27, no. 2.

Cadwallader Martin, 1992. *Migration and residential mobility.* Madison, The University of Wisconsin Press, 276 p.

Courgeau Daniel, 1972. Les réseaux de relations entre personnes. Étude d'un milieu rural, *Population*, vol. 27(4–5), p. 641–683.

Courgeau Daniel, 1973. Migrants et migrations, *Population*, vol. 28(1), p. 95–129. (Published in English: Migrants and migrations, *Population. Selected Papers*, 1979, 35 p.)

Courgeau Daniel, 1980. *Analyse quantitative des migrations humaines*. Paris, Masson, 225 p.

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel (ed.), 2003. *Methodology and Epistemology of Multilevel Analysis*. Methodos Series, no. 2. Dordrecht/Boston/London, Kluwer Academic Pulbisher, 236 p.

Courgeau Daniel, 1987. Constitution de la famille et urbanisation, *Population*, vol. 42(1), p. 57–82. (Published in English: Family formation and urbanisation, *Population: An English Selection*, 1989, vol. 1, p. 123–146.)

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, vol. 49(1), p. 7–26. (Published in English: From the group to the individual: what can be learned from migratory behaviour, *Population: An English Selection* 1995, vol. 7, p. 145–162.)

Courgeau Daniel (ed.), 2003. *Methodology and Epistemology of Multilevel Analysis*. Methodos Series, no. 2. Dordrecht/Boston/London, Kluwer Academic Publishers, 236 p.

Courgeau Daniel and Baccaini Brigitte, 1997. Analyse multiniveau en sciences sociales, *in*: Daniel Courgeau (ed.), *Nouvelles approches méthodologiques en sciences sociales: population*, vol. 52(4), p. 831–864. (Published in English: Multilevel analysis in social sciences, *Population: An English Selection*, vol. 10(1), 1998, p. 39–71.)

Courgeau Daniel and Lelièvre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, PUF, 268 p.

Courgeau Daniel and Lelièvre Eva, 1992. *Event history analysis in demography.* Oxford, Oxford University Press, 226 p.

Courgeau Daniel and Lelievre Eva, 2001. *Análisis demográfico de las biografías.* México, El Colegio de México, 305 p.

Courgeau Daniel and Meron Monique, 1995. Mobilité résidentielle, activité et vie familiale des couples, *Economie et Statistique*, vol. 290, p. 17–31.

Courgeau Daniel, Nedellec Vincent, and Empereur-Bissonnet Pascal, (1999). La durée de résidence dans un même logement. Essai de mesure à l'aide des fichiers EDF, *Population*, vol. 54(2), p. 333–342. (Published in English: Duration of residence in the same dwelling. A test of measurement using Electricity Utility Company records, *Population: An English Selection*, vol. 12(1), 2000, p. 335–342.)

Di Méo Guy, 1991. De l'espace subjectif à l'espace objectif: l'itinéraire du labyrinthe, *L'Espace Géographique*, vol. 4, p. 359–373.

Dupont Véronique and Dureau Franoise, 1988. *Renouveler l'approche de la dynamique urbaine par l'analyse des migrations?* Paris, CNRS, ORSTOM, 135 p. (Pratiques Urbaines no. 4, Inter-Urba, Tiers Monde.)

Dupont Véronique and Lelièvre Éva, 1993. La ville, antenne des villages, observations indiennes, *in*: AIDELF (ed.), *Croissance démographique et urbanisation. Politiques de peuplement et aménagement du territoire*, p. 117–130. Paris, PUF, 438 p. (Colloques, no. 5.)

Eisenstadt Shmuel N., 1953. Analysis of Patterns of Immigration and Absorption of Immigrants, *Population Studies*, vol. 7(2), p. 167–180.

Firebaugh Glen, 1978. A rule for inferring individual-level relationships from aggregate data, *American Sociological Review*, vol. 43(4), p. 557–572.

Fremont André, 1974. Recherche sur l'espace vécu. *Espace Géographique*, no. 3, p. 231–238.

Ginsberg Ralph B., 1972. Incorporating causal structure and exogenous information with probabilistic models: with special reference to choice, gravity, migration and Markov chairs, *Journal of Mathematical Sociology*, vol. 2, p. 83–103.

Gould Peter, 1975. *People in information space: the mental maps an information surfaces of Sweden.* Lund, CWK Gleerup, 161 p. (Lund Studies in Geography series B, no. 42.)

Guerin-Pace France, 1994. Présentation d'une enquête française sur la "population, l'espace de vie et l'environnement," *in*: Marco Bottai and Odo Barsotti (ed.), *Lo spazio e la sua utilizzazione*, p. 208–214. Milan, F. Angeli, 219 p.

Lelièvre Éva, 1990a. Point des connaissances quantitatives en matière de mobilité résidentielle, *in*: Catherine Bonvalet and Anne-Marie Fribourg (eds.), *Stratégies résidentielles*, p. 73–84. Paris, INED, PUF, 460 p. (Col. Congrès et Colloques, no. 2.)

Lelièvre Éva, 1990b. Bilan des connaissances sur la mobilité individuelle au cours du cycle de vie, *in*: Catherine Bonvalet and Anne-Marie Fribourg (ed.), *Stratégies résidentielles*, p. 49–64. Paris, INED, PUF, 460 p. (Congrès et Colloques, no. 2.)

Lelièvre Éva, 1992. L'étude des interactions entre phénomènes: dépendance unilatérale et causalité, in: AIDELF (ed.), *Démographie et Différences*, p. 523–530. Paris, PUF, 692 p. (Colloques no. 4.)

Lelièvre Ëva, Bonvalet Catherine, and Bry Xavier, 1997. Analyse biographique des groupes, les avancées d'une recherche en cours, *Population*, vol. 52(4), p. 803–830. (Special issue "Nouvelles approches méthodologiques en Démographie" edited by Daniel Courgeau. Published in English: Event history analysis of groups. The findings of an on-going research project, *Population: An English Selection*, vol 10, 1998, p. 11–37.)

Lelievre Éva, Vivier Géraldine, and Bonvalet Catherine, 2002. L'approche de la mobilité à partir des données de l'enquête Biographies et entourage, *in*: J. P. Lévy and Françoise Dureau (eds.), *L'accès à la ville. Les mobilités spatiales en question*, p. 383–398. Paris, L'Harmattan, 411 p. (Habitats et sociétés.)

L'Hospital Franck, 2001. *Les migrations internes en France. Estimation des paramètres du modèle "migrants-migrations" de Daniel Courgeau*, Service des Etudes Diffusion. Lyon, INSEE Rhône Alpes, 77 p.

Mangalam Joseph-J., 1968. *Human migration: a guide to migration literature in English 1955–1962.* Lexington, University of Kentucky Press, 194 p.

Poulain Michel, 1981. *Contribution à l'analyse spatiale d'une matrice de migration interne.* Louvain-La-Neuve, Cabay, 226 p. (Coll. Recherches Démographiques, Cahier no. 3.)

Puig Jean-Pierre, 1981. La migration régionale de la population active, *Annales de l'INSEE*, vol. 44, p. 41–79.

Robinson William S., 1950. Ecological correlations and the behavior of individuals, *American Sociological Review*, vol. 15, p. 351–357.

Rogers Andrei, 1967. A regression analysis of interregional migration in Carolina, *Review of Economics and Statistics*, vol. 49, p. 262–267.

Rossi Peter H., 1980. *Why families move*, 2nd ed. Beverly Hills, CA, Sage Publications, 243 p.

Sandefur Gary D., 1985. Variation in interstate migration of men across the early stages of the life cycle, *Demography*, vol. 18(3), p. 355–368.

Von Korff Michael, Koepsell Thomas, Curry Susan, and Diehr Paula, 1992. Multilevel analysis in epidemiologic research on health behaviour and outcomes, *American Journal of Epidemiology*, vol 135, p. 1077–1082.

P A R T

III

SPECIFIC APPROACHES TO TWO TYPES OF MIGRATION

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

In Chapter 22, Daniel Courgeau wrote, "The study of spatial mobility involves identifying the changes occurring over time in the relations between an individual or social group, and space."[1] This spatially focused definition of mobility sees individuals as continually mobile and focuses on the need to analyze all time-bound movements. This makes all forms of movement material, including repetitive, short-duration commuting movements and comparatively permanent return movements after a relatively long time spent far away from the place of departure. The time aspect enables a link to be made between mobility and other population and economic processes and to identify how all an individual's life events connect with one another.

Daily commuting, the subject of Chapter 64, is a key issue for migration analysts engaged in the vast field of spatial interactions, studying the spaces in which individuals live. In this chapter, Marco Bottai and Odo Barsotti give a detailed description of the theoretical and methodologic aspects of the phenomenon and analyze it through complementary macro and micro approaches.

On a different scale, return migration after a comparatively long stay abroad is a relatively unexplored but highly important field of research because of the size of the populations concerned. In Chapter 65, Jean Louis Rallu explains the instruments and methods of analysis and applies them to survey data.

[1] "L'étude de la mobilité spatiale consiste à mettre en évidence les changements au cours du temps, qui sont à l'oeuvre dans les rapports entre un individu ou un groupe social et l'espace."

CHAPTER

64

Daily Travel: Approaches and Models

MARCO BOTTAI AND ODO BARSOTTI

Dipartimento di Statistica e Matematica all' Economia, Università di Pisa, Pisa, Italy

Although daily travel merits its own category, it is not fundamentally different from other types of *spatial interaction*. We begin by defining what we mean by daily spatial interactions and describe how this type of spatial interaction is related to other forms of mobility. We present a macro-level perspective on daily travel before turning to a micro-level perspective. We conclude with a discussion of the challenges that face a unified theory.

I. SPATIAL INTERACTIONS AND LEVEL OF ANALYSIS

By *daily spatial interactions*, we mean the travel of individuals that ends within a single day, excluding all movements of merchandise, capital, data, and similar items.

Residential changes or migration are omitted by this definition. However, all these types of interaction are inextricably intertwined because daily travel depends on residential choices, and frequent round trips between the same destinations influence the decision to migrate. The link is so close that the approaches and models for studying daily *moves* are substantially the same as for other types of spatial interactions.

In particular, let us consider the spatial mobility of workers, which can be interpreted as an economic response (i.e., as a way to describe how workers adapt in their roles as factors of production and as consumers), and as a spatial equilibrium constraint (i.e., as a process constraining the overall localization pattern).

Because labor is mobile, workers can offer their labor services anywhere, and these offers can result in regular (daily) movements from residence to workplace or in changing the place of residence, or both.

Three processes of spatial adaptation in the individual's role as a factor of production can be identified (Termote, 1969):

1. Modification of the individual's spatial behavior as a result of a change in residence or workplace
2. Regular (periodic) movement from one locality to another
3. A combination of the two

Consider two geographic areas, A and B, with excess labor demand in A and excess supply in B. This situation can generate migration from B to A, commuting from B to A, or even migration from B to an alternative intermediate point C with associated commuting from C to A. Migratory and daily spatial interaction are linked by *spatial conjunction*.

After a period of commuting from B to A, a change of residence to A can follow, eliminating the commuting trips. In this case, migratory and daily spatial interactions are linked by *temporal conjunction*. There is also a hybrid case when a worker, after having moved to C, subsequently moves again to A, illustrating a *spatial-temporal conjunction* between permanent and daily moves (Termote, 1969).

Depending on the kinds of responses to spatial disequilibria, migrations and daily trips can be concurrent, sequential, or both concurrent and complementary.

Spatial adaptation of labor supply and demand can be formalized by the following matrix model:

Let M be an origin-destination matrix of order k (i.e., the number of regions in the spatial system) of migratory flows between time t and time $t + 1$ of the employed labor force, divided by the employed resident labor force in the region of origin.

Let L_t be the vector of the employed resident labor force in each locality at time t.

Then, $L_{t+1} = L_t \times M$ is the vector of the employed resident labor force in each locality at time $t + 1$.

If, in addition, N is the origin-destination matrix of commuters at time $t + 1$ divided by the employed resident labor force in each location of origin, then

$$D_{t+1} = L_{t+1} \times N$$

This is the vector of labor demand (jobs) among regions at time $t + 1$.

In essence, by using the two transition matrices M (for residential changes) and N (for commuting) for the employed labor force, we can measure over a specified period of time the combined effects of these two types of mobility in response to the demand for labor (Barsotti *et al*., 1976).

The model described above is oversimplified and should be viewed as only a framework because it is based on the classic view of labor force residential choice. Modern approaches view residential choices as a way to minimize work-residence distance and as a response to other criteria and constraints. An individual working and residing in location A may decide to move to B, thereby generating a daily trip from B to A where none previously existed. The classic model would not account for this, but current models recognize that the trade-off between work and residence locations is not entirely determined by the characteristics of the place of residence and that reducing the commute to work is not always the primary motivation for deciding where to live and how one uses the surrounding space.

An individual's place of residence is at the center of a widespread and interlinked network crossed by numerous pathways. Trips and travel along these paths are a way to use space to satisfy the diverse daily demands faced by each individual, and the resulting network defines a person's *life space*.

In analyzing daily travel, we often limit ourselves to trips such as to work or to school because of the relative abundance of data for these two types of travel compared with others. In the literature on gravitational analysis, shopping trips are occasionally taken into account, but travel based on free-time activities for leisure, social contacts, and health are generally ignored. Although the frequency of this type of travel may not be daily, it is becoming increasingly important because the number of such trips can exceed work or school trips and can be decisive in defining the life space and determining the quality of life.

A survey we took in Tuscany (Bottai and Barsotti, 1994) found that daily travel for work reasons represented barely 28% of the total number of trips. Adding trips to school raised the total proportion to 37%. These two types of travel formed only a little more than one-third of the volume of daily mobility. If we had only examined work and school travel, almost two-thirds of daily mobility would have been missed. Leisure trips (17%) and visits with family and friends (14%) form almost another one-third of the total, with the remainder (32%) attributed to shopping trips (15%) and, less frequently, travel for health, companionship, religion, volunteerism, and other reasons. If we distinguish between trips that must occur and those that are elective, we find that 40% of trips are for reasons of personal choice and not out of need. By ignoring these types of trips in studies of spatial interaction and territorial gravitation, we lose quantitative and qualitative information. It is as if we were ignoring the trips that most closely reflect spontaneous choices of mobility and use of space and tell us the most about individual attitudes.

The different spatial-interaction approaches can be split into two large classes at the macro level or micro level. At the macro level, attention is focused on territorial units (e.g., zones, communities, regions), and at the micro level, the units of observation are the individuals. The macro approaches comprise *gravitational models*; micro approaches are based on *behavioral models*.

In macroanalyses, spatial interactions are envisioned as the response to locational disequilibria between needs (the demand) and opportunities (the supply) under a cost function (distance), whereas in microanalyses, it is the individual's probability of movement that determines the volume of flows.

The choice of approach is often conditioned on availability of data. The greatest diffusion of macro approaches is probably due to the amount of information supplied by official sources, notably the censuses. In contrast, micro approaches most often use *ad hoc* surveys.

II. MACRO APPROACHES

1. Social Physics

Macro models are generally referred to as gravitational models. Historically, they antedate micro models. When positivism was at its zenith in the 19th century, social scientists competed to uncover hidden natural laws that guided human behavior in the field of mobility, as well as in other areas. They held the leibnitzian conviction that a natural law capable of determining the principles at the foundation of spatial interaction had to exist, and they believed that the laws of newtonian mechanics could be applied to human behavior.

They thought that spatial interactions were the result of demographic masses exerting gravitational attraction forces toward each other, as if variability in individual behavior was not determined by subjective choice, but rather that individual discretion was transcended by gravitational laws. In general, their assumption was that interactions (T_{ij}) between two nodes i and j were directly proportional to their sizes (P_i and P_j) and inversely related to the distance (d_{ij}) between them:

$$T_{ij} = P_i\, P_j\, f(d_{ij})$$

In this equation, $f(d_{ij})$ is a friction function of the distance between i and j.

From Quetelet and Carey to Reilly, proponents of social physics believed that they could apply the principles of newtonian astrophysics to determine the volumes of interactions between two nodes, as in the following equation:

$$T_{ij} = P_i\, P_j\, d_{ij}^{-2}$$

Recognizing that a strict analogy to physical laws may be too severe, a subsequent modification was made to allow flexibility by replacing the fixed exponent of 2 by an empirically estimated parameter β:

$$T_{ij} = P_i\, P_j\, d_{ij}^{-\beta}$$

Carroll and Bevis (1957) extended this gravitational model by adding a stochastic component (Isard, 1960, p. 494–499), but even in this case, important concessions had to be made: structural homogeneity in the sending and receiving "masses," friction that was independent of distance, and the modeling of individuals as particles without will. To each individual is attributed the same probability function for interactions that includes the probability of moving from node *i* (estimated by the ratio of all movements to the total population) and the probability of moving to node *j* (estimated by the ratio of the population at node *j* to the total population). Multiplying these two ratios together produces the expected number of interactions by any individual in the entire system with node *j*. Multiplying again by the population of *i* produces the total number of expected interactions from *i* to *j*, neglecting the impedance of distance:

$$pr(t_{ij}) = \frac{T}{P}\frac{P_j}{P}$$

Therefore,

$$T_{ij}{}^{*} = pr(t_{ij})\; P_i$$

In the previous equation, $pr(t_{ij})$ is the probability that a resident of *i* moves to *j* in the absence of friction due to distance; *T* is the number of overall commuting trips; *P* is the total "mass" of the system (i.e., the total population); and T_{ij}* is the number of trips expected between *i* and *j* in the absence of friction due to distance.

The ratio between the observed and expected interactions for each pair of nodes in the system is variable. Classic gravitational theory defines this variability as being a function solely of distance. By plotting each of these *ij* ratios and the distance between them on log-log axes, it is possible to find the function that best represents the friction due to distance. Generally, this function is assumed to be linear.

It can be shown that this approach is equivalent to the classic gravitational model:

$$\ln\left(\frac{T_{ij}}{T_{ij}^{*}}\right) = a - b \times \ln(d_{ij})$$

Therefore,

$$T_{ij} = \frac{\left(e^{a} \times T_{ij}^{*}\right)}{d_{ij}^{b}}$$

and

$$T_{ij} = \frac{e^{a}\,\dfrac{T}{P}\,\dfrac{P_j}{P}\,P_i}{d_{ij}^{b}}$$

Assuming that

$$G = e^{a}\,\frac{T}{P^{2}}$$

produces the following:

$$T_{ij} = G\,\frac{P_i P_j}{d_{ij}^{b}}$$

This pseudo micro approach leading to a macro model ignores individual behavior and treats individuals as homogeneous particles without preferences, undermining the usual point of micro models. A model cannot be considered as a micro model if it lacks the essential characteristic of allowing individuals to behave individually.

Isard, a disciple of the classic gravitational model, recognized the limitation and rigidity of using a law from physics as a model of trips. In reality, the exponent that modeled the friction of distance on the possible interactions combines too many different and non-negligible effects; the result is that the empirical data and the expectations based on the gravitational hypothesis are not close. Isard and geographers of the Chicago school sought a more flexible general law by introducing additional weights or coefficients to model mass, and exponents to model location-specific characteristics of accessibility and attraction. This becomes clearer when examining specific masses (i.e., groups of individuals) and their specific interactions; the friction of distance would be very different, depending on whether the groups were men or women, young or old, workers or retirees, and similarly, if work-related travel, leisure-related travel, shopping, or even weekends and weekdays are examined.

Isard's model presumes that the mass weights or coefficients and the exponents, estimated from empirical studies and exogenous to the model, vary according to the type of mass and the type of interaction; the distance exponent is estimated from the model:

$$C_{ij} = \frac{w_i P_i^a w_j P_j^b}{d_{ij}^{\beta}}$$

Bramhall (Isard, 1960, p. 543–544) hoped to sidestep this expansion of parameters:

> We prefer to maintain the point of view that the gravity model as defined by equation . . . applies to whole, integral masses; that is, to migration of all income, occupational, and racial classes and not to migration of Puerto Ricans alone; to passenger trip of all classes by all media over all distances and not to airline trips between New York and Miami; to all shipment of all commodities over all distances through all types of media and not to the shipment of Maine potatoes alone. To reiterate, the incorporation of all cases in the analysis cancels out the peculiarities, irregularities, and idiosyncrasies of interaction within a special sector or category of population or along a selected channel. Just as the brownian movements of any given microscopic particle are reduced to insignificance by the process of averaging over a multitude of particles, so are the capricious movements of any unit or small subgroup of units of population or other relevant social mass.

Because the masses of interest are population sizes, the weights should reflect a series of structural variables that influence the propensity to create or attract movements. It has not been clearly elucidated whether the interactive phenomena should be viewed globally or separately. If all movements, of whatever type, had to be viewed globally, the mass weights would be extremely complex because the variables that affect the different types of movements are multiple and diverse. Whichever we choose, it can only be arbitrary and approximate. Suppose in contrast that flows are decomposed into separate components (at least at the level of motivation or means of transport) and take, for example, work-related commuting. The total population is clearly too gross a measure of the actual population at risk. A young population with high activity rates will exhibit more travel than an older one with lower activity rates; a locality with large mass and a surplus of demand (i.e., jobs), compared with one with a surplus of labor, should have a larger weight.

The list of potential variables that characterize how a mass generates or attracts movements could be very long. This is why much more attention ought to be focused on describing a function that combines all of them into a single weight.

The sign and the magnitude of the mass exponents are unclear. Should they be parameters that depend only on specific characteristics of the mass (e.g., its geographic accessibility) or a simple control parameter permitting flexibility and a good model fit?

The measurement of distance also deserves attention. Imagine several reasonable measures:

Physical distance: distance as the bird flies, driving distance, travel time, the frequency of connections
Economic distance: transport cost
Social distance
Psychological distance

The hypothesis of a linear relationship in the logarithms of distance and the ratio of expected to observed interactions is often empirically unsupported. Other forms of functional relationship better represent the friction of distance. Several functions have been proposed for the distance component (Carrothers, 1956; Anderson, 1955). In particular, Carroll and Bevis (1957) found that a quadratic function seemed to fit well, which is consistent with more recent findings of our own (Bottai and Barsotti, 1994; Barsotti *et al*., 1990). Nonetheless, choosing a form for the distance function leads us away from the earliest models of gravitational physics without bringing us any closer to a theoretical alternative.

Even the most ardent supporters of the gravitational model admit imperfections in the theory and in its empirical performance, and it is not realistic to assume social masses behave according to the laws of physics. Particulate matter is devoid of will, and its behavior answers to an immutable law; human behavior and interactions answer to mechanisms that are far more varied, subtle, and complex.

2. Travel and the Structure of Opportunities

Another approach that models the distance-decay effect on interactions is Stouffer's (1940) theory of *intervening opportunities*. Travel from node i to node j is directly proportional to the volume of potential travelers in i and the opportunities in j that could satisfy their needs, and it is inversely proportional to the number of intervening opportunities available between i and j. Although distance is not an explicit variable in this model, it is obvious that as distance increases, so do intervening opportunities. Stouffer's model has rarely been applied because of the practical difficulties in collecting data on intervening opportunities. However, the model does have the virtue of moving beyond a strictly mechanistic view of the effect of distance, and it implicitly introduced the concept that opportunities to satisfy needs and preferences are not located at a single location but are geographically dispersed.

In effect, presuming that frequency of travel decreases with distance (with a maximum at the shortest distances) implies that destinations satisfying individual needs are uniformly distributed in space. In that case, the distance one travels depends only on the costs of travel; the probability of travel necessarily decreases with increasing distance. Consider a homogeneous space in which there is a constant distance (k) between destinations where needed services can be found (i.e., the supply), with an equally uniform but denser distribution of residences (i.e., the demand for those services); then trips to these destinations increase in frequency up to the distance $k/2$. In like manner, suppose that the demand is distributed in concentric rings around each supply point, with larger and larger radii. In the deterministic framework, this implies that the frequency of travel varies linearly from 0 to $k/2$ (with slope π), which determines the point of indifference between two points of attraction. The modal distance $k/2$ varies with how far apart the points of attraction are; this measurement depends on whether they are workplaces, schools, food or shopping opportunities, and so on.

The assumption of a homogeneous and uniform distribution of opportunities and residences is not realistic. Residences are clustered in some areas and absent in others, as are job opportunities. For each type of move, the distance k between pairs of destinations is variable, and the frequency of moves of a given type depends on numerous factors such as those described earlier.

3. Statistical Mechanics

Wilson (1968, 1971) refined the model by applying the concept borrowed from physics of *entropy*, which describes the tendency of a system to enter a more probable state. Suppose we have a regional system for which we can write down an origin-destination (O-D) matrix of flows (e.g., work commutes), for which rows represent destinations and columns represent origins, with fixed row and column marginal totals, and the added constraint of constant transport costs (or equivalently, travel distance). Our goal is to find the trip distribution that maximizes entropy (Tocalis, 1978, offers more details).

Wilson showed that the most probable pattern of trips between nodes i and j is directly proportional to the trips produced by i, to the trips directed toward j, to the ability of j to attract trips, to the ability of i to produce trips, and inversely to transportation costs (and therefore to distance) between i and j. This may seem no different from the gravitational model, but it is based on a model of statistical mechanics that provides a more solid fundamental theory than that of the gravity model.

Wilson showed that the trip matrix that maximizes entropy and produces the least-biased estimate of the trips between pairs of nodes is given by the following doubly constrained gravity model:

$$C_{ij} = A_i B_j O_i D_j f(d_{ij})$$

In this equation,

$$A_i = \left[\sum_j B_j D_j f(d_{ij})\right]^{-1}$$

$$B_i = \left[\sum_j A_j O_j f(d_{ij})\right]^{-1}$$

and, in general,

$$f(d_{ij}) = e^{-\beta d_{ij}}$$

In addition to its theoretical offering, the value of Wilson's model lies in its introduction of constraints that ensure consistency between expected trips and the marginal totals of the trip matrix. The introduction of constraints significantly improves the agreement between theoretical and observed trip flows compared with the unconstrained newtonian gravitational models.

4. Validation of Wilson's Model

To calibrate and estimate Wilson's model with actual data we must have a trip matrix between different nodes (O and D) and the corresponding distance matrix (Stillwell, 1991).

The main goal is to estimate the parameter β, which measures the friction of distance in the function $f(d_{ij})$. Because the parameters A_i and B_j are defined jointly and simultaneously, there is no closed-form solution, and we must resort to iterative methods of estimation.

We start with an initial arbitrary value for β and calculate the vector A after initializing B at unity; subsequently, we calculate values of B based on the estimated A and repeat the process until convergence. This gives the matrix of expected trip flows. Then we calculate the mean distances in the matrix of expected trips and check for convergence to those of the observed data. If, in absolute value, the difference is larger than the convergence threshold, we change the value of β and try again. At the end, we obtain an estimate of β and a trip matrix that maximizes entropy while maintaining constraints on the marginal totals and with a mean distance implied by the theoretical trip matrix not significantly different from that of observed trips.

We can also construct models constrained only for marginal row or column totals by replacing the omitted constraint by a proxy (i.e., population or workplaces). Relaxing one of the constraints results in a poorer fit to the data, but it allows the examination of alternative scenarios on the unconstrained variable.

5. Subsequent Models

The fundamental difference between Isard's model and Wilson's model is that Isard's poses only an overall constraint on total number of trips in the system and exogenously determines (i.e., by empirically calculated and arbitrarily chosen weights and coefficients) the pushing and pulling forces of each node. In contrast, Wilson's model binds the system to a fixed total cost (distance) and can include constraints on the vectors of trip flows and, by tracing these flows, can endogenously determine the push or pull forces.

Other variations based on the assumption that marginal row or column totals are allowed to vary within a given interval (i.e., "relaxed models") have been proposed as a way to add greater flexibility to the entropy model (Hallefjord and Jörnsten, 1985). Alonso (1978) proposed a general model from which it is possible to derive four typical variants on the Wilson model, ranging from the total absence of all constraints to the doubly constrained model. Alonso described his model as an integration of two submodels: a push-pull model of spatial interactions and a model of their distribution. This results from the role that systemic or relational variables (A and B) play as normalization factors and as explanatory variables.

Dorigo and Tobler (1983) have proposed a model similar to Wilson's, except for the fact that the relation between the normalization factors A and B is additive rather than multiplicative.

Ledent (1985) gives a more general formulation of doubly constrained models, in which an additional parameter provides the flexibility to allow for all variations of both the Wilson and Tobler models as special cases.

In an article, Tobler (1995) mused whether Ravenstein's migration laws, formulated more than a century ago, were still valid. He concluded that the analogy to physical laws remains useful for explaining spatial interactions. A fundamental continuity exists between successive macro models for which the gravitational principle is the common thread. However, there have been refinements that improve the predictive value of the models and, from this perspective, one can distinguish pre-Wilson from post-Wilson models: The latter models are able to integrate the internal structure of the trip matrix and to estimate all of its parameters endogenously. By doing so, those models are more coherent and consistent.

6. Trends in Social Context and Interpretive Paradigms

Gravitational models have played and continue to play an important role in the migration field. However, they have not been much used in the study of daily travel, especially for commuting.

That so little attention has been paid to daily trips seems less due to theoretical shortcomings or practical difficulties than as a kind of atonement for the original sin that stains the gravitational model. The gravitational model was born in an era (the second half of the 19th century) characterized by centralization in activities of production, which stimulated vast migratory flows toward cities, where residences and workplaces were closely situated. Work-to-home travel was not perceived as a problem. Work hours expanded, and individual differences and needs contracted. Freedom of choice was limited by structural constraints. The common man could be reduced to a molecule without will or volition.

With time, the rise of the capitalist system led to a more complex social structure. Progressive improvement in living standards and the transition to a postindustrial society have resulted in stratification in the social fabric, variations in consumption patterns, geographic differentiation, and rise of polycentric urban patterns.

This differentiation also applied to spatial use. Residential location is no longer dictated by job location. Other factors, often of a subjective nature, come into play. Structural constraints still hold sway, and individuals cannot dispense with them; sometimes, the constraints dominate, but in the end, it is the individual who, weighing social context and consequence in dialectic battle, must make the final decision.

An individual's life space expands with diversification and the multiplication of spatial interactions. Daily mobility has become increasingly important as a way to equilibrate labor supply and demand and as a way to satisfy personal needs and requirements, including personal care, developing a network of friends, leisure time, sport, shopping, studying, and voluntary activities.

When making the choice about where to live, individuals no longer view the job market and the place of work as monolithic constraints, but as part of a rich web of other constraints and opportunities. The housing market, the neighborhood environment, transportation, and communication networks are all evaluated for their ability to satisfy specific needs that depend on family composition and the stage of the life cycle.

7. Huff and *Homo economicus*

A clear weakness of gravitational models is that they ignore the effect of freedom of choice on more than one possible interaction (i.e., they cannot distinguish from among different ways to satisfy the same need). To address this other aggregated models have been developed to include theories of behavior.

Huff (1960, 1962), building on concepts drawn from behavioral psychology, developed a theory of individual choice based on trade-offs between the need to move (as a function of economic, social, cultural, and demographic characteristics) and psychological and physical barriers in the spatial structure of opportunities. His spatial interaction model was based on a behavioral approach in which individuals, to satisfy their needs, choose the location that maximizes expected utility. For each decision, individuals behave as if they perform a cost-benefit analysis (Berry, 1978, pp. 99–108).

This model could have led to a micro approach, but Huff's model remains a macro one. Because constructing utility matrices for each individual is difficult, Huff's approach was to combine individuals into the most homogeneous (from a socioeconomic perspective) groups possible. For example, in looking at shopping trips, principal factors in the model must include the type of merchandise offered, travel cost, income, and the size of the locality.

Each node in the system can be described by a series of macro variables that reflect an individual's utility and requirements, and the probability of spatial interaction for a particular node will depend on how this utility is achieved in comparison to other nodes.

The multinomial logit model followed the path blazed by Huff. That model rests on the assumptions that individuals exhibit economically rational behavior and that the utility functions that guide their rational behavior have random variable components that are independent and identically distributed. Refinements of the model have targeted the unrealistic assumption of independence; some alterations (Fotheringham, 1983; Gaudry and Degenait, 1979) relate to deterministic behavior, whereas others (McFadden, 1978) focus on risk.

Based on principles of behavioral science, this approach to the analysis of social interaction provided a way to escape the bonds of the gravitational model that so tightly linked individuals together. Interaction was no longer a mechanical response to structural disequilibria but rather the result of individual choices aimed at maximizing utility by avoiding costs and embracing opportunities afforded by macro-structural factors.

Although this step represents an important advance in the understanding of the process and the probabilistic approach adds predictive value, there are still some limitations. First, the underlying ideas of *discrete choice* and *rational expectations*, borrowed from microeconomics, presume abstract rationality of *Homo economicus*. Real-world behavior rarely conforms to that of *Homo economicus*. Economic rationality is only one of the behaviors (and rarely the most important) in a decision sequence that must weigh psychological, emotional, social, and cultural imperatives. It is only slight exaggeration to say that we have merely traded the bear trap of physical laws for the iron fist of economic rationality. Second, although the theoretical approach seems to fit within a micro-level model, it employs the macro-level instruments of ecologic variables and proxies for the way that individuals evaluate the value of differing locations.

In a complex society where situations and behaviors are so varied, we should not force behavior to match unrealistic and abstract theories. It is better to focus on less ambitious goals and to ask of our models only that they provide a true picture of the relationships that link the variables and a limited and current description of the interaction system, without pretensions to generality in time or space.

8. Applications of Macro Models

In the latest literature, gravitational models focus on the different factors that segment trips into different types of travel. There are two general approaches. The first applies separate models to homogeneous subgroups that share the most important defining characteristics; the other is based on classic linear models with additional explanatory factors. These models generally take a log-linear form.

The segmentation approach is used, for example, in a study by Cervero and Wu (1997). The authors considered four different combinations of the gravitational model by combining two worker categories (i.e., professional and nonprofessional) with two categories of work destination (i.e., traditional work destinations and newer, rapidly expanding employment zones). Their model included the usual variables: spatial impedance (i.e., travel time) and mass (i.e., labor supply at the origin and labor demand at the destination), and it included other explanatory variables, such as the destination's employment rate and the housing characteristics at both the origin and the destination.

The coefficients on these explanatory variables were statistically significant and of the expected sign, but overall goodness-of-fit was meager. This indicates that their model omitted key variables at the heart of the mechanism that generates daily trips, that the log-linear formulation was inappropriate, or that stochasticity dominated the decision-making process. This last point would mean that even within supposedly homogeneous subgroups, there is no universal behavioral response with respect to the use of space. However, we must not overlook the fact that the region studied (i.e., San Francisco bay area) has a polycentric urban structure that may have affected the results.

We may be in the middle of a transitional period between an historical phase of monocentric urban development in which work mostly determines travel and a new era characterized by polycentric urban patterns in which diverse demands for the daily usage of space influence the choice of where to live.

To evaluate the explanatory value of a gravitational model we must take into account the size of the geographic units for which data are available, because the larger the unit, the fewer the number of cross-boundary trips. This point, although critical, is often overlooked. In particular, it appears that the distance-decay parameter in gravitational models is not uniform with distance. Indirect support for this can be found in the positive coefficient on the index of urban contiguity in Baccaini's gravitational model (1997), which demonstrates that the friction of distance does not vary linearly with travel time (Barsotti and Bottai, 1994). More support can be found in a comparison of the Baccaini study with that of Cervero and Wu. Both are similar in that they update classic gravitational models by introducing explanatory variables reflecting polycentric urban patterns; however, only 33% of variance is explained by Cervero and Wu's model, whereas Baccaini's accounts for 96%.

Vickerman (1984) focused on the suburbanization processes in the London metropolitan area and observed that they led to an increase in the number of commutes. The direction of travel flows also changes as productive activities and jobs are dispersed toward the suburbs. Residential dispersion precedes job dispersion—or at least it occurs more quickly. Travel through the center of the urban system can increase even as the number of jobs located there decreases. City center residents are pushed out in a residential diaspora by rising housing prices, exacerbated by competing nonresidential land use, and by the increasing role of subjective and noneconomic factors in residential choice. Because of this, people no longer move to reduce home-to-work travel; instead, a residential move generates new commute trips.

The choice of the place of residence and type of housing are factors that strongly influence daily travel. Although this is obvious, the exact mechanism by which this operates is much less so. Both the housing market and urbanization policies have shown a weak ability to adjust to disequilibria between places of work and residence, which is important quantitatively and qualitatively. The housing market is increasingly segmented to the extent that almost everywhere (especially in Italy), it is dominated by the dream of owning one's own house, which tends to favor the migration of families with specific characteristics (e.g., income) to specific areas. If, in contrast, home-ownership status (i.e., renter or owner) and housing characteristics are satisfactory, migration incentives are greatly weakened, even in the face of job commuting. The housing market itself can raise barriers, as in the case of rent control. Such barriers can result in gentrification and segregation in urban and suburban centers.

It is difficult to imagine how we could disentangle the interdependencies of daily travel. Migration, residential location, job location, and travel costs are all factors jointly influencing daily mobility.

III. MICRO APPROACHES

The growing complexity of daily travel has given rise to the use of micro models. Even as macro models became more disaggregated and targeted, interest began turning to individual behavior as a key to understanding the travel behavior of groups. The goal of this approach was to highlight the ways in which structural factors are perceived and evaluated by individuals, depending on their demographic, social, occupational, and cultural characteristics. Surveys are the typical instrument needed to conduct this type of microanalysis.

As for all multivariate analyses, the most critical aspect is the selection of explanatory variables, because unexplained heterogeneity can bias the estimation of model parameters.

Gender is one of the factors most commonly studied. Gender differences in attitudes toward daily travel are so strong that it is often better to apply separate models to each sex than to introduce it as an explanatory variable in a single model (Baccaïni, 1997; Tkocz and Kristensen, 1994). It has been observed (Johnston-Anumonwo, 1992; White, 1977) that women are less likely to move long distances. This also appears to apply to daily travel; compared with men, women prefer shorter commutes.

The difference between single- and dual-worker families has also been studied. When both spouses work, family strategies seem to favor the husband's career, even if the resulting residential choice makes him commute longer. The wife then seeks employment closer to home, even if the resulting career opportunities are more limited. Even in the most advanced societies, there is asymmetry between male and female roles of production and reproduction (Madden, 1981).

The stage in the family life cycle is an intervening variable that can modify the effects of the two previous factors. The presence or absence of children in the household is commonly used as a proxy for the family life cycle stage. It has been shown that if the wife continues to work after the birth of children, she will try to bring workplace and home closer together. More precisely, she will try to minimize the distances between home, workplace, and the additional services (e.g., schools, shopping) needed to support the new family situation even though overall distance traveled and time taken may increase with the additional trips to school or shops (Camstra, 1996; Barsotti and Bottai, 1994; Rosenbloom, 1989).

Marital status seems to have the greatest influence on men's daily travel. Bachelors appear to commute shorter distances because it is easier for them to find housing in city centers or close to work (Baccaïni, 1997; Tkocz and Kristensen, 1994; Vickerman, 1984). However, the effect of marital status is not unambiguous and provides a good example of the way that the behavior of specific subgroups can differ. Although bachelors are more likely to live near work because of greater flexibility in housing size, characteristics, and cost, they are also less bound by familial or non–work-related constraints and therefore have the option to commute farther.

For adults, daily travel, like migration, tends to decline with age. According to some demographers, it also tends to be redirected toward urban centers. If one examines the entire age pattern and includes work and school travel (as perhaps we should, considering that they share the characteristics of both being compulsory and occurring on a daily basis), a graph of the relationship between age and daily travel resembles an inverted U (Barsotti and Bottai, 1994).

Taking income into account poses a number of problems. Should we consider household income or, especially when focusing on work travel, individual incomes of each family member? No matter how it is defined, commute distances and frequencies increase with income.

Socioeconomic status has considerable effect on daily travel. Home workers and farmers, for example, work close to or within their place of residence. The self-employed typically have low mobility. When controlling for other explanatory variables (e.g., sex, age, education, and home ownership status), the most mobile groups are supervisory white-collar workers. However, as shown by some of the studies we have discussed, blue-collar and office workers are among those with the longest commutes. This greater amount of daily travel (which has been rising in recent years) has resulted from the dispersion of residences to the suburbs.

Trends in daily commuting also depend on changes in home ownership status. Rising rates of home ownership reduces the propensity to change residence when changing the place of employment (Bottai, 1995).

Analysts can continue to add to the list of explanatory variables, including education (i.e., daily travel increases with the level of education) and modes of transport (e.g., cars, buses, trains, bicycles), which are determined by individual needs and by the existing transportation infrastructure and land use policy. All things considered, careful exploratory analysis is always well advised.

IV. THE CHALLENGES FACING A UNIFIED APPROACH

The need to reconcile the micro and macro approaches is clear. Two of the studies we have mentioned take two alternative paths in that direction.

Baccaïni (1997) supplements a micro model of individual-level variables with the addition of two variables that capture the regional context: one that takes into account spatial characteristics (geographic zone) and anotherthat measures the balance between labor supply and demand (i.e., employment rate) in the location of residence.

Tkocz and Kristensen (1994) build a polycentric spatial model in which the relationship between work-home distance and home-center distance is represented by a third-degree polynomial. In this model, the coefficients are replaced, following the method described by Casetti (1972), by a series of sociodemographic variables validated in a micro model.

Baccaïni's model begins as a micro model to which macro variables are added, and Tkocz and Kristensen's model begins as a macro model to which micro variables are added.

Courgeau (1999) carefully analyzes the biases that aggregated (macro) models can favor because they ignore how individual characteristics influence the likelihood of moves. These biases can be categorized as an ecologic fallacy arising from the reliance on large-scale empirical data. He also explores how biases arise in micro models by erroneous specification of individual-level variables and especially by discounting the influence macro variables have on the micro variables themselves. This bias is a type of atomistic fallacy. The probability of moving from one location to another for work reasons does not depend only on the type of job that an individual has (a micro-level characteristic) but also on the occupational structure of the area where the individual lives (a macro-level characteristic).

It seems more promising to tackle the problems of spatial interaction, as well as all similar problems of collective behavior, beginning from a micro perspective and adding macro variables of different levels of aggregation as needed. In any event, the macro-micro dichotomy seems too manichaean. In studying residential mobility, it may be better to begin at a point in the middle (i.e., with the household) because the choice of where to live concerns the household standing together rather than each individual standing alone. In this case, the problem, which several researchers are currently investigating, is how to handle characteristics of each member of the household (i.e., variables at a less aggregated level than the household itself).

The increasing preference for multilevel micro models implies a rejection of the typical macro approach of formulating laws that govern spatial interaction. The underlying reasons for this epistemologic evolution are the profound changes that have rendered modern and postmodern society more complex and differentiated. These changes have sparked the transition from monocentric spatial patterns to polycentric ones; an expansion of life space; a multiplication and increasing diversification of the reasons for using space; and the social fragmentation of mobility.

That we take recourse in microanalyses is as much due to the growing importance of individual choice as to the need to uncover by survey data the processes and mechanisms that give us insight into spatial behavior. We are witnessing a change of logic, and the transition from the cold deductions of macro models to the recognition that potential laws of interaction may only be discoverable by induction.

Acknowledgments

This chapter was translated from the French version by Robert Chung. Portions of this chapter previously appeared in the book by Tabutin and colleagues (1999).

References

Alonso William, 1978. A theory of movements. *in*: Niles M. Hansen (ed.) *Human settlements: international perspective on structure, change, and public policy*. Cambridge, MA, Ballinger, p. 197–211.

Anderson Theodore R., 1955. Intermetropolitan migration: a comparison of the hypothesis of Zipf and Stouffer. *American Sociological Review*, vol. 20(3), p. 287–291.

Baccaïni Brigitte, 1997. Commuting and residential strategies in the Île-de-France: individual behaviour and spatial contraints. *Environment and Planning A*, vol. 29, p. 1801–1829.

Baccaïni Brigitte, 1997. Les navettes des périurbains d'Île-de-France, *Population*, vol. 52(2), p. 327–364.

Barsotti Odo, Bonaguidi Alberto, and Bottai Marco, 1976. *Mobilità della Forza Lavoro e Zone di Programma, Regione Toscana*, Florence, Eurografica S.p.A., 111 p.

Barsotti Odo, Bottai Marco, and Costa Marco, 1993. Localisation centrale et potentiel démographique pour l'analyse des comportements différentiels de croissance urbaine: le cas de la Toscane, *in:* AIDELF (ed.), *Croissance démographique et urbanisation, Actes du séminaire international de Rabat*, 15–17 May 1990, p. 251–263. Press. Universitaires de France, Paris, 430 p.

Barsotti Odo and Bottai Marco, 1994. La ricerca "Spazio Utilizzato." Obiettivi e primi risultati, *in*: Marco Bottai and Odo Barsotti (eds.) *Lo spazio e la sua utilizzazione*, p. 16–60. Milan, F. Angeli, 219 p.

Berry Brian J. L. (ed.), 1978. *The nature of change in geographical ideas.* DeKalb, IL, Northern Illinois University Press, 167 p.

Bottai Marco, 1995. Migratory careers. An application and some suggestions, *in*: EAPS-IUSSP (eds.), *Evolution ou revolution dans la population européenne*, vol. 3. Milano, F. Angeli, p. 153–170.

Bottai Marco and Odo Barsotti (ed.), 1994. *Lo spazio e la sua utilizzazione.* Milan, F. Angeli, 219 p.

Bottai Marco and Barsotti Odo, 1999. Approches et courants explicatifs des mouvements pendulaires, *in*: Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant, and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie, Chaire Quetelet 1997*, p. 207–226. Louvain-la-Neuve, Academia-Bruylant, l'Harmattan, 669 p.

Bramhall David F., 1960. Gravity, Potential, and Spatial Interaction Models, *in*: Walter Isard (ed.), *Methods of regional analysis: an introduction to regional science*, p. 493–568. New York, J. Wiley & Sons, 784 p.

Camstra Ronald, 1996. Commuting and Gender in a Lifestyle Perspective, *Urban Studies*, vol. 33(2), p. 283–300.

Carroll J. Douglas and Bevis Howard W., 1957. Predicting local travel in urban regions. *Papers and Proceedings of the Regional Science Association*, vol. 3, p. 183–197.

Carrothers Gerald A. P., 1956. Historical review of the gravity and potential concepts of human interaction. *Journal of the American Institute of Planners*, vol. 22, p. 94–102.

Casetti Emilio, 1972. Generating models by the expansion method: application to geographical research. *Geographical Analysis*, vol. 4, p. 81–91.

Cervero Robert and Wu Kang-Li, 1997. Polycentrism, commuting, and residential location in the San Francisco Bay area, *Environment and Planning A*, vol. 29, p. 865–886.

Courgeau Daniel, 1999. De l'intérêt des analyses multi-niveaux pour l'explication en démographie, *in*: Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant, and Bruno Schoumaker (eds.) *Théories, paradigmes et courants explicatifs en démographie*, p. 93–116. Louvain-la-Neuve, Academia-Bruylant, l'Harmattan, 669 p. (Actes de la Chaire Quetelet, 1997.)

Dorigo Guido and Tobler Waldo, 1983. Push-pull migration laws. *Annals of the Association of American Geographers*, vol. 73, p. 1–17.

Fotheringham A. Stewart, 1983. A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A*, vol. 15(1), p. 15–36.

Gaudry Marc J. I. and Dagenais Marcel G., 1979. The dogit model. *Transportation Research*, vol. 13, p. 105–111.

Hallefjord Asa and Jörnsten Kurt, 1985. A note on relaxed gravity models. *Environment and Planning A*, vol. 17(5), p. 597–603.

Huff David L., 1960. A topographical model of consumer space preferences. *Papers and Proceedings, Regional Science Association*, vol. 6, p. 159–174.

Huff David L., 1962. Note on the limitation of intra-urban gravity models. *Land Economics*, vol. 38(1), p. 64–66.

Isard Walter (ed.), 1960. *Methods of regional analysis: an introduction to regional science*. New York, J. Wiley & Sons, 784 p.

Johnston-Anumonwo Ibipo, 1992. The influence of household type of gender differences in the work trip distance. *Professional Geographer*, vol. 44, p. 161–169.

Ledent Jacques, 1985. The doubly constrained model of spatial interaction: a more general formulation. *Environment and Planning A*, vol. 17(2), p. 253–262.

Madden Janice F., 1981. Why women work closer to home. *Urban Studies*, vol. 18, p. 181–194.

McFadden Daniel, 1978. Modelling the choice of residential location, *in*: Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull (eds.), *Spatial interaction theory and planning models*. Amsterdam, North Holland, p. 75–96.

Rosenbloom Sandra, 1989. Trip-chain behaviour: a comparative and cross-cultural analysis of the travel pattern of working mothers, *in*: Margaret Grieco, Laurie Pickup, and Richard Whipp (eds.), *Gender, transport and employment: the impact of travel constraints*. Avenbury, VT, Gower Publishing, 236 p.

Stewart John Q. and Warntz William, 1958. Physics of population distribution. *Journal of Regional Science*, vol. 1, p. 99–123.

Stillwell John, 1991. Spatial interaction models and the propensity to migrate over distance, *in*: John Stillwell and Peter Congdon (eds.), *Migration models: macro and micro approaches*. New York, Behaven Press, p. 34–56.

Stillwell John and Congdon Peter (eds.), 1991 *Migration models: macro and micro approaches*. New York, Behaven Press, 329 p.

Stouffer Samuel A., 1940. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, vol. 5, p. 845–867.

Tabutin Dominique, Gourbin Catherine, Masuy-Stroobant Godelieve, and Schoumaker Bruno (eds.), 1999. *Théories, paradigmes et courants explicatifs en Démographie, Chaire Quetelet 1997.* Louvain-la-Neuve, Academia-Bruylant, l'Harmattan, 672 p.

Termote Marc, 1969. *Migration et équilibre économique spatial.* Louvain, Université catholique de Louvain, Faculté des sciences économiques, sociales et politiques, nouvelle Série, no. 54, 256 p.

Tkocz Zygmunt and Kristensen Gustav, 1994. Commuting distances and gender: a spatial urban model, *Geographical Analysis*, vol. 26(1), p. 1–14.

Tobler Waldo, 1995. Migration: Ravenstein, Thornthwaite, and beyond, *Urban Geography*, vol. 16(4), p. 327–343.

Tocalis Thomas R., 1978. Changing theoretical foundations of the gravity concept of human interaction, *in*: Brian J. L. Berry (ed.), *The nature of change in geographical ideas*. DeKalb, IL, Northern Illinois University Press, p. 66–124.

Vickerman Roger W., 1984. Urban and regional change, migration and commuting. The dynamics of workplace, residence and transport choice. *Urban Studies*, vol. 21, p. 15–29.

White Michelle J., 1977. A model of residential location choice and commuting by men and women workers. *Journal of Regional Science*, vol. 17, p. 41–52.

Wilson Alan G., 1968. Models in urban planning: a synoptic review of recent literature. *Urban Studies*, vol. 5(5), p. 249–276.

Wilson Alan G., 1971. A family of spatial interaction models and associated developments. *Environment and Planning*, vol. 3, p. 1–32.

CHAPTER

65

Studying Return Migration: Data from Censuses, Surveys, and Registers

JEAN LOUIS RALLU

Institut national d'études démographiques (INED), Paris, France

Return migration has not been extensively studied, even though it can affect significant numbers of people. Although some data are available, there are difficulties in defining return migrants, and there is prejudice about most of them being retired and therefore of secondary economic interest. The issue of migrants' retirement pensions to be paid in their native countries, if they choose to return, is beginning to cause some concern. According to the data available, most return migrants are not retired, and many are middle-aged adults. As a consequence, return migration should not be defined as a move for good, but in a wider circulation perspective. A return migrant is a person who goes back to his or her starting point (or even to one of his or her previous residences) and who may or may not migrate again in the future.

The study of return migration is based on data from censuses, surveys, and registers. Because simultaneous censuses and surveys have been conducted in these places, the examples are taken from migration between Guadeloupe and Martinique, the French overseas departments in the West Indies, and mainland France.

I. CENSUS DATA

Using census data, return migration can be defined from information obtained by the questions on previous residence and birthplace. A *return migrant* is defined as a native of the country or region where she or he was interviewed at a census or survey, and has lived in a different country or region at a previous time. If the contribution of foreigners (i.e., of people who were born abroad) to migration is to be measured based on such information, then it seems necessary to establish two categories among the migrants who were spotted by means of the question on their previous residence: those who were born in the country or region (i.e., return migrants) and those born outside the country or region.[1] This applies particularly to countries with formerly high out-migration rates and therefore potentially high rates of return migration, such as Spain, Portugal, and Italy.

In the case of censuses, it is of limited interest to calculate the proportion of return migrants in the population by simply comparing the number of return migrants with that of the natives or of the total population, because migration can be accounted for only partly through the question on previous residence. The return of people who were away at the time of previous residence and had returned some time in the n years between that date and the census is measured, but the people who migrated and then came back in that period, although they are included in the census, are ignored. The limitation also applies to the measurement of migration by means of the same question—those migrants who came and went away again

[1] For the sake of simplicity, only countries are used subsequently in the text as entities.

between the date of previous residence and the census are ignored, but in that case, non-enumerated migrants do not appear in the census.

1. Return Rate

The return rate is the ratio of return migrants to the emigrant population at the time of previous residence.[2]

$$\text{Return rate} = R_{i\,j}^{t\,t-a} \;/\; P_{ij}^{t-a} \qquad (1)$$

In Equation 1, $R_{i\,j}^{t\,t-a}$ equals the migrants who were born in *i* and appeared in the census in *i* at time *t and* resided in *j* at the time of previous residence *t* − *a*, and P_{ij}^{t-a} equals the population who were born in *i* and were enumerated in country *j* at the time of previous residence. Because new migration between the date of previous residence and that of the census represents a bias, *a* should be quite small (ideally, about 1 year).

Mortality is a disturbing event and should be taken into account through an estimate of the survival rate between *t* − *a and t*. However, its effect is insignificant among young and middle-aged adults, although it is different at later ages, such as when considering migration about retirement age.

Age was not included in the notations to keep the formulas as light as possible; however, those rates should be calculated by birth cohort and consequently require detailed age data when censuses are not conducted at 5-year intervals. This results in unconventional age groups (Table 65–1).

Using Equation 1, it is possible to calculate the return rates for different *j* countries of previous residence, provided those denominators can be derived from censuses in the countries of previous residence if they took place at the date or at a date close to the reference date for the question of previous residence or, if that is not the case, from estimates of the native population who have emigrated to the relevant *j* countries.

2. Periodicity and Simultaneousness of Censuses

It appears that simultaneous censuses in the European countries make it possible to measure international (and internal European) migrations in much better circumstances than censuses conducted at different dates. Previous residence had best be measured at the date of the last census. However, in a closed or all but closed area, denominators can be calculated again based on the population enumerated in the census and on the migration that was observed (from the census or from a different source) during the same period. It then seems better to use, as some countries do already, several questions about previous residence 1 year before the census to calculate rates for 1 year and 5 years before, to get more significant numbers for the analysis of the social and demographic characteristics of the migrants, and if necessary, at the date of or at a date close to the last census, for an assessment during the period between the censuses. The further back in time the previous residence is sampled, the more unobserved moves there are.

TABLE 65–1 Rates of Return Migration from Mainland France to the West Indian French Overseas Departments According to Censuses

Age (years) in 1990	Guadeloupe		Martinique	
	Men (%)	Women (%)	Men (%)	Women (%)
18–22	8.0	9.3	6.6	9.3
23–27	8.7	13.4	10.0	13.8
28–32	14.0	15.7	15.2	16.6
33–37	13.2	13.1	12.0	12.7
38–42	12.5	11.4	10.8	10.4
43–47	10.4	8.1	9.3	7.6
48–52	8.8	7.7	8.1	7.0
53–57	8.5	8.9	9.6	8.0
58–62	10.6	12.1	11.6	12.4
63+	11.1	12.2	9.6	10.3

Calculated from the 1982 census in Mainland France and the 1990 census in the French Overseas Departments.

In the frame of simultaneous censuses, it is also possible to check the results of migration from the question on previous residence by using the method of expected population. The expected population in a given *i* country at *t* + *a* provides, by difference with the population counted in the census in *t*, an estimate of migration that can be compared with that obtained from the question on previous residence in the *j* country. If the migration estimated through the expected population method in the *i* country is higher than that observed in *j*, it is (if there are no variations in the completeness of enumeration of successive censuses) because migrants who left country *i* did not necessarily go to country *j* or later remigrated to other destinations. If the migration observed in *j* is higher than the estimate from data in country *i*, it is a sign that there were errors in the data; some statements about previous residence were wrong, or under-enumeration affects the first census conducted in *i*. Such errors may happen in one or the other country, and the results of those comparisons should be very carefully interpreted. Whenever data are available, the moves between *i* and all the other *j* countries should then be

[2] The rates are computed with the initial population as denominator and are actually probabilities.

considered. Where return migration is concerned, an estimate of the expected number of migrants originating from i in the j country is made and then compared with the returns observed in i.

3. Extensions of the Method

What has been studied hitherto is a single question about previous residence. Several countries have multiple questions about residence 1 year and 5 years before the census (e.g., Portugal) and about 10 years before the census (e.g., Spain). Those questions make it possible to study, based on the various previous residences, such facts as the circulation of migrants between the various foreign countries and where they stopped on the way home (including internal migration after their return), which are an important aspect of the migrants' re-localization—the towns where they choose to settle, depending on the latter's size, the region, and other factors. Those data should not be neglected, mainly because the cost is so low, because they are products of the census; however, the various situations are revealed only at various points—as in some registers, such as the PDS in France,[3] not complete biographies; as a result, data from surveys should be used preferably for this type of studies.

Another extension of the method consists in measuring second-generation migration. In that case, the parents' birthplace is enough to spot the children born to migrants in a host country—those who were residing abroad at the time of previous residence are second-generation return migrants. Then they do not exactly return, but migrate, individually or with their parents, to the country where the latter were born.

Some censuses include a question about the year of arrival in the country (or of the latest migration). The question, although more awkward to use in itself (see Chapters 22 and 23), makes it possible, by simultaneously using the flow data from the migration statistics offices, to calculate the return rates, or rather the rates of remigration; the census then provides the number of people who are still present by immigration cohort, excepting persons who were temporarily away and were not counted in the census (Péron, 1977). Based on a series of censuses, the intercensus probabilities of departure from the country of residence can be calculated, but the fact that some people return after they have left the country takes much of the significance away from those probabilities.

[3] Permanent Demographic Survey (INSEE) that includes census and civil registration data.

4. Limits and Use of Census Data

The main flaw about census data is that they do not take into account the migration that took place between the date of $t - a$ previous residence and that of census t. The return migration measured is then bound to be affected by unknown re-emigration (and so is the migration measured through the same question); besides, the two-way trips made between the two countries, between $t - a$ and t, are ignored. These represent visits that are usually shorter than those of the migrants counted in censuses, but there is a bias if the reintegration of people who have come back in the past 5 years, for example, is to be measured.

The theory of return migration has been little developed, as has the study of the phenomenon itself. Two main reasons for returning are distinguished in it: failure or success in migration. Based on that very rough framework, the characteristics of return migrants can be studied, using all the information made available during the censuses (and the surveys), such as sex, age, educational level, labor force participation, and occupation. By calculating the rates by sex and age, it has emerged that the rate of return migration is high among young adults, then it decreases, and it goes up again at retirement age (see Table 65–1). In a study of returns conducted in Portugal and Spain (Rallu *et al.*, 2000), it was revealed that return rates vary depending on the level of skills. The educational level of return migrants is usually higher than that of the native population as a whole,[4] although it appears on analysis that there are also groups of manual workers with little education behind them. Return migrants are usually working people, sometimes employers. This is a means of providing part of the answers to the following questions. Do return migrants create companies and jobs, and what part do they play in the economy and in development?

II. SURVEY DATA

Through surveys, unlike censuses, complete migratory biographies can be recorded. All migratory moves are then made available. It is possible to calculate the proportion of people by birth cohort who have been abroad at least once for a minimum stay of x months or years for a reason given as an answer to an ad hoc question (which may reflect the cause of departure, but

[4] When using census data, we prefer to compare return migrants with the total native population rather than with nonmigrants, because among the latter are included migrants who were not counted through the question on previous residence.

not the reason for the visit nor for extending it) or according to their actual occupation over there (e.g., studies, employment). Surveys are also useful because the number of visits abroad can be calculated along with their total duration and other aspects. In that case, it is important not to forget that in a survey conducted in a given country, only return migrants will be captured, whereas migrants presently abroad are ignored, although some may well have returned before their current migration. Likewise, in a survey conducted in the host country, the migrants who have returned or remigrated are ignored.

1. Return Rates by Cohort and by Rank Order

A survey makes it possible to know the departures to j per year and calculate the return rates by year (cohort) of departure:

$$T_{ij}^{t,a} = R_{ij}^{t,a} \;/\; E_{ij}^{a} \qquad (2)$$

In Equation 2, $R_{ij}^{t,a}$ represents the migrants who have returned from j to i in the year t and who went from i to j in the year a, and E_{ij}^{a} represents the number of migrants from i to j in year a (i.e., cohort of outgoing migrants). The intensity of the return migration from j to i among the cohort of year a of outgoing migrants is the sum of annual rates.

In a survey, it is possible to choose to calculate rates or probabilities in building attrition tables with the usual methods (Fig. 65–1). However, the intensity of returns by cohorts calculated does not provide the proportion of people who migrated to j in a and came back to i at the date of the survey for the cohort under study, because they may have re-emigrated. It is then possible to calculate the re-emigration and return rates by rank for every first-migration cohort or by latest previous migration (the latter may be re-emigration or return) (Fig. 65–2A), and therefore to get a measure of the circulation by combining the probabilities (Fig. 65–2B). It appears that although the migrants in the second half of the 1960s did frequently have a first return, many went away again later in the great wave of migration to mainland France in the early 1970s. It is different with the 1975–1984 cohorts; among them first returns are highest, and re-emigration is low compared with the previous cohorts. The last cohort observed is characterized by low circulation between the overseas departments and mainland France. The 1970–1974 cohort is peculiar in that it shows growing probabilities at every new migration, an effect of the selection of the most mobile people, which results in greater circulation. If the data are to be unbiased, migrants should be interviewed in their native (return) country and in the host country, with a view to securing a total picture of migration per year.

The wide array of opportunities offered by surveys in providing migratory biographies are revealed in this manner. However, the surveys must be conducted in both countries when the migration from a given country is mainly directed to another or several other countries. The cost of such surveys is high, but one large, well-designed survey is better than several imperfect ones or even a number of small-scale ones.

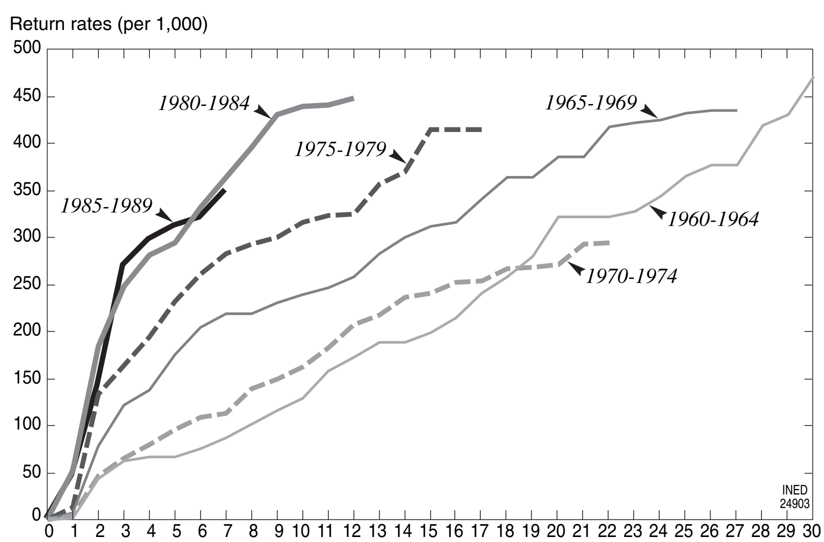


FIGURE 65–1 Total returns in the West Indian Overseas Departments by duration of stay and period of arrival in mainland France, from the probabilities of return (per 1000). (From INSEE survey on migration between the overseas departments and mainland France [Aziz and Lavertu, 1994].)

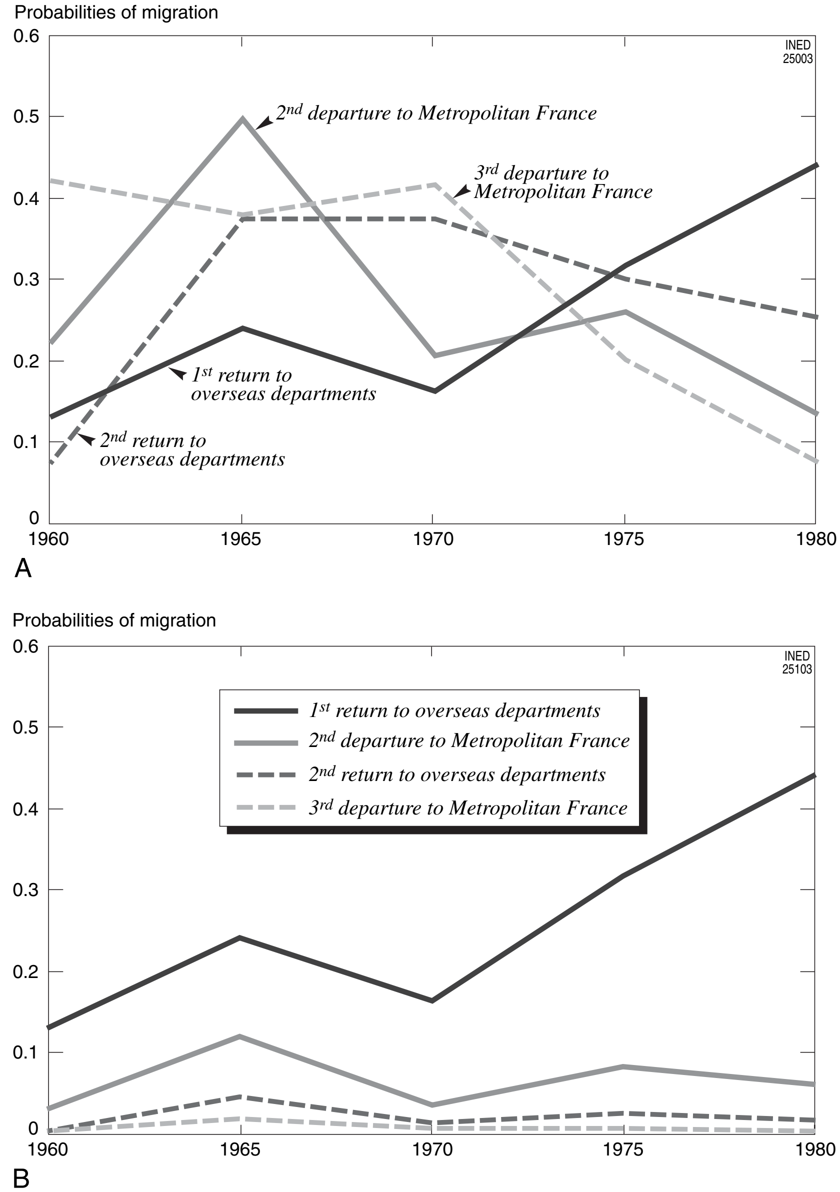


FIGURE 65–2 **A:** Probabilities (migrations n + 1/migrations n) of different types of migration (up to 9 years' duration) between the French West Indies and mainland France by cohort of first migration to mainland France. **B:** Probabilities of the different types of migration (migrations n/first migrations to mainland France) between the French West Indies and mainland France (at 9 years' duration) by cohort of first migration to mainland France. (From INSEE survey [Aziz and Lavertu, 1994].)

Simultaneous or quasi-simultaneous surveys have been conducted, such as the survey on the migration of natives of or residents in the overseas departments conducted by the French Institute of Statistics (INSEE) in 1991 to 1992 (Aziz and Lavertu, 1993, 1994) or the Research Network on Urban Migration in Western Africa (REMUAO) surveys by the ORSTOM (now IRD, French Research Institute on Development) on migration in West Africa.

2. Assessing the Productivity of the Sample

Though it is possible through surveys to measure migration over a long period, as with censuses, surveys may record such past migration as may be of limited interest for analyzing the present situation, particularly the social and economic aspects of migration. For example, a survey may yield information on migrants who returned when they were young adults several years or decades ago, who are presently older and now consider migration as only remotely linked, if at all, to the present situation.

The same applies to the study of the determinants of return migration. To compare the situation of a person who returned at age *a* in year *t*, it is necessary to know the situation of a person who stayed in the country of residence at the same age, in the same year, which implies that work biographies are available. Knowing the situation of the persons in the country of residence at the date of the survey makes it possible to compare the return migrants with the emigrants only in the year of the survey. It appears that the "productivity" of the sample may be considerably reduced if the information gathered is limited only to the current economic situation. Analyzing migratory biographies concurrently with work biographies provides all the resources offered by analyzing that type of data (see Chapter 23).

III. POPULATION REGISTERS

The study of return migration, as well as of migration in general, can be done using population registers. Such studies have been conducted, particularly in the field of historical demography (Tedebrand, 1976).

A distinction should be drawn between two types of registers: those that provide continuous observation of migratory moves—and are very much akin to biographic data—and those that provide truncated data or information relating to certain dates (e.g., the French Permanent Demographic Survey, which is based on censuses) or at the time of certain events (e.g., the respondents' personal address at the time of the vital events recorded by civil registration). Registers that provide biographic data can be analyzed by the methods of biographic data analysis. Registers that provide truncated data can be analyzed using methods similar to that presented for the study of censuses or, more systematically, using the particular methods appropriate for this type of incomplete data, such as have been presented in this book. Registers and surveys are used mainly for a longitudinal study of phenomena, whereas censuses offer mainly cross-sectional analyses. However, in the case of migration, censuses and certain registers provide a longitudinal fragment only, between the date of previous residence or arrival and the census, and that fragment is affected by the cross-sectional enumeration of the population in the census, in which the persons who were temporarily absent were excluded and the migratory moves between the dates under study were ignored.

CONCLUSIONS

The study of return migration must be freed from the idea that it is just a question of aged persons returning for good at retirement. It is part of the study of migration, defined as the circulation of persons between various places, in its various possible forms, including circular migration between two or several places or with more complex routes. In that sense, it is also a necessary element in measuring migration using the question about previous residence. In that circulation and through its age structure, it has significant effects on employment, skills of the labor force, and the economy.

For a complete, unbiased study of migration and of the migrants' characteristics, it is necessary to have data about all migratory moves, based on surveys or censuses simultaneously conducted in the various countries involved. The question about previous residence (usually 5 years before the census to limit the loss of information that may happen if the period is too long) should be asked, as is the case already in several countries, along with questions about the respondents' residence 1 year before the census and at the time of the last one. The countries involved would then have better data at their disposal to study migration.

Acknowledgment

This chapter was translated from the French version by Mireille Rabenoro.

References

Ahlburg Dennis A., 1994. Return migration from the United States to American Samoa: evidence from the 1980 and 1990 censuses, *Pacific Studies*, vol. 17(2), p. 61–84.

Aziz Jean-Samy and Lavertu Jacques, 1993. Enquête 1991–1992 sur les migrations des personnes nées dans les départments d'outremer. Guide d'utilisation. Paris, INSEE, 34 p.

Aziz Jean-Samy and Lavertu Jacques, 1994. Les projets de migration DOM-Métropole, *INSEE première*, vol. 311, p. 1–4.

Courgeau Daniel, 1997. Dynamique migratoire, Analyse démographique des biographies, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Démographie: analyse et synthèse: Causes et conséquences des évolutions démographiques. Actes du séminaire de San Miniato*, vol. 2, p. 117–128. Rome, Dipartimento di Scienze Demografiche and Paris, INED, 234 p.

King Russell and Russell Sharon (eds.), 1986. *Return migrants and regional economic problems*. London, Croom Helm, 276 p.

Lidgard Jacqueline M. and Bedford Richard D., 1994. Returning New Zealanders: an ignored dimension of international migration, *in*: *Proceedings of the 16th New Zealand Geography Conference*, p. 427–452. Christchurch, New Zealand Geographical Society, 619 p.

Péron Yves, 1977. L'analyse par cohortes de l'immigration définitive, *Population*, vol. 32, special issue, p. 69–78.

Rallu Jean-Louis, Munoz-Perez Francisco and Jose Carrilho Maria, 2000. Return migration from Europe to Spain and Portugal, *Studi Emigrazione/Migration Studies*, 37(139), p. 625–649.

Rallu Jean-Louis, Bedford Richard D., Sodter François and Baudchon Gérard, 1997. *Population, migration et développement dans le Pacifique Sud*. Paris, UNESCO, ORSTOM and INED, 224 p.

Rallu Jean-Louis, 1996. Migration. Work and projects of return migrants in Rarotonga (Cook Islands), *New Zealand Population Review*, vol. 22(1–2), p. 45–68.

Tedebrand Lars-Goran, 1976. Remigration from America to Sweden, *in*: Harald Runblom and Hans Norman (eds.), *From Sweden to America*, p. 201–227. Minneapolis, University of Minnesota Press, 391 p.

Index

B

E

G

H

J

K

N

O

P

Demography: Analysis and Synthesis

Demography: Analysis and Synthesis

Graziella CASELLI, Jacques VALLIN, and Guillaume WUNSCH

with contributions by

Daniel COURGEAU, Nico KEILMAN, Eva LELIÈVRE, James VAUPEL,
Anatoli YASHIN, and John WILMOTH

VOLUME III

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data
Caselli, Graziella.
[Dimographie. English]
Demography : analysis and synthesis : a treatise in population studies / Graziella Caselli, Jacques Vallin, and Guillaume Wunsch ; with contributions by Daniel Courgeau . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-12-765660-X (set : alk. paper)—ISBN 0-12-765661-8 (volume 1 : alk. paper)—ISBN 0-12-765662-6 (volume 2 : alk. paper)—ISBN 0-12-765663-4 (volume 3 : alk. paper)—ISBN 0-12-765664-2 (volume 4 : alk. paper) 1. Demography. I. Vallin, Jacques. II. Wunsch, Guillaume J. III. Title.
HB871.C37513 2005
304.6—dc22 2005019413

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 13: 978-0-12-765660-1
ISBN 10: 0-12-765660-X

For all information on all Elsevier Academic Press publications visit our Web site at www.books.elsevier.com.

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Contents
Volume III

CHAPTER

90

Household and Family Demography: Application to Developing Countries 435

MARC PILON

CHAPTER

91

Households and Families: Developed Countries 457

NICO KEILMAN

CHAPTER

92

Social Mobility: Then and Now 477

FRANÇOIS HÉRAN

CHAPTER

93

Gender and Demography 495

ANTONELLA PINNELLI

CHAPTER

94

Economic Activity and Demographic Behaviors 513

GIUSEPPE GESANO

CHAPTER

95

Demography and Ethnicity: An Ambiguous Relationship 531

JEAN-LOUIS RALLU, VICTOR PICHÉ, AND PATRICK SIMON

CHAPTER

96

Genetics and Demography 551

ENZO LUCCHETTI AND ALBERTO SOLIANI

Contents of Other Volumes

Volume I Contents

Volume II Contents

Volume IV Contents

Contributors

Amoakon Anoh, Ècole nationale supérirure de statistique et d'économie appliqué (ENSEA), Abidjan, Côte d'Ivoire.

Philippe Antoine, Institut pur la recherché et le developpement (IRD), Dakar, Senegal.

Isabelle Attané, Institut national d'études démographiques (INED), Paris, France.

Alexandre Avdeev, Université de Moscou, Moscow, Russia, and Université Marc Bloch, Strasbourg, France.

Maria Avdeeva, Departement de bibliographie et de l'information, Centre d'etudes demographiques, Faculte d'economie, Université de Moscou, Moscow, Russia.

Brigitte Baccaïni, Institut national d'études démographiques (INED), Paris, France.

Stefano Baldi, Permanent Mission of Italy to the United Nations, United Nations Plaza, New York, United States.

Magali Barbieri, Institut national d'études démographiques (INED), Paris, France.

Odo Barsotti, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Françoise Bartiaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Gijs Beets, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Giovanni Berlinguer, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Elza Berquo, Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil.

Francesco C. Billari, Istituto di Metodi Quantitativi, Università Bocconi, and Innocenzo Gasparini Institute for Economic Research, Milan, Italy.

Jean-Noël Biraben, Institut national d'études démographiques (INED), Paris, France.

Anna-Maria Birindelli, Dipartimento di Statstica, Università degli Studi Milano-Bicocca, Milan, Italy.

Alberto Bonaguidi, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Corrado Bonifazi, Instituto di Ricerche sulla Popolazione e le Politiche Sociali, Rome, Italy.

Marco Bottai, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Michel Bozon, Institut national d'études démographiques (INED), Paris, France.

Vittoria Buratta, Istituto Nazionale di Statistica (ISTAT), Rome, Italy.

Raimondo Cagiano de Azevedo, Facoltà di Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Gérard Calot, Institut national d'études démographiques (INED), Paris, France.†

Emmanuelle Cambois, Institut national d'études démographiques (INED), Paris, France.

Graziella Caselli, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Jean-Claude Chasteland, Institut national d'études démographiques (INED), Paris, France.

Jean-Claude Chesnais, Institut national d'études démographiques (INED), Paris, France.

Luciano Ciucci, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

†Deceased

Robert Cliquet, Population and Social Policy Consultants, Brussels, Belgium.

David Coleman, Department of Social Policy and Social Work, University of Oxford, United Kingdom.

Maria-Eugenia Cosio-Zavala, Universite de Paris X-Nanterre, Nanterre Cedex, France.

Daniel Courgeau, Institut national d'études démographiques (INED), Paris, France.

Jean Coussy, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Gianpiero Dalla Zuanna, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy.

Patricia David, Harvard School of Public Health, Boston, United States.

Lorenzo Del Panta, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna, Italy.

Bart de Bruijn, Netherland Interdisciplinary Demographic Institute (NIDI), Den Haag (La Haye), Pays Bas, Netherlands.

Jean-Michel Decroly, Laboratoire de Géographie Humaine, Université Libre de Brussels, Brussels, Belgium.

Ainhoa de Federico de la Rúa, Institut Federatif de Recherche sur les Economies et les Societes Industrielles (IRESI), Universite des Sciences et Technologies de Lille 1, Lille, France.

Arna Dellis, University of Hawaii, Manoa, Hawaii, United States.

Paul Demeny, Population Council, New York, United States.

Alessandra De Rose, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Paolo De Sandre, Departimento de Scienze Statistiche, Università di Padova, Padova, Italy.

Gustavo De Santis, Facoltà di Scienze Politiche, Università di Messina, Messina, Italy.

Martine Deville, Institut national d'études démographiques (INED), Paris, France.

Manon Domingues Dos Santos, Centre de recherché en economie et statistique (CREST), Malakoff, France.

Josianne Duchêne, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Viviana Egidi, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Vincent Fauveau, United Nations Population Fund, New York, United States.

Alexis Ferrand, Institut de Sociologies et d'anthropologie, Universite des Sciences et Technologies de Lille, Villeneuve d'Ascq, France.

Patrick Festy, Institut national d'études démographiques (INED), Paris, France.

Judith A. Fortney, Family Health International, Research Triangle Park, North Carolina, United States.

Luisa Frova, Instituto Nationale di Statistica (ISTAT), Rome, Italy.

Alexia Fürnkranz-Prskawetz, Vienna Institute of Demography, Vienna, Austria.

Hubert Gérard, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Giuseppe Gesano, Instituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy.

Piero Giorgi, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Valérie Golaz, Institut national d'études démographiques (INED), Paris, France.

Antonio Golini, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Rosa Gomez-Redondo, Universidad Nacional de Educacion a Distancia (UNED), Facultad de CCPP y Sociologia, Madrid, Spain.

Catherine Gourbin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Carl Haub, Population Reference Bureau, Washington, D.C., United States.

François Héran, Institut national d'études démographiques (INED), Paris, France.

Kenneth Hill, Department of Population and Family Health Sciences, Johns Hopkins University, Maryland, United States.

Serguey Ivanov, Population Division, United Nations, New York, United States.

Carol Jagger, Department of Epidemiology and Public Health, University of Leicester, United Kingdom.

Heather Joshi, Centre for Longitudinal Studies, Institute of Education, University of London, London, United Kingdom.

Nico Keilman, Department of Economics, University of Oslo, Oslo, Norway.

Shigemi Kono, Faculty of International Economics, Reitaku University, Chiba-ken, Japan.

Marlène Lamy, Institut de demographie de l'Universite de Paris (IDUP), Paris, France.

Jacques Légaré, Department of Demography, University of Montreal, Montreal, Canada.

Luc Legoux, Institut de démographie de l'université de Paris (IDUP), Paris, France.

Eva Lelièvre, Institut national d'études démographiques (INED), Paris, France.

Henri Leridon, Institut national d'études démographiques (INED), Paris, France.

Ron Lesthaeghe, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Thérèse Locoh, Institut national d'études démographiques (INED), Paris, France.

Michel Loriaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Enzo Lucchetti, Dipartimento di Biologia Evolutiva, Università degli Studi di Parma, Parma, Italy.

Dionisia Maffioli, Università degli Studi di Bari, Bari, Italy.

Godelieve Masuy-Stroobant, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

France Meslé, Institut national d'études démographiques (INED), Paris, France.

Alain Monnier, Institut national d'études démographiques (INED), Paris, France.

Karel Neels, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Annunziata Nobile, Dipartimento di Istituzioni politiche e Scienze sociali, Università degli Studi Roma Tre, Rome, Italy.

Alberto Palloni, Center for Demography and Ecology, University of Wisconsin, United States.

Sophie Pennec, Institut national d'études démographiques (INED), Paris, France.

Pierre Pestieau, Université de Liège, Liège, Belgium and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Victor Piché, Inter-university Centre for Demographic Studies, University of Montreal, Montreal, Canada.

Marc Pilon, Institut de recherché pour le developpement (IRD), Piagadpigpi, Burkina Faso.

Gilles Pison, Institut national d'études démographiques (INED), Paris, France.

Antonella Pinnelli, Department of Demographic Science, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Denise Pumain, Institut national d'études démographiques (INED), Paris, France.

S. Irudaya Rajan, Centre for Development Studies (CDS), Kerala, India.

Jean-Louis Rallu, Institut national d'études démographiques (INED), Paris, France.

Rosella Rettaroli, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna Italy.

Jean-Marie Robine, Démographie et Santé, Institut national de la sante et de la recherché medicale (INSERM), Montpellier, France.

Paul-André Rosental, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Silvana Salvini, Dipartimento di Statistiche, Università degli Studi di Firenze, Florence, Italy.

Antonio Santini, Dipartimento di Statistiche, Univrsità degli Studi di Firenze, Florence, Italy.

Francis Sartor, Departement d'epidemiologie-toxicologie, Institut scientifique de la Sante Publique, Brussels, Belgium.†

Bruno Schoumaker, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Catherine Sermet, Institut national d'études démographiques (INED), Paris, France.

Jolande Siebenga, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Patrick Simon, Institut national d'études démographiques (INED), Paris, France.

Lamberto Soliani, Dipartimento de Scienze Ambientali, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Salvatore Strozza, Dipartimento di Scienze Statische, Università di Napoli Federico II, Napoli, Italy.

Pierre Surault, Groupe d'études démographiques, economiques et socials (GEDES), Universite de Poitiers, Poitiers, France.

Dominique Tabutin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michael Teitelbaum, Alfred P. Sloan Foundation, New York, United States.

†Deceased

Marc Termote, National Institut of Scientific Research, University of Quebec, and Department of Demography, University of Montreal, Canada.

Laurent Toulemon, Institut national d'études démographiques (INED), Paris, France.

Tapani Valkonen, Département de Sociologie, Université de Helsinki, Helsinki, Finland.

Jacques Vallin, Institut national d'études démographiques (INED), Paris, France.

Celine Vandermeersch, Institut national d'études démographiques (INED), Paris, France.

Jean-Pascal van Ypersele, Institut d'astronomie et de geophysique G. Lemaitre, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

James Vaupel, Max Planck Institute for Demographic Research, Rostock, Germany.

Jacques Véron, Institut national d'études démographiques (INED), Paris, France.

Éric Vilquin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Paolo Vineis, Dipartimento di Scienze Biomediche e Oncologia Umana, Università di Torino, Turin, Italy.

Anatoly Vishnevsky, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Tania Vishnievskaia, Institut national d'études démographiques (INED), Paris, France.

Carolyne Wanja Njue, Population Council, Nairobi I, Kenya.

Christine Wattelar, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michel Willems, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

John Wilmoth, Department of Demography, University of California, Berkeley, California.

Guillaume Wunsch, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Anatoli Yashin, Max Planck Institute for Demographic Research, Rostock, Germany.

Sergei Zakharov, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Hania Zlotnik, Population Division, United Nations, New York, United States.

SECTION

I

POPULATION IN TIME AND SPACE

History of Population and Population Forecasts: Introduction to Volume III

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Volume I, Section I looked at the central study of the discipline–the in-built mechanisms of population dynamics. Volumes I, Section II, and Volume II, respectively, addressed the determinants of the three big forces of those dynamics (fertility, mortality and migration). This gives us the essential foundations for a better understanding of the historical and geographical forces driving population, and for a consideration of future prospects. It is this complex of issues that are the subject of this section, leaving the remaining text to examine population and society, population thought and policies, and methods of observation and corresponding tools for demographic analysis, as originally planned.

Between the time when the first *hominidae* appeared on Earth some 2 or 3 million years ago–or on a smaller timeframe, in the hundred thousand years since the most recent human species, *Homo sapiens*, entered the frame–and the mid-18th century, the history of the population of this planet has been a very long sequence of phases of slow expansion interspersed with crises and declines. A history in which the balance of births and deaths has been decisive for the future of each group, but also a history in which, for the most dynamic and ultimately successful groups at least, migrations have been a major determinant of human expansion across the continents. In the 18th century, however, the driving forces of that history changed radically. The industrial revolution led to unprecedented population growth, while relentless forces made the world population a reality as civilizations interacted with one another, and demographic behaviors in particular grew more uniform. Population growth has thus become a supremely global issue and its future, a big political issue.

The four chapters of Part I will attempt to give a broad outline of the history of the human population. Part II will look at the geography of population: the distribution of humankind across the surface of the Earth and the diversity of current population characteristics, for behind the converging trends, each population and group has its own history and differing timelines that give rise to a very wide range of point-in-time situations. Part III concludes the volume with an overview of the methods, and especially the hypotheses that offer avenues for an exploration of the future of a specific population, and ultimately the world population. It readily becomes clear that although the short and medium term future may be easily visualized, if not predictable, the same cannot be said for the long-term future, where the range of possibilities is very much wider.

PART

I

THE HISTORY OF THE HUMAN POPULATION

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

The history of humankind which Jean-Noël Biraben describes as "short," unfurling over "the last 3 million years at most" is indeed an eye-blink compared with the time that distances us from the first appearance of life on Earth, let alone that which elapsed between the formation of the universe and the emergence of life, which runs into billions of years. Even so, very little is known about most of those 3 million years, until the dawn of history with the invention of writing, probably around 4000 BC. What little we do know of the history of the human population, however, is enough for specialists in prehistoric demography to venture approximations of the possible sizes of the human groups who began to settle the planet even before we–*Homo sapiens*–emerged. It is this long period of prehistory in particular to which Jean-Noël Biraben devotes much of the opening chapter of this section (Chapter 66), painting a fascinating panorama which encompasses the human population "from the first beginnings to the present day." But although this chapter also deals fairly comprehensively with the historical period, it can clearly only skim the surface of the past 2 or 3 centuries–a time of unprecedented developments on the population front, which will be dealt with in greater detail in the following chapters.

Before that, however, David Coleman gives over much of Chapter 67 to a consideration of prehistory and ancient or medieval history in order to clearly delineate "the major role played by migration in the history of the human population." The population pattern of human groups has in fact long depended on their ability to exploit technical advances in order both to achieve natural increases through a surplus of births over deaths and to cope with the resulting pressure either through further technical progress or migration.

From the mid-18th century onward, however, humans–who had hitherto sought to adapt to the natural conditions of their environment and striven to tame nature for their benefit (such as, in the Neolithic era, turning from hunting and gathering to agriculture and stock rearing) had nevertheless remained wholly subjugated to the harsh equation of balance between their natural increase and the increase in available resources–finally began to liberate themselves from this ancient law and bring both the time of their death and the force of their fertility under their own dominion. In the space of 2 centuries, the demographic system, which had governed humankind for millennia, changed completely in Europe. It was what Adolphe Landry was to call the "demographic revolution," the *transition* from an old to a modern demographic regime (the pretransitional to post-transitional stages), with far-reaching consequences for European countries themselves in the first instance and for the rest of the world. The "European Demographic Transition" is considered by Jacques Vallin in Chapter 68.

The same author concludes this section on the history of the human population in Chapter 69, focusing on the half-century since the European demographic transition. This was a period of contrasts that saw the "theory of demographic transition," devised by writers in the late 1940s to account for the origins and consequences of the system change, come to fruition with the *globalization* of the vast historical movement initiated in Europe and the rise of uncertainties–first for Europe and its North American and Pacific Rim offshoots and then for the longer-term future of the other continents–so little does the "completion of transition" resemble the emergence of a stable and permanent demographic regime in which uncontrolled growth finally gives way to permanent stability.

CHAPTER

66

The History of the Human Population from the First Beginnings to the Present Day

JEAN-NOËL BIRABEN
Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

This is the most recent in a series of exploratory estimates begun in 1979 (Biraben 1979). These tentative studies, as the name suggests, make no pretence to give the final picture of the demographic history of humankind. It, as did its predecessors, aims merely to review current thinking and move a step forward in the unending quest for more knowledge of our demographic past. It is also more comprehensive in considering humankind from the beginnings to the present day.

Significant advances both in the sciences of prehistory and in historical demography over the past 20 years make the case for these ongoing endeavours. This exploration, as did those before it, is purely tentative in simply taking stock of the state of early 21st century knowledge.

I. THE FIRST HUMAN POPULATIONS

The short history of humankind—the past 3 million years at most in the long life of the planet Earth—is both very unlike that of other living and extinct species and very alike in terms of the evolutionary pattern of all other species. It is hard to pinpoint when humankind first appeared and depends on the criteria used. What most distinguishes its lifestyle from that of the other great anthropoid apes may be less the fashioning and systematic use of tools (from wood, flint, bone, etc., which their anthropoid progenitors also used at times) than the quality of their manufacture, and even more important is the existence of a sufficiently permanent encampment where the gathered or hunted food was brought back for consumption by the group. Three million years would seem the very outside for the origination of such human beings in east Africa.

Little skeletal evidence remains of these *Homo habilis*, taking them as the earliest human beings, and is concentrated in regions where fossil remains could be recovered. The traces of their earliest industries, by contrast, have been better preserved and afford some evidence of the *oecumene*, or scale of distribution of these assemblages. It extended over nearly 4 million square kilometers of tree and bush savannah between Ethiopia, the Central African Republic, and Zimbabwe, i.e., approximately the same surface area as the mainly forest region between Guinea and Congo occupied by chimpanzees in around 1930 before the encroachment of agriculture. The number of chimpanzees around 1930 was estimated at approximately 200,000 individuals divided between five main species. But this population, unlike that of the other great anthropoid apes, was also thought to have increased significantly since the turn of the century, especially on the fringes of banana plantations where they lived commensally with humans by plundering the fruit.

When he addressed this issue in an article, Janos Némeskéri (1974) estimated these first humans as numbering approximately 100,000 individuals. Assuming the savannah-dwellers to be more carnivorous in their diet than were the chimpanzees of 1930, their number could be conservatively estimated as between 100,000 and 200,000 individuals, already divided into different types of hominid.

At that time, however, the Earth's prevailing quite uniformly hot climate (no polar ice-sheets) began to cool down and entered a major glaciation comprising at least six main glacial advances and retreats, each lasting tens of thousands of years, as well as countless smaller pulses of glacial advance alternating with periods of retreat of no longer than a few millennia each. There have only been four of these major glaciations in the 4.5 billion-year-old history of the Earth. Each has lasted only a few million or tens of millions of years, but each has been accompanied by an evolutionary spurt, with many species disappearing and many more appearing during each stage. One such evolutionary spurt, based on all the evidence, occurred in the human species, whose brain, in particular, expanded by a full kilogram in under 3 million years—one of the most rapid evolutions known for an organ that initially weighed substantially less than 500 g. It is hard to resist the conclusion that the capacities of this new organ gave its possessors such a selective advantage as to accelerate their evolution, leading to the elimination of all their nearest rivals. The result was the extinction at each stage of the preceding less sophisticated species of the genus such as *H. habilis* and *H. rudolfensis*, which were supplanted by *H. ergaster* in Africa, spreading across Southern Asia and Europe as its close relative *H. erectus*, and later in Europe as *H. neanderthalensis* and in Indonesia as *H. javanicus*.

Prehistoric demography sets out to study three aspects of these populations through different investigative methods of varying accuracy:

Population measurement
Mortality
Causes of death

II. THE PREHISTORIC POPULATION AND METHODS OF MEASUREMENT

The total prehistoric population has been extensively researched by many interesting methods, none of them, unfortunately, wholly satisfactory. Research has followed two lines: (1) that into the total population of large unit areas—regions, countries, and even continents; and (2) that into the size of a localized group such as an encampment, village, or small community.

1. Four Estimation Models

Essentially, four main methods have been used to estimate total populations. Briefly, they are as follows:

1. For the earliest periods (those of *H. habilis*), estimation of the population is determined by estimation of *biomass*—either the direct biomass of the primates deduced from their interaction with their environment, or biomass indirectly determined from the mass of available and accessible edible big animals and plants.

Although this method may have some value for other mammals, it offers no more than a tentative approximation for even primitive humankind, for which cultural changes, be it hunting techniques or social organization, significantly and rapidly affected the relationship between a population and the accessible biomass.

2. By measuring the *oecumene*, that is, the size of the area inhabited by humans of a particular culture; the calculated surface area is then multiplied by the density of historically recorded populations with ostensibly similar natural and cultural living conditions.

This method cannot be used for any of the very long Lower Paleolithic Age, for which there is no recorded evidence of a population living in comparatively similar conditions.

It also presupposes that populations live in equilibrium with their environment, which is only the case after a protracted period of adaptation, and even nowadays, countries with comparable climates and living standards, such as England and New Zealand, have very different population densities.

Finally, it also predicates long-term stability of the ecological environment, which has been approximately the case for the past 10,000 years, but such a protracted period of comparative climatic stability has been very uncommon since the appearance of human beings—that is, the start of the current glacial period marked by frequent, sudden climatic fluctuations every 2000 or 3000 years as a rule in the preceding 120,000 years—and probably also during earlier periods for which less evidence is available.

These shortcomings notwithstanding, this is by far the most popular method and can, if used with caution, give a valid crude approximation of the prehistoric population of a vast area for a particular period. However, in more recent terms, it is akin to back-calculating the population of the 16th century Iberian peninsula based on a rough idea of that of 19th century Turkey.

3. Denis Peyrony (1949) devised a completely different method for estimating the total population of a regional unit area, even for very remote periods of history. It involves a fairly forensic, uniform exploration of all the archaeological layers for each period in the area of study before a careful investigation of the geological conditions of layering and erosion in the Quaternary Period in that area to assess what proportion of the strata may have vanished through erosion, agriculture, or earthworks (given that this process follows an approximately logistic law) to reconstitute the number of layers for each culture in that area. Finally, it requires a fairly exact knowledge of the chronology for calculating the relative number of stratified deposits per millennium for each ancient culture (or per century for the most recent ones). This per-millennium stratigraphy, it is argued, then gives a relative scale of the population in each period.

Although pleasing in theory, this method does not pass muster in practice. An exhaustive exploration over a very large area is not possible or desirable, and estimating vanished layers is always an uncertain business. The few known cases in which it has been used are limited in scope, but the method could be substantially improved by systematic or random sampling exploration, which would enable the population to be more forensically tracked than by other methods, even for cultures with no known contemporary equivalent.

4. For more recent periods (the Mesolithic, Neolithic, and Chalcolithic eras), the assumption was that, because the population was divided into small groups, the total population could be estimated by counting the number of groups, that is, encampments for nomadic peoples or villages for sedentary communities, multiplied by the average estimated size of each group.

As in the previous method, this is limited by archaeological exploration and estimating average group numbers, which can be very difficult given that the smallest groups, which far outnumber all others, are most likely to be missed by exploration.

Its accuracy could now be considerably improved, however, by using lognormal distribution of groups, which allows the number and total population of groups to be fairly accurately approximated from just the most easily identifiable groups—the largest. Obviously, that involves using the greater range and variety of group size measurement methods, which, as will be seen, yield results equally variable in quality as did the previous methods.

2. Eight Methods of Measurement

Eight methods for measuring a local prehistoric population group are reviewed.

1. The idea of estimating the size of a group—say a clan—by measuring its hunting ground came to Dr. Louis Pradel (1954) while he was studying the area of distribution of flint from a very small jaspoid deposit used only over a relatively short period by a small group of Neanderthals of the Mousterian plateau culture a little over 70,000 years ago. This hunting and fishing ground, not far from the confluence of the Vienne and Creuse rivers at Fontmaure-en-Vellêche, is only a hundred square kilometers in size, and seven approximately similar assemblages along the Vienne and Clain rivers have been identified for the same period, spaced approximately 15 to 30 km apart. It can be deduced from this that each group could have consisted of between 10 and approximately 30 individuals, but it would certainly be wrong to extrapolate these results areawide across a region that is far from being as favorable to hunting and fishing as are these valleys, and it only offers an approximation based on an educated guess as to the number of groups or number of inhabitants.

By using the same method, Henri de Lumley (1976) demonstrated from the dispersal fan of more than 100,000 objects found around the Caune de l'Arago cave site that the hunting ground of this *H. erectus* clan of the Tayacian culture, which inhabited Roussillon a little more than 250,000 years ago in the Mindel-Riss interglacial stage, was approximately 500 km^2 in area for a group that, year in, year out, could also have consisted of between 10 and about 30 individuals.

2. Livio Livi (1941) conjectured that genetic survival required a minimum size for these basic population groups and that small islands holding only one group assumed to be fairly strictly endogamous and so forming a true isolate in the real sense, would give the size of these groups. In fact, this population of seafarers, which often roamed far and wide in quest of wives, was very much exogamous, and the lower limit of these island populations was predominantly dictated by the economy of the time.

3. For more recent epochs, the Mesolithic Period in particular, attempts have been made to calculate the size of certain population groups from the size of

kitchen-midden accumulations apparently belonging to the same group over a specific period (Boone 1976). The volume of refuse divided by a presumed average individual daily consumption would give the number of individuals in the group. Such calculations have been attempted on the shell mounds of the Californian and Portuguese coasts, the Azilian on the Basque and Cantabrian coasts, the kjöken-möding in Denmark, and the Tunisian snail mounds.

Unfortunately, these calculations are little more than "best guesstimates." Assessments of what constitutes age- and sex-specific average daily consumptions differ widely, whereas the actual duration of site occupation is still more arbitrarily estimated. It is just about possible to identify seasonally-returning groups but not groups returning only occasionally or even sporadically, nor is it easy to say whether the spot was a feeding center for the whole group or reserved for part of the community, such as women or children.

4. For even more recent periods—such as the Neolithic Period and Chalcolithic, Copper, Bronze, and Iron ages—attempts have been made to estimate village populations by the conjectured yield of surrounding cropped areas in each period. Sherburne Cook and Woodrow Borah (1963) used this method to estimate the pre-Colombian population of certain Maya cities.

But this method of calculation is very tentative for production and consumption, as it disregards the contributions of gathering, hunting, and fishing, not to say that of trade, by presuming villages to be completely self-sufficient, surely a very uncommon occurrence.

5. In recent sedentary culture periods, extensive research into the living environment, its sizes, capacity, dispositions, and layout (see Petrequin *et al.* 1969) has made it possible to conjecture approximate, but only likely, population sizes, which certainly vary with time and within a fairly broad range. Aggregated for each dwelling in a village, it gives an estimated order of magnitude of the village population. It is this method, first propounded by M. Jindrich Matiegka in 1911 (Matiegka 1927), that was the first used in prehistoric demography.

With the caveat that certain buildings within sites may be used not for dwelling but other purposes (places of worship, cattle sheds, barns, storehouses, etc.), which must be established, this method can yield fairly good results.

6. The study of some major collective structures has led to suggestions that the size of the societal group responsible for them can be inferred from the estimated number of contributors involved in creating them (Nougier 1959). The huge Locmariaquer menhir, which when whole weighed 368 tons, or the giant statues of Easter Island are examples.

However, experiments to validate this in 1979 showed that the number of contributors was much less important than were the lifting and rolling techniques used, and that, even with very crude techniques such as using logs and saplings for leverage, no more than 200 people were needed to erect a monumental 150-ton standing stone (Rouzé 1982). The same applies to the largest tumuli that 200 people can build in a few days, some of which, moreover, were added to by successive generations.

7. Mesolithic burials found grouped together in large burial grounds or cemeteries afford a much more accurate approximation of populations. Two Hungarian authors, Acsadi and Némeskéri (1970), devised and implemented a formula for the direct reconstruction of a village population from burial evidence.

Let t be the duration of the burial ground, D the number of reference skeletons, e_0 the mean estimated lifespan based on age at death from skeletal evidence, and K the estimated number of skeletons no longer extant. If, for example, $K = 10\%$, then

$$K = 0.1 \times D.\ e_0/t,$$

and the population

$$P = K + D.\ e_0/t$$

So, a cemetery containing 500 skeletons, in use for 200 years, would represent a village with an average population of 69 inhabitants.

Although the theory is scientific enough, especially compared with preceding ones, the method is nevertheless marred by uncertainties, especially as regards the completeness of the death totals: some who may have died in fighting or hunting accidents far from the village may not have been buried in the communal burial ground, whereas estimating the number of nonextant skeletons is particularly hazardous, especially those of very young infants, who were generally buried close to the surface and whose remains may have long been ploughed over. A statistical estimate of these can be ventured by plotting the age-specific cumulative distribution curve of deaths and comparing it with that of the closest model life table, but a degree of uncertainty remains (Biraben 1969).

Claude Masset (1982) devised a statistical method by which to reconstruct the total number of individuals buried in a rock tomb or ossuary where the skeletal remains have been secondarily moved and mixed.

This method, too, although satisfactory on the theoretical level, has serious drawbacks in practical application, which can only partly be overcome by the application of significant material means.

8. In the case of an incineration ground, finally, accurate dating of the ashes in each burial urn could give a distribution of deaths by broad blocks of years, that is, a trend pattern of deaths over the period of use of the entombment ground. The method was devised by Odette Taffanel (1976) to study the Mailhac burial ground, for which she estimated the 30-year death trend over 170 years, from 870 to 700 BCE. Assigning to this population maximum and minimum life expectancies enables the trend range of change in the population of that village to be calculated in 30-year cycles over the period.

This method, ostensibly quite precise, likewise unfortunately suffers from uncertainties, because there can be no certainty that all individuals in the village were incinerated, and particularly because of the unresolved practical issues surrounding accurate dating of incineration ashes.

After this rapid overview of methods of counting prehistoric populations, a very tentative outline account can now be ventured of the key results achieved or, put differently, the orders of magnitude within which prehistoric populations may possibly have fallen.

III. ESTIMATES OF PREHISTORIC POPULATIONS

When dealing with the world population, only the second method is usable, and that is the one that will be used to postulate some estimates, or rather, as stated above, what are still tentative orders of magnitude.

1. From the Earliest Hominids to *Homo Sapiens*

The only extant evidence from the very long Paleolithic Era about developments in hunting or gathering technologies of the time is what may be deduced from the flint or wooden artifacts found. It may therefore be assumed that they evolved correspondingly slowly, and that the number of human beings was for long much more closely linked to the size of the *oecumene* (area of distribution and historical movement) and to climatic variations, which govern the biomass (i.e., the annual quantity of plants or animals likely to form a direct or indirect part of the human food chain), than to any improvement in hunting or gathering techniques.

It is, however, worth noting that changes in the *oecumene* at these remote eras are also arguably connected with the physical changes in humankind, which may themselves be related to the sweeping climate changes of the main glacial advances. So, the appearance of *H. ergaster* at least quadrupled the *oecumene* of its predecessors, *H. rudolfensis* and *H. habilis*, stretching across northern and southern Africa, south and east Asia, the Middle East, and southern and central Europe, found in its form of *H. erectus*.

Technological leaps, although few and far between, also occur. The taming of fire, almost half a million years ago, enabled the cooking of some foodstuffs that were inedible raw, significantly reduced the parasites in the food that was cooked, and kept predators away from encampments. Much later, Levalloisian flaking of flint gave much improved tool quality and a better output of cutting edge per kilogram of raw flint. It also enabled diversification, so a greater variety of stone tools developed. On the basis of this, the total number of *H. erectus* on the planet can be conjectured at between 500,000 and 700,000.

During the penultimate ice advance, the Riss Glacial Stage in Europe, at both ends of the *oecumene*, i.e., in Indonesia and Europe, different physical types gradually began to emerge that, long kept apart by the rising sea levels during the following much warmer Eemian Interglacial, caused three distinct hominid population groups to develop at the same time—Java man in Indonesia, Neanderthal man in Europe, and *H. sapiens*, who, probably making his first appearance in Ethiopia and the surrounding areas, slowly spread first toward southern Africa, then to the Arabian peninsula, Levant, Iran, India, Indo-China, and China, where he supplanted *H. erectus*.

Slow progress in hunting and gathering technologies, improving the productivity of subsistence activities (the development of the Mousterian culture in Europe), suggests the possibility of a fresh population spurt. A midrange assumption for the number of humans at this period of optimum climatic conditions may be 800,000 for *H. sapiens* in the Afro-Asian area, 250,000 for Neanderthal man in Europe, and 100,000 for Java man in Indonesia.

The last ice advance, known as the Würm glacial stage in Europe, beginning toward 72,000 BCE, changed the world order. The ice-sheet across the northern half of Europe and falling sea levels, which restored communications between Europe and Asia and between Indonesia and Asia, would enable vast population migrations to occur.

Neanderthal man first abandoned the North European Plain, and the harsh climate, combined with the very significant decline in plant and wildlife, would lead to a drastic fall in their numbers to perhaps 50,000, when analogized from the established numbers of Eskimo or Chukchi populations living in the same climatic conditions in the 18th and 19th centuries but with a markedly more developed cultures. Then, from 70,000 BCE onward, they very rapidly spread east into as-yet-unpopulated Turkestan and southeast through Asia Minor, made accessible by the falling sea level, as far as to Palestine, where they lived alongside *H. sapiens*. At the same time, in southeast Asia, *H. sapiens* entered Indonesia, then part of the Indo-Chinese peninsula, squeezing out Java man within some 10 thousand years.

In a second stage, from 50,000 BCE perhaps, *H. sapiens* crossed the Timor Strait from Indonesia into then-uninhabited Australia and New Guinea (reunited by the falling sea level).

In a third stage, from about 42,000 BCE, the Middle Eastern branch of *H. sapiens* in Asia Minor gradually spread across Europe via the Caucasus and the Balkans, where a new culture—Proto-Aurignacian—developed. Despite the attempts of the Mousterian culture (i.e., Neanderthal man) to emulate it, this was to give *H. sapiens* a decisive superiority. Around 30,000 BCE, they spread through the Danube valley, Po River plain, and Provence to occupy all of Central Europe, Italy, most of France, and the Cantabrian mountains area of Spain. The last traces of the Mousterian culture offer presumptive evidence that the last Neanderthal men died out in northwestern France and in England around 29,000 BCE and in Andalusia about 28,000 BCE.

It was also probably somewhat before 40,000 BCE that humans crossed the Bering Strait and spread across North American and then South America. These vast lands, which, similar to Australia, were not everywhere habitable, nevertheless very soon added at least 500,000 individuals to humankind.

Twice—first around 23,000 BCE, then around 18,000 BCE during the last two lowest sea levels, *H. sapiens* migrated from Sicily (then attached to Italy, as were Malta and Pantellaria) to Tunisia via the less than 20-km-wide strait that separated them, rapidly ousting the very scattered people of the Aterian culture descended from *H. erectus*, who were still living there and relatively cut off by the desert. Their rapid advance westward would take them to the Canary Islands; their eastward move, into Egypt and Arabia. In the Maghreb, the population may have risen from 10,000 in the Aterian period to 50,000 for the Oranian culture and 70,000 for the Capsian culture.

In the same era, others left the Ukrainian and Russian plains, spreading eastward to become the Palaeosiberian peoples, and on into the Yenisei and Lena basins, Kamtchatka, Chukotka, and Japan (Ainu). They were very few in number at first, and even several millennia afterward, these vast spaces were probably inhabited by fewer than 50,000 people.

Later, around 12,000 BCE, a population drift from eastern Asia would again cross the Bering Strait, swamping the indigenous populations and driving them to seek refuge in a handful of sites, this becoming the majority population of America south of the Rio Grande River. The Eskimos would comprise the third population migration from Asia, where a very few can still be found. But another final, later-occurring and more miscegenous Palaeosiberian population movement was to invade the great plains and Rocky Mountains of North America, driving the occupants south of the Rio Grande River.

The vast forested tracts of sub-Saharan Africa and the Congo basin excluding the Sahel, which was relatively populous between 9000 and 4000 BCE (conjecturally, between 200,000 and 400,000 people), remained very sparsely populated. The earliest population of Pygmy inhabitants living in scattered groups was joined perhaps much later (perhaps 15,000 or 10,000 BCE) by a black population arriving in several waves, the first of which now remain only in a handful of refuge-habitats (similar to the Dogons in their rocky hills), and the last of which seem to be the Fulani (or Peul) people.

It is worth noting in this broad-brush picture that even before the development of navigational technology, humans seem to have crossed stretches of ocean that were not too wide—70 or 80 km at most—and that a systematic investigation into the periods at which channels such as the Bering Strait emerged add little to population research. Crossing conditions, by contrast, are more decisive than is width, and the strong current of the Strait of Gibraltar explains the late and still infrequent crossings of the Early Neolithic era.

2. The Prehistoric Population of Europe

In the work *Histoire de la population européenne* (Bardet and Dupâquier 1998), the authors described the method used to estimate the population of the European continent—the best explored from an archaeological standpoint—at between 140,000 and 170,000 inhabitants in the earliest period of habitation, that of *H. sapiens*, between 32,000 and 16,000 BCE, rising perhaps to nearly 200,000 at the height of the Magdalenian and Epigravettien cultures between 10,000 and 8700 BCE. It should also be noted that nomadism did not by any means produce a uniform pattern of human distribution: they wandered in small groups but almost always in the same regions, which were presumably most suited to their diet and activities and where almost all of the archaeological sites relating to them are now found.

The lognormal distribution of these groups suggests that both small (2 to 8 people) and large encampments (50 to 250 people) are uncommon, with most falling between 10 and 50 people. With some slight adjustment, this distribution is found among all populations in all parts of the world. Such a structure, even allowing for traditional exogamy (i.e., where marriages are prescribed with a nearby group), favors genetic drift. Where, by chance or design, a group is isolated from the outset, as it is always too small to be representative of the general population from which it originates, there is a significant founder effect, and within the space of a dozen generations, the genetic, cultural, and linguistic drift is so significant that, coming back into contact with the general population, it appears alien. This is the cause of differentiation and diversification in human types and ethnic groups, some of which are very old and many others comparatively recent, at least in prehistorical terms.

The sudden climatic warming which occurred in Europe between 8700 and 8650 BCE (when the average temperature of western and central Europe is estimated to have risen by 13°C in 50 years!) cut short the then pan-European Magdalenian and Epigravettien cultures. The southern steppes, hitherto teeming with herds of large animals, became overgrown by dense forest inhabited now only by packs of smaller animals and the odd solitary animal (wild boar, deer). Adaptation to the new plant and animal resources was initially reflected in a significant population decrease. By contrast, the northern tundra became forested, but less densely, interspersed with vast clearings where large herds of aurochs and bison continued to roam, and in Scandinavia, the reindeer population increased, as did the human population, under the effect of the climatic amelioration. Broadly, it may be conjectured that after a slight drop to perhaps 180,000 inhabitants around 8000 BCE in response to the sudden climate change, the population learned to adapt, began to rise again, and doubled around 7000 BCE to something approaching 400,000 inhabitants.

At this time, people from Anatolia established a first Neolithic village settlement in Greece, near Thessaloniki, growing hand-hoe—cropped cereals and pulses; rearing dogs, goats, sheep, oxen, and pigs; and living in stone-built, timber, and cob houses. This proto-Sesklo culture would rapidly evolve into the Sesklo civilization around 6800 BCE, making pottery and living in brick-built houses. Two courses developed from this.

In the south, along the Mediterranean coast, a seaborne course—the Cardial culture—characterized by impressed ware decorated by patterns made with a cockleshell, usually *Cardium edule*. Although not a colonizing people, they were culturally expansionist and moved from the seaboards inland to invest the Mesolithic (Epipalaeolithic) populations, each of which evolved in their own way from seminomadic hunter-gatherers into sedentary pastoralist and farming peoples. The civilization reached the Strait of Gibraltar before 6000 BCE and, after a pause, moved on to the Atlantic seaboard around 5000 BCE, where from 4 200 BCE the civilization ended up on the southern coasts of England, separated even then from the continent for 2500 years by the rising sea level (the civilization also, with difficulty, sporadically crossed the Strait of Gibraltar and established a few scattered settlements strung out along the coast of the Er-Rif mountain range and the Oranie region).

Somewhat later, around 6000 BCE, a course thrust northward across land, settling in the central Balkans, a civilization characterized by fine-painted ceramic ware, and diversifying into a wide variety of local cultures. About 5500 BCE, in Hungary, the Starcevo culture underwent a surprising population surge and moved up the Danube to occupy all the European hinterland from the Paris basin to Poland and Moldavia, driving back or absorbing the nomadic Tardenoisian hunters and then meeting and merging with the seaborne course around 4000 BCE in central France or in the Alps.

When another Neolithic course, which around 5000 BCE spread along the east coast of the Black Sea, the Kuban river, and the Ukraine, is added, then by around 4000 BCE almost the whole of Europe had moved into the Neolithic period with a population that may have stood at 2 million.

3. The Prehistoric Population of the Middle East and the Rest of the World

Among extra-European populations, the only one known with comparative certainty is that of the Middle East, where the Neolithic stage of development was attained around 10,000 BCE in the Fertile Crescent with cereal farming (mainly corn) and animal domestication—goats, sheep, and oxen. It was consolidated around 8000 BCE, with the appearance of pottery, and rapidly spread to Egypt, Anatolia, the Iranian plateau, and thence to the Caucasus, the plains of Punjab, Nubia, Ethiopia, and Yemen. The population for the entire Palestine, Lebanon, Syria, and Iraq region may have increased from 200,000 around 10,000 BCE to 500,000 in 8000 BCE and 1.5 million by 4000 BCE.

India was to be far and away the most populous region, but its population size is unfortunately very difficult to determine, because it seems to have been ignored by prehistorians and studied only by historians who have deduced utterly impossible totals for these ancient times from the purple passages of legendary tales. So, Pran Nath (1929) calculated the size of the armies engaged in the Mahabharata war—placed by Chandra (cited in Pran Nath, 1929) as at around 3100 BCE—at 1.97 million soldiers, which infers a total population size of between 29.6 and 39.4 million inhabitants! This number seems overstated, given that only the northwest of the country (present-day Pakistan) had attained the Neolithic culture, and the entire east and south of the country remained at the Mesolithic stages of development. If the population of India in these periods absolutely had to be conjectured, then reasonable orders of magnitude might be from 600,000 around 4000 BCE, 2 or 3 million around 3000 BCE, and an absolute maximum of 20 or 25 million around 2000 BCE, when virtually all the peninsula—India, Pakistan, Bangladesh, and Sri Lanka—had attained the Neolithic culture.

Around 8000 BCE, a Neolithic culture also developed in the plains of the Huang Ho River basin, rapidly developing to produce the glorious Chinese civilization. Originally confined to an area of just 600,000 km^2, it spread rapidly, first eastward to the China Sea and then southward, where corn gave way to rice cultivation. The population of China was of the same order of magnitude as that of India, perhaps a little faster-growing, with 800,000 inhabitants around 4000 BCE, 3 or 4 million around 3000 BCE, and an absolute maximum of between 15 and 20 million around 2000 BCE.

Somewhat later, another Neolithic culture in the Tehuacan valley in Mexico began to grow maize, tomatoes, and beans and to domesticate dogs and turkeys; yet another culture on the high plateaus of Peru grew potatoes and domesticated llamas and alpacas.

Elsewhere, in Neolithic cultures at various stages of completion, pottery appeared from 10,000 BCE in Japan and from 8000 BCE in a huge swathe of the African Sahel, followed by the Middle East, India, China, and Europe, and finally Mexico, Ecuador, and Peru.

So, by around 4000 BCE, not only in the Middle East and Europe but also in the African Sahel, Sudan, Ethiopia and Yemen, northwest India, China, Japan, and the high plateaus of the two Americas, the populations had evolved from hunter-gatherers to agro-pastoralists. With these eight Neolithic population centers, the world population had risen from just 6 or 7 million at 2000 years before to nearing 30 million and, 2000 years after that, when the hand-hoe was supplanted by the swing-plough, which vastly increased yields, it would be close to 100 million.

In seeking to chart the long rise of the human population, authors have long stopped short at estimating populations at a given date in Prehistory, Antiquity, the Middle Ages, and, more often still, the Renaissance or modern times and the contemporary era. This point-in-time approach has tended to give the impression of more-or-less uninterrupted exponential growth. Of the many published works on this subject, those of John Durand (1974) seem to have the most cogently argued case.

Prehistory, however, more suggests sharp growth spurts interspersed with long periods of stagnation or even decline. Does that paradigm still hold good for historical periods? The only way we have of knowing is to calculate the world population at a sufficiently large number of dates close enough together to plot the trends. Arguably, this idea is lent credence by the long, wide fluctuations in the long population curves of regions as populous as China or Europe, which according to the period each account for a fifth or sixth of humankind.

Until recently, humankind had remained segregated in large, comparatively hermetic cultural spaces, which Pierre Chaunu (1974) has dubbed "enclaved universes." Conceivably, the population within the same cultural space presents overall changes that reflect boom-bust cycles, but there is no circumstantial evidence from which to conclude that the populations of discrete or even far-apart cultural spaces evolve at the same pace, except by pure chance.

Similar to any plant or animal population, humankind is concentrated at points of the *oecumene* that are most suited to its subsistence and activities. So, it is clear that from Antiquity, China, India, and the European and Mediterranean countries accounted for between two-thirds and three-quarters of humankind between the three of them; in other words, the evolution of people in these areas determined that of all human populations.

IV. FROM ANTIQUITY TO MODERN TIMES

The table compiled (Table 66–1) shows that, significant divergences aside, there has been a correspondence of some favorable and unfavorable long periods in all three regions since Antiquity (although obviously not just between these three regions, since it may obtain generally almost worldwide). To show this

TABLE 66–1 World Population by Broad Regions at Different Dates

Region	–400	–200	JC	200	400	500	600	700	800	900	1000	1100	1200
China	19	40	70	60	25	32	49	44	56	48	56	83	124
India, Pakistan, Bangladesh	30	55	46	45	32	33	37	50	43	38	40	48	69
Southwest Asia	42	52	47	46	45	41	32	25	29	33	33	28	27
Japan	0.1	0.2	0.3	0.5	1.5	2	4	5	6	7	7	7	6
Rest of Asia (excluding the Soviet Union)	3	4	5	5	7	8	11	12	14	16	19	24	31
Europe excluding the Soviet Union	19	25	31	44	36	30	22	22	25	28	30	35	49
Soviet Union*	13	14	12	13	12	11	11	10	10	11	13	15	17
North Africa	10	13	13	16	13	12	11	9	10	10	10	8	8
Rest of Africa	7	9	12	14	18	20	17	15	16	20	30	30	40
North America	1	2	2	2	2	2	2	2	2	2	2	2	3
Central and South America	7	8	10	9	11	13	14	15	15	13	16	19	23
Oceania	1	1	1	1	1	1	1	1	1	1	1	2	2
Total world	152	223	250	255.0	204	205	211	210	227	227	257	301	399

Region	1250	1300	1340	1400	1500	1600	1700	1750	1800	1850	1900	1950	1970	2000
China	112	83	70	70	84	110	150	220	330	435	415	558	774	1273
India, Pakistan, Bangladesh	83	100	107	74	95	145	175	165	190	216	290	431	667	1320
Southwest Asia	22	21	22	19	23	30	30	28	28	31	38	75	118	259
Japan	6	7	7	8	8	12	28	30	30	31	44	83	104	126
Rest of Asia (excluding the Soviet Union)	31	29	29	29	33	42	53	61	68	78	115	245	386	653
Europe excluding the Soviet Union	57	70	74	52	67	89	95	111	146	209	295	395	462	492
Soviet Union*	14	16	16	13	17	22	30	35	49	79	127	180	243	290
North Africa	8	9	9	8	8	10	9	10	9	13	23	44	70	143
Rest of Africa	49	60	71	60	78	104	97	94	92	90	95	167	266	657
North America	3	3	3	3	3	3	2	3	5	25	90	166	228	307
Central and South America	26	29	29	36	39	10	10	15	19	34	75	164	283	512
Oceania	2	2	2	2	3	3	3	3	2	2	6	13	19	30
Total world	413	429	439	374	458	580	682	775	968	1243	1613	2521	3620	6062

*Community of Independent States (CIS) + Baltic states.

clearly, however, the points have been increased by aggregating the population of all countries, first in blocks of 2 centuries at a time between 400 BCE and 400 CE, then in single-century blocks up to the 17th century, and in half-century blocks from 1700 to the present day; finally crucial dates that were watershed moments for humankind's population development are indicated: in 1250 (just after the Mongol invasions) and in 1340 (just before the Black Death).

None of these numbers lay any claim to accuracy: even for the year 2000, as for 1950 and 1900, international yearbooks (chiefly those of the International Statistical Institute, League of Nations, and United Nations) are uncertain and may vary between editions; they are probably accurate for total world populations to within 5%. For periods further back, estimates have to be even more approximate, say 7% or 8% around 1700 and 10% around 1500. Further back than this, the scant, almost always indirect, statistics available enable one to work only within very uncertain ranges. The key thing is that the direction of variations and magnitude of fluctuations tend to more than compensate for the uncertainties.

1. Asia

The data for China are taken from Michel Cartier (Cartier 1973; Cartier and Will 1971), adjusted for current borders. The vicissitudes of China's history will not be rehearsed again in detail here, apart from one arguably key episode in which Chinese censuses report a loss of nearly one-third of the population in 49 CE. From this date until 73 CE, China called a halt to the reconquest of its territory, embroiled in battle as it was with the Yueh-chih from the west. The probable key to this puzzle is offered in the writings of a Chinese doctor of the time, Ko-Hong, in the form of the first description of smallpox, which he reports to be a new disease brought by the Yueh-chih horsemen in 49 (Ming Wong 1974). The same disease was similarly to ravage the West, i.e., the Roman Empire in 168 (where it was known as the Antonine plague) (Bakfalouni 1983, 1984) and the Aztec Empire in 1521 (Biraben 1992).

The data for India are taken from Ajit Das Gupta (1972), except for the pre-16th century era, where he merely argues that the population of India could have reached 100 million at the zenith of the Maurya empire (between 321 and 185 BCE), during the Gupta empire (320 to 470 CE), and under the reign of Harsha (612 to 627 CE). This very round figure of 100 million obviously requires qualification. It must be remembered that in all these periods, vast tracts of the interior were still tribal lands, governed by a very antiquated economic system (some still a Mesolithic culture almost until modern times), certainly with very low population densities. It was therefore thought preferable to estimate populations of 50 to 60 million for these periods, which are still within the bracket accepted by John D. Durand (1974), who argues for 75 million, but they could have been put 10 or 20 million higher without affecting the results. A later detailed study by Shireen Moosvi (1984) using two different methods estimates the population of India at approximately 145 million in 1595, tallying with Ajit Das Gupta's estimate taken for 1600 in preference to the 100 million conjectured by Moreland with little supporting evidence.

Southwest Asia includes the Arabian peninsula, Jordan, Palestine, Israel, Syria, Iraq, Turkey, Iran, and Afghanistan (Bachi 1974). Studies on the population in Antiquity exist for several of these countries; all of the countries are covered for the 19th and 20th centuries (Benham and Amani 1974). However, there are few (Israel, Syria, Turkey, Iraq) estimates for the Middle Ages (Bakfalouni 1983) or the Renaissance period (Barkan 1958, 1969; Erder 1979; Panzac 1981; Karpat 1985; Lacquerer 1987). The trend has therefore been plotted by reference to the flourishing and turbulent periods of their history, which, although of doubtful validity, broadly lines up with trends in which ignorance less outweighs knowledge. Additionally, it was only in Antiquity that the population of the Middle East influenced that of the world to any significant extent.

The figures for Japan follow the revised estimates of Akira Hayami (1987), reported in an article in *Population* (Biraben 1993).

Pre-19th century data for the rest of Asia is scant, except for Korea, Vietnam (Cartier 1973), Kampuchea (Meng-Try 1980), and Thailand in 1700 (Vallin 1976). Although several other countries in the region—Sri-Lanka, Burma, Champa, and Indonesia—reached apogees of civilization in the Middle Ages, their demography remains almost an unknown quantity and all that can be offered is pure conjecture. But the small population sizes of these countries made little contribution to the world total.

2. Europe

The data for Europe are drawn partly from volume 1 of the *Histoire des populations de l'Europe* (Bardet and Dupâquier 1998; although this arguably relies too heavily on McEvedy 1978, which is not the best authority) and partly from unpublished personal records. Broadly, rapid population growth in Antiquity was interrupted by smallpox (Antonine plague in 168) and then measles epidemics, giving way to a population decline. This decline gathered pace after the appearance of St. Cyprian's plague (a severe epidemic of an unknown type) and deepened with the successive waves of Germanic invasions in 251, 276, 375, and especially 407–412, when the economy was crippled by a general decline in trade. Although the late 5th and early 6th centuries seem to show an upturn, the Justinian plague of 542–543 and the reappearance of smallpox 10 years later completed the population loss. This was followed by stagnation and a very slow return to growth owing to the insecurity created by the Slavic, Avar, and Bulgarian invasions in the east and the Saracen, Norman, and Hungarian raids in the west. With these tribulations at an end, a strong population surge occurred in the 10th century, continuing through until the early 14th century, with only a slight blip caused by the Mongol expansion in 1242–1243. The Black Death, which appeared in 1347–1352 and turned into a new epidemic, caused a population loss of at least 20% in the space of a century. With the plague vanquished, another population surge occurred between 1450 and 1550. The late 16th, mid-17th, and early 18th centuries were also periods of stagnation or loss owing to wars, religious strife, and climatic deterioration, but most striking of all is the spiraling population growth from the mid-18th to early 20th century, which abated under the sharp check delivered by two world wars and slowed even further by fertility decline accentuated by the economic recession from 1973. In 2000, almost all of Europe was in a phase of population stagnation or even decline, despite continuing life expectancy gains, except in Eastern Europe, where life expectancy has even shortened slightly.

For the countries of the former Soviet Union, there is sparse evidence from the 13th century and relatively copious data from the 17th century to the modern day.

3. Africa

North Africa from Egypt in the east to the Maghreb in the west experienced Middle East—like periods of great prosperity in Antiquity, but the large body of research has produced no overall data or data too discrete to carry conviction owing to the continuing great uncertainties even about the best-researched periods. This chapter produces a tentative synthesis, which arguably best summarizes the achievements of the past 20 years.

The rest of Africa poses major problems, because only point data exist for most countries before the 20th century. An attempt has been made to estimate the population based on what is known of its economic and political history, outlined in the author's first publication in 1979.

4. America and Oceania

Although there are no statistical data for America predating the 16th century, archaeological research is beginning to yield a large body of much older local evidence. Also, close examination of the Spanish census documents from Peru and Bolivia showed that the population recorded by the quipu-camayoc in the Inca period and kept on the quipus (accounting apparatus) had often been recorded by the Spanish agents. The population decline seems not to have been as sharp as has been claimed on the basis of work skewed by a faulty methodology. I concur with John Durand (1974) that the population is likely to be somewhere between the two extremes argued for, at around 40-odd million in 1500.

It should also be noted that, after a slight retreat, the population grew rapidly from the mid-14th century, in stark contrast to the population of the Old World, on which the Black Death had inflicted serious losses at the same time.

For Oceania—excluding Australia, which has been populated since prehistory—population growth has only been hypothesized from the 10th century onward, when the Polynesians began to occupy the Pacific islands and, much later, New Zealand.

5. World Total

Turning now to the world population as a whole, it is clear that, notwithstanding the sometimes crude approximations used, the human population has not experienced the uninterrupted exponential growth that some studies affirm as an accepted truth. The more probable explanation is successive phases of growth or decline varying in length and rate. Before the recent rapid growth phase from the 16th to 20th century, it had been in decline for a good 100 years after the sudden emergence of the plague in the mid-14th century and a series of economic tribulations and wars. This peak around 1340 had followed a trough in the 5th and 7th centuries, itself preceded by a peak in the 2th century. There is a suspected trough around the 9th century BCE, preceded by a peak in the 11th century and so on.

The falls are approximately 10% to 20% of the preceding peak and, apart from the last which was only a century in length, may last up to three, four, or five centuries. The peak/trough cycle is very irregular, varying between 8 and 12 centuries, so an extrapolation, even based on a multicentury trend, would give no pointer as to the future world population. Calculated on the 4th and 3rd centuries BCE, it would have given over 20 billion inhabitants in 2000, but projected from the 3rd and 4th centuries CE would have given a 35 million decrease!

Variations of this magnitude seem to have occurred in all periods of history and prehistory, but, as was seen, some have been of such extraordinary magnitude as to suggest sudden, sharp, stepwise population growth. These step increases were first natural, through the evolution from one species of humankind into another, from *H. habilis* to *H. erectus* and then *H. sapiens*, but also occurred to an increasingly cultural extent (e.g., the taming of fire), becoming almost entirely cultural, that is, much more rapid and closer together than in all other species. Upper Palaeolithic technologies—clothing, bow and arrows, spear throwers, etc.—are one sharp step-change; the Neolithic era, with farming and sedentarism, another.

We are living through the industrialization step, the hesitant, localized beginnings of which are to be found in the Renaissance (see Chapters 68, 69) and will probably come to an end in the 21st century, as will be seen in Chapter 77.

Figure 66–1 summarizes the long history of humankind's population of the Earth.

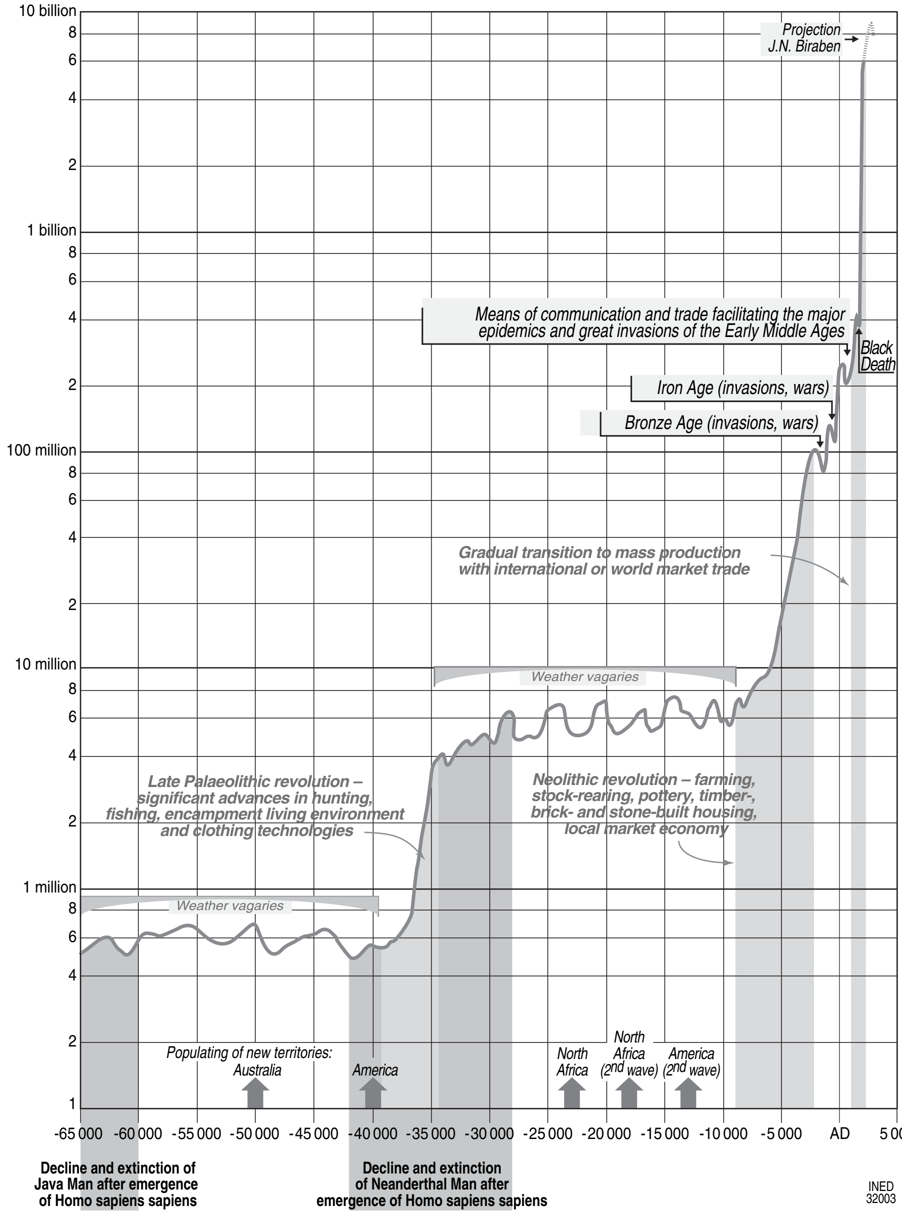


FIGURE 66–1 The rising numbers of humankind since prehistory.

References

Acsádi George Y. and Nemeskéri Janos, 1970. *History of human life span and mortality.* Budapest, Akadémiai Kiadó, 346 p.

Bachi Roberto, 1974. *The population of Israël.* Jerusalem, CICRED, 428 p.

Bakfalouni Tarek, 1983. *Quelques observations sur l'idée densité-population. Le cas des villes syriennes au Moyen-Âge et à l'époque ottomane.* Paris, EHESS, 31 p. [Unpublished manuscript].

Bakfalouni Tarek, 1984. *Les effets démographiques des épidémies, des guerres et des disettes au Proche-Orient du VIIe au XIXe siècles.* Paris, EHESS, 762 p. [Doctoral thesis].

Bardet Jean-Pierre and Dupâquier Jacques (eds.), 1997–1999. *Histoire des populations de l'Europe.* Paris, Libraire Arthème Fayard, 2099 p.

Barkan Ömer Lûtfi, 1958. Essai sur les données statistiques des registres de recensement dans l'Empire ottoman aux XVe et XVIe siècles, *Journal of the Economic and Social History of the Orient*, no. 1, p. 7–36.

Barkan Ömer Lûtfi, 1969. Contribution à l'étude de la conjoncture des pays méditerranéens au cours du XVIe siècle, *in:* UIESP (ed.), *International Population Conference, London 1969/Congrès international de la population, Londres, 1969*, vol. 4, p. 3309–3314. Liège, IUSSP, 3050 p.

Beloch Karl-Julius, 1908. *La popolazione dell'Europa nell'Antichità, nel Medio evo, e nel Rinascimento.* Turin, Biblioteca dell'Economista, 941 p.

Benham Djamchid and Amani Mehdi, 1974 *La population de l'Iran..* Paris CICRED, 79 p.

Biraben Jean-Noël, 1969 Durée de vie de la population de Columnata (épipaléolithique oranais), *Population*, vol. 24(3), p. 487–500.

Biraben Jean-Noël, 1979. Essai sur l'évolution du nombre des hommes, *Population*, vol. 34(1), p. 13–26.

Biraben Jean-Noël, 1992. *La population de l'Amérique pré-colombienne. Essai sur les méthodes d'études*.- Paris, INED, 30 p. [Paper to the Conference on the Peopling of the Americas, Vera Cruz, Mexico, 1992, IUSSP].

Biraben Jean-Noël, 1993. Le point sur l'histoire de la population du Japon, *Population*, vol. 48(2), p. 443–472.

Boone Yvonne, 1976. Le ramassage des coquilles, *in:* CNRS (ed.), *La préhistoire française*, p. 703–707. Paris, Centre national de la recherche scientifique, 853 p.

Cartier Michel, 1973. Nouvelles données sur la démographie chinoise à l'époque des Ming (1368–1644), *Annales ESC*, no. 6, p. 1341–1359.

Cartier Michel and Will Pierre-Etienne, 1971. Démographie et institutions en Chine: contribution à l'analyse des recensements de l'époque impériale (2 ap. J.-C.-1750), *Annales de démographie historique*, p. 161–245.

Chaunu Pierre, 1974. *Histoire, science sociale: la durée, l'espace et l'homme à l'époque moderne.* Paris, SEDES, 437 p.

Cook Sherburne F. and Borah Woodrow W., 1963. *The aboriginal population of central Mexico on the eve of the Spanish conquest.* Berkeley, University of California Press, 159 p.

Corradi Alfonso, 1865–1894. *Annali delle epidemie occorso in Italia dalla prime memorie fino al 1860*. Bologna, Ristampa Anastatica, 5 vol., 3960 p.

Das Gupta Ajït, 1972. Study of the historical demography of India, *in:* David V. Glass and Roger Revelle (ed.), *Population and social change*, p. 419–435. London, Edward Arnold, 520 p.

Durand John D., 1974. *Historical estimates of world population: an evaluation.* Philadelphia, Population Studies Center, University of Pennsylvania, 72 p.

Erder Leila, 1979. Population rise and fall in Anatolia 1550–1620, *Middle Eastern Studies*, vol. 15, p. 322–345.

Hayami Akira (ed.), 1987. *Population trends in Tokugawa, Japan: 1600–1868.* Voorburg, IIS, 17 p. [Forty-sixth International Statistical Institute session, Tokyo, 1993].

Karpat Kemal H., 1985. *Ottoman population 1830–1914. Demographic and social characteristics.* Madison, The University of Wisconsin Press, 243 p.

Laqueur Hans-Peter, 1987. Osmanische Grabsteine als demographische Quelle, *in:* Jean-Louis Bacqué-Grammont (ed.), *Les provinces arabes à l'époque ottomane*, vol. 2, Paris, Istanbul, Leyden.

Livi Livio, 1941. *Trattato di demografia. I fattori bio-demografici nell'ordinamento sociale*, Padova, CEDAM, Casa Editrice dott. Antonio Milani, 268 p.

Lumley (de) Henri, 1976. Les civilisations du paléolithique inférieur en Languedoc méditerranéen et en Roussillon, *in:* Henri de Lumley (ed.), *La préhistoire française. Tome. 1: Les civilisations paléolithiques et mésolithiques de la France*, p. 852–878. Paris, Centre national de la recherche scientifique (CNRS), 1532 p.

Masset Claude, 1982. *Le recrutement d'un ensemble funéraire: dénombrement.* Toulouse, CNRS, 406 p. [Report to a symposium on physical and archaeological anthropology: methods for burial ground study, Toulouse, 4–6 November 1982].

Matiegka M. Jindrich, 1927. L'idée d'une démographie préhistorique, *in:* IIA (ed.), *Troisième session de l'Institut international d'anthropologie, 20–29 septembre 1927. Deuxième section*, Paris, Librairie E. Nourry.

McEvedy Colin and Jones Richard, 1978. *Atlas of world population history*. Harmondsworth, Penguin, 368 p.

Meng-Try Ea, 1980. *Histoire de la population khmère*. Paris, Université de Paris V, 495 p. [Doctoral thesis].

Ming Wong, 1974. *La médecine des Chinois*. Paris, Hachette, 256 p.

Moosvi Shireen, 1984. Une estimation de la population de l'Inde en 1601, *Population*, vol. 39(1), p. 9–26.

Nemeskéri Janos, 1974. *Esquisse d'histoire du peuplement humain* [hand-written translation by I. Hunyadi of an article published in Hungarian], *MTA Biol. Hözt.*, no. 17, p. 9 and following.

Nougier Louis-René, 1959. *Géographie humaine préhistorique.* Paris, Gallimard, 325 p.

Panzac Daniel, 1981. La population de l'Empire ottoman et de ses marges du XVIe au XIXe siècles, *Revue de l'occident musulman et de la méditerranée*, vol. 31(1), p. 119–137.

Petrequin Pierre E., Urlacher Jean-Pierre and Vuaillat Dominique, 1969. Habitat et sépulture à l'âge du bronze final à Dampierre-sur-le-Doubs, *in:* CNRS (ed.), *Gallia Préhistoire. Fouilles et monuments archéologiques en France métropolitaine, Tome XII, Fascicule 1*, p. 1–35. Paris, CNRS, 271 p.

Peyrony Denis, 1949. *Le Périgord préhistorique: essai de géographie humaine. Suivi des listes des stations, gisements, monuments divers connus, avec leur bibliographie.* Périgueux, Société historique et archéologique du Périgord, 92 p.

Pradel Louis, 1954. *Les gisement paléolithiques de Fontmaure*. Toulouse, Université de Toulouse, 93 p.

Pran Nath Chopra, 1929. *A study in the economic condition of ancient India.* London, Royal Asiatic Society, 172 p.

Rouzé Michel, 1982. Le dernier des mégalithes était en béton, *Science et vie*, no. 773, p. 32–36.

Taffanel Odette, 1976. Chronologie du bronze final d'après la nécropole du Moulin à Mailhac (Aude), *in:* International Union of Prehistoric and Protohistoric Sciences (ed.), *Colloque XXV. Les Champs d'urnes dans le midi de la France*, p. 11–20. Gap, Imprimerie Louis Jean, 55 p. [IXth Congress of the International Union of Prehistoric and Protohistoric Sciences, Nice, September 1976].

Vallin Jacques, 1976. La population de la Thaïlande, *Population*, vol. 31(1), p. 153–175.

CHAPTER

67

Migration as a Primary Force in Human Population Processes

DAVID A. COLEMAN

Department of Social Policy and Social Work, University of Oxford, United Kingdom

INTRODUCTION

This chapter will review selected migration processes in the longer term. It will seek to place modern migration patterns into a wider range of the migration behavior that human populations have experienced in their history and prehistory. It will emphasize that the forms of migration that we see today, and on which we judge the demographic effects of migration, are only a subset of possible migratory behavior. Other forms of migration, some now extinct, have had much more substantial effects upon the numbers, distribution, and characteristics of human populations than do most current forms of migration, even the recent substantial international movements to Western Europe and the United States.

Migration is defined here by a change of dwelling-place by individuals or peoples usually (but not always) at a distance greater than that of day-to-day mobility and for a substantial period of time, usually at least a year. In modern terminology, the chapter will concentrate on international more than on internal migration within countries. However, during much of the period considered, national boundaries for countries did not exist and the concept of regulated migration across them was far in the future. Indeed, during most of human prehistory, human populations had no permanent location but moved continually as hunters and gatherers or as pastoral nomads between sites within a roughly defined larger territory. In the absence of a permanent settlement, everyone may be regarded as a permanent migrant, and the distinction between migration among sites or to new territories, as well as daily mobility around a (temporarily) fixed site, becomes blurred but does not disappear. Such continual movement dominated the demographic environment, and the risk environment, in which the human species has developed. These issues will become apparent in this chapter, in which the term "migration" will have to encompass a wider variety of human behavior than it is usually called upon to do.

I. THE STATUS OF MIGRATION IN DEMOGRAPHY

1. Migration as a "Weak Sister"

Until recently, migration has typically been regarded as the "weak sister" of modern demography. Although its effects have always been preeminent on local populations, at the larger level of aggregation that demographers traditionally study, especially the national level, its statistical effects in peacetime upon size, growth rates, and age-structure have usually been small until recently (Voets *et al*. 1995). In recent historical times, immigrant populations have usually comprised a relatively small component of established populations and have not tended to displace them. Postwar immigration has not much retarded the process of ageing in mature societies (Kuijsten 1995). That which is caused primarily by trends in fertility cannot be put right by trends in migration. This marginal demographic status may be changing, however. In recent years new migratory flows, especially from the Third World, have been making increasingly large demographic contributions to Western European populations and to those of North America and the Antipodes. By the beginning of the 21st century, net immigration had become the major component of remaining population growth in most Western European countries. Where the death rate exceeds the birth rate (e.g., Sweden, Germany, Italy), net immigration has prevented population decline (see Council of Europe 2003). Major ethnic consequences in the host population can follow from immigration if permanent migration chains develop among immigrant populations with extended family and kinship obligations and high fertility. Based on current and projected trends, population projections suggest that between 20% and 30% of the population of a number of Western European countries, and of the United States, are likely to be of first- or second-generation foreign descent by midcentury (Coleman 2004). This pattern is reestablishing migration as a primary demographic factor in demographic processes, after a long period of relative unimportance. This chapter, however, is not concerned in detail with this recent renaissance of international migration, which is reviewed elsewhere, but with its varying significance in past times and long-term influence on human population characteristics.

2. Why Most Demographers Do Not Like Migration

The process of migration itself is fundamentally different from the vital processes with which demographers have traditionally concerned themselves. That is one of several reasons why migration is often thought to be a subject more fit for geographers than for demographers. Births and deaths are well defined as irreversible beginnings and ends, but migration is continually repeatable. International movement varies in its intensity, duration, and frequency from permanent exile to border commuting and, in the latter case, can hardly be considered to be "migration" at all. Migration has no firm statistical definition shared by the different countries that exchange migrants (see United Nations 1998). Even an apparently clear category such as "labor migration" is shown to be ill-defined when scrutinized (Salt *et al*. 2004). Migration generates statistics only insofar as it is controlled. So in the modern world, there are as many migration definitions as there are systems of migration control. These laws define migration for different purposes, durations of stay, and so on (Salt *et al*. 1993). However, labor migration under work permit, asylum seeking, and family re-union migration—some of most significant components of modern migration flows into the industrial countries—are potentially more comparable. Migration can be estimated indirectly by population statistics on birthplace or nationality from censuses or, more usefully, from registration systems. Indeed most of the "migration" statistics presented in the annual SOPEMI report (OECD 2003) are derived indirectly from the latter sources, not the least because emigration data are absent or defective. "Internal" migration may share many of the processes and patterns of international migration, but the relation between the two is little studied. Migration and mobility in historical and prehistoric times gives us further problems of measurement but also presents us with forms of migration that, once dominant, are now extinct or nearly so and that are not included in modern typologies.

3. Migration Paramount

The relative unimportance of migration except at the small scale was a relatively recent phenomenon, which now, as we saw above, is proving to have been a transient episode. The particular range of the forms of migration that we nowadays take for granted as being the norm is also a recent phenomenon. We should abandon any notion of a "principle of uniformity" in the study of migration. By no means can all forms of human migration be observed in contemporary populations. Some aspects of migration are new, and many others that have radically shaped human settlement are now extinct. In the past, migration dominated human populations. It often involved the movement of whole peoples, whose cultural entity was much more permanent than their shifting homeland, which indeed often had recently belonged to someone else. For example, it was not until the late 10th century that the "Kings of the English" began to describe themselves as the "Kings of England" (Wormald 1994). The ancient impermanence of human attachment to a homeland in many populations is too easily forgotten.

In the beginning, and for all but the last 10,000 of the 5 million years or so of human prehistory and history, all human groups have been in a state of continual migration at the local level, moving around their home range to hunt and gather food. That mobility, at a small scale and on a daily basis, determined the mortality and fertility risk regimes to which all human populations had to accommodate in prehistory, and to which a few still adhered until recently (see, e.g., Blurton-Jones 1986). Human species—specific patterns of natural fertility and of age-related senescence evolved under such conditions of mobility. At the other end of the scale is the long-range migration by which the world was initially populated and by which peoples subsequently moved, displaced others, and were displaced in their turn. Such past movements have powerfully influenced human distribution and fertility and mortality patterns and created, as well as moderated, the genetic and linguistic diversity of human populations (Cavalli-Sforza 2000). The impact of current migration patterns, which attract so much interest and concern at present, appear trivial by comparison. However, it is important, before we examine the events of the remote past, not to forget that time becomes compressed at long range. Changes that appear sudden, even brutal, to us are likely in reality to have taken place over a long period of time, so that their unfolding could have been perceived by a contemporary observer to have been very gradual. The development of the structure of the population consequent upon the direct and indirect effects of the current migration toward Western Europe and to North America is of this kind.

II. TYPES OF MIGRATION ANCIENT AND MODERN

It might be helpful to illustrate the complexity of the process by briefly reviewing the diverse array of the forms of human mobility. A simple list is presented below of various forms of migration grouped according to chronology, transience or permanence, and motivation. It includes old and extinct forms of migration as well as persistent and new ones. Most contemporary attention is devoted, not surprisingly, to contemporary migration, which is complex and diverse enough. Much more detailed typologies and analyses of the processes behind contemporary migration will be found in Salt (1989), Massey *et al*. (1993), and elsewhere, but such treatments rarely attempt to incorporate ancient or historical forms of migration into their schemata. Some of these will be considered in more detail in the sections that follow. The more modern forms of migration will not be discussed further; they are dealt with in other chapters of this section.

1. Local Mobility and Short-Range Migration

1. *Local migration and mobility*—The list is long and varied: daily food foraging and hunting by individuals and by small groups in their home range or territory; raiding neighbors; traveling for the exchange of goods, gifts, and trade; commuting for work, local house-moves for work; traveling for education. These are the essential movements of everyday life, prerequisites for survival, subsistence, advancement, and reproduction and their essential concomitants. The earlier forms will be as old as the human species. Some are almost extinct. Few constitute "migration" in the sense discussed here. Their motivation requires no new or special explanation.

2. *Migration associated with particular stages of the life cycle*—This includes marriage migration and retirement migration; the former is probably almost as old as humanity, fundamental to the kinship relationships and intergroup contact and cohesion of traditional societies. The latter, however, dates only from the invention of retirement and in the ancient world was a privilege solely of the elite.

These forms of mobility will not be considered further. Instead, attention will be concentrated on movement at longer range and in the longer term, and with various of its consequences.

2. Movement at Longer Range and in the Longer Term

1. *Initial world colonizing migration*—This probably arose in two stages, the first around 1.5 million years BC, the second around 100,000 years BC. By about 8000 BC, the process was mostly complete, with the exceptions noted below. Both episodes involved movement out of Africa to most parts of the globe; probably not by "purposeful" migration but by the spread of settlement by population growth at the margins, as discussed below under the concept of the "wave of advance."

2. *Regular longer-term movement associated with nomadism, pastoralism, and early forms of "unsettled" agriculture*—These include regular movement of whole groups to different parts of their home range to exploit new resources; nomadic pastoralism; transhumance; and swidden, or slash-and-burn agriculture, rotating around the same very large territory in cycles of 25 years or more (Allan 1967).

3. *Völkerwanderung*—This term is given to the movement of whole peoples, or large parts of them, with (some) means of subsistence and with all their impedimenta, over large geographical areas as a directional, opportunity, or plunder-seeking migration or as a refugee movement after displacement by other stronger peoples. This form of migration had overwhelming consequences for neighboring settled populations if the wandering people proved to be

militarily superior, as they often did, especially if they were mounted pastoral nomads. Such movements are recorded from the earliest times but peaked in Eurasia between the end of the first millennium BC to the beginning of the second millennium, a time often denoted as the *Völkerwanderungszeit* (Diesner 1978).

McNeill has proposed a fourfold classification to include the migratory movements of the ancient world once initial colonization was completed, some of which include elements of this phenomenon (McNeill 1984:4). These are presented in modified form below:

Displacement of one population by another by force (invasion/extermination).

Conquest of one population by an invader, leading to coexistence of two previously separate populations (creation of new ruling class and/or incorporation into a larger empire).

Infiltration by outsiders with some degree of acquiescence from existing populations, without displacing rulers (trade, special skills, mercenaries, missionaries, refugees).

Movement by force of individuals or whole communities from their initial homeland by slave raiders or traders (forced labor migration, especially for domestic and sexual services; North Africa and Arabia) and also for agriculture or manufacture (Rome, West Indies, Southern United States, and South America) or administration and military service (such as the *devshirme* of the Ottoman empire, a tribute of Christian children taken to form the armies of the Janissaries). Also forced movements within empires for labor or security reasons (the *corvée* in China, slave labor in Nazi Germany, Ottoman rearrangements, deportations and the *Gulag* in the Soviet Union)

Further forms of migration are more characteristic of money-based market economies and more advanced forms of division of labor:

4. *Importation by invitation*—Individuals or groups keen to increase permanent population, diversify their economy, or expand their labor force. In earlier centuries these were typically minorities with "private" special skills who might often require state protection and privileges specific to their communities. Examples include the encouragement by Peter the Great and his successors of German settlers in Russia, whose descendants became the *aussiedler* of modern times, as well as the attraction of Flemish weavers to England by Edward III. Mercenaries might become permanent settlers through the same process or even displace their employers as did, according to tradition, Hengist and Horsa in the Kingdom of Kent (449).

5. *Conventional voluntary international labor migration*—With the growth of markets and the convenience of modern transport, the previous category matures into something more general. Modern labor migration occurs between economies of equal rank, as within the first 15 countries of the European Union (EU15) countries, but large-scale movement is typically driven by wage differentials between economies and by economic growth that is rapid by historical standards. At the present time, differentials in living standards, of the order of 10-fold between the modern economies and the Third World, are the widest ever. In monetized and internationalized economies of the 20th and 21st centuries, labor migration has often been promoted by recruitment policies and has been managed and limited by work-permit arrangements. Such relatively formal movements began before the 20th century but reached a new high from the 1960s to the mid 1970s during the postwar "guest worker" migration, mostly from southern Europe and the Third World into Western Europe. That movement was expected to be temporary, and indeed some workers did return, but for the most part that movement has perpetuated itself on a large scale through the chain migration of spouses, dependants, and relatives.

Labor migration is a global phenomenon (Stalker 1994), in Europe (Salt *et al*. 2004), in North America, in the growing economies of the Pacific Rim (e.g., Taiwan, Korea, Malaysia, even Japan) (see Salt 1989), and, often less formally, in the Southern Cone of Latin America and throughout Africa. Over historical time, its formalization was developed following the consolidation of nation-states and the control and counting of movement across their borders. Modern labor migration into rich countries is focused particularly upon the active recruitment of persons with specific modern technological skills, often from specific countries. Much highly skilled movement is managed through "intracompany transfers," though which multinational companies shuffle their own employees between countries.

6. *Migration for settlement* (peaceful variants of displacement, conquest, and movement by force)—There is a substantial division today between the major "countries of immigration" outside Europe (the United States, Australia, Canada, and New Zealand), which welcome population for settlement, and others that wish to recruit temporary labor but not necessarily a permanent population. Although the countries of immigration justify their positive migration policies on economic grounds, only a small proportion of their immigrants—under one-fifth in the United States—enter under workforce preferences; the rest enter under various family preference schemes. These coun-

tries of immigration usually recognize a formal category of "permanent settler migration" unknown elsewhere, although in fact many such "permanent" immigrants return to their origins as do participants in all other migration streams.

7. *Entitlement migration*—These streams are forms of chain migration, in particular as a consequence of postwar labor migration to the developed world, in a moral and legal climate that has recognized, unlike in previous times, entitlements to family unification in the receiving country and an increasing equality of rights of immigrants with those of citizens. After the 1970s, postwar migration involving the developed countries has been dominated by the continuation of migration from poorer countries into the richer ones. These flows consist primarily of the dependents and relatives of those labor migrants and their new spouses and, increasingly, the spouses of their children in the growing populations of immigrant origin. These persons, not workers, comprise the majority of regular immigrants to Europe in recent decades.

8. *New irregular avenues of migration*—A substantial flow of asylum-claiming immigration and of illegal immigration, sometimes from the same poor countries as those that provided early labor migration, has arisen since the early 1980s. Asylum claiming, as a large-scale migration stream, depends specifically on rights to asylum formally specified only in the 1951 Geneva Convention and to other benefits of welfare economies. Information on these entitlements is now globalized, together with familiarity about the welfare and other provisions that accompany them. In the majority of countries that receive asylum seekers, only a minority are accepted as qualifying as "refugees" (although most remain anyway). The process is as much one of settlement as of protection and appears to overlap substantially with a more general "betterment" migration, much of it irregular or clandestine. There have been over 6 million claims of asylum in Europe since 1980.

In the discussion that follows, particular attention is given to the earlier forms of migration that are less familiar today and that stand in the greatest contrast to those with which we are most familiar. The characteristics of contemporary migration streams at the end of the 20th and the beginning of the 21st centuries, labor migration toward Europe and North America, family reunion and formation, asylum claiming, and illegal immigration are considered in other chapters, particularly in Volume II, but also in the chapters of Volume III devoted to the demographic transition (Chapters 68, 69) and to the projection of population (Chapters 76, 77).

III. THE INITIAL COLONIZATIONS OF THE WORLD: OUT OF AFRICA

The further back into prehistory we look, the more different—and the more demographically and evolutionarily salient—are the migration processes of the human species. Much controversy about migration in prehistory exists, both about facts and about processes. But it is clear that something quite new migrated out of Africa about 1.5 million years ago and then again about 100,000 years ago—us or, at least, our ancestors. The colonization of the world by the human species began by migration—probably twice—out of Africa. It seems established that hominids first evolved there. No remains of the earliest kinds of hominid—the genus or group of genera *Australopithecus* dating from about 5 million years to about 2 million years ago—are to be found anywhere outside that continent. What first came out of Africa was something more advanced than were the Australopithecines but not yet like modern man. About 1.5 million years ago, *Homo erectus* (now often called *H. ergaster*) had the unique opportunity of colonizing the 75% of the world's remaining land surface that had absolutely no other human beings on it at all. In the end—through processes still unclear in detail—no part of the world came to be free of the human species. We are thus faced with at least three migration questions: Why did *H. erectus* and, later on, early *H. sapiens* escape from Africa when *Australopithecus* did not; what kind of migration should we consider this to be; and why are human beings now found more or less everywhere?

An enhanced capacity for adapting to new environments might have been a first essential. *H. erectus/H. ergaster* had better tools and weapons than those that were available to his predecessors. But even more important, it is claimed, were fundamental human properties, such as the ability to act socially, to collaborate in groups, and to gossip (Dunbar 1996). The demands on intelligence and memory arising from complex human social life are now thought to be a more potent source of evolutionary pressure for enhanced brain development than are the simpler demands of tool-making, which are by no means confined to the human species. According to Gamble (1995), human migration beyond the local area had to await human social organization and possibly speech. It is also noteworthy that the time of the first migration out of Africa coincided with the rise of a type of human whose sexual dimorphism in stature and estimated weight was much less prominent than it had been in previous hominids (McHenry 1994). That argues for a more complex, co-operative attitude between the sexes, away from the dominance of a substantially polygamous harem social structure, which is inferred to have been characteristic of earlier hominids.

Debate continues between two rival explanations of the peopling of the world. Fossil remains of *H. erectus* are found outside Africa, from Java to Wales, but not in the Americas or Australasia. Somehow *H. erectus* were replaced by populations of anatomically modern man. Did *Homo* evolve from the *erectus* level to the *sapiens* level *in situ* in several places throughout the Old World (the "polygenic" or "multiregional" hypothesis) (Wolpoff 1989)? Or were earlier forms, including the Neandertals, who are believed to be descended from earlier *H. erectus*, replaced by a new migratory stream of anatomically modern man out of Africa in much more recent times, beginning about 100,000 years ago? According to the latter "monogenic," "recent African origin," or "out of Africa" hypothesis, the colonization of Europe by anatomically modern man was only complete in Europe by about 35,000 years ago. Anatomically modern man may have colonized Australia and New Guinea as early as 60,000 years ago (Stringer 1989), although evidence in China is lacking before 30,000 years ago, and the characteristic anatomical features of modern Chinese populations are only clearly evident in specimens dated to about 10,000 years ago (Stringer 2001). In this later migration, the human species reached the Americas (via northeast Asia) and Australasia for the first time. This proposed migration coincides with the significant improvement in tools between about 100,000 and 50,000 years ago. Suggestions that the major colonization was also associated with significant improvements in language remain conjectural. However, it seems certain that migration into the dry cold of northeast Asia, and thence to the Americas, would require clothing and other cultural adaptations to climate that were not available to the earlier migrants. And even though the sea voyages would have been much shorter in the lower sea levels at the end of the ice age, man could not have reached Australia without navigation.

The polygenic or multiregional hypothesis would bestow a much greater antiquity upon the genetic and racial differences observed between modern human populations. Accordingly, it requires some migratory contact between separated groups over long periods of time, otherwise the hypothesis would not be compatible with the relatively modest genetic differentiation observed between contemporary human populations. The monogenic or out of Africa hypothesis implies that contemporary human racial differences all arose within the past 100,000 years, without denying the possibility of gene flow with the previous populations whom they replaced or displaced. The low density of fossil remains and the problems of dating them maintains this controversy. But biochemical evidence favors the latter view, suggesting that modern human divergence is more likely to be the result of a migratory radiation in the past 100,000 to 200,000 years, with (until recent times) only restricted subsequent gene flow (by migration) between the separated human groups and between earlier groups and their successors. For example, a DNA comparison of contemporary European populations with fossil Neanderthal material revealed no significant genetic admixture that could be attributed to the earlier Neanderthal populations (Krings *et al*. 1999).

Estimates of genetic distance based on numerous genetic loci, most of which are not thought to be affected by adaptive processes relating to climate (and therefore to geographical location), usually place African populations at a considerably greater distance from European, Asian, or native American populations than the latter are from each other. An African cluster is almost always identified by statistical processes as being primarily distinct from all the others (Cavalli-Sforza *et al*. 1988). Methods based on the analysis of genetic polymorphisms are superior to measures based on biometrical characteristics (body and skull size, shape, and color). Some of the latter are thought to be adaptive with respect to climate and therefore convergent between tropical populations (e.g., skin color in Africans and Melanesians). The trees of genetic distance created from analysis of these populations cannot of itself produce a migratory path with an origin. But connections between the populations by minimum migratory paths fit the supposed migration from Africa to Europe and separately to Asia, and from Asia separately to the Pacific and the Americas.

More detailed patterns will emerge from the analysis of the DNA of the complete human genome. In the meantime, further information bearing on early migration comes from the complete sequencing of another type of DNA, that found in the mitochondria of the cell cytoplasm (mtDNA). Mitochondria are transmitted entirely through the female line. This mtDNA is much shorter than is the nuclear DNA (about 1/200,000 of the size) and is therefore much easier to sequence, although by no means necessarily representative of the whole genome. Results support the theory of recent African origin, suggesting that about 200,000 years has elapsed since the time of the oldest African mtDNA ancestor whose mitochondrial descendants have survived to the present time. This is based on assumptions about the constancy of mitochondrial mutation rates and the nonselective nature of the variations observed (Stoneking and Cann 1989), which does not prevent the actual fission of populations taking place much later. Unless the assumed mutation rate is wrong by an order of magnitude, this evidence is generally held to rule out a multiregional model of human origins and migration, and to favor a later migration and radiation throughout the world for the immediate ancestors of modern man.

Recent studies of genetic distance based on 100 DNA polymorphisms (but based on a restricted group of populations) have suggested that the genetic distance of European origin populations is anomalous in that it is closer to African origins than expected (Bowcock *et al*. 1991). Another, somewhat contradictory study, based on 186 genetic loci, suggests that European populations are closer to Asian than to African populations (Nei and Livshits 1989). Admixture (arising from migration) between African and Asian populations in west Asia about 30,000 years ago, after the initial dispersal from Africa, may have led the ancestors of European populations to become intermediate between the two before their expansion from their west Asian homeland into Europe. Later reinforcements of genes from west Asia would be expected from the much later movements noted below.

IV. POPULATION GROWTH RATES, COLONIZATION, AND THE "WAVE OF ADVANCE"

After the second of the two expansions described above, two new continents were colonized for the first time. Only modern man has been able to migrate to and through the dry cold areas of northeast Asia and thence across the Baring Strait to colonize the New World, acquiring marked racial features in doing so. The orthodox view is that no human remains from the New World are older than about 30,000 years or possibly 20,000 years. However, the colonization of Australasia—by very different kinds of people—occurred somewhat earlier, possibly helped by changes in sea level. In both areas, however, the precise dating of the first human settlement is continually being revised by new archaeological discoveries.

What kind of migratory process produced this global distribution? Textbooks commonly depict these migrations with arrows, as if they were military advances, purposeful and rapid. But there is no need to assume that the migrations were directed, conscious or fast. The slowly moving diffuse fronts of the weatherman might be a better model. Simple population growth, rather than migratory waves, would be enough to push a human frontier of first settlement at a respectable pace across previously untenanted areas. This is the "wave of advance" model. Thus, if annual population growth rates of up to 1% can be allowed at the margins of an area of human colonization, then the initial peopling of the Americas and of Australasia could have been accomplished by that process alone (maintaining population density by settling into marginal new land) within between 1000 and 3000 years (Carneiro and Hilse 1966). That is nothing in the light of the long time-span involved. In the Palaeolithic, in the very long term, average rates of population growth must have been very slow indeed. Only an order of magnitude can be specified, inferred from the obviously small increase of population over a very long period of time. A long-term rate of about 0.001% per year is much more likely than a rate of about 0.01% per year. However, in the short run in individual populations, vital rates would probably have been quite erratic, both for stochastic reasons and in response to changing local encouragements to population growth or decline. At the rate of 0.001% per year, it would take 2 centuries, on average, to add an average of one imperceptible extra person to a band of 500 people.

During parts of the Neolithic, however, when human carrying capacity was raised by an order of magnitude by the development of agriculture, average growth rates may have risen to 0.1% per year (Birdsell 1957; Hassan 1973; Martin 1973) and locally and occasionally to 1% or more. In the short run, population and vital rates are likely to have fluctuated considerably. Population losses from high death rates arising from epidemics, adverse weather, and raids would have to be repaired by a robust capacity for growth, at least in the short term, otherwise populations would simply have become extinct. Some no doubt did become extinct. Such episodes are recorded in the early ethnographic literature and even in historical time, and such was the fate of many early European colonies in the New World. Some populations are too small to be viable, depending on the type of economy being practiced and the standards of living deemed acceptable (Sutter 1951). In recent years some small settlements in inhospitable environments have been abandoned as unsustainable. For example the entire population (diminished to 35) of the remote Scottish island of St. Kilda was evacuated to the mainland in 1930, at the islanders' request.

Among animal species, temporary migration is a key response to adverse conditions of every sort and is also a crucial way in which local population loss or extinction can be repaired from neighboring populations. No doubt the same was true of early man. Most models of early human population envisage local population exchange with neighboring groups, particularly through marriage (Wobst 1974, 1976). That would have had a homeostatic effect upon local populations by increasing effective size.

By about 8000 years ago, this colonizing phase of migration was complete in respect of the larger landmasses of the globe, although lesser colonization of unpeopled land has continued to the present day. The latter includes agriculturally marginal land colonized only during periods of population pressure, for example, uplands and hill slopes in mediaeval Europe. The expansion of the human settled area by agriculturalists continues in parts of the Third World today, for example, on the slopes of the Himalayas and on the thin soils of tropical rain forests. In some of those cases, the agriculturalists displace pastoralists or the few remaining hunters. Other examples are oceanic islands and zones of extreme climate that require modern survival equipment for year-round residence, as in the Antarctic. Many of the Pacific islands near Indonesia and the Philippines were first colonized by about 1500 BC; others, such as Hawaii, New Zealand, and Easter Island, were not colonized until historical times (in the case of New Zealand, 800 to 1100 AD) (Bellwood 1989). This second expansion required developments in navigation, especially the outrigger canoe. Later expansion, mostly European, brought humans to the most remote islands for the first time: Pitcairn was not occupied until 1790, and in the south Atlantic, the Falkland Islands had no recorded landing before 1690 and no permanent population before 1833. Really hostile environments such as the seabed, space, and the other planets in our system still lack permanent residents, although they are likely to gain them in the coming century. But for most practical purposes, the process of new colonizing migration is long over. Migratory newcomers everywhere now have to deal with established residents.

In the Palaeolithic, human populations once established in a new area were invariably locally nomadic; with rare exceptions nomadism was an essential prerequisite for a hunting and gathering existence. Even after the development of agriculture, populations shifted their fields or gardens on a regular basis to maintain soil fertility as swidden agriculturalists. What was once universal in the human species is still practiced by a few remaining hunters and gatherers, swidden agriculturalists in arid areas, and a larger number of pastoral nomads and transhumants. On a local level, even daily life was continuously or episodically migratory. But such populations also began to interact with their neighbors and others further afield through various media: the routine connections of trade and exchange and marriage (which are familiar in modern guise today) and, more radically, those of directed migration, conquest, and replacement.

V. PREHISTORIC MIGRATION OF PEOPLE OR MIGRATION OF CULTURES?

Constant small-scale local migratory exchanges between neighboring human groups can be taken for granted. What is so surprising about early migratory behavior on the larger scale, insofar as we can reconstruct it, is that it tends to develop in particular directions. Judging from artifacts, living floors and settlements, remains of cultivated plants, and so on, successive "waves" of migrants appear to manifest themselves repeatedly, displacing previous peoples and cultures over large areas and long periods of time. Generally, the movements in Eurasia west of the Urals have proceeded from east to west, movements within the Americas and in south and east Asia and in Africa, from north to south. Such a succession of waves of migrants, so easily caricatured and apparently implausible, seems to be the incontrovertible conclusion from abundant archaeological and linguistic evidence, although a different story may be told by the genes. These processes were very different from the migratory processes with which we are familiar today, although they continued in sub-Saharan Africa up to the 19th century.

Peoples are named primarily from the artifacts found at type sites, which are much more distinctive and numerous than are the small number of poorly preserved human remains, which may not be so readily distinguished from each other. However, we do not make such inferences about the migration of peoples from the presence of Japanese televisions in Paris, for example. It may be that cultures, as identified by artifacts, cereal grains, and other evidence of the cultural traditions of the past, may also have traveled quite independently of their original begetters, and their spread could have proceeded without population replacement. That is so for individual objects of value, such as jewels and coins used as grave goods in cultures that clearly did not produce them, with no other evidence of general cultural change. Thus the spread of agriculture, dated at successively more recent places from the Fertile Crescent to its eventual arrival in the British Isles and Scandinavia, has been interpreted by some as evidence of the spread of techniques and artifacts rather than of people. However, others (e.g., Ammerman and Cavalli-Sforza 1985) claim that evidence from gene frequencies in European populations point to a process of demographic diffusion and (partial) replacement of population through a demographic wave of advance, proceeding at about 1 km/year on average. According to this view, barley and wheat traveled across Eurasia in 5000 years primarily because the people who grew them spread from east to west in that time.

Other, more recent work (e.g., Sykes *et al*. 1999), however, suggests that genetic contributions of Middle Eastern origin (and therefore net demographic contributions) to modern European populations comprise between 10% and 15% of the total, and that overall, most current European ancestry derives from earlier populations. Further analysis of mtDNA from European and Middle Eastern samples, which attempts to take reverse migration into account, has suggested that less than 10% of mtDNA lineages in contemporary European populations date from the first colonization by anatomically modern humans; about 20% is derived from Neolithic colonizers, and the rest is from several waves of migrants in the Upper Palaeolithic (Richards *et al*. 2000). The populations known to archaeologists from the Cro-Magnon period, from about 35,000 years ago, were initially hunters and gatherers, and in this view, their descendants learned agricultural techniques and were not mostly displaced during the Neolithic. If that is so, then the wave of advance was more cultural than demographic. The much later spread of the use of the stirrup, however, appears by contrast to have been more or less instantaneous by the standards of prehistory. This debate remains lively and unsettled, although more recent investigations that use detailed biochemical analysis of human remains suggest that considerable population replacement did take place at least in some areas, for example, in England between the sixth and the eight centuries (Weale *et al*. 2002), reinforcing older views based on language change.

The biochemical genetics of modern human populations permits some reconstruction of human migratory patterns and even some dating of the divisions between major human groups. Much prehistoric "diffusion" of techniques and/or of peoples would have taken place before the organization of states and involved small populations combined, if at all, only relatively loosely. The problems of defining migratory populations on a cultural basis with imperfect knowledge of their human biology is exemplified by the difficulties involved in defining the Celts. Everyone knows what is meant by the word "Celts," who as Gauls gave the name to Gaul, to Gallipoli in Turkey, to Galicia in Poland, and to Galicia in Spain. Native speakers of Celtic languages, however, are nowadays only found in the western fringes of Europe. But everyone appears to know a different Celt, at once described as tall and blond, short and dark, and every other combination; even the Celtic artifacts have little coherence of style over their whole range (Chapman 1992).

When most of the habitable world was already occupied, early movements of whole known populations often seem to have been set in motion by pressure from others. According to this Dark Age domino theory, such pressure inevitably created a cascade of repercussions as the migrants then collided with other migrant and settled peoples. Thus, the arrival of the Sarmatians is "explained" by pressure from the Scythians, that of Goths and of Alans by Huns, and so on. These movements dominate the history of Eurasia for several centuries, and their origins require a special explanation, one that is so far lacking.

VI. THE EFFECT OF MIGRATION ON DEMOGRAPHIC REGIMES

In the past the process of migration has dominated the magnitude of human population levels and the patterns of fertility and mortality, which have to adapt to the conditions of mobile and migratory regimes over the greater part of human existence. Nomads may need long birth intervals, especially those among whom the women, as well as the men, travel frequently on foot both to gather food and to change campsites. !Kung Bushman women would have to carry impossible loads for much of their lives if they could not space their children such that two would never need to be carried at once (Lee 1972). A subtle environment/nutrition interaction with characteristic "shallow and frequent" breast feeding patterns may be the possible mechanism whereby their unusual birth intervals are achieved (Howell 1986). Whatever the process, birth intervals in the Bushmen are known to have been very long, and completed family size are correspondingly modest (Blurton-Jones 1986). Perhaps surprisingly, abortion and infanticide are believed to be rare (about 5%) in that population. But they are strikingly prevalent among many other nomadic and shifting populations—aboriginal Australian hunters and gatherers, the Yanomamö Indians of Brazil (Chagnon 1974; Chagnon and Irons 1979), the garden cultivators of Papua New Guinea (Schiefenhovel 1984), and others—to the extent that up to 30% of births, especially female births, are thereby eliminated (Hausfater and Blaffer Hrdy 1984; Harris and Ross 1987). Other studies, however, do not show a clear relationship between economic type and demographic regime (Campbell and Wood 1988).

The influence of migration or its absence on mortality has been even more potent. When human populations migrated for the first time out of Africa and more generally out of the tropics, they probably gained in survival terms because they may have left behind many of the dangerous parasitic organisms, especially those with insect vectors, that depend on a tropical African habitat. Furthermore, nomadic human populations, necessarily relatively small scale, cannot harbor some major causes of death that can become prominent in settled populations; for examples, nomads are typically free of diseases of contaminated water supplies as typhoid and cholera, of the parasites of contaminated soil such as hookworm, or of nest-pests such as fleas. They do not usually stay in one place long enough to complete the life cycle of parasites carried in local populations of insect vectors of human diseases such as malaria. They are unlikely to be able to harbor tuberculosis; without domestic animals, ancient populations would develop neither tuberculosis nor influenza. Small population size would protect them from smallpox and measles. Such diseases could not evolve in humans until populations had settled and grown in size (Fiennes 1978; McKeown 1988); many require a large minimum population size (about 500,000 in the case of measles) for their existence. Contact with settled people who had developed such diseases, however, would be all the more disastrous as nomadic populations would typically lack any acquired or genetic immunity to crowd diseases (Cockburn 1963).

On the other hand, the migration of small-scale societies would take them into areas—especially new and hitherto unpopulated areas—where contact with wild animals and their insect parasites would expose them to serious zoonoses such as bubonic plague, scrub typhus, and other rickettsias; leishmaniasis; and other serious acute or chronic diseases (Pavlovsky 1966).

The most dramatic example of the new burden of disease brought by migration is the destruction of population in North America, in especially Central and South America, on numerous Pacific Islands, and elsewhere following the arrival of Europeans. The aboriginal populations of those hitherto isolated areas, never having experienced the extensive armory of Old-World diseases brought to them for the first time by transoceanic migrants, conquistadores, settlers, and traders succumbed to them with extraordinarily high mortality rates. That was perhaps the biggest proportional depopulation ever arising (indirectly) from migratory processes, although there are many other examples arising from the "globalization" of diseases after human migration (McNeill 1977). Exactly how big the decline was is almost impossible to state. Estimates of the precontact population sizes of North and South America are highly contentious and vary by an order of magnitude (Thornton 1980; Henige 1992; Zambardino 1997). Conservative numbers are given in Table 67–2 below.

Early agriculturalists from about 6000 BC started to contract diseases from their domestic animals, which in many cases then evolved into specifically human parasites (Stanley and Joske 1980). The same problem would have affected pastoral nomads, although the nomads rise is thought to have been later than that of agriculturalists (about 3000 BC) and depended on the spread of genetic polymorphisms for lactose tolerance, which enabled adults to digest substantial quantities of milk. Subsequent migratory exchanges between settled populations—for trade, slavery, conquest, exploration, or whatever—would tend to increase the sum total of human disease mortality. Through migration at all levels of magnitude, all settled populations eventually acquired the diseases of all other populations without being able to lose any of their own. The historical timetable whereby populations acquired for the first time new diseases that had evolved in other areas of the world is reasonably well known for the last two millennia (McNeill 1977). The process of circulation of disease has, if anything speeded up and intensified in the age of universal mobility with new diseases such as AIDS, new variants of cholera, or antibiotic-resistant strains of other bacteria.

The transition from more mobile nomadic human populations to less mobile and larger settled populations of swidden farmers and intensive cultivators is believed to have reduced average human life expectation by several years (Cohen and Armelagos 1984) through the heavier burdens of infectious disease noted above. In the recent past at least, hunting and gathering societies in Africa seem to have enjoyed lighter mortality than did simple cultivators. For example, the nomadic !Kung of the Kalahari and the Hadza of Tanzania were estimated to have had life expectations of 32 years and 31 years, respectively (Dyson 1977; Howell 1986), compared with about 24 years for simple cultivators in a village in the Gambia in the 1950s (Billewicz and McGregor 1981). Earlier Palaeolithic populations, however, may have suffered higher mortality, according to some skeletal evidence (Acsádi and Nemeskëri 1970). Some hunting and gathering environments were intrinsically very dangerous.

VII. MIGRATION, WARFARE, AND ORGANIZED STATES

In prehistory patterns of migration must be inferred from the relics of genes, bones, artifacts, and contemporary language and place names. These leave much room for argument. History starts with and comes from organized states with writing. At long range at least, its account is dominated by expansion, trade, control, warfare, and destruction. Much of this is connected with migration. From the beginning of written records, usually first attributed to the Sumerians of Uruk from about 3000 BC, high levels of migration and population displacement are at once apparent and almost invariably associated with warfare and conquest and with the extinction of some political entities and the coalescence of others. Some of these shifts of population occur between organized states, usually centered on defensible cities. Sometimes conquest leads to the destruction of those cities and, it may be assumed, of their populations. Most of the first Sumerian cities of Mesopotamia were eventually burned, as were the cities and palaces of Crete. The warlike cities of Assyria expanded their harsh control over the area of modern Iraq and beyond in to Asia Minor and Lower Egypt until they ceased winning in about 703 BC. Conquered populations paid tribute if they submitted, under centrally appointed officials, and were destroyed or transported if they rebelled, as were the northern Hebrews under Sargon II (reign, 721–705 BC). In an Empire without natural defenses, constant victory is the price of survival. In 612 the capital Nineveh, the population of which may have been the first to exceed 100,000, was itself destroyed by its former subjects in Babylon and their Scythian nomad allies, never again to be inhabited and leaving neither cultural vestiges nor mourners. Babylon in turn was to be captured 75 years later by Cyrus the Achaemenid. A later Achaemenid, Xerxes (reign, 486–465 BC), set new standards for the size of mobile armies, possibly taking half the Empire's total levy of 360,000 men for his successful invasion of Greece.

Quite often, however, in the relations between settled populations, the permanent migration of peoples was not an important component of political change. Conquered populations were not necessarily displaced or destroyed, although this may have been more usual after conquest by nomads. Often political control and the recipients of taxes changed, rather than the people who generated the taxes, as befits the spirit of mercantilism. The Achaemenid Persians, their successors the Parthians and Sassanids, and the later neighboring Roman empires behaved in this way: aggressive and absorptive (especially the Roman) and explicitly multicultural and multifaith. As equipotent neighbors they remained incapable for centuries of making serious permanent inroads into each others' territory until the final reduction of the Sassanids by the emperor Heraclius in 628.

1. Migrant Peoples

The expansion or migration of population of warlike nomads or transhumants into territory occupied by other people is a distinctive ancient historical phenomenon, which is unknown to modern Europe although it still occured in parts of sub-Saharan Africa into the 19th century. Nomads moved into lands occupied by other nomads, but the nomads more significantly (and more often recorded) moved from marginal areas into more central areas occupied by denser, settled agricultural populations who could write. As a result of expanding their area or in moving *en masse*, they frequently caused the previous resident population to disappear as a recognizable culture, although the genetic inheritance might continue. On other occasions, as with the Mongol and Manchu invasions of China, the nomads replaced only the ruling class, albeit destroying much of the population in the process. These movements constitute one of the most potent forces affecting the population density and the vital rates of neighboring populations in the past, nomadic and settled, helping to damp world population growth for a millennium. The picture that we can present of them undoubtedly foreshortens events and combines diverse peoples into rather oversimplified groupings, but some major events do seem to be established beyond much controversy.

One of the world's first empires, that of Sargon of Akkad (Mesopotamia), was put to an end by invasion of Indo-European tribes ("Guti") from the Zagros Mountains in modern Iran around 2200 BC. Another example are the Indo-European Hittites, who expanded southward out of Anatolia around the same time (late third millennium BC). They appear to have arrived at the edge of Mesopotamia by moving around the north of the Black Sea and through the Balkans. In the second millennium BC (and later), one of the chief foci of the emigration of mostly pastoral people was the area east of the Black and Caspian seas. Indo-European speaking pastoral nomads, popularly "Aryans" or "Iranians" generated a number of population movements. Aryans moved east into modern Pakistan and then India, permanently destroying *en route* the early isolated centers of literacy at Harappa and Mohenjo-daro around 1600 BC. Others, in the form of the Mitanni, moved into northern Mesopotamia; yet others moved into central Asia. The Dorian Greeks and Phrygians moving south out of the Balkans around 1200 BC overthrew the Achaean Greek and Hittite kingdoms of Southern Greece, Crete, and Asia Minor. Other marginal areas also provided notable examples of expansion and conquest. For example, the Semitic Amorites moved north from Arabia and, with others, overthrew the Sumerian empire of Ur at the end of the 21st century BC. Semitic Hyksos "shepherd kings" (possible also Amorites) moving from the east across Palestine established by force their own dynasty in Egypt in the early second millennium BC (around 1700 BC), possibly at the time of Joseph. A continual series of such incursions is recorded in early history (see, e.g., McEvedy 1967), as well as more "conventional" rivalries and warfare between settled states (often themselves the descendants of previous invaders or ruled by their dynasties.)

A formidable impression is given of an alarming transience of political systems, life, and property in these areas, arising partly out of migratory processes, although the effect is exaggerated by the foreshortening effect of the remoteness of time. Naturally some areas on migratory crossroads were more prone to invasion than were others. Recovery when it occurred could be very fast. For short periods, human populations might have grown at up to 1% per year in recovering from previous even sharper declines, rather than at the tiny average growth rates typical of longer periods. Mesopotamia, Central Asia, and to a lesser extent the borders of India are littered with the mounds of abandoned cities (some destroyed and rendered uninhabitable by local climatic change or by deforestation, overgrazing, soil exhaustion, or salination after intensive cultivation). Part of the key to the success of the later nomads is their use of the horse, first to draw chariots and then, later and more important, for riding, which appears to have been first developed by the Scythians around the seventh century BC (Ferrill 1985).

2. Völkerwanderung—The Coming of the Barbarians

The most dramatic manifestation of these migratory pressures are those in the so-called "migration period," or "*völkerwanderungszeit*," which occurred from perhaps the 2nd century BC to the 16th and which had its European peak in the 4th to the 11th century. This was a period in which migratory activity, despite all the uncertainties and distortions of history, was clearly of a different kind to anything experienced since. Over that long period, the history of the Later Roman and Byzantine Empires, China, and India, as well as other areas of settled population in between, are to a depressing extent the stories of their attempts to defend themselves against such incursions. Accounts of this "great migration of peoples" read as something not short of fantasy or of a caricature of history memorably satirized by Sellar and Yeatman (1931). We are invited to believe that successive "waves" of people emerged from a few focal points in marginal areas—the Baltic shores, the steppes of Central Asia, the Arabian peninsula, Mongolia, and Manchuria—and were able to extend their control over huge areas of western Eurasia, central and western Asia, China, and India. In the course of this migration, much of the settled populations and settlements in their path were allegedly destroyed, the rest overruled or enslaved, and previous migrant settlers or conquerors replaced. But, however implausible, much of it appears to be true.

The migrations appears to have involved demographic processes of a kind whose free operation and consequences are today unknown. Unfortunately, their origins are to be found in remote continental areas, which are devoid of historians and offer few archaeological relics and from which no data will ever come. Writings and linguistic evidence are likewise sparse. All we usually know are the consequences and records of their outward expansion, when they touch on the civilized world. At serious risk of oversimplification; two basic groups of migratory invaders can be recognized. The first are groups of cultivators primarily from Scandinavia, Germany, and other parts of Europe west of the steppes (Teutons, Vikings, and others), whose migratory routes were varied—sometimes taking them with wagons or boats further east to the Black Sea before colliding with the western and eastern parts of the Roman Empire. For example, Scythians and Sarmatians, associated with an Eastern origin, nonetheless appear to have spoken Indo-European languages. Further east, pastoral steppe nomads from central Asia and further east of very different languages and culture (Huns, Mongols, Turks, Manchus) pressed more on north China and the northern fringes of India, although extended their power west to Eastern Europe, the plains of Hungary, and the Middle East. Only the Arabs had a southern origin. It is widely supposed that population pressure is more likely to be episodically generated from the most remote and marginal areas of simple cultivation or pastoral nomad subsistence: the unproductive agriculture of northern Europe; the Arabian deserts and its fringes to the south, the steppes to the north of the Caspian, and further eastward to the East. From such places, all roads led to more prosperous areas and eventually to settled areas of cultivators, as one set of migrants or nomads displaced another. The Asian populations mostly spoke a variety of Ural-Altaic languages: Altaic includes Mongol (a language family including Kirghiz, Tatar, and Turkish); Ural includes Magyar, Finnish, Estonian, and Lapp. In many cases (Huns, Avars, Hsiung-nu) we do not know what languages were spoken. In this area many assertions rest on thin foundations: central Eurasia has been described as the "last of the happy hunting-groups of the dilettante historians," where the "first quality needed by the historian is distrust" (Sinor 1977:113–114).

These nurseries of nomads were the areas least equipped to sustain any population growth, and from which any outward migration would by definition encounter more attractive and easier southern and western environments. By the fourth century, the growing weakness of the Roman Empire, especially its western part, was sufficiently well known to be a further attraction. Precarious and marginal lands can most readily be exploited by mounted nomads, who may require the habit of raiding as a necessary insurance against risk and disaster to herds, as well as a potent contributor to them. Neighboring agricultural populations (e.g., the Persian and Tadjik neighbors of the Turkomans and Kirghiz) typically suffer continuous slave raiding and servitude. Outward expansion from these centeres would exert pressure of neighboring peoples, who would themselves migrate. Thus, pressure from the Huns mobilized Iranians, Goths, and many others; the Avars appear to have been driven into Europe by pressure from the Turks, and the Turks in turn were driven out from 1225 by the Mongols to Armenia and Asia Minor, as were the Cumans to Hungary. Some of these movements were promoted by coalitions with a proper leadership structure, such as the early Turkish Khanate. The Mongols of the 13th century, unified politically in 1206 AD, had become a mobile organized state, with a well-developed and highly successful exploitative military strategy (Sinor 1977:XVI).

Satisfactory explanations are lacking to show why large sections of the usually fissiparous peoples from these areas should be able from time to time to unite themselves in large numbers and depart bent on conquest (or at least more than usually extended and systematic plunder) as well as raiding. Some population was usually left behind in the steppes, unless the whole tribe was being driven out by others. It also remains to be explained why nomads in harsher environments, characterized by unpredictable extremes of heat and cold, were able to generate natural increase. It may be that intrinsic mortality rates in such environments were lower than were rates among densely settled populations, for reasons outlined earlier. On some occasions unusual climatic change, or reverses by settled populations at the other edge of their range, may have mobilized them. It may be recalled that pastoral nomads have three elements of potential instability in their ecology—that of people, herds, and the plants that they eat—not just the two (people, plants) that characterize agricultural populations. Herds may become too numerous to be sustained, plant productivity may fall, or human numbers increase; all requiring new grazing grounds (J. Landers, personal communication).

The consequences on settled populations at the end of this chain of dominos was severe. For a long time, the nomads gained considerable advantage from their mastery of the horse (most others fought on foot), mobility (all they needed was readily portable), and their hardiness from familiarity with severe conditions and long migrations (the Kirghiz annual transhumance was about 1000 miles), to which might be added the unique long-distance strategy, strict military timetabling, and fierce discipline of the Mongols. Relatively small numbers of migrants appeared able to overcome more numerous settled people or to live on tribute from them over long periods. For example, the population of Han China, about 54 million, was for long periods of time unable to defend itself effectively from the Hsiung-nu, who numbered perhaps 1 million at most. It was difficult for settled populations to acquire the lifelong skills of lightly armed nomad horse archers or to concentrate or maneuver fast, as mobile people could. On the other hand, the nomad mode of fighting was an "ethnic technology" limited geographically by its need for substantial supplies of grazing and remounts. That helped to explain why the Hunnish center remained on the Hungarian plain. Their possible relatives, the Hsiung-nu, appear to have exploited their mobility with great subtlety in order to maintain a long-term policy of economic extortion and predation upon the settled Chinese. Deep mobile strikes would retreat before Chinese retaliation could be arranged. This violent raiding had the twin aims of loot and terrorism, the latter to persuade Chinese rulers to accept temporary, more favorable peace treaties through which more tribute could be exacted by what was in essence a huge protection racket. The Hsiung-nu usually declined to occupy Chinese territory even after victories, relying upon the Chinese to build them up again so that they could be looted once more in due course. The Mongol empire, too, appears to have depended primarily on tribute rather than on natural resources (Barfield 1989:49–50).

Sometimes, however, even without gunpowder, the empires could strike back. For example, the Chinese responded to the northern nomads (known to them as the "five barbarian tribes": Hsiung Nu, Chieh, Hsien Pei, Ti, and Chiang) not only with extraordinary defensive measures such as the Great Wall begun by Ch'in Shi Huang (reign, 221–206 BC) but also with active campaigns, such as the unusually successful ones undertaken by the Early Han emperor Wu Ti (reign, 140–87 BC) to knock out the western Hsiung-nu in 44 and 36 BC. Later Chinese resistance is believed to have diverted some Hunnish interest further West, to the discomfiture of Europe. It was not really until the age of gunpowder that settled populations could become more secure from the depredations of nomads on the fringe of settlement (Landers 2003) and could start to push them back to their heartlands.

One of the most consequential of the ripples of pressure arose from the destruction by the Huns of the Ostrogothic empire on the Dniester in 375. After this disaster the Goths obtained the permission of the Roman Emperor to settle within the Roman Empire in Thrace—a kind of mass asylum claim from which a certain element of threat was not absent. Through developments too complex to relate here (see Heather 1991), the Goths eventually inflicted one of the most decisive victories ever over Roman arms at Adrianople (Edirne) in 378. That victory established them permanently within the empire. That migration was eventually to contribute to the end of the western Roman Empire in 475 and the establishment of the Goths in many parts of the Western Europe (Ferrill 1983)—the Ostrogoths primarily in Italy and the Visigoths in France and Spain—albeit in part through negotiations with their nominal Roman overlords, which established them and other invaders as "allies" *(foederati)*.

The Huns, who originally set in motion this sequence of migration, themselves invaded Western Europe later on. By the fifth century, they were no longer mere pastoral nomads. United under Attila from his "palace" in the Hungarian plain (more a shanty town of wooden buildings, tents, and wagons), where Latin and German were allegedly spoken more than was Hunnish, they ruled large numbers of both Slavic and Germanic peoples and bordered both halves of the Roman Empire. As with all the barbarian tribes, relations with the Romans were complex and included periodical alliances, including the use of Hunnish mercenaries by Rome and ruinous tribute extracted from the eastern Empire. Unlike many of the other barbarians who entered Europe (e.g., Magyars, Bulgars, Burgundians, Lombards), the Huns under Attila did not become a settled people, to be attacked in turn by new invaders. They remained on horseback; their subsistence was derived almost entirely from Roman tribute as the subsistence of others derived substantially from plunder. Active conflict with Rome began on the pretext of the Roman acceptance of what we would now term asylum-seekers, who would undoubtedly, under those circumstances, have entertained a well-founded fear of persecution. Checked by almost the last gasp of western Roman arms in alliance with Burgundian and other (barbarian) allies at Châlons-sur-Marne (451), the Hunnish Empire fell apart on the death of Attila (453, following his stag-night binge) and, unlike most of the other invaders, featured no further in Western European history.

Perhaps the most remarkable of these mass migrations of whole peoples is that of the Vandals. They, along with many other Teutonic tribes, breached the Roman defenses on the Rhine (406) and established themselves in France and Spain, partly through force and partly through Roman appeasement, treaty, and partition of the land with local landowners. When established in Spain under their king Gaisarix, the Vandals looked, as the Goths under Alaric had also done, to the richest province of the Roman Empire in North Africa, the key to the food supply for Italy. In 429 they crossed to North Africa. Initially established as *foederati* after early military successes, they soon (442) became *de jure* as well as de facto rulers of the western part of North Africa, around Carthage, and no longer regarded themselves as part of the Roman Empire. Previous inhabitants, although dispossessed and regarded as a subject people, do not appear to have been generally slaughtered. In 455 the Vandals invaded Italy and sacked Rome, removing most the objects left behind by the Visigoths in 410 but did not settle. By 476 the empire of Gaisarix (died, 477) included all the Roman province of Africa, Corsica, Sardinia, and Sicily and was raiding the coasts of Greece. But even this established state could not defend itself against the reprisals of the eastern Empire, and its culture was eventually swept away entirely by the Arab Muslim invasions of the seventh century. The Arab invastions eliminated a "European" cultural and political presence in former Roman North Africa and, shortly after, took over most of the Vandal's former territory in Spain itself, for the next six centuries.

The litany of such migrant peoples is perhaps usefully, if inevitably very partially, recalled in a table (Table 67–1). Their (very approximate) dates are those when they first come to the notice of people with writing. The numbers claimed are sometimes surprisingly high. Contingents of 30,000 are regularly mentioned. Confederations of barbarians could raise armies estimated by contemporary sources to be up to 300,000 men (e.g., a union of Ostrogoths and Visigoths defeated by the Emperor Claudius II on the Danube in 269 AD). Even if this represents all the men of military age, it still seems to be a very large figure, about the maximum size of the Roman army at its height, and not then surpassed until mass armies were raised for the first time in modern Europe from the time of Louis XIV (Landers 2003; Tables 13–1, 13–2). How such a large force could have been sustained at that time,

TABLE 67–1 Some Migrant Peoples of Eurasia, 400 BCE–1600 CE

Peoples	Time	Origin	Settlement
Teutons			
(Suevi, Marcomanni, Cimbri, Ambrones etc.)	4th BC	West Baltic	Germany, France
Other members of Teutonic confederacy			
Alemans	3rd AD	Elbe	France
Franks	3rd AD	Weser-Rhine	France
Vandals	3rd AD	Jutland, Norway	France, Spain, north Africa, Italy
Ostrogoths (East Goths, east of Dniester)	3rd AD	Scandinavia/Black Sea, Italy	Germany
Visigoths (West Goths, west of Dniester)	3rd AD	Scandinavia/Black Sea, Italy	France, Spain
Angles	3rd AD	Schleswig-Holstein	England, lowland Scotland
Saxons	3rd AD	lower Elbe region	England, lowland Scotland
Jutes	3rd AD	north of Rhine	England, lowland Scotland
Burgundians	3rd AD	Oder/Vistula	France
Lombards	6th AD	south Scandinavia	Italy
Huns	4th AD	Steppes	central/east Europe, north India
(Hsiung-nu / Xiongnu)	3rd BC	Steppes	North China
Bulgarians	7th AD	Steppes	Balkans, Italy
Magyars (Hungarians)	8th AD	Steppes	Hungary
Mongols (Tatars)	12th AD	Manchuria	Middle East, Russia 13th–16th century, Yüan dynasty (China) 1276–1368
Moguls	16th AD	central Asia	Indian dynasty 1527–1857
Manchus (Ch'ing)	17th AD	Manchuria	Ch'ing dynasty 1644–1908
Scythians	7rd BC	central Asia	Balkans, Caucasus, northwest India
Sarmatians	3rd BC	central Asia	Transcaucasia, south Russia
Alans	1st AD	central Asia	Caucasus
Turks			
Seljuk	11th AD	central Asia	north India
Western	6th AD	Mongolia	central Asia
Eastern (Uighurs)			central Asia
Ottoman	13th AD	central Asia	Balkans, north Africa, Arabia,
Asia			Asia Minor, Egypt
Arabs	7th AD	south Arabia	Spain, north Africa, Persia, central Asia
Vikings	8th AD	Scandinavia	Normandy, east England, Iceland

This list is far from complete; no list can hope to do justice to a complex process covering two continents and a millennium. Migration histories were usually much more complex than given here; for example, the Germanic Goths began, or were first recognized, in the area of the Vistula and moved south to the area of the Dniestr before moving west.

Some supposedly related peoples are grouped by indentation, but many relationships or origins are unknown or vague at best. The dates and locations can only be approximate.

even by plundering all the countryside through which it moved, is not clear; such estimates are unlikely to have been based on any kind of actual counting. The contemporary (professional) Roman army on all fronts of the empire about that time is known to have been about 250,000 strong (Ferrill 1994). A more modern estimate of the numbers of all the warriors in all the Gothic hosts in the late 4th century (which were never unified) suggests about 60,000 men and therefore a total population of perhaps 300,000 in all the Gothic groups together. Estimates of the Thracian Ostrogoths alone in the late 5th century suggest a population total of between 30,000 and 50,000 (Burns 1978, Heather 1991).

VIII. THE DEMOGRAPHIC CONSEQUENCES OF THE "MIGRATION PERIOD"

The consequences of this form of migration on most of the peoples visited by it appears to have been almost entirely negative in demographic terms and probably in almost all other terms as well. The visiting peoples added little in terms of population and typically nothing in terms of cultural and technical advance (except possibly in warfare). Through destruction of people, habitations, and the infrastructure of cultivation (especially in the dry East), they caused a perceptible and often long-term diminution of human numbers. Estimates of the numbers of the migrants are obviously difficult and are likely to have been exaggerated by contemporaries. On the other hand, from the first century AD onward, we do at least have somewhat firmer bases for estimating regional populations, from remnants of Han and Augustan censuses and a much more intensive archaeological knowledge of the number and extent of settlements.

According to the estimates of Jean-Noël Biraben, discussed in the previous chapter (Table 67–2), the world population is thought to have diminished from about 250 million in 1 AD to 205 million by 500 AD, not recovering to its previous level until about 1000 AD. (It can only be repeated that all such estimates are tentative and uncertain; figures for 1 AD vary from 169 to 297 million) (Caldwell and Schindlmayr 2002: Table 4). Much of such loss can be attributed to the disruption of settled populations. The recovery to 1000 AD did not include Europe, whose population fell from 35 million in 1 AD to about 30 million in 500 AD and in 1000 AD, nor did the recovery include the population of North Africa, which fell from 14 million to 9 million over the same millennium. European population by 1500 was little larger than it had been in 1250. China and south Asia suffered worse from the invasions of Huns and their relatives and from Aryans and Muslims from the north and west. China's population fell from 70 million in 1 AD to 32 million in 500 AD, recovering to only 56 million by the end of that millennium. The period between the capture of the Tsin dynasty capital (then at Loyang on the Yellow River) in 316 AD by the Hsiung-nu until the reestablishment of complete Chinese control over the North by the Sui dynasty in 589 AD is regarded as China's Dark Age. The capital had to be removed to Nanking on the Yangtse over 300 miles further south, with the consequent migration of many of the survivors from the Yellow River valley, which was effectively abandoned, to that of the Yang-tse. Exact numbers are unknown, but the losses and migrations together must have involved millions of people. India's population is estimated to have fallen from 46 million to 33 million to 500 AD, then only recovering to 40 millon by 1000 AD (Biraben 1979). A basically similar picture is given by other estimates (e.g., those summarized by Durand 1974; McEvedy and Jones 1978), all of which must be regarded as being rather precarious.

Most parts of Eurasia also had to endure the impact of the bubonic plagues, which reached Europe in 1347 from the Crimea and had afflicted China earlier. China in particular, however, from the mid 13th century, also suffered from immense destruction inflicted by the Mongols, who also enslaved Kiev (1210), much of Russia, and parts of eastern Europe but who failed to reach Western Europe. The expeditions and depopulations through slavery instituted by Tamurlane (1336–1405) are claimed to have caused the deaths of up to 17 million people; an opinion impossible to verify, of course. The Mongols may have reduced the population of China by up to 35 million people during their protracted conquest in the 13th century, in an invasion reminiscent of that of the Hsiung-nu (sometimes identified, on thin evidence, with the Huns) in the 4th century. However the Hsiung-nu remained essentially raiders, whereas the Mongols, having eventually abandoned their preference for depopulating China in favor of exploiting its riches, founded a new (Yüan) dynasty in 1271. North China (then the most populous part) is thought to have lost about three-quarters of its population (McEvedy and Jones 1978): the Chin census of 1195 enumerated 50 million in North China; that of the Mongols in 1235–6, admittedly in a time of turmoil, counted just 8.5 million. As a result of massacre and forced migration, the demographic center of gravity in China seems to have shifted to the south, as maps based on contemporary censuses appear to show (Population Census Office 1987:70–71). By 1500 the population of China was still

TABLE 67–2 Some Estimates of Regional Population 1 BC–1500 AD (millions)

Population	1 BC	500 AD	1000 AD	1250 AD	1500 AD
China	70	32	56	112	84
South Asia	46	33	40	83	95
Russia	12	11	13	14	17
Europe	35	29	30	57	67
North Africa	14	11	9	8	8
North America	2	2	2	3	3
Central/South America	8	11	14	23	34
World	250	205	257	413	458

Note: South Asia comprises India, Pakistan, and Bangladesh. (From Biraben 1979) less than it had been in 1250: 84 million compared with 112 million. These figures cannot be exact, but similarly large estimates of population loss appear in all estimates that have been made. They are at least based on some of the few (albeit discontinuous) registration systems of the time, and the losses are in accord with the contemporary reports of the ferocious genocide practiced by the Mongols in the process of their conquest of north China.

As an earlier commentator noted

> This is the role of nomadism in the history of the world: countries too distant from its base it could only ravage transitorily, with robbery, murder, fire and slavery. But the stamp it left upon the peoples which it directly dominated or adjoined remains uneffacable. The Orient, the cradle and chief nursery of civilization, it delivered over to barbarism; it completely paralyzed the greater part of Europe . . . had Germany or France possessed steppes like Hungary, where the nomads could also have maintained themselves and thence completed their work of destruction, in all probability the light of West European civilization would long ago have been extinguished, the entire Old World would have been barbarized and at the head of civilization today would be China (Peisker 1911:359).

As a slight counterpoint, it should be noted, following earlier allusions, that settled and migratory peoples also engaged in trade and military alliances with each other. Thus Pope Innocent IV opened correspondence with the Mongols (some of whom later became Christian for a while) in 1245 in the hope of deflecting their attacks on Europe, securing their aid to restore the Holy Land, and converting them to Christianity. Although ultimately fruitless, this enterprise continued when, for example, Philip the Fair, King of France, corresponded with his ally the Mongol Ilkhan Arghun in 1289 on a joint project to capture Damascus from the Moslem Mamelukes, and Prince Edward of England (later Edward I, 1274–1307) successfully sought an alliance with the Ilkhan Abaga when on crusade at Acre in 1271 (Sinor 1977: IX) although it eventually came to nothing.

It is these processes, the weakness and instability of settled populations in the face of new populations of migrants, that perhaps marks the greatest contrast between the world of the prehistoric and remote historic past and the present day. In the past, settled populations or their elites often could not withstand the migrants. As time as gone on, the settle populations have been more successful, and it is the less settled populations who have been marginalized. In the course of that process, migratory forces, at least of that kind, have ceased to exercise their once dominant role on human population. Instead, variation in vital rates arising from other more endogenous causes, notably the onset of the "demographic transition," has exercised the predominant effect on human population trends. That in turn has affected migration rates.

IX. EUROPEAN EXPANSION AND ITS CONSEQUENCES

The more recent European expansion is much better documented and needs little further description here. Much of it has been summarized in Cohen's magisterial encyclopedia (Cohen 1995) as well as in earlier works (e.g., Segal 1993; Chaliand et al. 1994). Although essentially another once-off process that has run its course, in small ways it grades into the present day. In some areas it is now reversed, with the return to Europe of settler populations ousted, often by violence or its threat, from Angola and Mozambique (Portuguese), North Africa (French, Italians, and Spanish), Central Asia, and the Baltic (Russians); a question mark hangs over the white populations of southern Africa. Other changes in population distribution and density arising from the European expansion are irreversible, notably the 300 million people mostly of European origin now settled in North America, Australia, and New Zealand, as well as many of the ancestors of a further 300 million people in Central and South America. Some of its earliest manifestations of European expansion, of Teutonic Knights into the region of the Baltic and Eastern Europe and of the return of Europeans to the shores of the Middle East and of North Africa in the 12th, 15th, and 19th centuries—with success that endured for a similar length of time—were the familiar form of armed invasion of relatively populous but (in the latter case) less-developed regions. This expansion coincided with, and was in some respects a component of, the rolling back of the nomadic conquests of previous centuries. In Spain it culminated in the final reduction of Granada, the last Moorish stronghold in Spain, in 1491. At the other end of Europe, in Russia, the removal of Tatar (Mongol) rule was confirmed by the capture of Kazan and Astrakhan from the Mongol Khans by Ivan the Terrible in 1552 and 1556. That was just the beginning of an arduous process of conquest and colonization that reached the shores of the North Pacific by the end of the 17th century but that did not finally occupy the central Asian homelands of the (former) nomads until 1873 (Linedy 1994).

Other aspects of European expansion, however, were of a quite different kind with little previous parallel (Scammell 1989; Pohl 1990), such as the transport over hitherto impossible sea distances, thanks to advances in navigation, of small but technically proficient populations, at times explorers, soldiers, adventurers, pirates, traders, settlers, refugees, and missionaries (Crosby 1986). The only partial comparison was with the Vikings of the 7th to 11th centuries and some earlier Mediterranean seafarers. But the latter, mainly interested in plunder and trade, respectively, did not enjoy overwhelming technical superiority against the peoples they encountered. Neither could they depend on reliable navigation. Their transatlantic adventure to Greenland and to "Vinland" failed. In Greenland, a long-term deterioration of climate during the Middle Ages may have been partially responsible for their eventual extinction. However, their transatlantic and global successors were able to overcome and displace simple societies living at a low population density and others of a more advanced and densely populated culture. In the earlier ventures to the New World from the 15th century, culture shock and above all the impact of the new diseases that came with the immigrants enabled them to displace aboriginal populations in part or whole, as well as to establish new populations of their own. In some cases (e.g., Tasmania, Argentina) a deliberate policy of extermination was initiated.

Once colonial or independent states were established in the Americas and the Antipodes, peaceful and regular migration out of Europe to these colonial New Worlds became the most salient global migration stream from the late 18th to the early 20th centuries. About 54 million people are believed to have migrated from Europe to the Americas from 1815–1930 alone (Baines 1991), although it must also be recalled that about one-third probably returned to Europe. This strand of migration continues, although at a much reduced level, and is overshadowed by the migration of populations from the Third World into North America and Western Europe. Meanwhile, in the "Wild East," substantial European population movements continued in an opposite direction. Migration to the "empty lands" of Siberia had only been encouraged from 1889, and freedom of "internal colonization" was granted only in 1906. Between 1906 and 1915, about 3.5 million Russians emigrated to Siberia. Elsewhere, the processes of voluntary and of forced migration within the Tsarist and then the Soviet Empires had created, by 1989, a position whereby 25 million people of Russian "nationality" lived in those parts of the former Soviet Union outside the Russian Federation; that is, not counting the migrants to Siberia, which is within that Federation. The return of some of these migrants or their descendants, some of it forced, had turned Russia in Europe's biggest net receiver of migrants by 1994.

In India, China, and other parts of the Orient, settled populations were already dense, and the main motive of initial European contact was trade (Emmer and Mörner 1992). By the 18th century, military superiority in weapons, discipline, and tactics enabled small European forces to overcome oriental dynasties that could field huge but ineffectual feudal armies. But despite conquest, no settlement was usually attempted in Asia or Africa except in Indonesia, central Asia and Siberia, southern Africa, and a few other places where climatic or demographic circumstances seemed propitious. In the tropical Old World, disease organisms generally favored the native, not the invading, populations (Curtin 1989, 1990). New navigation combined the ancient habit of forced labor migration through slavery brought a new component of population in the New World and an additional element of population destruction for Africa, in addition to the slave trade already organized for centuries to other parts of Africa and the Middle East.

Estimates of the number of slaves transported across the Atlantic vary from 3.5 to 25 million. A careful analysis suggests that 9.6 million were landed in the Americas between 1451 and 1870 (Curtin 1969), although the population loss to Africa must be inflated by the mortality at sea (variously estimated at 13% to 33%) and the direct and indirect mortality involved in the slave-taking process itself. This is only part of the damage to Africa as a whole, although the magnitude of the effects of slavery on African economy and society, previously taken for granted as highly damaging, has been questioned (Eltis 1991).

Slavery to other parts of Africa and the Middle East had continued from the earliest times and was formally suppressed progressively up the early 20th century. In the late 20th century, slavery was reported to be reviving in the Sudan and elsewhere. In the case of the transatlantic trade, for the first time since Roman times, slavery was primarily devoted to agriculture and manufacture. For the first time, it formed the basis of major intercontinental flows of money and goods (Lovejoy 1986; Solow 1991).

X. THE RISE OF "MODERN" FORMS OF MIGRATION

In the beginning, there was no labor migration. The nonmarket subsistence agriculture of traditional peasant economies did not need it. Specialist skills would later increasingly be concentrated in villages and towns, which implies migration. And once agrarian economies developed sufficiently to produce a division of labor greater than that between priests, rulers, warriors, and peasants, groups of people with special but often alien skills, in metalwork, textiles, and later finance were encouraged to enter the domain, often giving an ethnic component to the division of labor. In early economies, merchants too often had to be migrants, at least in the short term, but over long distances. Mercenaries, often migrants, are a special case, sometimes accepted into the society (as in the later western Roman Empire) in order to avoid having to fight them and at the risk of their overthrowing the regime they were paid to support (as in fifth century Rome and in seventh century Britain). Slaves might occasionally do the same, for example, the Mamelukes in Egypt (1250–1517) and the blacks in Haiti (1790–1804).

Most states in the past, and some to the present day, have been pronatalist and mercantilist in their approach to population. Explicit responses to that included encouragement of the birth rate, the inhibition of emigration, and the selective encouragement of immigration, including permanent immigration for settlement. This was above all the response of many of the European settlements in sparsely populated areas of the world overseas in the latter part of the second millennium, especially by countries in South America, North America, and Australasia. Only the latter three, and Israel for special reasons, continue such a policy today, although France has in the past had a more favorable attitude to the demographic consequences of immigration than has its European neighbors.

Agricultural labor migration remains undeveloped in an early subsistence economy, with its locally produced and sustained labor force, especially if the economy is not yet monetized. Voluntary labor migration becomes common as markets in labor develop on a seasonal and longer term basis within market-based agricultural economies (Coleman 1984). Mobility became the norm in the early modern period in Europe through such pervasive Western European practices as husbandry service (local not international) (Kussmaul 1981), through the hiring of adult labor among an increasingly landless population (Moch 1992), and as "internal" rural-urban migration (Wrigley 1967; Clark 1979) promoted by a more specialized and differentiated society. Elsewhere, especially where forms of serfdom were practiced that tied peasants to the land, the mobility revolution (Zelinsky 1971) comes much later. Labor movement does not become generally international until a much later date, particularly during the Industrial Revolution, although large populations left Europe in increasing numbers from the 18th century in pursuit of more general aspirations for a better life in a less rigid world—about 54 million crossed the Atlantic between 1830 and 1914. The growth and survival of cities, especially the larger cities and those in more recent periods, depended on constant positive in-migration streams to make good the natural loss of population arising from the usual excess of urban deaths over births (Landers 1987).

CONCLUSIONS

Human (international) migration has generally been conceptualized in modern demographic studies as being of relatively small numbers of people moving for reasons of labor and personal improvement of prospects, family connection, or asylum-seeking between or, at least into, established states. It has generally been considered to be a minor force in demographic change compared with vital processes. Although generally true at the present time, except for its increasing scale, that picture gives a partial and restricted view of the full range of human migratory possibilities. A much wider variety of forms of migration has existed over time, some of which are now extinct. A few were once-off and unrepeatable events. Although at the present time migration is usually regarded in the demography of large-scale populations as being a second-order and separate process compared with the prime demographic engines of fertility and mortality, this has not always been so. And it may not be so in the future in view of the potential transformation of low-fertility developed populations by immigration from third-world countries.

At various times in the past, particularly in the remote past and during the "migration period," migration has been primary in determining human population size and composition over large areas of the world's surface. The demands and conditions of a migratory life determined the disease and risk environment of human populations—for all of prehistory, of all human populations—and therefore the level and pattern of mortality and fertility of populations over than immense period of time. This is the risk environment to which the human species has adapted in the course of its evolution. The effects of the interaction between settled and nomadic populations for the past 4000 years has had substantial and usually with negative effects on the populations of the latter, particularly at the beginning of the period of urban settlement and later between the 3rd and the 14th centuries. Frequently during that time, the depredations of mobile populations on settled peoples put human population growth substantially into reverse over large areas of the world's surface. And although expansion of European populations out of the Old World eventually led to much larger populations in the areas in which they settled, the effects of the transported new diseases on the aboriginal populations were initially disastrous and the political, cultural, and social consequences endure to the present. Although migration again engages the attention of the industrial world's political systems, and for good reason, modern concerns are of a different, and in many respects of a lesser kind, than are those of the past.

References

Acsádi George Y. and Nemeskéri Janos, 1970. *History of human life span and mortality.* Budapest, Akadémiai Kiadó, 346 p.

Allan William D., 1967. *The African husbandman.* Edinburg, Oliver and Boyd, 506 p.

Ammerman Albert J. and Cavalli-Sforza Luigi Luca, 1984. *The Neolithic transition and the genetics of population in Europe.* Princeton, Princeton University Press, 176 p.

Baines Dudley, 1991. *Emigration from Europe 1815–1930.* London, Macmillan, 89 p.

Barfield, T.J. 1989. *The perilous frontier: nomadic empires and China.* Oxford, Basil Blackwell.

Bellwood Peter S., 1989. The Colonization of the Pacific: some current hypotheses, *in:* Adrian V.S. Hill and Susan W. Serjeantson (ed.), *The Colonization of the Pacific: a genetic trail*, p. 1–59. New York, Oxford University Press, 298 p.

Billewicz W.Z. and McGregor Ian A., 1981. The demography of two west African (Gambian) villages 1951–1975, Journal of Biosocial Science, vol. 13, p. 2192–2140.

Biraben Jean-Noël, 1979. Essai sur l'evolution du nombre des hommes, *Population*, vol. 34(1), p. 132–136.

Birdsell Joseph B., 1957. Some population problems involving Pleistocene man, *Cold Spring Harbor Symposia in Quantitative Biology*, vol. 22, p. 476–479.

Blurton-Jones Nicholas G., 1986. Bushman birth spacing: a test for optimal interbirth intervals, *Ethology and Sociobiology*, no. 7, p. 911–905.

Bowcock Anne M., Kidd Judith R., Mountain Joanna L., Herbert Joan M., Carotenuto Luciano, Kidd Denneth K., and Cavalli-Sforza Luigi Luca, 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms, *Proceedings of the National Academy of Science, USA*, vol. 88, p. 8398–8343.

Burns Thomas S., 1978. Calculating ostrogothic population, *Acta Antiqua*, vol. 26, p. 4574–4563.

Caldwell, J.C. and Schindlmayr T., 2002. Historical population estimates: unraveling the consensus. *Population and Development Review*, vol. 28(2):183–204.

Campbell Kenneth L. and Wood James W., 1988. Fertility in traditional societies, *in:* Peter Diggory, Malcom Potts, and Sue Teper (ed.), *Natural human fertility: social and biological determinants*, p. 396–399. London, Macmillan, 201 p.

Carneiro Robert L. and Hulse Daisy F., 1966. On determining the probable rate of population growth during the Neolithic, *American Anthropologist*, vol. 68, p. 1771–1781.

Cavalli-Sforza, L.L. (2000) *Genes, peoples and languages.* London, Allen Lane, The Penguin Press.

Cavalli-Sforza Luigi Luca, Menozzi Paola, and Piazza Alberto, 1994. *The history and geography of the human genes.* Princeton, Princeton University Press, 518 p.

Chagnon Napoleon A., 1974. *Studying the Yanomama.* New York, Holt, Rinehart and Winston, 270 p.

Chagnon Napoleon A. and Irons William (ed.)., 1979. *Evolutionary biology and human social behavior: an anthropological perspective.* North Scituate, Duxbury, 623 p.

Chaliand Gérard, Jan Michel, Rageau Jean-Pierre and Petit Catherine, 1994. *Atlas historique des migrations.* Paris, Le Seuil, 140 p.

Chapman Malcolm, 1992. *The Celts: the construction of a myth.* Basingstoke, Macmillan, 342 p.

Clark Peter, 1979. Migration in England during the late seventeenth and early eighteenth centuries, *Past and Present*, vol. 83, p. 579–580.

Cockburn Aidan, 1963. *Infectious diseases, their evolution and eradication.* Baltimore, Johns Hopkins, 420 p.

Cohen Robin (ed.), 1995. *The Cambridge survey of world migration.* Cambridge, Cambridge University Press, 570 p.

Cohen Mark N. and Armelagos George J. (ed.), 1984. *Paleopathology at the origins of agriculture.* New York, Academic Press, 615 p.

Coleman David A., 1984. Marital choice and geographical mobility, *in:* Anthony J. Boyce (ed.), *Migration and mobility*, p. 195–196. London, Taylor and Francis, 378 p.

Coleman, David A., 2003. Mass migration and population change. *Zeitschrift fär Bevölkerungswissenschaft* vol. 28(24), p. 1832–1815.

Coleman, David A. 2004. Partner choice and the growth of ethnic minority populations. *Bevolking en Gezin* 33 (2), p. 7–33.

Coleman, David A. and Rowthorn R., 2004. The economic effects of immigration to the United Kingdom. *Population and Development Review* 30(4), p. 5796–5722.

Council of Europe, 2003. *Recent demographic developments in Europe.* Strasburg, Council of Europe.

Crosby Alfred W., 1986. *Ecological imperialism: the biological expansion of Europe 9001–900.* Cambridge, Cambridge University Press, 368 p.

Curtin Philip D., 1969. *The Atlantic slave trade: a census.* Madison, University of Wisconsin Press, 338 p.

Curtin Philip D., 1989. *Death by migration: Europe's encounter with the tropical world in the nineteenth century.* Cambridge, Cambridge University Press, 251 p.

Curtin Philip D., 1990. *Disease and imperialism before the nineteenth century.* Minneapolis, Associates of the James Ford Bell Library, University of Minnesota, 16 p.

Diesner, H.-J., 1978. *The great migration.* Leipzig, Edition Leipzig.

Dunbar, R., 1996. *Grooming, gossip and the evolution of language.* London, Faber and Faber, 72 p.

Durand John D., 1974. *Historical estimates of world population: an evaluation.* Philadelphia, Population Studies Center, University of Pennsylvania, 72 p.

Dyson Tim, 1977. The demography of the Hadza-in historical perspective, *in: Seminar on historical demography (African historical demography)*, p. 1391–54. Edinburgh, Center of African Studies, University of Edinburgh, 2 vol., 468 p.

Eltis David, 1991. Precolonial Western Africa and the Atlantic economy, *in:* Barbara L. Solow (ed.), *Slavery and the rise of the Atlantic system*, p. 971–919. Cambridge, Cambridge University Press, 355 p.

Emmer Peter C. and Mörner Magnus (ed.), 1992. *European expansion and migration. essays on the intercontinental migration from Africa, Asia and Europe.* New York, Berg, 312 p.

Ferrill Arther, 1985. *The origins of war from the stone age to Alexander the Great.* New York, Thames and Hudson, 240 p.

Ferrill Arther, 1986. *The fall of the Roman Empire: the military explanation.* New York, Thames and Hudson, 192 p.

Fiennes Richard, 1978. *Zoonoses and the ecology of human disease.* London, Academic Press, 196 p.

Gamble Clive, 1994. *Timewalkers: the prehistory of global colonisation.*— Cambridge, Havard University Press, 309 p.

Harris Marvin and Ross Eric B., 1987. *Death, sex and fertility.* New York, Columbia University Press, 227 p.

Hassan Fekri A., 1973. On mechanisms of population growth during the Neolithic. *Current Anthropology*, vol. 14, p. 5355–5342.

Hausfater Gleen and Blaffer Hardy Sarah (ed.), 1984. *Infanticide comparative and evolutionary perspectives.* New York, Aldine, 598 p.

Heather Peter J., 1991. *Goths and Romans 3324–89.* Oxford, Clarendon Press, 378 p.

Henige, D., 1992. Standards of proof and discursive strategies in the debate over Native American population at contact, *in:* International Union for the Scientific Study of Population [IUSSP] (ed.), *The peopling of the Americas. proceedings of the IUSSP Conference Veracruz 1992*, p. 17–46. Veracruz, International Union for the Scientific Study of Population.

Howell Nancy, 1986. Feedbacks and buffers in relation to scarcity and abundance: studies of hunter-gatherer populations, *in:* David Coleman and Roger Schofield (ed.), *The state of population theory. Forward from Malthus*, p. 156–157. Oxford, Basil Blackwell, 311 p.

Krings Matthias, Geisert Helga, Schmitz Ralph W., Krainitzki Heike, and Pääbo Svante, 1999. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen, *Proceedings of the National Academy of Sciences*, vol. 96(10), p. 5581–5585.

Kuijsten Anton, 1995. The impact of migration streams on the size and structure of the Dutch population, *in:* Saskia Voets, Jeannette J. Schoorl, and Bart de Bruijn (ed.), *The demographic consequences of international migration*, p. 283–305. The Hague, NIDI, 430 p.

Kussmaul Ann, 1981. *Servants in husbandry in early modern England.* Cambridge, Cambridge University Press, xii + 233p.

Landers John, 1987. Mortality and metropolis: the case of London 1675–1825, *Population Studies*, vol. 41(1), p. 59–76.

Landers, J., 2003. *The field and the forge.* Oxford, Oxford University Press, 656 p.

Lee Richard B., 1972. Population growth and the beginnings of sedentary life among the Kung Bushmen, *in:* Brian Spooner (ed.), *Population growth: anthropological implications*, p. 329–342. Cambridge, MIT Press, 425 p.

Lincoln W. Bruce, 1994. *The conquest of a continent: Siberia and the Russians.* London, Jonathan Cape, 500 p.

Lovejoy Paul E. (ed.), 1986. *Africans in bondage: studies in slavery and the slave trade: essays in honor of Philip D. Curtin on the occasion of the twenty-fifth anniversary of African Studies at the University of Wisconsin.* Madison, African Studies Program, University of Wisconsin, 378 p.

Martin Paul S., 1973. The discovery of America, *Science*, no. 179, p. 969–974.

Massey Douglas S., Arango Joaquín, Hugo Graeme, Kouaouci Ali, Pellegrino Adella, and Taylor J. Edward, 1993. Theories of international migration: a review and appraisal, *Population and Development Review*, vol. 19(3), p. 4314–4366.

McEvedy Colin, 1967. *Penguin atlas of ancient history.* Harmondsworth, Penguin, 96 p.

McEvedy Colin and Jones Richard, 1978. *Atlas of world population history.* Harmondsworth, Penguin, 368p.

McHenry Henry M., 1994. Behavioural, ecological implications of early hominid body-size, *Journal of Human Evolution*, vol. 27(13), p. 77–87.

McKeown Thomas, 1990. *The origins of human disease.* Oxford, Basil Blackwell, 238 p.

McNeill William H., 1977. *Plagues and peoples.* Oxford, Basil Blackwell, 85 p.

McNeill William H., 1984. Human migration in historical perspective, *Population and Development Review*, vol. 10(1), p. 11–18.

Moch Leslie P., 1992. *Moving Europeans: migration in Western Europe since 1650.* Bloomington, Indiana University Press, 257 p.

Nei Masatoshi and Livshits Gregory, 1989. Genetic relationships of Europeans, Asians and Africans and the origin of modern *Homo sapiens*, *Human Heredity*, vol. 39, p. 276–281.

OECD, 2003. *Trends in international migration: continuous reporting system on migration. SOPEMI 2002.* Paris, OECD, 372p.

Pavlovsky Evgeny N., 1966. *Natural nidality of transmissible diseases, with special reference to the landscape epidemiology of zooanthroponoses.* Urbana (Illinois), University of Illinois Press, 261 p.

PCO China, 1987. *The population atlas of China.* Hong Kong, Oxford University Press, 217 p.

Peisker Théodore, 1911. The Asiatic background, *in:* John B. Bury, Henry M. Gwatkin, and James P. Whitney (ed), *The Cambridge mediaeval history. Volume 1: The Christian Roman Empire and the foundation of the Teutonic kingdoms*, p. 323–359. Cambridge, Cambridge University Press, 754 p.

Pohl Hans (ed.), 1990. *The European discovery of the world and its economic effects on pre-industrial society, 1500–1800.* Stuttgart, Franz Steiner Verlag, 330 p.

Richards Martin, Côrte-Real Melna, Forster Peter, Macauley Vincent, Wilkinson-Herbots Hilde, Demaine Andrew, Papiha Surinda, Hedges Robert, Bandelt Hans-Jürgen, and Sykes Bryan, 1996. Paleolithic and Neolithic lineages in the human mitochondrial gene pool, *American Journal of Human Genetics*, vol. 59, p. 185–203.

Salt John, 1989. A comparative overview of international trends and types, 19508–0, *International Migration Review*, vol. 23(3), p. 431–457.

Salt John, Singleton Ann, and Hogarth Jennifer, 1994. *Europe's international migrants. data sources, patterns and trends.* London, HMSO, 223 p.

Salt John, Clarke J., and Wanner P., 2004. *International Labour Migration.* Strasburg, Council of Europe, Population Studies no. 44.

Scammell Geoffrey V., 1989. *The first imperial age: European overseas expansion 1400–1715.* London, Unwin Hyman, 281 p.

Schiefenhovel Wulf, 1984. Preferential female infanticide and other mechanisms regulating population size among the Eipo, *in:* Nathan Keyfitz (ed.), *Population and biology*, p. 169–192. Liège, Ordina, 316 p.

Segal Aaron, 1993. *An atlas of international migration.* London, Hans Zell, 233 p.

Sellar W.C. and Yeatman R.J., 1930. *1066 and all that.* London, Methuen, 115 p.

Sinor, D., 1977. *Inner Asia and its contacts with medieval Europe.* London, Variorum Reprints, 392 p.

Solow Barbara L. (ed.), 1991. *Slavery and the rise of the Atlantic system.* Cambridge, Cambridge University Press, 355 p.

Stalker, P. 1994. *The work of strangers: a survey of international labour migration.* Geneva, International Labour Office, 327 p.

Stanley Neville F. and Joske R.A. (ed.), 1980. *Changing disease patterns and human behaviour.* New York, Academic Press, 666 p.

Stoneking Mark and Cann Rebecca L., 1989. African origin of human mitochondrial DNA, *in:* Paul Mellar and Chris Stringer (ed.), *The human revolution: behavioural and biological perspectives on the origins of modern humans*, p. 17–30. Princeton, Princeton University Press, 800 p.

Stringer Chris B., 1989. The origin of early modern humans: a comparison of the European and non-European evidence, *in:* Paul Mellar and Chris Stringer (ed.), *The human revolution: behavioural and biological perspectives on the origins of modern humans*, p. 232–244. Edinburgh, Edinburgh University Press, 800 p.

Stringer Chris, 2001. The evolution of modern humans: where are we now?, *General Anthropology*, vol. 7(2), p. 1–5.

Sutter Jean and Tabah Léon, 1951. La notion d'isolat et de population minimum, *Population*, vol. 6(3), p. 481–498.

Sykes Bryan, 1999. The molecular genetics of European ancestry, *Philosophical Transactions of the Royal Society of London. Series B Biological Sciences*, vol. 354(1379), p. 131–138.

Thornton R. 1997. Aboriginal North American population and rates of decline ca. A.D. 1500–1900. *Current Anthropology*, vol. 38(2), p. 310–315.

United Nations 1998. *Recommendations on statistics of international migration revision 1. statistical series M, no 58 rev. 1.* New York, United Nations.

Voets Saskia, Schoorl Jeannette J., and de Bruijn Bart (ed.), 1995. *The demographic consequences of international migration*. The Hague, NIDI, 430 p.

Weale M.E., Weiss D.A., Jager R.F., Bradman N., and Thomas M.G., 2002. Y chromosome evidence for Anglo-Saxon mass migration, *Molecular Biology and Evolution*, 19(7), p. 1008–1021.

Wobst H. Martin, 1974. Boundary conditions for Palaeolithic social systems: a simulation approach, *American Antiquity*, vol. 39, p. 147–178.

Wobst H. Martin, 1976. Locational relationships in Paleolithic society, *in:* Richard H. Ward and Kenneth M. Weiss (ed.), *The demographic evolution of human populations*, p. 49–58. New York, Academic Press, 158 p. [Reprinted in *Journal of Human Evolution*, 1976]

Wolpoff Milford H., 1989. Multiregional evolution: the fossil alternative to Eden, *in:* Paul Mellar and Christ Stringer (ed.), *The human revolution: behavioural and biological perspectives on the origins of modern humans*, p. 62–108. Edinburgh, Edinburgh University Press, 800 p.

Wormald Patrick, 1994. Engla Lond: the making of an allegiance, *Journal of Historical Sociology*, vol. 7, p. 1–24.

Wrigley Edward A., 1967. A simple model of London's importance in changing English society and economy, 16501–750, *Past and Present*, vol. 37, p. 44–70.

Zambardino R.A., 1980 Mexico's population in the 16th century: demographic anomaly or mathematical illusion? *Journal of Interdisciplinary History 11*, p. 1.

Zelinsky Wilbur T., 1971. The hypothesis of the mobility transition, *Geographical Review*, vol. 61(2), p. 219–249.

C H A P T E R

68

Europe's Demographic Transition, 1740–1940

JACQUES VALLIN

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

The Industrial Revolution had many antecedents, whose analysis lies outside the scope of this chapter: they include cultural emancipation and scientific revival in the Renaissance, the formation of powerful centralized states, the saga of great discoveries (the compass, printing, gunpowder, America, etc.), the accumulation of merchant wealth, and 18th century progress in agriculture. These economic and social changes in Western Europe also triggered another revolution: the *demographic transition*.[1]

For several hundred thousand years—as Jean-Noël Biraben shows in Chapter 66—mortality and fertility had been governed largely by natural factors. In the late 18th century, they came within the sphere of human control. Medical progress, as well as changes in attitudes and behaviors, radically transformed the old demographic regime. These new factors made it possible to combat disease and death efficiently and to control fertility. At the same time, they stimulated a massive demographic expansion. But, as they first occurred in Europe,[2] the Industrial Revolution and the demographic transition also caused a radical shift in the planet's geopolitical balance. They gave the European continent both the technological means and the demographic impetus that would enable it to dominate the world (Vallin 1986).

One of the last witnesses to the old demographic regime, as well as one of the first inventors of a general theory of population, was Thomas Malthus. Struck by the harshness of natural constraints on mankind, he believed that humans, as all living beings, tended to increase the population of their species beyond the level sustainable by the quantity of food that they were able to procure or produce (Malthus 1798). Thomas Malthus believed that economic development stimulated fertility, leading to excess population growth and inevitably a food crisis, hence excess mortality. To prevent such drastic adjustments, he proposed controlling the number of births by means of continence. To say the least, his suggestion has been overtaken by events, although the neo-Malthusian school admittedly gave new luster to his arguments later on. In the 2 centuries from 1750–1950, the population of Europe (with first the Russian Empire and then the Soviet Union included) rose from 146 million to 573 million; moreover, it contributed massively to demographic

[1] The first author to have described the process in depth named it the "demographic revolution" (Landry, 1934). However, it is now universally referred to as the "demographic transition," a term borrowed from postwar U.S. authors (Davis 1945; Notestein 1945).

[2] Albeit not throughout the continent or in a consistent pattern in all European countries. As discussed below, this major historical process actually began in northwest Europe (England, France, Scandinavian countries, Netherlands) before spreading to central Europe (Germany, Austria, Northern Italy) and then finally to Mediterranean Europe (Spain, Portugal, Italy, Greece) and eastern Europe (Poland, Romania, Russia, etc.).

TABLE 68–1 Growth of World Population by Major Region, 1500–1950

Region	1500	1600	1700	1750	1800	1850	1900	1950
Population size (million)								
Europe and ex-USSR	84	111	125	146	195	288	422	573
North America	3	3	2	3	5	25	90	172
Latin America	39	10	10	15	19	34	75	167
Oceania	3	3	3	3	2	2	6	13
Africa	86	114	106	104	101	103	118	221
Asia (not including Soviet Union)	243	339	436	504	646	791	902	1377
Total	458	580	682	775	968	1243	1613	2522
Proportional distribution (%)								
Europe and ex-USSR	18.3	19.1	18.3	18.8	20.1	23.2	26.2	22.7
North America	0.7	0.5	0.3	0.4	0.5	2.0	5.6	6.8
Latin America	8.5	1.7	1.5	1.9	2.0	2.7	4.6	6.6
Oceania	0.7	0.5	0.4	0.4	0.2	0.2	0.4	0.5
Africa	18.8	19.7	15.5	13.4	10.4	8.3	7.3	8.8
Asia (not including Soviet Union)	53.1	58.4	63.9	65.0	66.7	63.6	55.9	54.6
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Sources: Biraben [see Chapter 66]; United Nations 2001.

growth in North America and Australia–New Zealand, whose combined population went from 6 million to 185 million in the same interval (Table 68–1). Nor should we forget Europe's more modest but not negligible contribution to the settlement of many other regions of the world, most notably Latin America.

True, the first shock that followed the conquest of America was, if anything, negative. Latin America, in particular, had suffered devastating population losses owing not only to the massacres but also to the microbial shock[3] and the collapse of pre-Columbian civilizations. Authors differ,[4] but, according to the estimate given by Jean-Noël Biraben in Chapter 66 and reproduced here, the population of Latin America was divided by 4 between 1500 and 1600, decreasing by almost 40 million to 10 million. Although far more sparsely inhabited, North America underwent a roughly similar fate in the 17th century. America did not regain its 15th century population size until the second half of the 19th century. Yet, thanks to European emigration (see Chapter 58, Volume II), it has since registered an extraordinary boom.

In Africa, population decline owing to the European shock and, in particular, to the slave trade persisted until the 19th century. From 1600–1800, the population fell an estimated 10%, from 113 million to 102 million. The numbers started rising again slowly until the mid-20th century.

In contrast, Asia (in fact, mostly China) experienced robust population growth between the 16th and 18th centuries. Its share of world population expanded from 53% to 64%. Since then—with European domination a later and more incomplete phenomenon than in Africa and America—the Asian population has grown steadily but at a slower pace. From 1750–1950, it rose from 500 million to 1.377 billion.

In the same period, the demographic equilibrium between continents shifted unquestionably to the benefit of Europe. In 1750, Asia held two-thirds of the world population (65%) and Europe less than one-fifth (19%). Africa had most of the rest (13%), because America and Oceania were very sparsely inhabited at the time, with less than 3% of the total.

By 1900, a century and half later, Asia's share had dropped to 55% and Europe's share had climbed to 26%. North America, a true outgrowth of Europe, now represented 5.6%, and Latin America, also heavily dependent on European emigration, accounted for 4.6%. In contrast, Africa's share had fallen to 7.3%.

What had happened? First, Europe had experienced an unprecedented long-term mortality decline, which paved the way for steady, accelerating population growth. True, the mortality downturn had been followed by a restriction of fertility that eventually

[3] The importation of new germs by the Spanish conquistadors devastated the American Indian population, cut off from the rest of the world for millennia.

[4] Colin McEvedy and Richard Jones (1978), for example, estimated that the American Indian population fell by only 25% and not by almost 75%, the figure chosen by Jean-Noël Biraben.

slowed population growth and led to the emergence of a new demographic regime with low mortality and low fertility. In most cases, however, the timing difference between the mortality and fertility downswings has generated unprecedented excesses of population over a variable but usually rather long period. This large-scale historical trend radically altered Europe's position vis-à-vis the rest of the world. Initially named the "demographic revolution" by Adolphe Landry (1934), it spawned an abundant literature that ultimately gave birth, after World War II, to the theory of "demographic transition"—a major explanatory framework for the macro-demographic phenomena of the second half of the 20th century.[5]

Although this momentous historical trend clearly originated in Europe in the mid-18th century, its point of arrival is harder to identify (assuming it even exists, as we shall see in the chapters that follow). However, the present chapter discusses only the European transition. Its spread to the rest of the world is examined in the next chapter, and, for clarity's sake, I prefer to give precedence to the historical division and conclude this chapter where the following one begins, using World War II as the cutoff point. Indeed, it was largely after World War II that the transition phenomenon spread to the rest of the world; it is also fairly clear that the postwar baby boom marked a turning-point in the process in Europe. Consequently, although all the changes described were far from complete by 1940, all the charts illustrating the phenomenon here will stop (provisionally) in that year.

I. AN UNPRECEDENTED MORTALITY DECLINE

Humans have always sought to avoid disease and death, but for millennia, their struggle for survival was relatively inefficient and their life expectancy at birth very short. As far back as evidence exists for their behavior, humans seem to have consistently made a rather conscious effort to vanquish disease and delay death. Initially, of course, they did so by trying to make arrangements with the gods or the forces of evil; with the accumulation of knowledge, however, they gradually turned to seeking medications that would improve the efficiency of prayers and spells—and eventually replace them. Did this millennial quest have any effect on the mortality level before the 18th century? Presumably not, as Jean-Noël Biraben notes in Chapter 66. In contrast, the second half of the 18th century saw the start of an unprecedented rise in life expectancy in northwest Europe.

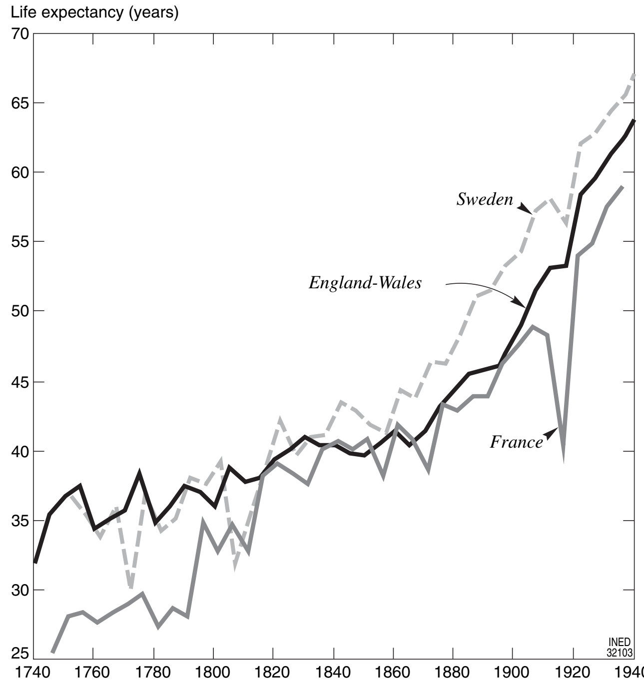


FIGURE 68–1 Change in life expectancy at birth in Sweden, England-Wales, and France from the mid-18th century to the eve of World War II (by 5-year period). (From: England-Wales, Wrigley and Schofield 1981; Sweden, Holmberg 1970; France, Vallin and Meslé 2001.)

1. Two Centuries of Spectacular Progress

Figure 68–1 plots life expectancy at birth over 2 centuries, from 1740–1940, in Sweden, England-Wales, and France, countries for which we have the longest complete statistical series on the subject.[6] Life expectancy at birth nearly doubled from 25 to 35 years in 1740 to 55 to 65 years in 1940.

The starting point of this momentous change is partly in dispute. It is much lower for France than for the two other countries in Fig. 68–1. In fact, the Swedish data are based solely on recorded deaths: however high the quality of the Swedish registers even at the start of the period, we can assume that a percentage of deaths—particularly of infants—went unrecorded. For England-Wales and France, on the contrary, we have time-series reconstructions by historical demographers. The English estimates also incorporate a backward projection from later data by Edward Wrigley and Roger Schofield (1981), whereas

[5] It cannot, however, claim to explain everything. In particular, the baby boom of the 1940s to 1960s was no doubt largely driven by other factors.

[6] Thanks to early recording of current statistics in Sweden and historical reconstructions in France (Henry and Blayo 1967; Blayo 1975) and England (Wrigley and Schofield 1981).

the French estimates are based on the late-1950s study by Louis Henry of a national sample of parish registers (Blayo 1975). It is surprising to find so wide a divergence in the results of these studies for two countries that shared so many common economic, social, and political features at the time. The studies suggest a life expectancy of only 24.8 years in France in 1740–1749, versus 35.3 years in England-Wales in 1746—a gap of more than 10 years. France did experience several years of food shortages in the 1740s.[7] Indeed, in the far more clement 1750s, French life expectancy rose to 27.9 years, but this figure remained well below that of England-Wales, which also increased to 37.3 years in 1756, preserving a differential of nearly 10 years. There are two conflicting explanations of this (Vallin 1989c, 1992). Some authors—most notably those of the English reconstruction—argue that health conditions genuinely differed between the two countries. True, the climate may have played a role. England's prevailing oceanic climate may have given it an advantage not extended to France, whose climate is more continental and Mediterranean. Jean-Noël Biraben and Louis Henry (1957) have shown, for example, that infant mortality was higher in the Mediterranean countries than in the Scandinavian countries because of the risk of infection owing to the hotter summers, which promoted the development of germs in children's food. One can also argue that France was more densely inhabited than was England-Wales in the 18th century and thus closer to the maximum population compatible with its carrying capacity. Although this situation may have prompted couples to limit their births, it initially, no doubt, contributed to worsening the population's health conditions.[8] But some of the difference is also likely owing to conflicting estimation biases. The English estimate probably underestimates mortality.[9] Louis Henry adjusted France infant and child mortality heavily to allow for underrecording of infant deaths, whereas British authors seem to have paid less attention to this issue (Henry and Blanchet 1983); they may also have overlooked the size of the migration-induced bias in the backward projection.[10] Conversely, the adjustment of French data for underrecording of infant deaths may have been exaggerated. It is thus highly probable that life expectancy has been underestimated in France and overestimated in England-Wales. This does not in itself, of course, invalidate the thesis of a real gap. Other examples of pretransition differences have been observed between European countries and, most importantly, between regions. In the early 19th century, even before the start of the large-scale mortality decline, life expectancy in Italy (Breschi and Livi Bacci 1986; Breschi et al. 1994; Del Pant 1998) and Spain (Dopico and Reher 1998) exhibited sizable regional differences of about 10 years. It is usually argued that the lesser demographic pressure in some regions enabled them to enjoy higher life expectancy. One of the likely effects of this lesser pressure was that mortality crises were less frequent and less severe there, as shown by the comparison between France and England in Fig. 68–2. Yet life expectancy at birth was no doubt of about 30 years in the 1740s, probably somewhat less in France and somewhat more in England-Wales.

A few decades later, the three countries (Sweden, England-Wales, and France) actually had far closer life expectancy values of 41.7, 39.5 and 40.7 years, respectively. Sweden and England-Wales subsequently outpaced France. By 1935–1939, French life expectancy, at 58.9 years, lagged almost 8 years behind Sweden's 65.6 years, with England-Wales in an intermediate position at 62.3 years. However, as we shall see, the explanation lies in the countries' different histories.[11] Yet, despite these notable distinctions, the three countries—and probably also a few others whose past is less precisely documented, such as Denmark, Norway, and the Netherlands—spearheaded the major advances in health conditions that began in the mid-18th century and gradually spread to all of Europe.

2. Attenuation of Fluctuations

This historical progress involved several phenomena. At the outset, the most salient was certainly the attenuation of wide fluctuations. We lack sufficiently precise data to track earlier annual variations in mortality through the life-expectancy indicator. However, the crude death rate makes a perfectly adequate proxy, given the minimal change in the age distribution from year to year (see Chapters 3, 15 of Volume I). Figure 68–2 shows just how abrupt those swings still were. In France mid-18th century, for example, the annual

[7] In this context, a life expectancy of 25 years is hardly implausible. Compare this with the average 20 years estimated by Kingsley Davis (1951) for 19th century India.

[8] "France was as full as an egg and its inhabitants poorly nourished" (Chesnais, Chapter 119 of Volume IV).

[9] See, in particular, Ronald Lee's critical analysis (1993) of the backward extrapolation method used for England.

[10] "From the early 16th to late 18th centuries, English trans-Atlantic emigration totaled about 7000 a year, out of an average population of 4 million" (Livi Bacci 1997)—an outflow that, over the long run, can have sizable effects on population and mortality estimates.

[11] Special emphasis should be placed here on the rise of alcoholism in France and the spread of tuberculosis.

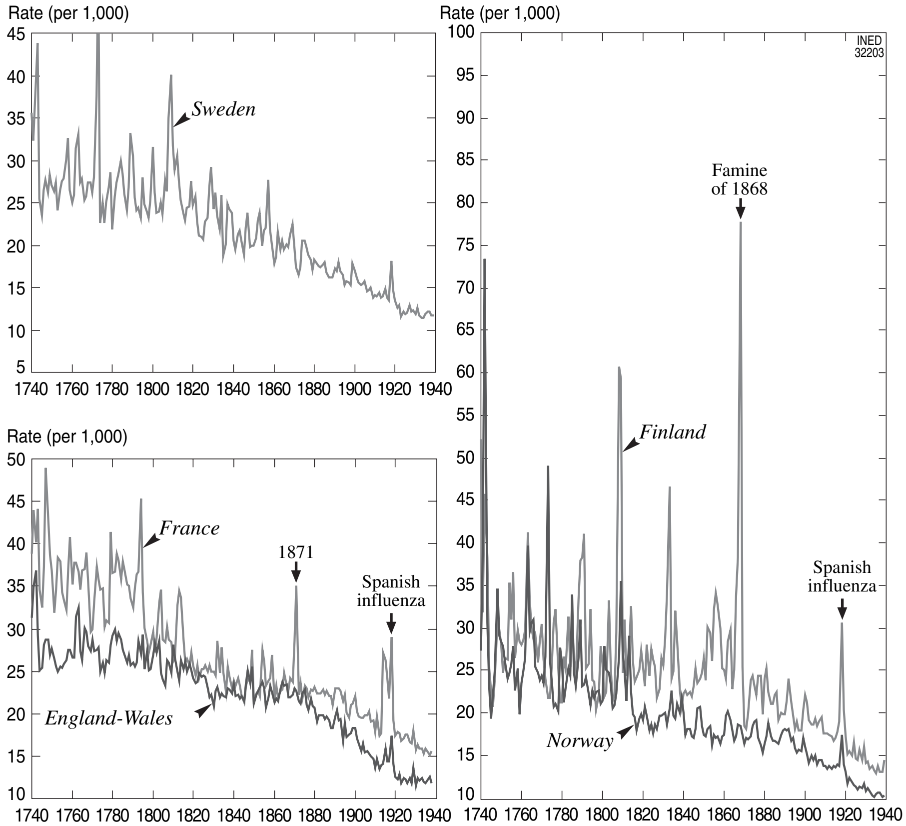


FIGURE 68–2 Annual change in crude death rate in Sweden, England-Wales, France, Norway, and Finland, 1740–1939. (From Mitchell 1975; Wrigley and Schofield 1981.)

death rate varied between 30 and 50 per 1000. The fluctuations were smaller in England-Wales. Should this be seen as a corroboration of the effects of a milder and more regular climate or as the sign that England-Wales managed to control its major crises even earlier than did France? Or—a third alternative—is the difference an artificial consequence of the reconstruction method? The issue cannot be settled conclusively. In contrast, it is clear that, whatever the initial size of the fluctuations, they had been substantially dampened in less than a century. By the mid-19th century, they were very modest in both countries; had France not experienced the Franco-Prussian War of 1870 (aggravated by the Paris Commune and the siege of the capital), World War I, and the 1918 Spanish flu to revive their memory, the fluctuations would have almost entirely disappeared after the mid-19th century. In less populous Sweden and Norway, the fluctuations were even sharper in the 18th century. In the worst years, the crude death rate could rise to 60 to 70 per 1000. In both countries, the easing of the crises in the 19th century was all the more spectacular as a result. The attenuation was all the more pronounced as neither country took part in the Franco-Prussian War of 1870 or World War I, and both were far less severely hit by the Spanish flu than was France.

The case of Finland is slightly different, offering partial evidence of the size of the timing lags between European countries in their march toward better health. Finland long remained a poor agrarian country, supplying labor to the Swedish economy. Its mortality was very high until the end of the 19th century. As late as 1868, it suffered a major famine (Pitkänen 1993) that drove up the crude death rate to 80 per 1000—a fairly late reminder of the severity of earlier large-scale crises. The four other countries experienced food shortages rather than famines in the 18th century and after: the severe famines of the past were not repeated, but it has never been possible to measure the exact impact on mortality. At a time when its national death register was already almost comprehensive, Finland shows us the potential consequences of such crises: the loss of nearly one-tenth of the population in a year! However, in Finland as elsewhere, albeit with a lag of at least half a century, the fluctuations ceased after this last major crisis; by comparison, the 1918 influenza 50 years later seemed just a pale replica (although Finland was harder hit than were Sweden and Norway).

Thus, between the mid-18th and mid-19th centuries, northwest Europe rid itself almost entirely of two leading causes of short-term mortality crises: epidemics and famine. As we saw, full-scale famine had nearly vanished (Finland, on this criterion, clearly fit the eastern European pattern) and the most recent food shortages were less damaging. These had been conclusively eliminated thanks to progress in agriculture, better transportation and communication, and the abolition of internal borders (most notably, a freer circulation of grain). As regards epidemics, although the plague had been almost totally contained by the early 18th century, it was soon replaced by cholera, particularly in the 19th century, which in turn was superseded by influenza in the 20th. But, whatever their infectious agents, the severity of health crises steadily diminished. Step by step, before even identifying the true causes and action mechanisms of the crises, the health authorities learned to take increasingly effective protection measures (such as quarantine) and to stop epidemics ever more swiftly. Influenza, the last great epidemic disease (before AIDS), never killed as many people—even in 1918—as did the plague and cholera in earlier times.

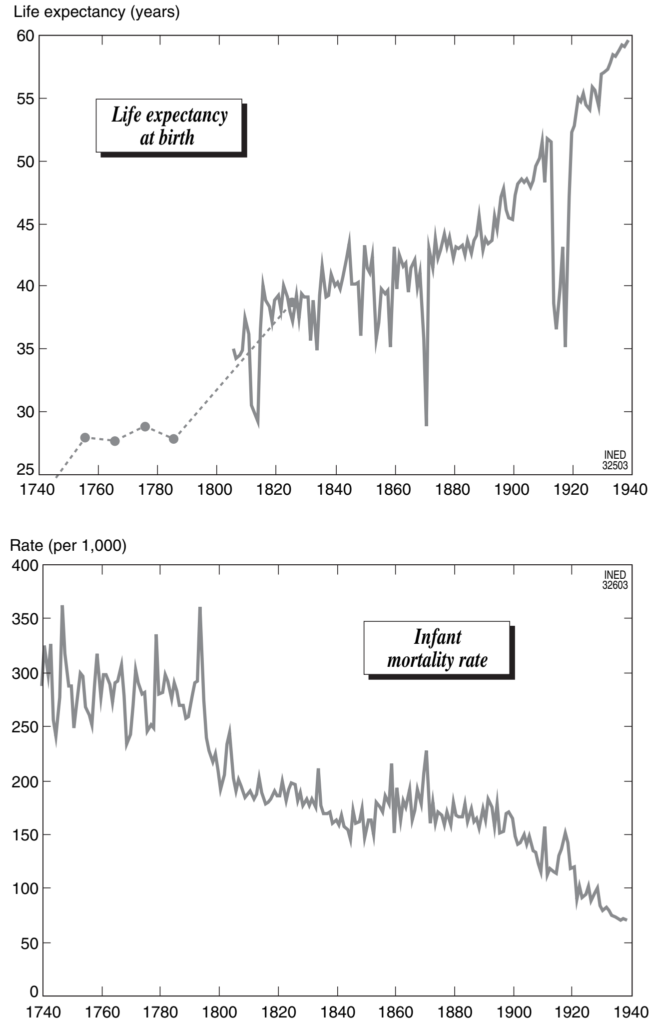


FIGURE 68–3 Annual changes in life expectancy at birth and infant mortality in France, 1740–1939. (From: 18th century, survey by Louis Henry [Blayo 1975]; 19th-20th centuries, Vallin and Meslé 2001.)

3. Progress Phases and Stagnation Phases

Despite its sizable contribution to reducing the average death rate, the smoothing of fluctuations would not have sufficed to generate the spectacular uptrend in life expectancy discussed at the start of this section. There were virtually no further occurrences of excess-mortality peaks in crisis periods. But mortality in prosperous years declined substantially as well. In France, for example, in the best years of the 18th century, the crude rate exceeded 30 per 1000. In 1769, the best year of the Ancien Régime, it was 29.5 per 1000; it did not dip slightly under 28 per 1000 until 1798. By the 1930s, it had fallen to 15 to 16 per 1000. But it should be recalled that the French population had aged sharply in the meantime, and this powerfully inhibited the fall in the crude rate (see Chapter 4 of Volume I). Annual estimates of life expectancy for France are available only from 1806 on (Vallin and Meslé 2001). Nevertheless, whereas life expectancy (for both sexes) was still only 37.4 years in 1810—the best of the earliest documented years of the 19th century—it had reached 59.5 years by 1939. This represents a gain of over 22 years in 130 years, not including the already considerable progress achieved before 1810 (Fig. 68–3). The magnitude of the gains in non-crisis periods is even more visible if we look at infant mortality rates, available on an annual basis since the mid-18th century. Although these rates reflect only one aspect of total mortality, they are independent of age distribution. In 1769, the best year of the Ancien Régime, infant mortality was still at 232 per 1000 (Fig. 68–3). By the late 1930s, it had fallen more than threefold to under 70 per 1000. Because of population aging, the crude death rate had been only halved.[12]

However, the drop in mortality was not achieved in a single step. In France, over the 2 centuries that have seen demographic transition in the classic sense, two stages of major progress were separated by a phase of stagnation, if not regression (Fig. 68–3). This phenomenon, already perceptible in the curve of the crude death rate, is very clear from the data on life

[12] However, aging does not explain the entire difference, which is also owing to the change in the age distribution of mortality itself.

expectancy at birth and infant mortality (Fig. 68–3). The first phase of decisive progress runs from the 1780s to the 1840s: life expectancy rose from under 29 years to more than 40 (disregarding the gain recorded between the 1740s and 1750s, attributable mainly to the fact that the first decade saw more food shortages than did the second, rather than to a genuine mortality decline outside of crisis periods).

Taking a closer look, we see that this gain, in turn, was achieved in two stages. First, excluding the darkest years of the Napoleonic wars (1812, 1813, and 1814), the French Revolution and the First Empire brought dramatic life-expectancy gains.[13] By 1816, life expectancy stood at almost 40 years. History textbooks generally emphasize the heroic or deadly aspects of the period. They are less likely to discuss the huge economic, social, and cultural progress accomplished in the same years, which largely contributed to the improvement in health conditions for the French. This was followed by a brief interlude of stagnation, or even mild decline, during the Restoration, before life-expectancy gains resumed under the July Monarchy. By 1845, the mean length of life peaked at over 43 years.

More than 40 years would then be needed before the start of a second major phase of progress in the late 1880s. Meanwhile, life expectancy had lastingly stalled or even declined, in particular during the Second Empire. France underwent a series of changes that anchored the country to the industrial era but had a broadly negative impact on health conditions. The 1850s, 1860s, and 1870s saw new crises, particularly outbreaks of cholera and, of course, the Franco-Prussian War of 1870 (above all, the siege of Paris and the repression that followed the Commune). But, even in the good years, life expectancy declined slightly. In 1867, for example, a rather positive year for the decade, life expectancy was under 42 years, almost a year and a half less than in 1845. Despite a mild increase at the fall of the Second Empire, life expectancy in 1887, at 43.5 years, had just barely caught up with its 1845 level.

[13] One of the major drivers of this improvement came, in fact, from England: "vaccination," i.e., immunization against smallpox through the inoculation of exudate of vaccine (a cow disease similar to smallpox), invented by Edward Jenner in 1796. Inoculation—borrowed from the Chinese—was a slightly older method but rather dangerous as administered with smallpox itself. Vaccination spread fairly quickly in France and explains some of the rise in life expectancy in the early 19th century, but the method was not extended to other diseases until the late 19th century, after the work of Louis Pasteur. This role of medicine in the initial fall in mortality has been vigorously challenged by Thomas McKeown (1976). But McKeown's arguments have, in turn, sparked extensive controversy (Vallin 1989c; see Chapter 57).

As discussed earlier, the fall in infant mortality played a major role in improving life expectancy in these various stages (see Chapter 57, Volume II). The second chart in Fig. 68–3 effectively shows the close fit between the successive periods of rising life expectancy at birth and the declines in infant mortality. However, some variations in infant mortality were more pronounced than were those in life expectancy. In particular, the long break in life-expectancy gains in the period around the Second Empire was clearly a regression as regards child survival. Setting aside the terrible year 1871, the infant mortality rate climbed steadily between the 1840s and 1870s—from 144 per 1000 in 1845 to 190 per 1000 in 1867 and almost 180 per 1000 in 1880. Despite undeniable economic progress, child-raising conditions worsened for many French families in the second half of the 19th century (Rollet 1990, 1997). With industrialization, a growing share of the population left the countryside to swell the ranks of the urban proletariat, whose working, housing, and living conditions undermined their life expectancy. This period saw urban mortality overtake rural mortality more than ever before. In addition, lack of time to raise their own children forced many working-class mothers to place them with nurses, which correspondingly diminished their chances of survival. This phenomenon is the main cause of the resurgence of infant mortality during the Second Empire.

Decisive progress did not resume until Pasteur's discoveries on the microbial origin of infectious diseases and the mechanisms for inhibiting their effects. From the late 1880s to the eve of World War II, life expectancy rose from 43 years to almost 60. These brisk gains were halted only by World War I, whose impact was heavily compounded by the Spanish flu in 1918. But Pasteur's discoveries alone would not have sufficed; they occurred in an economic and social context that promoted their dissemination and application. During the Third Republic, the social struggles against the perverse effects of industrialization began to bear fruit. Work laws were introduced, and the Villermé Report (1840) fostered greater concern for public health. In particular, this was the period in which large-scale sanitation infrastructure projects were realized in the major cities—from the improvement of sewer systems to the construction of drinking-water distribution networks. The late 19th century also saw the introduction of compulsory, secular, and free public schooling with the Jules Ferry education laws. Schoolteachers—the "hussars of the Republic," in the phrase of writer Charles Péguy (1913)—played a key role in furthering not only formal instruction and human and civic training but also hygiene and health education. This widespread diffusion of basic knowledge (from medical congresses to compulsory schooling and information conveyed in the media) was one of the main reasons why the Pasteurian revolution was so easily incorporated into personal habits and stimulated the life-expectancy gains observed between the 1890s and the eve of World War II.

The rise in life expectancy did not cease in 1940. On the contrary, it resumed vigorously just after World War II. After the first vaccines and the sulfonamide vogue of the 1930s, France was able to take Pasteur's ideas to their logical conclusion thanks to the discovery of antibiotics and the establishment of general social-insurance systems. Indeed, on the mortality side, there was no intrinsic change in the drivers of life-expectancy gains until the late 1960s (see Chapter 57 of Volume II). It is the developments on the fertility side that offer a more decisive explanation of why the profile of European demographic transition given here stops in 1940. Suffice it to say that in France, as in the other Western countries, life expectancy is still a long way from peaking (see Chapters 48, 57 of Volume II), although its recent trends are now shaped by factors that lie outside the scope of the most classic model of demographic transition, to which the present chapter is deliberately restricted.

4. Timing Differences

The above reminder of the main stages in the improvement of French public health is just one of several examples. Each country has had its own history, with accelerations and slowdowns linked to economic and social developments. But even more striking than the distinctive features of each national profile is the steady spread of this health revolution throughout Europe, which resulted in major timing differences between countries.

As noted earlier, the three pioneering countries—Sweden, England, and France—did not make gains at the same pace. France began by closing what was its partly genuine gap with the first two countries. It was then overtaken again, most vigorously by Sweden but also by England-Wales. More generally, it took at least a century for the progress initially achieved in northwest Europe to spread to the central, southern, and eastern parts of the continent. Figure 68–4 charts life expectancy at birth and infant mortality in selected countries. Owing to the lack of data, not all the curves go back to the 18th century—far from it—but the leads and lags are obvious.

In four of the countries shown—Netherlands, Italy, Spain, and Russia—the known life-expectancy path

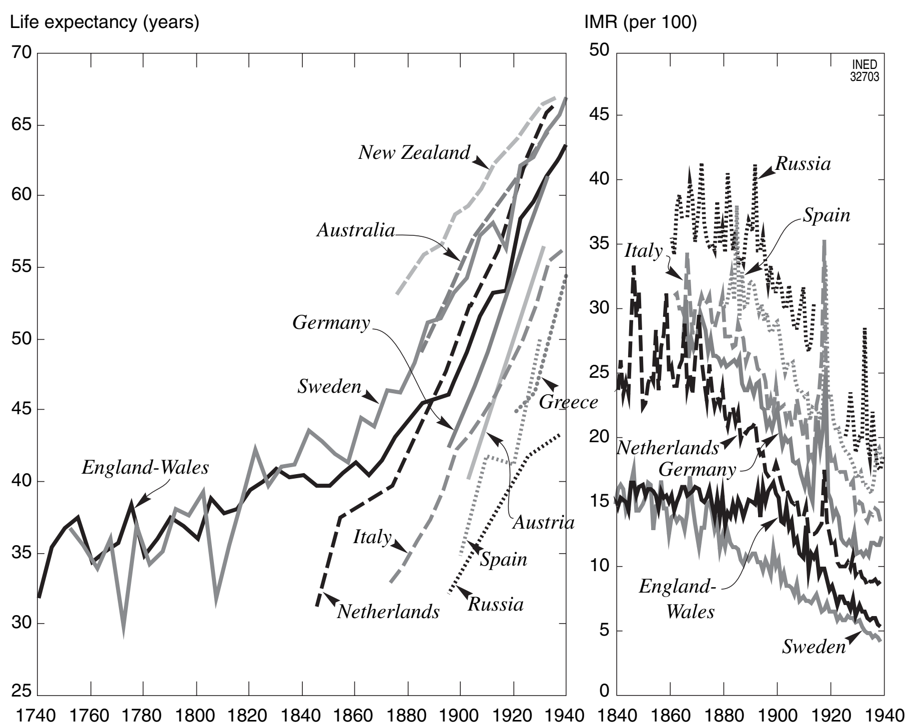


FIGURE 68–4 Changes in life expectancy at birth, 1740–1939, and in infant mortality rate, 1840–1939, in selected European countries. (From Vallin 1989b, 1991; Mitchell 1975.)

starts at about 30 years, which is comparable to the level prevailing in the three pioneering countries in the mid-18th century. If we consider the beginning of these paths as the starting points of the health revolution, the latter appears to have begun at about 1840 in the Netherlands, about 1880 in Italy, and about 1900 in Spain and Russia. From the more partial curves for Germany, Austria, and Greece, we can easily surmise that their starting points must have been interspersed with those of the previous four. This is confirmed, incidentally, by the section of Fig. 68–4 on infant mortality, for which we have a greater number of complete series since the mid-19th century.

We also note that, in many cases, gains were achieved faster in the countries where they began later. The most obvious case is the Netherlands, whose life expectancy exceeded Sweden's on the eve of World War II. Likewise, Germany caught up with England. More modestly, Italy, for example, closed the life-expectancy gap with Sweden from 15 years in about 1880 to 11 years in the late 1930s. Russia did not fit this pattern in the period charted in Fig. 68–4, but it did so after World War II. In fact, as described in Chapter 55 of Volume II, all European countries converged to very similar levels by the mid-1960.

5. Conjunction of Positive Factors

This convergence of European countries toward new health conditions characterized by a massive decline in infant mortality and infectious mortality (see Chapter 57 of Volume II) results from a conjunction of several positive factors. As we have seen in the case of France, these were linked to the dramatic changes brought by the Industrial Revolution in European countries. Although medical advances were critical, they were far from being the only driver. Despite the controversies over the issue, medical progress before Pasteur's discoveries, although significant, does not appear to given much momentum to the improvement of survival rates. Initially, it was by improving their political and administrative organizations that European countries were able to contain major epidemics and—if not to cure the victims—at least to protect the uncontaminated and prevent the spread of diseases. Famine and food shortages were eliminated thanks to progress in agriculture and to more efficient commercial distribution of foodstuffs through better transportation and storage and the abolition of obstacles to free circulation. Lastly, the initial decline in ordinary infectious mortality, i.e., in noncrisis years—was mainly owing to the construction of urban sanitation infrastructure (sewers and drinking-water networks).

That is no reason to deny the decisive role of the Pasteurian revolution, which, in turn, enabled many earlier medical advances to achieve their full effects. For example, surgery, which had taken huge technical strides forward since the age of Ambroise Paré (about 1509–90), often remained powerless for lack of resources to prevent postoperative infection. The discovery of the principles of asepsis dramatically broadened surgery's scope of action and its chances of success. More generally, vaccines, sulfonamides, and antibiotics paved the way for a near-eradication of infectious mortality. But here again, as we saw with France, medicine alone would probably have had scant impact if social change had not provided the momentum needed to disseminate its benefits: the growth of education systems, the implementation of social-insurance systems, better living standards and housing conditions, environmental control, and so on.

II. A STEEP FALL IN FERTILITY

The modernization of European societies not only revolutionized health conditions and caused a massive drop in mortality but also led to a near-total control of fertility. The decrease in mortality, most notably infant mortality, allowed more children to survive until the reproductive ages and eliminated the need to maintain high fertility for generational replacement. This created an opportunity for a parallel decline in fertility wholly consistent with macrodemographic logic. However, the decision to have children is not directly determined by official authorities. It is taken by individual couples, who act in accordance with their outlook and views about the family and procreation, even if these are inspired by collective norms and are informed by environmental constraints. It is therefore equally possible to imagine situations in which couples continue to have as many children despite lower mortality or, on the contrary, restrict their fertility for reasons other than the mortality decline. These behavioral differences have momentous consequences on populations, as we shall see later. Let us simply note here, at the outset, that most European countries experienced a fertility decline after a mortality decline.

1. Decline in Total Fertility Rate and Completed Fertility

Figure 68–5 plots the total fertility rate (TFR)[14] observed since the mid-18th century in the three coun-

[14] The TFR, the sum of the age-specific fertility rates, is an indicator of the period level of fertility (see Chapter 14 of Volume I).

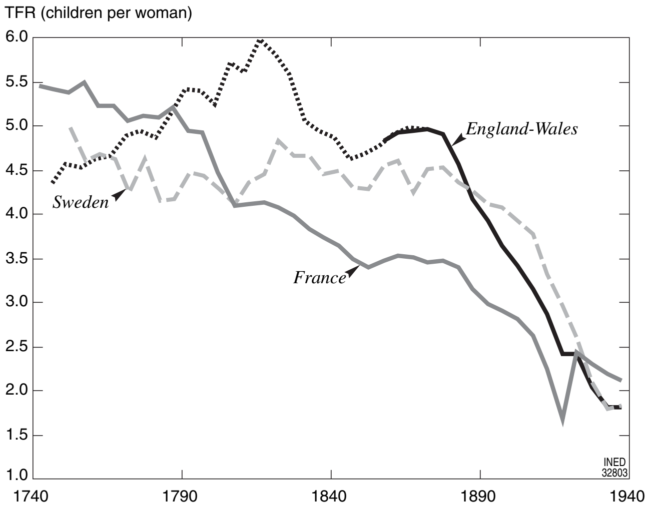


FIGURE 68–5 Changes in total fertility rates in England-Wales, France, and Sweden, 1740–1940.

tries mentioned earlier that first registered a mortality decline: England-Wales, France, and Sweden.

In France, Louis Henry's survey results suggest a fertility of about 5.5 children per woman in the first half of the 18th century.[15] The rate seems to have been significantly lower in Sweden (about 5 in 1750–55) and England-Wales (4.3 in about 1746). Again, one can debate whether these measures are fully comparable, but England does appear to have registered a genuine increase in fertility in the second half of the 18th century, as agricultural progress and the new jobs created by the early Industrial Revolution enabled young people to marry sooner and thus to have more children (see Chapter 119 in Volume IV) (Deane and Cole 1964; Kindleberger 1964). In any event, from this relatively high level, France also experienced the earliest and most sustained fertility decline. Its TFR had dropped to 4 children per woman in the early 19th century (3.96 in 1805–10) and had fallen below 2.5 on the eve of World War I (2.23 in 1910–15). In Sweden as well, fertility apparently started to fall by the second half of the 18th century, incidentally matching French fertility in the early 19th century; however, it then rebounded almost to its initial level in 1820–25 (4.82 children per woman) and stayed fairly high until the late 19th century (4.51 in 1875–80). By then, the gap with France stood at more than one child per woman. In fact, it was only in the late 19th century and even more so in the early 20th century that Swedish fertility dropped sharply; by the eve of World War II, it had fallen below that of France, at 1.79 versus 2.09 for the period 1935–40.

Almost the same pattern was repeated in England-Wales, where, as in Sweden, fertility fell much later but also much faster than in France. In the 1870s, it was still nearly 5 children per woman (4.88 in 1875–80), but then it plunged to exactly the same level as Sweden's on the eve of World War II (1.79 in 1935–40).

In fact, France is a fairly exceptional case. It is the only major country where the long-term fertility decline began in the 18th century and persisted without a break. Everywhere else, this trend never truly started until after 1870, as in Sweden and England, or, even more often, in the early 20th century (Fig. 68–6a). The probable starting points are the 1890s in Denmark, the 1900s in Norway and the Netherlands, and the 1910s in Finland.

However, it is hard to give a full picture of these trends, as the distribution of births by mother's age—an element needed to compute TFRs—has not been available for a sufficiently long period. In the 1870s, births by mother's age were compiled only in Scandinavian countries (for France and England, they have been reconstructed); in many European countries, they did not begin until the 1920s or 1930s (Festy 1979). We have fuller information on completed fertility, thanks to census questions on women's past fertility. By using the data developed by Patrick Festy (1979), we obtain Fig. 68–6b for completed fertility, which is fuller than is Fig. 68–6a.

Belgium and Switzerland rank between Sweden and England-Wales, but in almost all the other countries the fertility downtrend began later. This is especially true of the two Mediterranean countries represented: Italy and Spain. The decline was delayed even further in the central/east European countries (data not shown). In these agrarian countries on the Mediterranean and eastern rim of the Industrial Revolution heartland, fertility did not start to decrease until later.

In sum, although the fertility decline (as the mortality decline) was common to all European countries, it did not occur everywhere simultaneously. Even if we overlook France's exceptionally early start, the downtrend began at dates ranging from the 1870s to the 1920s. On this criterion as well, the demographic transition was slow in spreading to all European countries.

2. Immediate Determinants of the Fertility Decline

Demographers conventionally distinguish between two fertility regimes: *natural fertility* and *controlled fertility*, the latter being defined by the use of methods

[15] No reconstruction for all of France has yet been published, but Henri Leridon (personal communication) estimates the TFR at 5.56 in 1700–1740. An indirect estimate by Patrick Festy from birth rates gives 5.43 for 1740–1744.

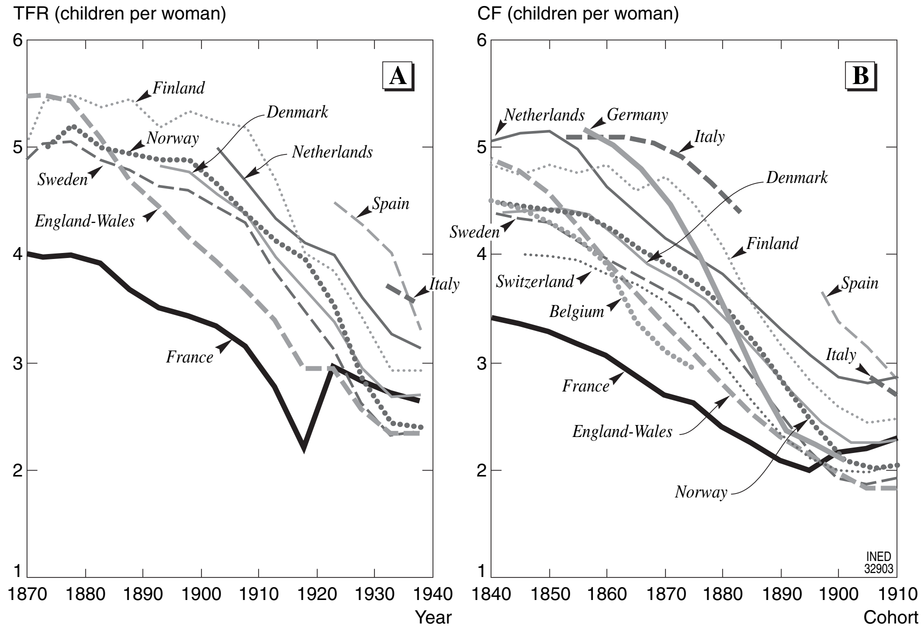


FIGURE 68–6 Changes in total fertility rate (TFR), 1870–1940 (A), and completed fertility (CF), 1840–1910 (B), in selected European countries. (From Festy 1979.)

explicitly aimed at restricting births. It is the shift from the first to the second that is largely responsible for the long-term decline in European fertility. However, the determinants of natural fertility, which are not exclusively biological, are themselves susceptible to change and contribute to the trend shifts. The extraordinary paradox of the fertility decline is precisely its occurrence at a time when the evolution of these natural-regime determinants should have triggered a steep rise.[16]

a. Increase in Reproductive Capacity

But that should not come as a surprise. A fundamental determinant of the fertility downtrend—albeit not the only one—was the decline in infant and youth mortality. This was achieved thanks to the improvement in children's living and survival conditions, i.e., progress in nutrition and health. Unsurprisingly, these same advances had a positive impact on the decline in causes of sterility and in fetal mortality. But beyond these effects of medical progress itself, other factors—including behavioral patterns—are also responsible for the rise in reproductive capacity.

Assuming a woman were to live in a union from puberty to menopause and refrain from all contraception, five biological factors would determine the number of her children: the duration of the *childbearing period* between puberty and menopause, the couple's *fecundity* (reproductive capacity), *fecundability* (probability for a woman living in a union of conceiving during each menstrual cycle), *intrauterine mortality* (miscarriages and stillbirths), and *nonsusceptible periods* (pregnancy and time before resumption of ovulation) (see Chapters 30, 31 of Volume I). At least four of these factors have moved in a direction that promotes an increase in reproductive capacity.

First, apparently because of gains in nutrition, the mean age at puberty has fallen from about 16 years in around 1750 to approximately 13 years today in Western societies (Chapters 30, 31 of Volume I). In the same interval, the mean age at menopause has risen significantly, from about 40 to 47 to 48 years. The theoretic childbearing period has thus presumably lengthened from 25 to 35 years, a 40% increase! At the same time, thanks to better healthcare and hygiene, as well as a reduction in the frequency and seriousness of spousal diseases, both sterility (primary or acquired) and intrauterine mortality have fallen. Lastly, and perhaps most important, although the length of pregnancy has not changed, the mean period of postpartum temporary infertility has dropped sharply. The two basic reasons for this are very different but are both related to changing behavior. The growing use of bottle-feeding has reduced the influence of temporary sterility owing to breast-feeding, whereas the weakening of sexual taboos has allowed an ever-earlier resumption of sexual intercourse after a birth. The

[16] As we shall see in Chapter 69, some developing countries have registered such uptrends at the start of their fertility transitions.

duration of the postpregnancy nonsusceptible period has shorted considerably—from 1 to 2 years in premodern Europe to 1 or 2 months in the industrialized countries. We know less about the fifth factor, fecundability, which depends on the frequency of sexual intercourse, its closeness to ovulation, and the quality of the ovum–sperm encounter. It may have changed hardly at all from its earlier value of about one chance in four in each cycle.

On balance, the theoretical reproductive capacity has increased significantly. Putting together the available estimates for the five parameters, the imaginary woman living in a union from puberty to menopause would have had roughly 10 children in around 1750. Today, she would have had 15—but, in Europe, the gain was nearly achieved by the eve of World War II. However, there is more to the story.

b. Lengthening of Childbearing Period

In premodern Europe, *fertility outside marriage*[17] was statistically negligible owing to social conventions and strict social control of female sexuality. As a result, this theoretical reproductive capacity effectively depended on the actual duration of marriage. Its complete realization required very early marriage (before puberty) and its nontermination by divorce or widowhood before menopause. The fulfillment of either condition was exceptionally uncommon.

In mid-18th century France, for example, age at first marriage for women was relatively late, at about 26 years, and a significant proportion of women—about 10%—never married. Although divorce was forbidden, de facto separation was not; most important, widowhood was very frequent, and remarriage was neither the rule nor very rapid. As John Hajnal has shown (1965, 1982), this model of late marriage with a high proportion of never-marrieds was typical of premodern Western Europe. In some population groups, the rate of exclusion from sexuality was as high as 50% of women of reproductive age. The period in which fertility could express itself was, on balance, rather short. However "natural" it may have been—in the sense of being uninhibited by effective contraception—fertility was rather limited. A woman surviving to at least age 50 had only 5.5 children on average, compared with the 9.8 she would have had if married from ages 15 to 50. In other words, the mere fact that women in that period spent a significant part of their reproductive years outside marriage nearly halved the number of children they could bear. This vividly illustrates the extent to which marriage and widowhood, combined with strict control of female sexuality, could regulate the number of births in a natural-fertility regime.

By way of comparison, in otherwise fairly similar conditions (no contraception, no fertility outside marriage), the fertility of marriages in Algeria in about 1970 (9.5 children per married woman between ages 15 and 50) was more or less the same as in France in 1750, but overall fertility was sharply higher at 7.5 versus 5.5 children per woman. The reason is that women in Algeria first married at a much younger age (18 years, on average) and the proportion of never-marrieds was far lower (1%); at the same time, despite many terminations through repudiation, the time lost outside marriage was much shorter thanks to the lesser frequency of widowhood (lower mortality) and the speed of remarriages. In consequence, the natural potential was reduced by only one-quarter rather than one-half.

On the eve of World War II, if they had not voluntarily limited their births, French women would have had an even higher overall fertility. Although not as precocious as their Algerian counterparts of the 1970s, they did marry much sooner[18] than their 18th century ancestors, and their risk of widowhood before age 50 was almost negligible. Although fertility outside marriage was still socially condemned and uncommon, the average number of children would have been close to 9—almost as much as an 18th century woman could have expected had she lived in an uninterrupted marriage between the ages of 15 and 50.

c. The Start of Transition: Control through Marriage

In reality, the restrictive marriage regime prevailing in 18th century France—characterized by a late mean age at marriage and a high proportion of never-marrieds—was a recent innovation. At the start of transition, it replaced the previous traditional pattern of early and near-universal marriage in France sooner than elsewhere, but it eventually spread throughout Europe. As a result, fertility control was initially achieved through marriage control. One could call this the Malthusian phase of transition, for although Thomas Malthus (1798) had forcefully advocated birth control, he never recommended contraception.

This first stage of the fertility regime, intuited by Adolphe Landry (1934) in the interwar period, was confirmed by the large-scale study by the Office of Population Research at Princeton, under the leader-

[17] We use this term in the manual in preference to the classic term illegitimate fertility (see Chapter 29 of Volume I).

[18] For example, the French female cohort born in 1931 married at an average age of 23.1 years (see Chapter 25 of Volume I), compared with 26 for women born in the mid-18th century.

ship of Ansley Coale (Coale and Watkins 1986). To analyze the change in fertility in 700 European regions and provinces, the authors developed three now famous indicators: *If*, *Ig*, and *Im*.

The method is benchmarked to the fertility of the Hutterites, a Central European Protestant sect that had taken refuge in North Dakota and still refrained from contraception in the 1950s. The following indicators were developed:

If—Index of overall fertility, the ratio of total observed births N to the number that would be observed by applying to women F_x of age x the age-specific rates of Hutterite marital fertility[19] f_x:

$$If = \frac{N}{\sum f_x F_x}$$

Ig—Index of marital fertility, the ratio of observed births in marriage Nl to the number that would be observed by applying to women M_x of age x the age-specific rates of Hutterite marital fertility f_x:

$$Ig = \frac{Nl}{\sum f_x M_x}$$

Im—Index of proportion married, the ratio of (1) births that would be observed by applying Hutterite marital fertility to married women only to (2) births that would be observed by applying Hutterite marital fertility to all women:

$$\mathrm{Im} = \frac{\sum f_x M_x}{\sum f_x F_x}$$

If there are no births outside marriage, $N = Nl$ and so, by definition, $If = Ig \cdot Im$. If fertility outside marriage was not negligible, one could introduce an adjustment to allow for the variable.

The values of the three indices obtained for Russia, France, and Ireland in the early 20th century are summarized in Table 68–2.

At the time, Russia was still in a typical pretransition stage in which high marital fertility and high nuptiality combined to produce an even higher overall fertility. In France, in contrast, the age at marriage had fallen substantially—albeit not to Russian levels—and

TABLE 68–2 Indices of Overall Fertility, Marital Fertility, and Nuptiality in Russia, France, and Ireland in the Early 20th Century

Country	*If*	*Ig*	*Im*
Russia, 1897	0.546	0.773	0.687
France, 1901	0.238	0.383	0.543
Ireland, 1901	0.235	0.706	0.325

Source: Coale and Watkins 1986.

therefore had a less restraining effect on overall fertility. However, the latter was twice as low as in Russia, for contraception was already widespread, sharply reducing marital fertility. In Ireland, instead, overall fertility nearly matched France's by then, whereas the very low nuptiality rate kept marital fertility almost as high as Russia's. Ireland found itself in very special circumstances, because it had shifted to a new fertility regime after the Great Famine of 1845–48; in the early 20th century, it adopted an extreme version of the principle of fertility control through marriage that had prevailed throughout most of Europe at the start of the fertility transition but had later been abandoned rather quickly. France, in contrast, had by then fully converted to the controlled fertility regime—even in marriage—thanks to contraception.

d. Controlled Fertility: Abortion and Contraception

By the second half of the 18th century in France, later in the Scandinavian and Anglo-Saxon countries, toward the end of the 19th century in the other European countries, and even more recently in the central/east European countries and especially in Russia, fertility control through marriage gave way to control through contraception and abortion: *Malthusianism*[20] was followed by *neo-Malthusianism*. This new school of thought, founded by Francis Place, drove the 19th century rehabilitation of practices hitherto reserved for prostitutes, as well as the development and dissemination of new methods such as the condom and the diaphragm—most notably in the Scandinavian and Anglo-Saxon countries. In France, however, where these developments occurred far earlier, the spread of contraception relied exclusively on the method least dependent on technological progress: coitus interruptus (withdrawal). This remained the only method effectively in use until the advent of the pill and the loop in the 1960s.

[19] Although the authors use the terms legitimate fertility or legitimate births—in keeping with the universal practice of the time—to denote fertility and births in marriage, we prefer in this manual to speak of fertility and births in marriage (as indicated in footnote 17, and as in Chapter 29 of Volume I). Although arguably less concise, the term is more respectful of the rights now granted to unwed mothers and their children.

[20] Apart from late marriage, the only birth-control method contemplated by Thomas Malthus (a Protestant minister) was sexual continence, also accepted by the Catholic Church. But it was never actually practiced.

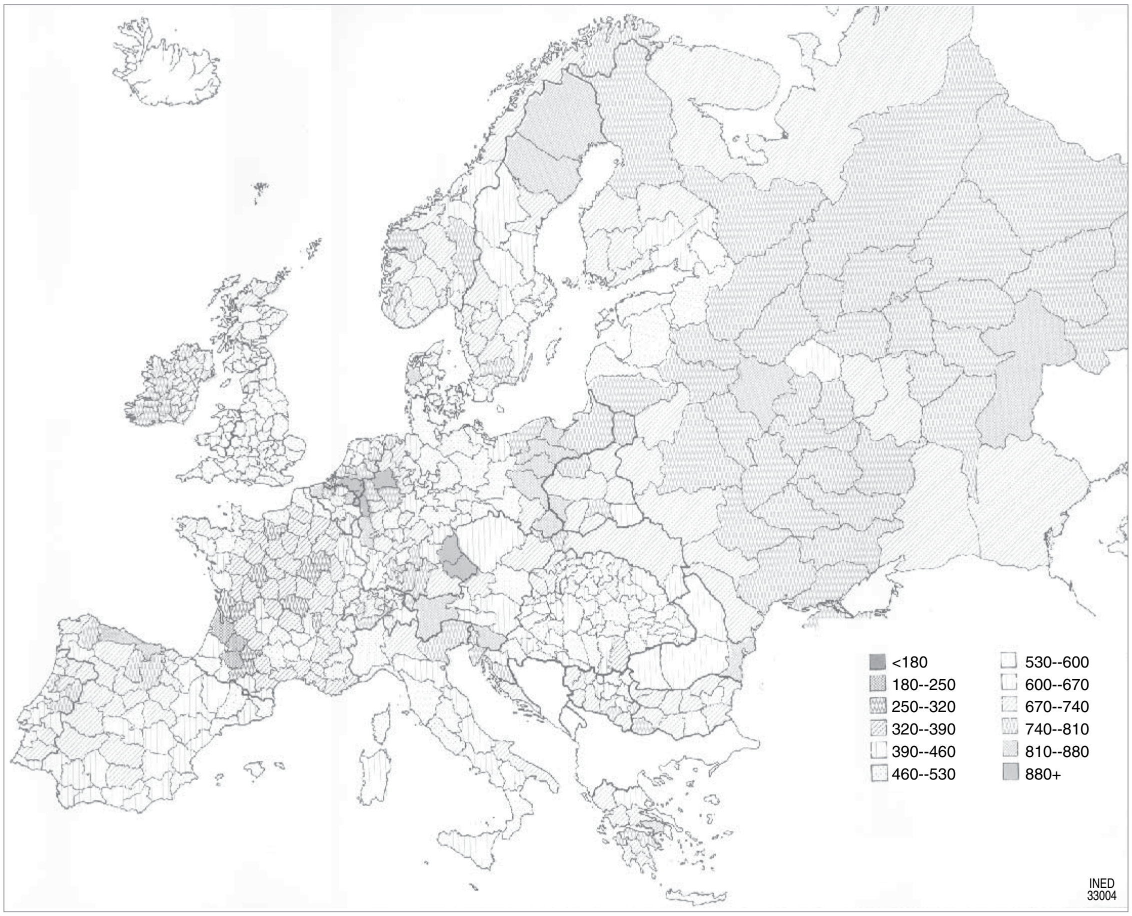


FIGURE 68–7 Values of *Ig* index (marital fertility) by European provinces,around 1900. (Source: Coale and Watkins 1986.) (Figure 68–7 also reproduced in color plate section.)

Abortion, an age-old technique, was the focus of moral reprobation and religious prohibition for even longer than was contraceptives. Indeed, until very recently, it was a penal offense in almost all countries. Although this never prevented clandestine abortions, they were severely restricted by the difficulties and risks involved, not only in the legal sphere but also in medical terms: illegality sometimes encouraged quack doctors and often meant appalling medical conditions. The first country to legalize abortion was the Soviet Union in 1920.[21] It was followed after World War II by the now Communist countries of central/eastern Europe and, in 1948, by Japan, which was struggling with the economic problems created by its defeat. Hardly surprisingly, these are the countries in which abortion—even today—is the most common birth-control method. Elsewhere, even after legalization, abortion is only a last resort to remedy the failures of other, increasingly effective contraception techniques.

[21] At least until 1936, when Stalin abruptly back-tracked on this liberalization.

Each country thus shifted to the new controlled-fertility regimes at its own pace and following a distinctive path (Fig. 68–7 illustrates this diversity by showing the values of the *Ig* index in the early 20th century in Europe's different regions). The new regimes drastically reduced the number of children per woman. Control via marriage had enabled French women to cut their biological potential from about 10 to 5.5. In the 20th century, birth control allowed them to reduce a realistic potential of about 9 to 2, based in turn on a biological potential of nearly 15.

3. Key Factors in the Fertility Decline

Although birth control has followed different courses according to time and place, its dissemination has never been primarily determined by the development of reasonably efficient techniques. The best proof of this, no doubt, is provided by French couples. They were the first to embark on this adventure, with coitus interruptus as their only resource. The key factor is couples' intentions, as effectively demonstrated by the pioneering Indianapolis survey[22] (Kiser and Whelpton 1953). Identifying how these intentions were expressed does not suffice to understand why they have changed.

The basic idea in the transition theory formulated by Adolphe Landry (on which more is below) is that the decision to limit the number of births is a response to the decrease in infant mortality, which enables a growing proportion of newborns to survive and thus to form the family. This reaction is lagged for two reasons. First, the individual awareness of the lasting decline of a phenomenon as random as infant mortality is not immediate. Second, and more important, the adoption of the new reproductive behavior runs counter to an age-old prejudice in favor of high fertility—a prejudice so strongly anchored in people's minds as to be the object of a cult in nearly all traditional societies. Taking this postulate as the starting point, the transition process can apparently move ahead unimpeded.

As always, reality is more complex. To begin with, the precedence of the decline in infant mortality is not an absolute rule. In France, in particular, the fertility decline was more or less concomitant[23] with the mortality decline. But the Princeton study cited earlier (Coale and Watkins 1986) also showed that a more detailed breakdown by region identifies cases in which fertility fell before mortality. In contrast, the economist Sydney Coontz (1957) suggests an additional cause for the lag. He argues that industrialization, by triggering greater labor demand, kept fertility high in areas where the drop in infant mortality should have made fertility decline sooner.

But, more generally, the downtrends in mortality and fertility take hold amid a modernization of society with many elements that are liable to influence both trends. The economic factors were emphasized in particular by Ansley Coale and Edgar Hoover (1958), who took up the arguments of Frank Notestein and other authors: the change in production structures leads to (1) the decline of the family as unit of production, (2) the emergence of impersonal systems of job allocation, and (3) the participation of women in economic activity outside the home. These phenomena tend to reduce the economic advantages of large families and to increase the need for mobility, which is easier to satisfy with a small family. The child, ever less capable of being incorporated into a productive family activity, becomes an ever greater burden.

At the frontier between the economic and social spheres, John Caldwell (1982), after other authors (see, in particular, Bertillon 1911), would generalize the latter notion by identifying the changes in parent–children transfers as a key factor in the fertility decline in Third World countries.

But cultural transformations, largely induced by economic change, probably had a more decisive effect. The diffusionist model (see Chapter 39 of Volume I) was described by Dudley Kirk (1946, 1996), who illustrated his theory with detailed maps of Europe and explained its mechanisms; it has since been the topic of many research studies. There is an abundant literature on the influence of education and women's emancipation—whether to explain the differences currently observed in the developing countries or the trend in European fertility in the 19th and in early 20th centuries. For ultimately this is a sector in which the diffusion of ideas played a central role. And it is certainly no coincidence that the new reproductive behaviors appeared in France in the century of the Enlightenment, the *philosophes*, and the Revolution. Some researchers have, in fact, attempted to measure the extent to which the diffusion of contraception was informed by the *secularization* of society, in other words, the emancipation of individuals from religious authorities and the moral imperatives imposed by the latter. In France, for example, the most Catholic regions continued for decades to practice the old regime of late marriage combined with near-natural fertility (as did Ireland), whereas birth control and earlier marriage were quickly adopted by the de-Christianized regions (Bardet 1988). Elsewhere in Western Europe, other authors have sought to relate the fertility decline in the late 19th century to a second wave of secularization, in which the social programs of new political groups faced a late offensive from traditional religion (Lesthaeghe and Wilson 1986). On this point, the reader will find it useful to refer to Chapter 38 of Volume I.

Not only is the fertility decline shaped by multiple factors but, as Étienne van de Walle writes (1998):

> The analysis of the Princeton project monographs enables us . . . to glimpse, in each country's "provinces," one or more specific factors apparently linked to the emergence of birth

[22] Survey conducted in Indianapolis in 1941 on the psychological and social determinants of fertility. Analyzed in 33 reports, of which the final one, published in *Population Studies* in 1953, summarizes the findings of the entire series.

[23] The concomitance was not absolute, because mortality gains had been achieved by the early 18th century. Jean-Claude Chesnais (1986), for example, dates the start of the mortality decline to 1715 and that of the fertility decline to 1780. In France, however, the gap between mortality and fertility rates has never been very significant.

control: the pattern of employment and female economic activity (Germany, Switzerland, Belgium), urban residence (Italy), education (England), real-estate ownership (France), etc. There is no unique combination of characteristics for modernization, but sets of factors that are different and sometimes contradictory from country to country, reflecting their particular circumstances. However, while socio-economic factors and fertility do not always interact in the same way, the fertility decline is a general phenomenon.

I would add that one always finds couples' intentions, not technology, at the center of the process—an observation that was not negated in the posttransition era with the advent of modern contraceptives.

4. A Phase of Exceptional Growth

As mortality and fertility declined between the 18th and 20th centuries, European populations shifted to a new demographic regime. The old combination of high mortality (life expectancy of about 30 years) and high fertility (TFR of about five children per woman) gave way to a new regime with low mortality (life expectancy of over 60 years) and low fertility (TFR of about two children per woman). This large-scale historical movement was first called *demographic revolution* by Adolphe Landry, then *demographic transition* after World War II. Its first consequence, as noted in the "Introduction," was a massive increase in the European population owing to the time lag between the mortality and fertility declines. This lag and its effect have been modeled, and the resulting diagram serves as the initial basis for the expression "demographic transition." The theory that bears the same name seeks to rationalize the terms of the model. But, before the theory is even applied, the model itself comprises variants. Not only does the timing differ from one population to another, but so do the scale of the gap (between the declines in mortality and fertility) and the length of the transition. This naturally entails different consequences on population dynamics.

5. Birth Rate, Death Rate, and Rate of Population Increase

To grasp this, we need to move away from specific measures of mortality and fertility such as life expectancy and the TFR. We must return to the effective incidence of these phenomena on population growth via the crude death and birth rates. Although it represents in many ways a borderline case in Europe, we can take the example of England, where the changes in these parameters between 1760 and 1940[24] illustrate the demographic-transition diagram perfectly (Fig. 68–8).

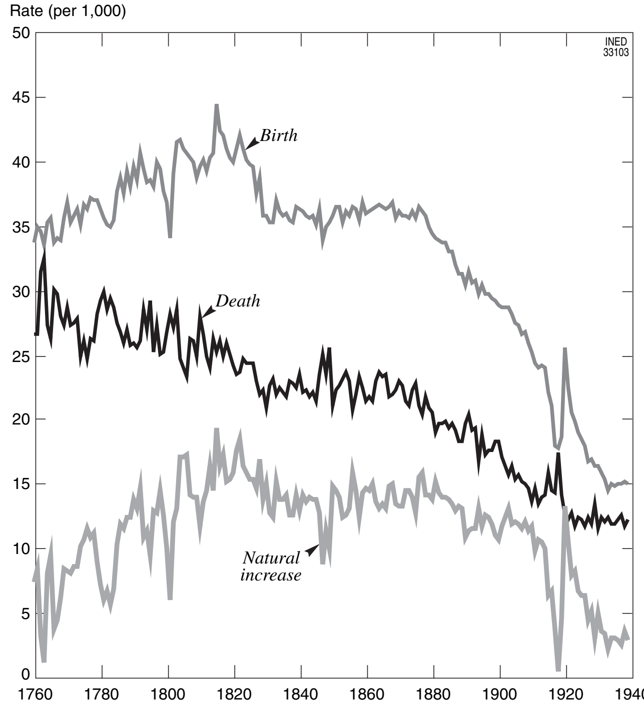


FIGURE 68–8 A typical example of demographic transition: crude birth rate, death rate, and natural increase in England-Wales, 1760–1940. (From: to 1860, Wrigley and Schofield 1981; since 1860, Mitchell 1975.)

Toward 1760, ignoring small fluctuations, the birth rates and death rates converged closely at a high level of about 30–35 per 1000. This resulted in a weak annual natural increase of around 5 per 1000. Throughout the 1760–1940 period, the death rate fell almost continuously (disregarding short-term fluctuations). In contrast, the birth rate rose from 1760 to 1820 for various reasons—most notably the decline in the age at marriage[25]—after which it plunged to its 1760 level; however, it stayed there for half a century (1830–80) before the start of its lasting decrease, driven by the historic decline in fertility. By 1934, the birth rate had sunk below 15 per 1000, a level at which—in those days—it seemed bound to remain.[26] The figure was barely higher than the level reached contempora-

[24] However, in England-Wales before 1760, the changes in birth and death rates were anything but monotonous.

[25] Anthony Wrigley and Roger Schofield (1981) estimate the mean age at first marriage for women at 26.5 years in 1650–99, 26.2 years in 1700–49, 24.9 years in 1750–99, and 23.4 years in 1800–40. The turning point of the 1800s was therefore still characterized by a long-term downtrend in the age at marriage.

[26] We know today that events turned out quite differently, but this is, of course, what an observer in the early 1940s would have projected.

neously by the crude death rate (slightly above 12 per 1000). Consequently, by the late 1930s, the natural increase had slipped again to a mere 3 per 1000 or so.

Meanwhile, however, England-Wales had enjoyed nearly a century and a half of outstanding demographic growth. The rate of natural increase accelerated swiftly in the late 18th century, first from 5 to 10 per 1000 in the 1780s, then to 15 in the 1800s, and even approaching 20 per 1000 in the early 1820s. More generally, it remained almost consistently above 10 per 1000 from 1775–1914 (barring a few exceptional years) before retreating rapidly in the interwar years. For over a century (1790–1900), it even exceeded 13 per 1000 almost without a break—no doubt for the first time in history. It is this long period of vibrant growth that enabled the population of England-Wales not only to rise from 6.1 million in around 1750 (Wrigley and Schofield 1981) to almost 42 million on the eve of World War II[27] but also to export large contingents of emigrants overseas, particularly to North America. Indeed, if we applied the annual natural increases for the total 1760–1939 period to an initial population of 6.1 million, we would obtain 45.8 million for 1940 rather than 41.5 million. Admittedly, this is just a rough approximation, given the complex effects of migrations on long-term demographic trends. First, a large majority of emigrants are often working-age males. This can cause a reduction in female nuptiality, which—in 19th century England—significantly inhibited fertility. Second, emigration is selective and may have comprised the healthier and more fertile individuals. This would influence not only fertility but also mortality. In other words, emigration is very likely to reduce the birth rate and raise the death rate of the population remaining in the home country. Nevertheless, this approximation indicates that in less than 2 centuries, without migrations, the population of England-Wales would have grown by a factor of at least 7.5 as a result of the demographic transition. Historians agree that England's extraordinary demographic dynamism stems from the speed of its industrial takeoff, characterized by a plentiful supply of jobs and a massive influx of young people into payroll employment, which facilitated family formation.

6. Length of Transition and Growth Intensity: the Transition Multiplier

In sum, the effect of transition on demographic growth depends on two factors: (1) the size of gap between the crude birth rate and the crude death rate, which yields an exceptional rate of natural increase; and (2) the length of the period in which this exceptional gap prevails. Jean-Claude Chesnais (1986) suggests a description of the phenomenon by means of what he calls the *transition multiplier*. In theory, the indicator is obtained as above for England-Wales, but Chesnais simplifies the computation by taking only the average annual natural increases over fairly long periods and combining the potential increases for each period. He performs the calculation for only five European countries: Sweden, Germany, Italy, the Soviet Union, and France (Table 68–3).

Chesnais' proposed simplification of the computation is entirely acceptable. For Sweden, for example, the complete annual computation on the lines of the one performed earlier for England-Wales would yield almost exactly the same result (3.84). In contrast, the dating of the transition period is a far more debatable choice and weighs heavily on the result. In particular, it is rather questionable to set the end of the transition at 1960 for Sweden, 1965 for Germany, and even 1970 for France. We know that fertility after World War II registered a broad uptrend followed by a downtrend: this pattern is almost entirely unrelated to the large-scale historical movement that gave birth to transition theory. For those three countries, I prefer—as in this chapter—to regard the transition as ending on the eve of World War II. Italy and the Soviet Union are different. It seems to me that the transition should be seen as complete when the TFR approaches two children per woman after a roughly constant decline over a long period. Sweden, Germany, and France had met this condition by the eve of World War II; Italy and the Soviet Union had not. The Soviet Union is such a patchwork in demographic terms that its profile is not very meaningful in itself: the central Asian republics experienced a transition far closer in time to that of the developing countries discussed in the next chapter than to the transition in European countries. Italy is a more interesting case. Admittedly, the fertility decline—as defined here—did not end before the 1970s since Italy's TFR did not reach 2.1 for the first time in history until 1976.

One can also debate the validity of Chesnais' choice of transition start dates. The transition obviously began with the start of the long-term downtrend in mortality, but how do we identify that starting point? As we saw in the opening pages of this chapter, our information on the subject is often fragmentary or imprecise. To go back to the example of Sweden, I believe the transition began some time between 1760 and 1800, but that 1810 is certainly too late (Fig. 68–1). Table 68–4 gives the values obtained for the transition multiplier by choice of transition start and end dates. The multiplier varies from 3.4 to 4.4 depending on whether we use Jean-Claude Chesnais' hypotheses or

[27] That is, 41.5 in 1939 (CSO 1952).

TABLE 68–3 Measure of Transition Multiplier for Five European Countries after Method of Jean-Claude Chesnais (1986)

Country	Demographic transition period	Main phases	Average annual natural increase	Index (start of phase = 100)	Transition multiplier*
Sweden	1810–1960	1810–1910	1.05	284.2	3.83
		1911–40	0.61	119.9	
		1941–65	0.59	112.5	
Germany†	1876–1965	1876–1910	1.32	157.2	2.11
		1911–40	0.60	119.8	
		1941–65	0.45	111.9	
Italy	1876–1965	1876–1910	1.03	143.1	2.26
		1911–40	0.88	130.0	
		1941–65	0.78	121.6	
Soviet Union	1896–1965	1895–1910	1.73	129.3	2.05
		1911–50	0.55	124.5	
		1951–65	1.62	127.3	
France	1785–1970	Not relevant	0.26	Not relevant	1.62

*Product of indices (divided by 100) in previous column.
†West Germany only after 1946.

TABLE 68–4 Sensitivity of Transition Multiplier at Selected Transition Start and End Dates: The Case of Sweden

End of demographic transition	Start of demographic transition			
	1780	1790	1800	1810*
1940	3.90	3.75	3.46	3.37
1960*	4.44	4.27	3.94	3.84

*Jean-Claude Chesnais's assumption.

other plausible ones. In particular, if we take 1940 rather than 1960 as the end date, the multiplier is reduced from 3.8 to 3.4, but if we also take 1780 instead of 1810 as the start date, the multiplier rises back to 3.9. Admittedly, this value is not all that different from Chesnais' choice, but it is, no doubt, obtained with assumptions more consistent with the conventional transition model. In contrast, if we took the entire 1780–1960 period into account, we would obtain a much bigger multiplier (4.4); again, however, I do not find it very relevant to include the fertility increase of the 1940s and 1950s in the transition period.

For Germany, Jean-Claude Chesnais' choice of start date is certainly too late. By 1876, the crude death rate had already fallen to 26.4 per 1000 for a crude birth rate of 40.9 per 1000. The peak mortality of 31 per 1000 observed in 1866 is probably not the best starting point. However, the earlier data—all of which indicate lower values—are very likely biased by an underrecording of deaths that increases as one moves back in time. In 1817, the first year available, the crude death rate was only 26.7 for a birth rate of 39.5 per 1000. In these conditions, it is very difficult to estimate the transition multiplier. If we took 1866 as the start date and 1940 as the end date, we would, however, already obtain a multiplier of 4.4—well above the 2.1 chosen by Jean-Claude Chesnais, but almost certainly below the real value. If, instead, we took 1817 as the start date, we would obtain 15.8, a result certainly highly exaggerated by the underrecording of deaths.

For Italy, the estimate is even more difficult as the earliest data available are from 1862. They do indicate, however, that 1876 is surely too late, because the crude death rate (29 per 1000) was already more than 10 points below the birth rate (40) and, most important, it was already declining in 1862–1876. Assuming an effective transition start date of around 1845 and an average annual natural increase of 9 per 1000 between 1845 and 1862, we obtain a transition multiplier of 2.9 (instead of Chesnais' 2.3) or even 3.2 if we take 1976 (not 1965) as the transition end date, as suggested earlier.

For France, Chesnais finds a multiplier of only 1.6: this value is overstated as it includes the postwar fertility upswing, which was particularly robust in France. Taking 1940 as the end date, we obtain less than 1.4. Admittedly, if we assume the transition began in 1750 (rather than 1785, Chesnais' choice), the value rises to 1.5.

Despite these uncertainties, we can clearly see the wide differences between European countries with regard to the outstanding growth they were able to achieve thanks to the demographic transition. Had there been no migrations, the French population would have enjoyed extra growth of about 40%, the Italian population could have increased more than threefold, the population of Sweden could have multiplied by 3.8, that of Germany, no doubt, by nearly 5 or even more, and that of England-Wales by 7.5! Actually, England-Wales and France seem to have been fairly exceptional borderline cases. Although it is hard to estimate all the transition multipliers with precision, we can reasonably assume a value of about four in most of Europe. Indeed, that is very roughly the order of magnitude suggested by the comparison between the population estimates given for Europe and its North American extension around 1750 (150 million) and on the eve of World War II (720 million).[28]

III. STRUCTURAL CONSEQUENCES: POPULATION AGING AND MIGRATIONS

This phase of exceptional growth is one of the most striking effects of the demographic transition, but it is not the only one. Naturally we cannot discuss all the direct and indirect consequences here, but we must at least mention two crucial corollaries: population aging and the build-up of exceptional migration flows.

1. Population Aging

One of the main effects of the demographic transition is a radical change in the age distribution. We already saw in Chapter 4 of Volume I how the dynamics of a population are driven by the interaction between movement factors (fertility, mortality, and migration) and age distribution. A variation in any single factor entails a change in the age distribution; reciprocally, any change in the age distribution influences the rate of population growth, because—assuming stable fertility, mortality, and migration—it modifies the crude birth, death, and migration rates. Let us concentrate in this section on the first aspect of this interaction—the impact of the transition on age distribution—by taking up the comparison between France and England begun in Chapter 4. The two countries were not picked at random. France had a transitional multiplier of 1.4, and England-Wales had a value of 7.5. By choosing two extremes, we cover the entire field of possible paths for the European demographic transition. In essence, we are comparing a situation of aging and little or no growth with a situation of aging after steep growth (Fig. 68–9).

The age distributions of Fig. 68–9 are plotted as absolute values in order to display both the population growth and the shift in the population pyramid. The French case is actually rather untypical of the demographic transition, because the latter generated only a minimal population growth, little different from what could occur in the boom years of the pretransition regime. The total surface occupied by the age groups in the pyramid barely changed between 1750 and 1936 (the last census before World War II). In contrast, the profile of the French pyramid indicates population aging by the early 19th century. From 1750 to 1800, the decrease in infant and youth mortality, which tends to rejuvenate the distribution from the bottom up, was almost completely offset by the fertility decline already under way. The pyramid was practically unchanged (apart from the impact of the Revolutionary and Napoleonic wars, which impacted on young working males). But the pyramid base was visibly narrowing as early as 1800–51; the phenomenon gained momentum up to the 1936 pyramid, narrower at the base than in the adult age groups. England-Wales registered brisk population growth because fertility declined well after mortality. As a result, the decline in infant and youth mortality displayed its rejuvenating effects very clearly: between 1751 and 1801, and even more so between 1801 and 1851, the pyramid base widened.

Only in the second half of the 19th century did this phenomenon begin to be offset by the fertility decline, but this occurred at a far quicker pace than in France and, from that time on, the English population aged more quickly as well. By 1931 (last census before World War II), the shape of the English pyramid almost matched that of its French counterpart. But there was one difference: the French pyramid displayed a deep "gash" caused by the huge birth deficit in the cohorts depleted by World War I, a phenomenon almost imperceptible in England-Wales.

As a result, in the French transition, the proportion of young people aged under 15 stagnated from 1750–1800 and then diminished steadily from 1800–1950. In England-Wales, it rose sharply from 1750–1820 and later declined, fairly slowly until 1900 then more quickly (Fig. 68–10). Conversely, the proportion of persons aged 60 or older rose continuously during both centuries in France, whereas it declined slightly in England-Wales until 1900 and surged abruptly thereafter.

[28] In 1940, Europe without the Soviet Union had 375 million inhabitants (Livi Bacci and Tapinos 1999); the Soviet Union, 193 million (Blum and Darskij 1999); and the United States and Canada, slightly over 150 million (Biraben 1979). At the time, however, a sizable part of the North American population came from Africa, Asia, and Latin America.

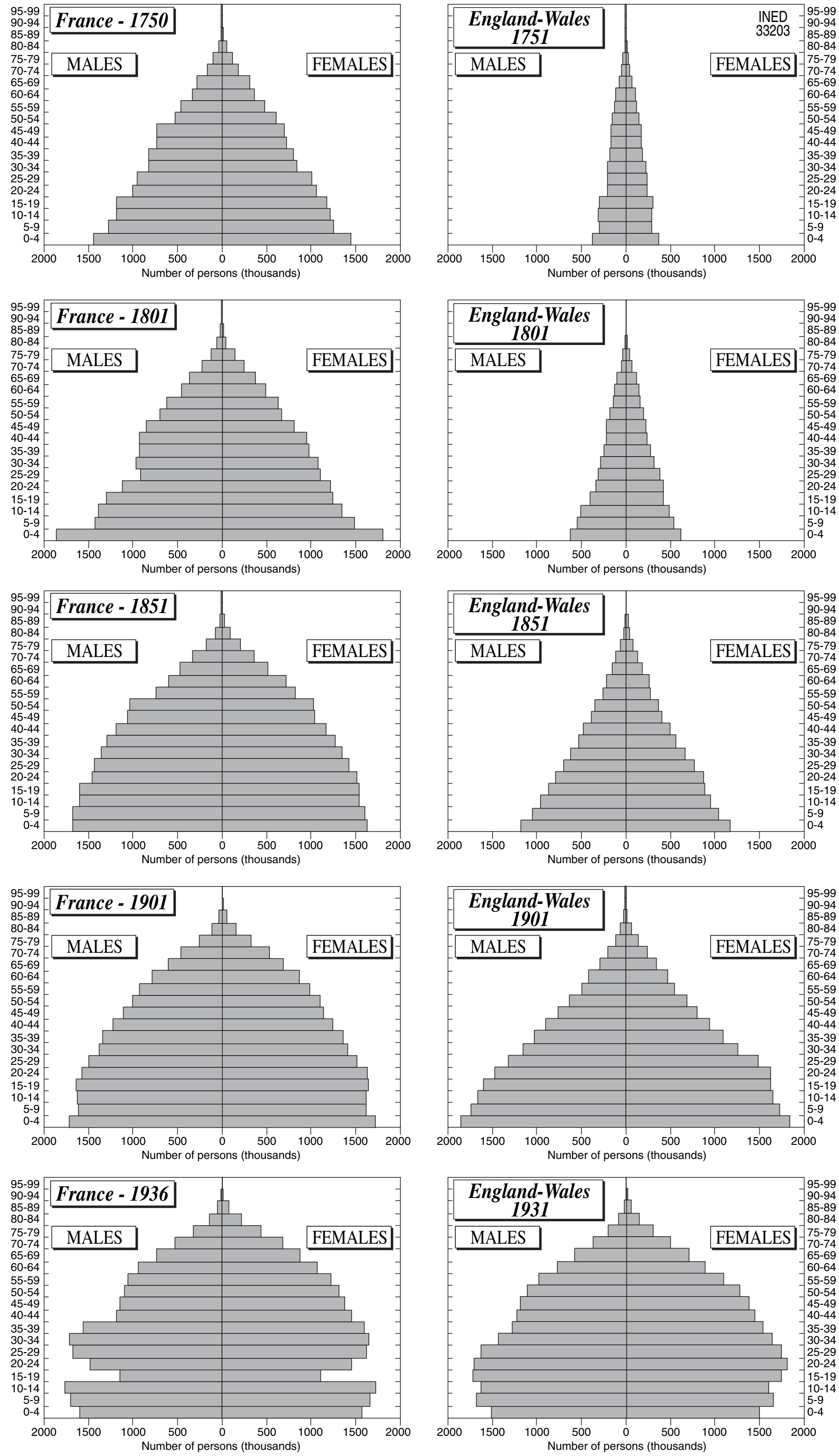


FIGURE 68–9 Age distribution in two borderline cases of demographic transition: aging without growth (France) and aging with strong growth (England-Wales). (From: France 1750–1850, Henry and Blayo 1975; France 1901 and 1936, censuses; England-Wales 1751, 1801: Wrigley and Schofield 1981; England-Wales 1851: Mitchell 1988; England-Wales 1901 and 1931: censuses.)

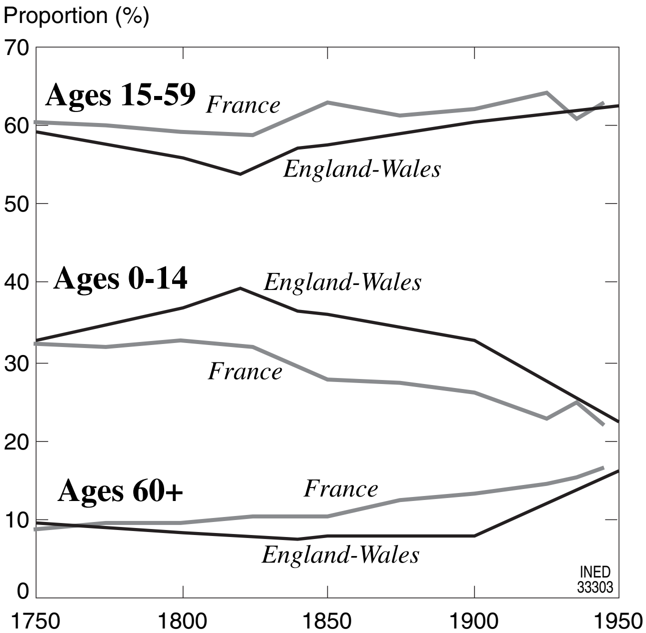


FIGURE 68–10 Population distribution by broad age group in France and England-Wales, 1750–1950.

2. A Dual Migration Flow

Migration is certainly not an original feature of the European transition period. Since time immemorial, humans have migrated across territories and frontiers, chiefly in search of better resources for their livelihood and survival (see, in particular, Chapter 67). In Europe, however, external and internal migrations climaxed during the "extended" 19th century that stretched from the second half of the 18th century to the first half of the 20th—a period in which Europe witnessed the steady build-up of industrialization. In addition to seasonal, temporary migrations, often on a large scale (Poussou 1998), two types of flows dominated Europe in that period (Zelinsky 1971): urbanization and emigration to the rest of the world. It is worth recalling the key features of both.

a. Overseas Emigration

Fueled by sharp natural demographic growth, European emigration to other continents took on exceptional proportions. From 1850–1930, "51.7 million departures were recorded . . . in other words, in 80 years, ten times as many as before" (Poussou 1998)—not counting Russian emigration to Asia, within the frontiers of the Empire, then the Soviet Union. Four European countries topped the list of this tremendous massive emigration overseas: Great Britain, with 11.4 million departures (22% of the total), Italy with 9.9 million (19.1%), Germany with 9.8 million (18.9%), and Ireland with 7.3 million (14.1%). Ireland is a very special case, on account of the Great Potato Famine of 1845–48. In contrast, in strictly proportional terms, France was a modest exporter of migrants (2.4 million) because of its weak transition multiplier, and would probably have been far less so without the need to maintain its rank as colonial power. Unlike its neighbors, France was also a net importer of migrants throughout the period. Meanwhile, migrations from Spain and Portugal resumed vigorously. On balance, the flow of emigrants from all over Europe swelled continuously until World War I, particularly in the last decade of 19th century and the first decade of the 20th.

The emigrants' main destination was the Americas, particularly North America. The United States obviously attracted the greatest number, with 28 million European immigrants between 1850 and 1930 (Poussou 1998)—an overwhelming 63%. Argentina came a distant second with 6 million (12.3%). It was followed by Canada with 5.2 million (but the country was often just a stepping-stone to the United States), Brazil (4.6 million), and Australia (3.5 million), with Algeria and South Africa lagging far behind. The rest of the colonial empires ultimately played a very minor role in Europe's global demographic expansion, except as witness to the geographic diversity of the destinations. No barrier truly stood up against this tidal wave. However, the predominantly English-speaking first waves focused on North America, whose vast expanses were weakly populated and already very heavily Anglicized. Apart from the Spanish and Portuguese—historically tied to Latin America—most of the rest of Europe followed the same trail, which sounded so promising. Admittedly, with rare exceptions, the other regions already under European domination were subjected to a colonial system more concerned with exploiting resources (and the local population) than with developing "settlement colonies" there.

As an entire chapter is devoted to the topic in the previous section, there is hardly any need to say more here about European emigration overseas. However, for the sake of completeness, we should add that the phenomenon was not the only migratory outlet for the European countries with the largest transition multipliers. By way of confirming its distinctive historical position, France, too, was a major destination for almost all its neighbors and even beyond: first Belgians, Germans, Swiss, and Austro-Hungarians, then Poles and Russians, Italians, and Spaniards. France experienced labor shortages in several industries by the mid-19th century. Its foreign population surged from 400,000 in 1851 to 3 million in 1931—not counting, of course, all those who had been naturalized in the meantime. In the same period, many Germans went east, gradually Germanizing the eastern fringes of the Empire and beyond, as far as the Volga regions (in so doing, they amplified a flow that began in the Middle Ages—most notably toward Transylvania—and revived in the 18th century, particularly toward Catherine the Great's Russia). Lastly, under the fierce pressure of the famine, many Irish began by trying to seek refuge and sustenance in Britain itself, even as the country was exporting its population surplus.

These population transfers naturally had sizable effects on the demographic trends at the main points of arrival. They affected not only the total population size there, which literally exploded, but also the mortality and fertility patterns, which closely mirrored Europe's transition stages. In terms of mortality, some overseas countries even outperformed their home countries. New Zealand, in particular, quickly achieved a record life expectancy that long remained the world's highest by a wide margin (Oeppen and Vaupel 2002). The European pattern was also copied for fertility. As in the Old World, of course, the timing and intensity varied, but there was ultimately a close convergence, at least in the English-speaking overseas countries.

b. Internal Migrations: Rural-Urban Migration and Urbanization

Interacting with the Industrial Revolution, the demographic transition also spawned a massive transfer of rural dwellers to towns and cities. Changes in farming methods from the 18th century onward, combined with the modernization of production techniques, reduced farm-labor demand. At the same time, the rise of manufacturing and, shortly after, of the service industries created new types of employment. This inevitably spurred urban growth, but the phenomenon was intensified by the fact that the demographic transition created overpopulation in rural areas, where most people initially lived. The only choice was to leave, mainly for cities. To oversimplify, we could say that international emigration actually consisted of the residual flows that cities were incapable of absorbing. At all events, in quantitative terms, it was dwarfed by the masses displaced from the countryside to the towns.

Of Europe's 146 million inhabitants[29] in 1750, only 23 million were urban dwellers and 132 million lived in rural areas (Bairoch 1998).[30] By 1950, of a total 573 million, 300 million were urban dwellers versus 272 million rural dwellers. If the rural population had grown at the same pace as the total population (a 370% increase between 1750 and 1950), it should have reached 486 million in 1950 instead of 272 million. In other words, the rural-urban migration can be estimated very roughly at over 200 million (486 − 272 = 214).

True, cities were not born with the Industrial Revolution. They date from the Neolithic and the discovery of agriculture, which promoted the growing settlement of the human population. This also generated surpluses that allowed economic specialization and the emergence of urban centers, which gave rise to the great civilizations of Antiquity. But the Industrial Revolution and the demographic transition radically altered the nature of cities and completely reversed the demographic equilibrium between urban centers and rural areas.

Table 68–5 draws on Paul Bairoch's estimates (1998) to plot urbanization rates in the main European countries from 1750–1913. The last column, for 1950, is based on United Nations estimates (2001).

Once again, we can only underscore the diversity of timing and paces of change. Although there is no doubt as to the overall trend, each country participated in it consistently with its specific history, sometimes with huge lags but also with distinctive variations. For example, let us take Spain, Italy, and, generally speaking, the Mediterranean countries, which have had a strong urban tradition since Antiquity. Their demographic transition came later; urbanization rates were already rather high in 1700 but rose rather slowly thereafter in comparison with the United Kingdom or Germany. In Italy, for instance, the urbanization rate rose from 23% in 1700 to 54% in 1950, whereas in the United Kingdom it grew from 12% to 84%. Once again, England's singularity deserved to be emphasized: the establishment of enclosures—which ended free access to commons—triggered a massive rural-urban migration, and the proliferation of manufactures offered attractive wages in town. Both developments quickened the pace of urbanization.

In almost all countries, the urbanization trend accelerated sharply at a time when transition was triggering a demographic boom, that is, at fairly different dates but seldom before 1850, with the sole exception of the United Kingdom (Fig. 68–11). Interestingly, the same happened in France: although the demographic transition generated only very modest overpopulation, the change in the urbanization rate diverged very little from the average. This was largely thanks to immigration, even though its first destination was not

[29] Including the Asian part of Russia.

[30] Urban environment is not easy to define (see Chapter 61 of Volume II). Paul Bairoch (1998) has estimated the probable real level of urbanization in 18th century Europe.

TABLE 68–5 Urbanization Rate (%)*in Europe, 1700–1950

Country	1700	1750	1800	1830	1850	1870	1890	1900	1913	1950
Austria-Hungary	4.8	7.1	7.8	8.0	9.7	12.1	19.7	25.6	29.4	36.5
Germany	7.7	8.6	9.3	10.0	15.0	24.5	34.5	42.0	51.0	71.9
Balkans†	9.7	8.8	6.8	7.6	10.5	11.8	12.5	19.8	22.6	26.1
Belgium	30.6	22.2	21.7	25.0	33.5	38.8	48.0	52.3	58.0	91.5
Spain	20.4	21.4	19.5	17.5	18.0	22.5	30.5	34.0	39.3	51.9
France	12.3	12.7	12.9	15.7	19.5	24.3	30.7	35.4	39.5	56.2
Italy	22.6	22.5	21.9	19.0	23.0	25.0	31.0	35.5	41.5	54.3
Netherlands	39.0	36.3	34.1	35.7	35.6	38.5	46.3	47.8	51.3	82.7
Portugal	18.5	17.5	15.2	15.0	15.0	15.0	15.3	15.7	15.6	19.2
United Kingdom	11.8	17.3	20.8	27.5	39.6	53.3	64.0	67.4	69.7	84.2
Russia	4.6	4.6	4.8	6.0	7.2	9.2	11.6	13.2	14.6	44.7
Scandinavia‡	4.8	6.2	7.6	7.5	7.9	11.2	17.3	21.1	24.3	56.2
Switzerland	5.9	7.7	6.9	7.5	11.9	17.5	24.5	30.6	39.3	44.3
Europe	11.0	11.6	11.5	12.8	16.3	20.6	26.4	33.2	33.6	52.4

*1700–1913, share of population of cities of more than 5000 inhabitants; 1950, share of urban population as defined by United Nations.
†Bulgaria, Greece, Romania, and Serbia.
‡Denmark, Finland, Iceland, Norway, and Sweden.
Sources: 1700–1913, Bairoch 1998; 1950, United Nations 2001

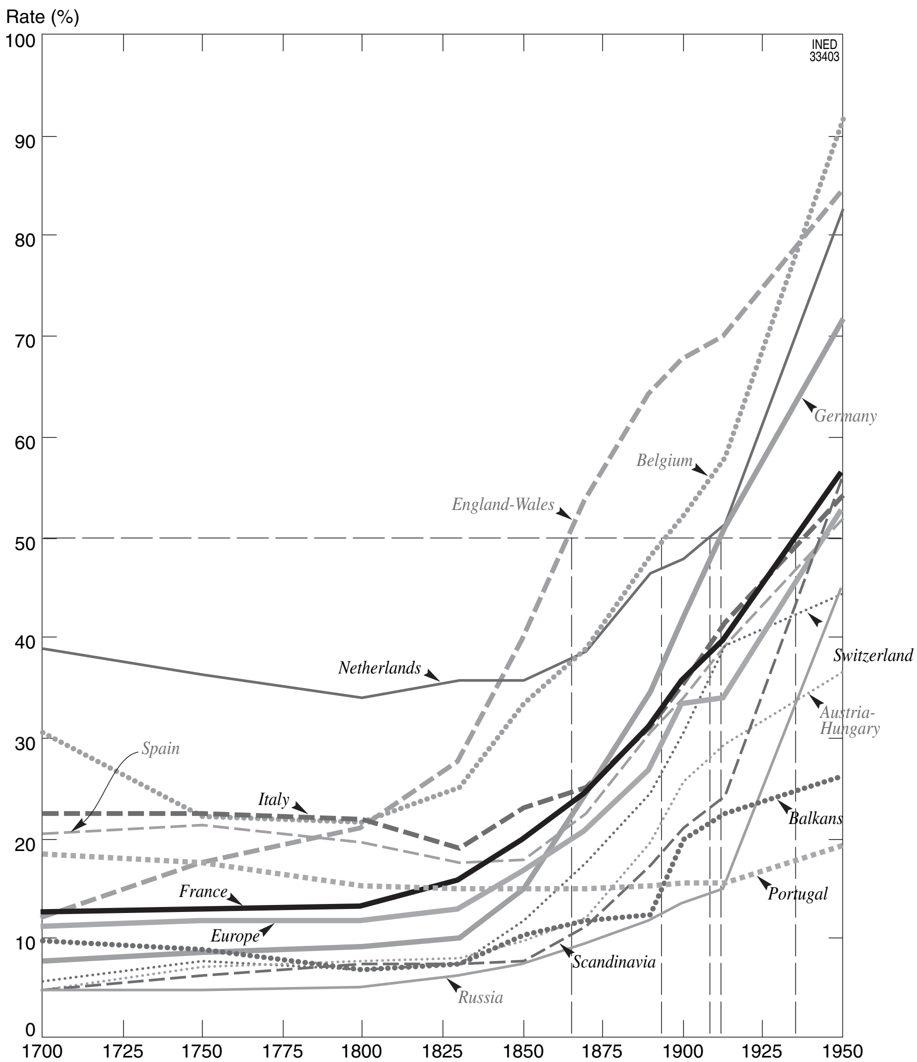


FIGURE 68–11 Urbanization rate in Europe, 1700–1950. (Data from Table 68–5.)

always the city. But another reason was the preservation of a rather archaic form of subsistence farming and heavy pressure on land. This stimulated rural-urban migration such as the Breton emigration to Paris.

An interesting indicator of the diversity of urbanization timing can be obtained by identifying the date at which the urban population exceeded 50% of the total population. The switchover occurred by the mid-1860s in the United Kingdom but in the early 1890s in Belgium. It should be noted that Belgium (particularly Wallonia) was the first country to industrialize after England. The Netherlands had long been heavily urbanized, but the rate leveled off to nearly 40% between 1700 and 1850. It crossed the 50% threshold, as did Germany, shortly before World War I. France and Italy did so in the mid-1930s; Switzerland and Spain, on the eve of World War I. Russia and Austria-Hungary did not catch up until the late 1940s; Portugal and the Balkan countries, much later still. In fact, the process was far from over by 1950: at that date, the urbanization rates varied from less than 20% in Portugal to over 90% in Belgium.

IV. BIRTH OF A THEORY

Adolphe Landry (1934) gave the name *demographic revolution* to the shift observed in Europe from an old demographic regime with high mortality and fertility to a new regime in which healthcare resources and new behaviors drove both variables to very low long-term levels. Although Landry confined the term to the description of observed facts, he nevertheless sought to provide a general explanation for it in the shape of a *contemporary theory*. This would replace the two population theories that, according to Landry, had contended for the explanation of past demographic regimes: Joseph Townsend's theory and Richard Cantillon's theory[31] (Landry 1909, 1934). As did all their contemporary peers, Townsend and Cantillon agreed that the population is determined by food resources and that a given quantity of food corresponds to a sustainable maximum population. However, Townsend (1786) noted, people "*multiply like mice in a barn*" whenever their means of subsistence enable them to do so; as a result, that maximum depends directly on the quantity of food, and all populations would aim for the maximum. Richard Cantillon (1755), instead, argued that the maximum only partly depended on the quantity of food; it was determined by people's way of life—which varied considerably from one place and period to another—and in particular by whether they were willing to live in poverty or sought to improve their living conditions. If a group wants to eat meat, it cannot afford to be as large as if it is content with grain. The two theories differ not only over the role of the population/food ratio but also over the way in which the ratio is established. In Townsend's theory, the adjustment is performed only through mortality; in Cantillon's theory, fertility also plays a part, most notably through control of sexual activity via marriage.

For Adolphe Landry, Townsend's theory "*applies only to the most backward societies*," whereas Cantillon's was valid for "*most of the societies that history tells us about*" and even still at the beginning of the transition to the new regime. Yet, Landry continued, Cantillon's theory ceases to apply to the contemporary demographic regime. That is why he offered to replace it with a *contemporary theory* that diverged from the previous two over one key aspect: the available quantity of food ceased to have any effect whatsoever on the population size. Owing to economic, social, and medical progress, mortality has fallen more sharply than ever, but population sizes have not necessarily grown as a result (even less so in France than elsewhere). This is because fertility has diminished well beyond the level that would have been dictated by the subsistence law—even in Cantillon's theory. Landry believed that the spectacular increase in life expectancy would soon encounter the insurmountable biological barrier.[32] Interestingly, however, he set no minimum level for fertility, merely arguing that it would be very unlikely to drop to zero. Landry's theory therefore does not incorporate the notion that modern populations would converge toward the stationary state. On the contrary, he argued, the near-equilibrium observed between fertility and mortality in the 1930s could well be short-lived, and he feared *depopulation*. The earlier theories viewed *populationist* models as fallacious. Landry, instead, stressed that the new demographic regime paved a wide path for

[31] It is worth noting that Adolphe Landry accused Thomas Malthus of sowing confusion between the two theories: "We may legitimately regret . . . that these theories have been confused as often as they have. The author who displays this confusion most vividly is at the same time the one who has focused his attention most powerfully on the issue of population and who owes his reputation to the book he wrote on the subject: I am referring to Malthus." (Landry 1934:176).

[32] "Will one want to set at an average of 85 years the end-point at which human life, in the best conditions, should cease? The actual mean length of life will never reach that point." It is rather striking to note that this endpoint of 85 is identical to the one recently put forward by the believers in a biological limit to human longevity, such as James Fries and Jay Olshanski (see Chapter 48 of Volume II).

population policies capable of stabilizing population size at an *optimum* level by promoting behavior that could at least maintain fertility at the generational-replacement level. In fact, all the ingredients of the transition theory that came on the scene after World War II—most notably in the work of Kingsley Davis (1945) and Frank Notestein (1945)—were already in place. This is all the more remarkable as the main focus of Adolphe Landry's research was France, a country that never experienced strong population growth in the transition period. In other words, the French situation was the opposite of the one that prevailed after World War II in the Third World, whose demographic explosion would become the key issue in transition theory.

References

Bairoch Paul, 1998. Une nouvelle distribution des populations: villes et campagnes, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. II. La Révolution démographique*, p. 193–229. Paris, Librairie Arthème Fayard, 648 p.

Bardet Jean-Pierre, 1988. La chute de la fécondité: les incertitudes de l'explication, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire de la population française. III: de 1789 à 1914*, p. 364–378. Paris, Presses universitaires de France, 554 p.

Bertillon Jacques, 1911. *La dépopulation de la France, ses conséquences, ses causes, mesures à prendre pour la combattre*. Paris, Alcan, 346 p.

Biraben Jean-Noël, 1979. Essai sur l'évolution du nombre des hommes, *Population*, vol. 34(1), p. 13–26.

Biraben Jean-Noël and Henry Louis, 1957. La mortalité des jeunes enfants dans les pays méditerranéens, *Population*, vol. 11(4), p. 615–645.

Blayo Yves, 1975. La mortalité en France de 1740 à 1829, *Population*, vol. 50 (special issue, Démographie historique), p. 123–142.

Blum Alain and Darskij Leonid E., 1999. Le modèle soviétique (1917–1991) et ses devenirs, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998*, p. 659–698. Paris, Librairie Arthème Fayard, 792 p.

Bois Jean-Pierre, Lequin Yves, and Troyansky David G., 1998. La progressive mutation des structures, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. II. La Révolution démographique*, p. 161–191. Paris, Librairie Arthème Fayard, 648 p.

Breschi Marco and Livi Bacci Massimo, 1986. Effets du climat sur la mortalité infantile: résultats pour la Savoie, le Piemont et la Ligurie en 1828–1837, *Population*, vol. 41(6), p. 1072–1074.

Breschi Marco, Pozzi Lucia, and Rettaroli Rosella, 1994. Analogie e differenze territoriali nella crescita della popolazione italiana, 1750–1911, *Bollettino di demografia storica*, no. 20, p. 41–94.

Caldwell John C., 1982. *The theory of fertility decline.* London, Academic Press, 386 p.

Cantillon Richard, 1755 (1979). *Essay de la nature du commerce en général*. Manuscript text of the Rouen municipal library. With the text of the first edition of 1755 and a bibliographical study by Takumi Tsuda. Tokyo, Kinokuniya Book-Store Co, 453 p. (The Institute of Economic Research Hi Totsubashi University Economic research Series) (Also published in 1931 by Henry Higgs, London, Macmillan, 394 p).

Chesnais Jean-Claude, 1986. *La transition démographique, étapes, formes, implications économiques.* Paris, INED and PUF, 580 p. (Travaux et Documents, Cahier no. 113).

Coale Ansley J. and Hoover Edgar M., 1958. *Population growth and economic development in low income countries.* Princeton, Princeton University Press, 389 p.

Coale Ansley J. and Watkins Susan Cotts, 1986. *The decline of fertility in Europe*. Princeton, Princeton University Press, 484 p.

Coontz Sydney, 1957. *Population theories and the economic interpretation.* London, Routledge & Kegan Paul Ltd, 208 p.

Central Statistical Office [CSO] 1952. *Annual abstract of statistics no. 88, 1938–1950.* London, Central Statistical Office, Her Majesty's Stationery Office, 332 p.

Davis Kingsley, 1945. The world demographic transition, *Annals of the American Academy of Political and Social Science*, no. 273, p. 1–11.

Davis Kingsley, 1951. *The population of India and Pakistan*. Princeton, Princeton University Press, 263 p.

Deane Phyllis and Cole William A., 1962. *British economic growth, 1688–1959, trends and structures.* Cambridge, Cambridge University Press, 348 p. (University of Cambridge, Department of Applied Economics Monographs, 8).

Del Panta Lorenzo, 1998. Costruzione di tavole di mortalità provinciali abbreviate 1881/82, *Bollettino di demografia storica*, no. 29, p. 61–69.

Dopico Fausto and Reher David-Sven, 1998. *El declive de la mortalidad en España, 1860–1930.* Saragosse, Associacion de Demografia Historica (ADEH) and Prensas Universitarias de Zaragoza, 168 p. (Monografía, 1).

Festy Patrick, 1979. *La fécondité des pays occidentaux de 1870 à 1970.* Paris, INED, Presses universitaires de France, 399 p. (Travaux et Documents no. 85).

Hajnal John, 1965. European marriage pattern in perspective, *in:* David V. Glass and David E.C. Eversley (eds.), *Population in history. Essays in historical demography*, p. 101–143. Chicago and London, Aldine Publishing Company, 682 p.

Hajnal John, 1982. Two kinds of preindustrial household formation systems, *Population and Development Review*, vol. 8(3), p. 449–494.

Henry Louis and Blayo Yves, 1967. Données démographiques sur la Bretagne et l'Anjou de 1740 à 1829, *Annales de démographie historique*, p. 91–172.

Henry Louis and Blayo Yves, 1975. La population de la France de 1740 à 1829, *Population*, vol. 50 (special issue, "Démographie historique"), p. 71–122.

Henry Louis and Blanchet Didier, 1983. La population de l'Angleterre de 1541 à 1871, *Population*, vol. 38(4–5), p. 781–826.

Holmberg Ingvar, 1970. A study of mortality among cohorts born in the 18th and 19th century, *in:* Statistiska centralbyran (ed.), *Cohort mortality of Sweden. Three studies describing past, present, and future trends in mortality*, p. 71–86. Stockholm, Statistiska centralbyran/National Central Bureau of Statistics, 86 p. (Statistiska Meddellanden/Statistical reports 1970).

Kindleberger Charles P., 1964. *Economic growth in France and Britain 1851–1950*. Boston, Harvard, 378 p.

Kirk Dudley, 1946. *Europe's population in the interwar years*. New-York, Gordon and Breach Science Publishers, 307 p. [Also published in Geneva with a preface by A. Rosenbourg, League of Nations, 1946, 303 p.].

Kirk Dudley, 1996. Demographic transition theory, *Population Studies*, vol. 50(2), p. 361–387.

Kiser Clyde V. and Whelpton Pascal K., 1953. Résumé of the Indianapolis Study of social and psychological factors affecting fertility, *Population Studies*, vol. 7(2), p. 95–110.

Landry Adolphe, 1909. Les trois théories principales de la population, *Scientia*, vol. 6(11), p. 3–29.

Landry Adolphe, 1934. *La révolution démographique. Étude et essai sur les problèmes de la population.* Paris, INED, 227 p.

Landry Adolphe, 1945. *Traité de démographie.* Paris, Payot, 646 p.

Lee Ronald D., 1993. Inverse projection and demographic fluctuations, *in:* David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 7–28. Oxford, Clarendon Press, 426 p. (International Studies in Demography).

Lesthaeghe Ron J. and Wilson Christopher, 1986. Modes of production, secularization, and the pace of the fertility decline in Western Europe, 1870–1930, *in:* Ansley Coale and Susan C. Watkins (eds.), *The decline of fertility in Europe*, p. 261–292. Princeton, Princeton University Press, 484 p.

Livi Bacci Massimo, 1997. *A concise history of world population*, 2nd ed., Oxford, Blackwell, 250 p.

Livi Bacci Massimo, 1998. *La popolazione nella storia d'Europa*. Rome, Bari, Laterza, 300 p.

Livi Bacci Massimo, 1999. *La population dans l'histoire de l'Europe.* Paris, Seuil, 302 p. (Transl. from Italian by Jean-Pierre Bardosok; preface by Jacques Le Goff, "Faire l'Europe" series).

Livi Bacci Massimo and Tapinos Georges P., 1999. Économie et population, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998*, p. 93–124. Paris, Librairie Arthème Fayard, 792 p.

Malthus Thomas R., 1798. *An essay on the principle of population.* Londres, Macmillan/J. Johnson, 396 p. (See also French edition by Éric Vilquin, Paris, INED, 1980.)

McEvedy Colin and Jones Richard, 1978. *Atlas of world population history.* Harmondsworth, Penguin, 368 p.

McKeown Thomas, 1976. *The modern rise of population.* London, Edward Arnold, 169 p.

Mitchell Brian R., 1975. *European historical statistics 1750–1970.* London, Macmillan, 446 p.

Mitchell Brian R., 1988. *British Historical Statistics.* Cambridge, Cambridge University Press, 836 p.

Notestein Frank, 1945. Population, the long view, *in:* T. Paul Schultz, *Food for the world*, p. 36–57. Chicago, University of Chicago Press, 367 p.

Oeppen Jim and Vaupel James W., 2002. Broken limits to life expectancy, *Science*, vol. 296(10), p. 1029–1031.

Péguy Charles, 1913. *L'Argent.* Paris, Gallimard, 93 p. (Cahiers de la Quinzaine: Sixième cahier de la quatorzième série.)

Pitkänen Kari J., 1993. *Deprivation and disease. Mortality during the great Finnish famine of the 1860s.* Helsinki, Hakapaino, 176 p.

Poussou Jean-Pierre, 1998. Migration et mobilité de la population en Europe à l'époque de la révolution industrielle, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. II. La Révolution démographique*, p. 231–285. Paris, Librairie Arthème Fayard, 648 p.

Rollet Catherine, 1990. *La politique à l'égard de la petite enfance sous la III^{e} République.* Paris, INED, PUF, 594 p. (Travaux et Documents, Cahier no. 127.)

Rollet Catherine, 1997. Childhood mortality in high-risk groups: some methodological reflections based on French experience, *in:* Carlo A. Corsini and Pier P. Viazzo (eds.), *The decline of infant and child mortality. The European experience: 1750–1990*, p. 213–225. Dordrecht, Martinus Nijhoff, 258 p.

Tabutin Dominique, Gourbin Catherine, Masuy-Stroobant Godelieve, and Schoumaker Bruno (eds.), 1999. *Théories, paradigmes et courants explicatifs en démographie.* Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 672 p. (Chaire Quetelet, 1997.)

Tabutin Dominique, Eggerickx Thomas, and Gourbin Catherine (eds.), 1995. *Transitions démographiques et sociétés.* Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 692 p. (Chaire Quetelet, 1992.)

Tabutin Dominique, 1980. *Problèmes de transition démographique.* Louvain-la-Neuve, UCL, Département de démographie, Cabay, 267 p. (Document pédagogique no. 1.)

Townsend Joseph, 1786 [1971]. *A dissertation on the poor law.* Berkeley, University of California Press, p. 23–24. (Preface by Ashley Montagu and afterword by Mark Neuman; available on the Web: http://socserv2.mcmaster.ca/~econ/ugcm/3ll3/townsend/poorlaw.html.)

United Nations, 2001. *World population prospects: the 2000 revision. Vol.1: Comprehensive tables.* New York, United Nations, Population Division, 746 p.

Vallin Jacques, 1986. *La population mondiale.* Paris, La Découverte, 128 p. (Repères, no. 45, latest ed.: 1995.)

Vallin Jacques, 1989a. *La population française.* Paris, La Découverte, 128 p. (Repères no. 75, latest ed.: 2001.)

Vallin Jacques, 1989b. La mortalité en Europe de 1720 à 1914: évolution à long terme et changements de structure par âge, *Annales de démographie historique*, p. 55–77. English translation, 1991. Mortality in Europe from 1720 to 1914: long-term trends and changes in the patterns by age and sex, *in:* Roger Schofield, David Reher, and Alain Bideau (eds.), *The decline of mortality in Europe*, p. 38–67. Oxford, Oxford University Press.

Vallin Jacques, 1989c. Théorie(s) de la baisse de la mortalité et situation africaine, *in:* Gilles Pison, Mpembele Sala Diakanda, and Étienne Van De Walle (eds.), *Mortalité et Société en Afrique au sud du Sahara*, p. 399–431. Paris, INED, PUF, 446 p. [Travaux et documents, Cahier no. 124]. (English translation, 1992: Theorie(s) of the declining of mortality and African situation, *in:* Étienne Van De Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in Sub-Saharan Africa*, p. 405–436. Oxford, Clarendon, 450 p.).

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX^{e} et XX^{e} siècles et projections pour le XXI^{e}.* Paris, INED, 102 p. + CD-ROM (Données statistiques, no.4–2001).

van de Walle Étienne, 1998. Nouvelles attitudes devant la vie: la limitation des naissances, *in:* Jean-Pierre Dupâquier and Jacques Bardet (eds.), *Histoire de la population de l'Europe. II: la révolution démographique, 1750–191*, p. 131–159. Paris, Librairie Arthème Fayard, 648 p.

Villermé Louis René, 1840. *Tableau de l'état physique et moral des ouvriers employés dans les manufactures de coton, de laine et de soie*. Paris, Jules Renouvard et Cie, 2 vol. reprint 1989: Paris, Études et documentation internationales (EDI), 670 p. (Preface by Jean-Pierre Chaline and Francis Démier.)

Wrigley Edward A. and Schofield Roger S., 1981. *The population History of England 1541–1871.* London, Edward Arnold, 780 p.

Zelinsky Wilbur T., 1971. The hypothesis of the mobility transition, *Geographical Review*, no. 2, p. 219–249.

CHAPTER

69

From the Globalization of the Transition to the Return of Uncertainty (1940–2000)

JACQUES VALLIN

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

Industrialization, expansion, European imperialism, and colonization: for 2 centuries Europe led the way. But in so doing, it also started a process leading once more to unprecedented growth rates in the rest of the world. After the devastating era of colonization, the effects of which were particularly severe for pre-Columbian America but also were of sad memory for Africa bled by the slave trade, European domination was accompanied by medical progress that was extremely rapid because techniques that were already tried and tested could be exported. Between the two world wars, some countries in Latin America and Asia had already experienced a significant decline in mortality. But particularly after World War II, this progress was accelerated with the development of simple methods to combat infectious and parasitic diseases and the development of concerted programs notably through the World Health Organization (WHO). For example, Sri Lanka and Mexico gained as many years of life expectancy in a decade or two as Sweden did in a century. As in Europe, however, the response to medical progress was slower for fertility, the gap between birth rates and fertility rates increased rapidly, and demographic growth accelerated up to rates significantly higher than those seen in Europe during the 14th century. This was the population explosion of the Third World (Sauvy, 1952; Balandier, 1956).[1]

Very rapidly after World War II, the international political community was concerned by the question of how far this growth would go, as well as what economic, social, political, and ecological consequences it would have. Was it necessary to develop global policies capable of controlling it? Obviously this question also mobilized the scientific community and notably encouraged the development of data collection and demographic research at all levels. From the end of the 1940s, the need to understand the acceleration in growth occurring in all the countries of the south incited a certain number of researchers to base their studies on Adolphe Landry's work (1934) on *demographic revolution* (*Révolution démographique*; see Chapter 68) and to propose a new theory, that of the "demographic transition" (Davis, 1945; Notestein,

[1] The expression Third World (Tiers-Monde) was created by Alfred Sauvy to signify the emergence of a third world between the West and the East. He used this term for the first time in an article that appeared in the French weekly *L'Observateur*, on August 14, 1952, which he concluded, "as finally, this Third World which is ignored, exploited, and despised like the Third State also wants to be something." He thus transposed Sieyès' famous phrase about the Third State during the French Revolution. He sometimes jokingly added that the capitalist world could be assimilated with the Nobility and the communist world with the Clergy.

1945; Chesnais, 1986). Under demographic transition, with the help of economic and social development, all societies and notably those of the Third World would eventually go from a primitive demographic system to a new system, already established in Europe, in which fertility and mortality would become more or less balanced but at very low levels.

Simply stating such a theory was obviously not sufficient to end the debate between the neo-Malthusians, according to whom only a strong birth-control program could re-establish the balance, and Third-World activists, for whom the best contraceptive was development. In their initial world population projections, United Nations experts avoided exploring the future beyond the year 2000, for fear of not being able to correctly translate the theory into a concrete forecast (see Chapter 77). The main uncertainty concerned the schedule that future fertility trends might follow: when and at what rate fertility would begin to decline. However, at the end of the 1970s the facts were there. The second phase of the transition (fertility decline) had started in nearly all the southern countries, and the growth rate of the world population was already calming down. It was therefore possible to envisage and then ascertain the generalization of the demographic transition.

Could the forecast of a general stabilization of the world population be trusted nevertheless (Poursin, 1989)? Although growth rates have certainly been decreasing continually for 30 years, the general stabilization (still present even today in the United Nations' forecasts) in the strictest sense suggests growth rates are moving toward a convergence towards zero, combining a peak in life expectancy imposed by biology with the stabilization of fertility at 2.1 children per woman. This would result in the convergence of collective policies and individual decisions based on the simple idea of only ensuring the replacement of generations. The world population would then become permanently *stationary*, in the strict sense of the word (see Chapter 20 of Volume I). It is indeed probable that all populations will approach this balance. But will they remain there? Already Europe has largely committed itself to trends that infallibly lead to negative growth, toward population decline. What reasons could be proposey to certify that this is only an accident and that the rest of the planet will not follow the same example? The transition theory gives no clues on this point. Better still, the paradigm disappears at the same time as it occurs, for it does not tell us anything about what comes next.

Consequently, announcing a general stabilization of the world population as a result of reaching a point of permanent balance is not to trust in the outcome of a theory but in an act of faith that is no doubt imprudent. Without a new theory capable of completely understanding the new trends, we are indeed in the long-term future faced with an uncertainty that at least equals the one that prevailed after the World War II (see Chapter 78). In other words, although the fear of the Third World explosion is dissolving in the generalization of the demographic transition, the very realization of the latter ruins the theory and increases uncertainties.

I. THE THIRD WORLD POPULATION EXPLOSION AND THE GENERALIZATION OF THE DEMOGRAPHIC TRANSITION

I have already mentioned the first negative consequences of colonization for the populations of Africa and America (see Chapter 68). In its final phase, colonization produced the opposite effect. With the 20th century, the population of Africa again experienced positive growth, comparable to that of the Asian population, which had not experienced a decline, whereas the populations of the two Americas and Oceania increased more than ever with the mass immigration of prolific Europeans (see Chapter 58 of Volume II). In 1900, 37% of the world population was either in Europe[2] (26.2%), North America (5.6%), Latin America (4.6%), or Oceania (0.4%). In 1950, this proportion remained largely unchanged, but Europe's proportion[3] had diminished by 3 points, falling to 22.7%, whereas America's increased from 10.2% to 13.4%. Furthermore, since the beginning of the 20th century, the American expansion has been occurring differently in the north and south. The United States and Canada, totally committed to the industrial revolution of which they would soon take the leadership, also advanced in their demographic transition and followed quite closely the European model of a natural growth of 1% to 1.5% per year, although doubled by very high immigration. Latin America, on the contrary, only took one shaky step toward industrialization and only started a mortality decline, with fertility remaining high and even increasing in some countries. Natural increase rapidly rose more than in North America and was reinforced by a significant migratory contribution. From 1900–1950 North America's proportion of the world population increased from 5.5% to 6.6%, whereas that of Latin America increased from 4.6% to 6.6%. The first closely followed Europe, with the second falling into

[2] Including Russia.

[3] Including the Soviet Union.

what will henceforth be known as the Third World. In the opposite direction during the same period, Japan caught up with the Western countries, the industrial model of which it copied brilliantly, on the path to demographic transition. The first half of the 20th century thus marked the division between the developed countries and the developing countries, which was to dominate the second half of the century.

Thus on the eve of the World War II, the premise of the population explosion in the Third World that would rock the demographic-political balance of the planet was already set: a decline in mortality without an immediate response in fertility. The same scenario that characterized the European expansion during the preceding century was to be reproduced in the rest of the world with increased rapidity, opening the way to unprecedented growth rates.

1. Mortality Decline in the Southern Countries

In the wake of World War II, the decline in mortality, which had already started in several southern countries in the 1920s and 1930s, experienced significant acceleration. In the course of a few decades, the mass use of cheap but efficient methods developed in the industrial countries to fight infectious and parasitic diseases quickly increased the chance of survival in most Third World countries. At the beginning of the 1960s, the United Nations (1962) published an eloquent graph comparing the rapid increases in several Third World countries during the 1940s and 1950s compared with the secular evolution of Swedish life expectancy (Fig. 69–1). One of the most striking examples was that of Ceylon (since renamed Sri Lanka). In 8 years, from 1945–1947 to 1954, owing to a campaign to eradicate malaria through mass spraying of dichloro-diphenyl-trichloroethane (DDT) organized under the auspices of WHO, life expectancy at birth increased from 46 to 60 years![4] In Sweden, 60 years of progress in health had been necessary from 1865–1925 to obtain the same increase of 14 years of the expectation of life. The case of Sri Lanka is not an isolated one, or the most spectacular; according to the above-mentioned graph, Mauritius during the same period gained 18 years of life expectancy in 8 years. Trinidad and Tobago gained nearly 25 years of life expectancy

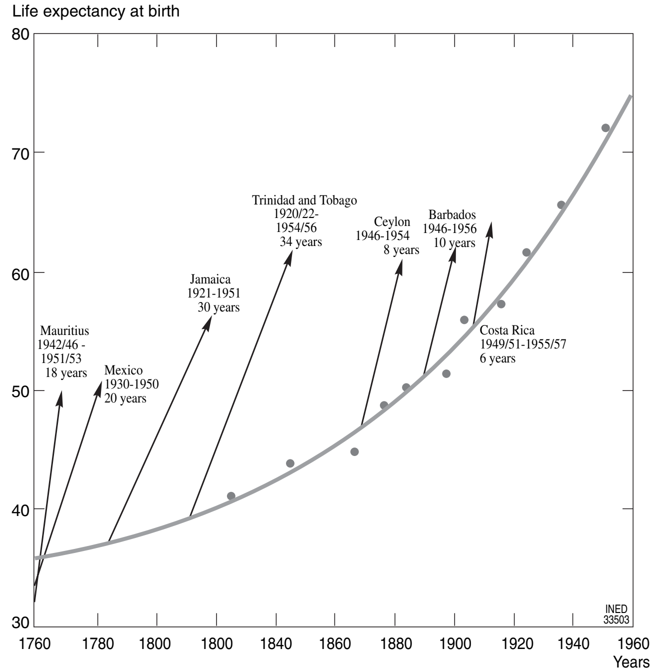


FIGURE 69–1 Increases in life expectancy in some Third World countries after World War II compared with that of the Swedish secular evolution. The curve is a freehand adjustment of the cloud of points representing Swedish life expectancies at birth during different periods. Each arrow represents the trajectory of a Third World country between the two indicated dates, and its starting point is placed on the Swedish curve at the point where it reaches the starting level for Sweden. (Data from United Nations, 1962.)

in 34 years (between the 1920s and the 1950s). But these increases are not limited to small insular countries. In 20 years (1930–1950), Mexico gained 17 years in life expectancy, as much as Sweden did in 150 years (between comparable levels).

Generally, especially during the first decades after World War II, a significant convergence in life expectancies was observed, as large regions of the Third World reduced the gap between them and the developed countries (Fig. 69–2).

From 1950–1955 to 1995–2000, the average life expectancy for the developing countries as a whole increased from 41 to 62.5 years, an increase of nearly 22 years, and at the same time, that of the developed countries increased by less than 9 years, rising from 66.1 to 74.8 years (Table 69–1). Thus, the overall gap between the two, which was 25 years at the beginning of the 1950s, was only 12 years on the eve of the 21st century. If we put aside sub-Saharan Africa, where progress is slower than in other regions of the south and where the AIDS epidemic is a threat, then the convergence is even more remarkable. The average life expectancy of all the developing countries, without including sub-Saharan Africa, reached nearly 65 years

[4] The role of this eradication campaign in the decline of mortality was the subject of great controversy (Molineaux, 1985). In 1961 Harald Frederiksen even went as far as to state that this increase in life expectancy obtained in Sri Lanka could be explained only by the improvement in the standard of living, but this hypothesis was easily contradicted by Paul Newman (1965).

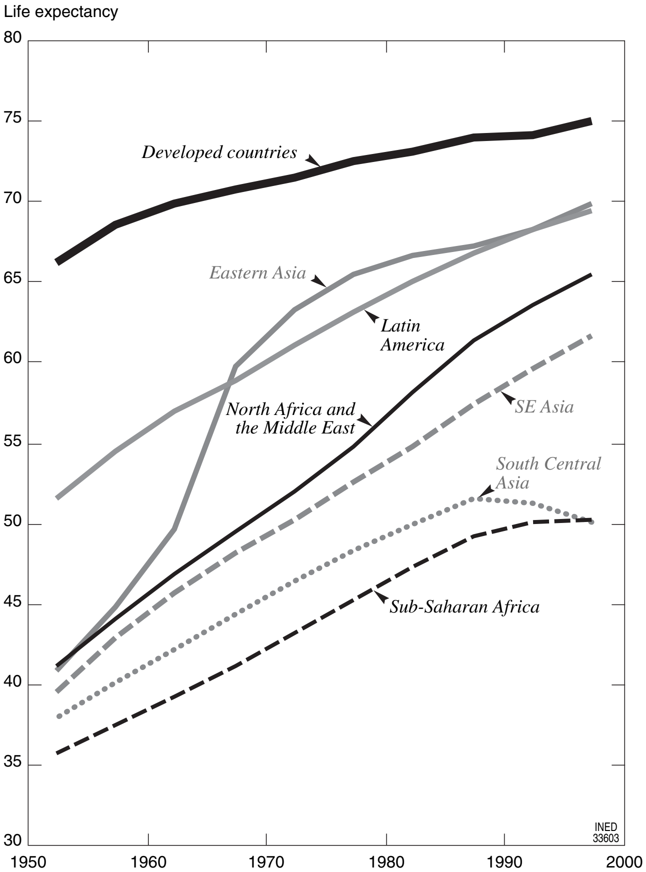


FIGURE 69–2 Evolution since 1950 of life expectancy at birth by large Third World regions compared to developed countries. (Data from United Nations, 2003.)

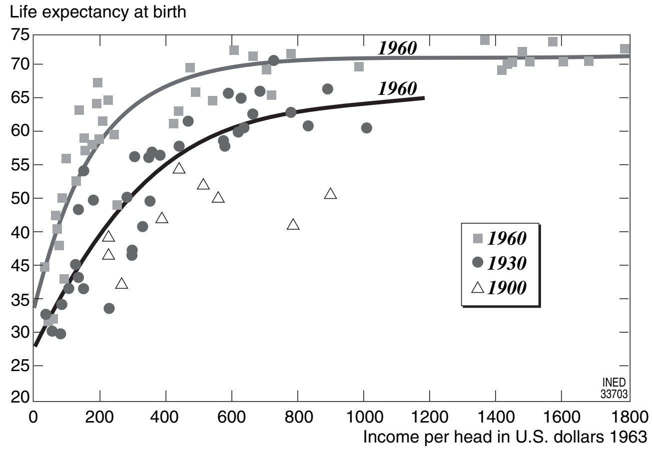


FIGURE 69–3 Life expectancy at birth in relation to income per head in 1900, 1930, and 1960. (Data from Samuel Preston, 1975.)

in 1995–2000, the gap with the developed countries being less than 10 years (Table 69–1).

East Asia's recovery is particularly spectacular[5] and is dominated by China with a current life expectancy of nearly 70 years, equal to that of Latin America and separated by less than 5 years from that of the developed countries (against over 26 years at the beginning of the 1950s). The recovery is also very distinct for southeast Asia and north Africa and the entire Middle East, which have a gap of 8 to 9 years with the developed countries. Considered individually, many countries that previously belonged to the Third World now have life expectancy levels close to those of the developed countries. Even if we put aside the microstates of the Caribbean and the Pacific and only consider countries with more than 1 million inhabitants, 8 developing countries (South Korea, United Arab Emirates, Panama, Jamaica, Puerto Rico, Argentina, Chile, and Uruguay) have a life expectancy of 74 to 76 years, close to the average level in the developed countries; 20 others are between 70 and 74 years (Mauritius, China, Malaysia, Sri Lanka, Azerbaijan, Armenia, Georgia, Syria, Kuwait, Jordan, Lebanon, Palestine, Saudi Arabia, Oman, Libya, Tunisia, Trinidad and Tobago, Mexico, Venezuela, Colombia); and 5 are just over 76 years (Hong Kong, Singapore, Israel, Cuba, Costa Rica).

The determinants of this mortality decline have already been discussed abundantly in Volume II. In Europe's case, medical progress depended generally on economic and social progress, but in the developing countries they are not so closely linked. The progress already made in the developed countries meant the means of fighting most infectious diseases already existed, and health policies created at national and international levels have made the decline in mortality possible in even the poorest countries. Figure 69–3 from Samuel Preston (1975) combines life expectancy at birth and income per head in 10 countries in 1900–1910, 38 countries in 1930–1940, and 57 countries in 1960–1970. These groups of countries were selected according to the quality of data available and are obviously of unequal value. The first, in particular, is not very representative of the world situation, as it only consists of countries that are today considered developed. The significant deformation of the adjustment curve between the 1930s and the 1960s nevertheless clearly illustrates the evolution in the association between mortality and standard of living from the 1930s to the 1960s. By comparing the average global life span that would have resulted in the 1960s if the structure of this association observed in the 1930s had been maintained and, reciprocally, the life expectancy that would have been produced in the 1930s with the application of the structure of the 1960s,

[5] Not including Japan.

TABLE 69–1 Evolution of Life Expectancy at Birth from 1950–1955 to 1995–2000

Regions	Life expectancy (in years)			Increase (in years)		
	1950–1955	1975–1980	1995–2000	1950–1955/ 1975–1980	1975–1980/ 1995–2000	1950–1955/ 1995–2000
Sub-Saharan Africa[a]	36.7	46.8	47.1	10.1	0.3	10.4
North Africa and the Middle East	43.5	57.2	66.3	13.7	9.0	22.8
North Africa	41.8	53.9	64.6	12.0	10.7	22.7
Middle East[b]	45.2	60.5	67.9	15.3	7.4	22.7
South central Asia	39.3	52.5	61.5	13.1	9.0	22.1
India	38.7	52.9	62.1	14.2	9.2	23.4
Western Asia (excluding Japan)	39.9	65.1	69.6	25.2	4.5	29.6
China	40.8	65.3	69.7	24.6	4.4	29.0
Southeast Asia	41.0	54.6	65.3	13.7	10.7	24.4
Latin America	51.4	63.0	69.3	11.6	6.3	17.9
Oceania, excluding Australia and New Zealand	39.4	52.2	60.1	12.8	7.9	20.8
Developing countries total	41.0	56.8	62.5	15.7	5.8	21.5
Total excluding sub-Saharan Africa	41.5	58.0	64.9	16.5	6.9	23.3
Australia and New Zealand	69.6	73.3	78.4	3.7	5.1	8.9
North America	68.9	73.4	76.7	4.5	3.3	7.8
Western Europe[c]	66.5	73.0	77.3	6.5	4.3	10.7
Eastern Europe[d]	64.3	69.4	68.2	5.1	–1.2	4.0
Japan	63.9	75.5	80.5	11.5	5.1	16.6
Developed countries total	66.1	72.3	74.8	6.2	2.5	8.7
World total	46.5	59.8	64.6	13.2	4.9	18.1

[a] Consisting of four regions in the United Nations' sense: west Africa, central Africa, east Africa, and southern Africa.
[b] The United Nations western Asia region (including the countries of the Caucasus).
[c] Consisting of three regions in the United Nations' sense: northern Europe, western Europe, and southern Europe.
[d] Including Russian Asia.
Data from estimates adopted by the United Nations for their 2002 population projections (United Nations, 2003).

Samuel Preston (1975) estimated the proportion of the increase in life expectancy that was owing to the increase in income per head and the proportion owing to the modification in the association between mortality and standard of living. He concluded that, "Factors exogenous to a country's current level of income probably account for 75% to 90% of the growth of life expectancy for the world as a whole between the 1930s and the 1960s. Income growth per se accounts for only 10% to 25%." Preston continued by stating that the same type of results appear when income per head is replaced by a population's literacy level or by the average number of calories consumed per person. Although these three factors make a real contribution, they alone do not explain the rapid decline in mortality observed after World War II.

This relative independence of medical progress to economic development results from the fact that after World War II, numerous Third World countries benefited from the technical progress already achieved in the most developed countries. The association nevertheless exists and varies with the level of development of each country. Jean Bourgeois-Pichat (1966) noted during the 1960s that in the most advanced Third World countries where the evolution of mortality was known with enough precision, this evolution occurred in three stages. First of all, there was a slow increase in life expectancy followed by a sudden jump before it finally returned to a slower growth rate. This evolution was very distinct in the case of Mexico, studied by Jean Bourgeois-Pichat, and was also observed in many Third World countries (Vallin, 1968). This hypothesis corresponds to the *health transition* as Abdel Omran (1971) conceived it (see Chapter 57 of Volume II), in which economic and social progress and an intermediate-phase reduction of infectious diseases help populations move from an *age of plague and famine* when progress is very slow to an *age of degenerative and social diseases* when, with a higher level of life expectancy, progress is again very slow. We know that this theory can no longer explain recent development, taking particularly into consideration the progress made since then in the successful treatment of cardiovascular diseases (Olshansky, 1986; Vallin, 1993; Meslé and Vallin, 2003; Vallin and Meslé, 2004), but it corresponds well enough to the phase of medical progress in the eradication of infectious diseases. Figure 69–4 gives an illustration of it by comparing the progress in life

FIGURE 69–4 Increases in life expectancy between 1950–1955 and 1980–1985 according to the level already reached in 1950–1955. (Data from Vallin, 1989.)

expectancy observed from the beginning of the 1950s up to the beginning of the 1980s, with the initial level already reached in 1950–1955. Where the level of life expectancy is low at the beginning, progress has been modest, and it has also been modest in the countries that were initially the most advanced. On the other hand, in the countries that have just started progress, the increase is very rapid. At the lowest levels, the start of medical progress necessitates a relatively expensive global development effort; when it has already begun, the effectiveness of health policies is far greater and far cheaper as the medical tools have already been developed in the advanced countries. Finally, in the more advanced countries, progress is again more difficult as efficient methods of fighting on new fronts must now be invented.

This is how, during this period, a number of developing countries bridged a significant part of the gap with the developed countries (Fig. 69–5a,b). Although at the beginning of the 1950s life expectancies of the Third World were homogenous and low, with a mode of around 40 to 45 years, this changed at the beginning of the 1980s, with two distinct modes appearing, one at 65 to 70 years and near to that of the developed

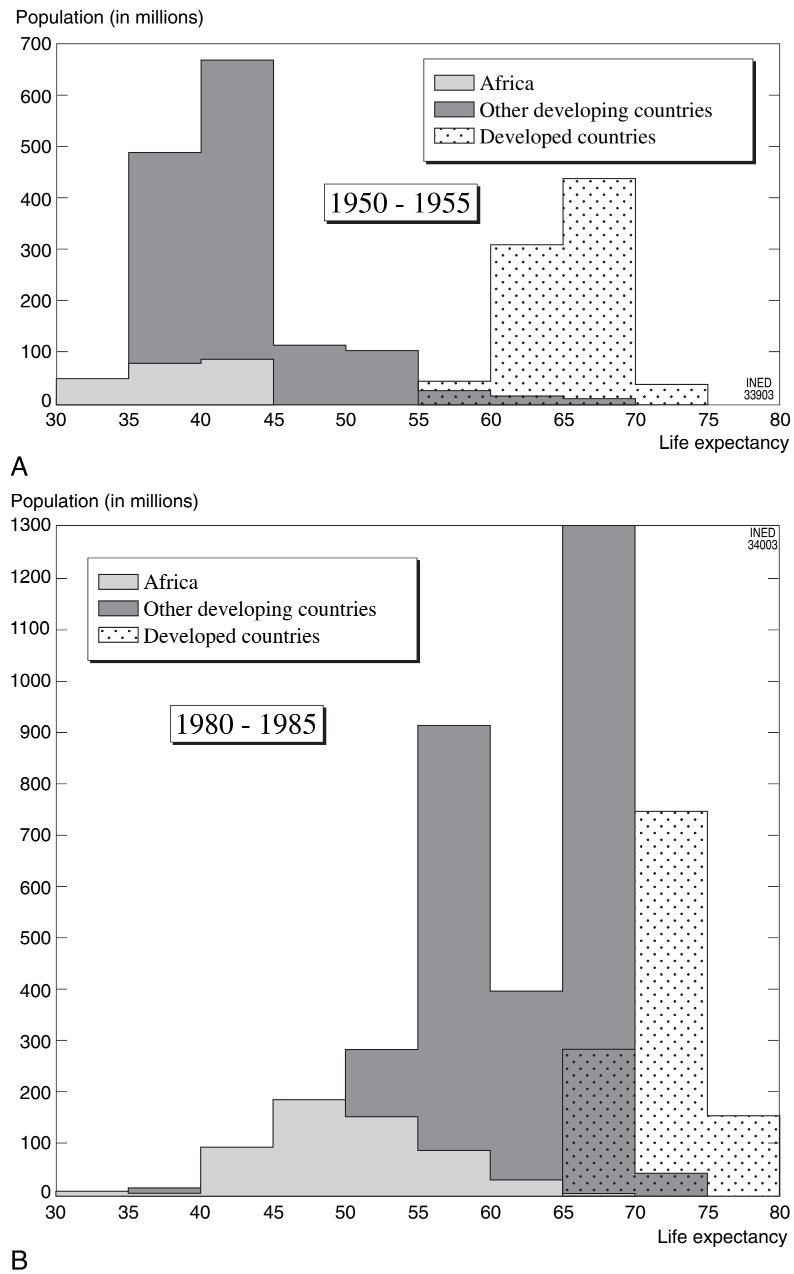


FIGURE 69–5 **(Top)** Distribution of the different countries of the world (weighted by their population), by 5-year sections of life expectancy in 1950–1955. (Data from Vallin, 1989.) **(Bottom)** Distribution of the different countries of the world (weighted by their population), by 5-year sections of life expectancy in 1980–1985. (Data from Vallin, 1989.)

countries and the other 10 years lower, at 55 to 60 years. All the southern countries had made progress, but some much more rapidly than others, leaving sub-Saharan Africa in particular resolutely lagging behind.

2. The Evolution of Fertility

Faced with mortality decline, fertility (as was previously the case in Europe) remained high for a certain period of time. It did not decline before the 1970s in hardly any of the large Third World regions (Fig. 69–6). In some countries it even increased quite significantly as a result of social and cultural changes associated with the modernization of society. In some cases it was as a result of current events. For example, in Tunisia, fertility increased rapidly in 1964 owing to a peak in marriages, which itself was provoked by an increase in the legal age at marriage: the new law having been

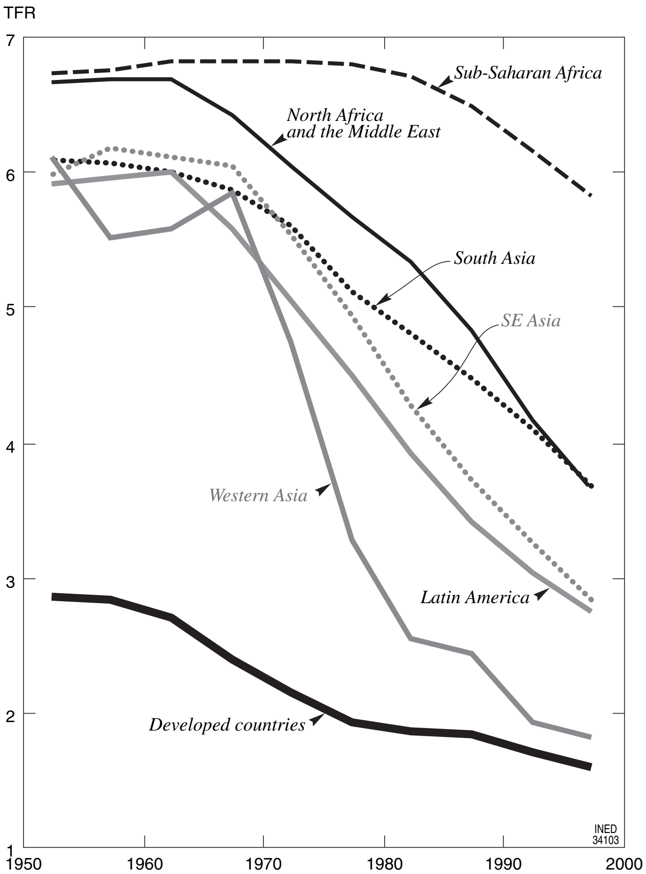


FIGURE 69–6 Evolution since 1950 of the total fertility rate by large Third World regions compared with that of the developed countries. (Data from United Nations, 2003.)

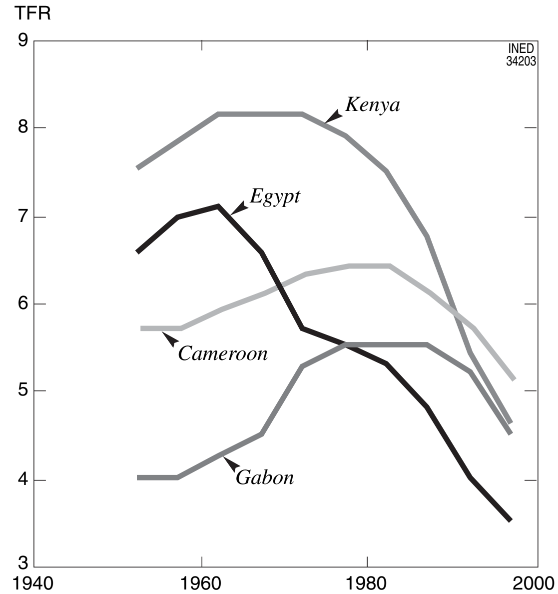


FIGURE 69–7 Evolution since 1950 of the total fertility rate in four countries that had experienced a significant increase in fertility at the beginning of the transition. (Data from United Nations, 2003.)

announced beforehand, families rushed to conclude the marriages of very young girls that would have been prevented by it (Vallin and Lapham, 1969), and the total fertility rate (TFR) increased from 7 in previous years to 7.9 children per woman during that year (Waltisperger *et al*., 2001). In the same way, in Algeria the period after independence provoked a return to traditional values, which resulted notably in a lowering of the age of marriage for girls and consequently in a sharp rise in fertility. Although the TFR had an average of 7.2 between 1960 and 1967, it increased rapidly in 1968–1969 to reach 8.1 in 1970 (Ouadah and Vallin, 2000).

But more fundamentally than these increases associated with very specific events, fertility has often tended to increase in developing countries with the premise of the modernization of society. Several factors worked in this direction, whether it was the progressive abandoning of practices that favored long birth intervals in traditional societies or progress made in the treatment or prevention of sterility and infertility. In numerous traditional societies, and notably in Africa, the duration of breast-feeding could on average be very long, from 18 months to over 2 years, and the fertility reduction resulting from this could be increased by a quite strict observance of *postpartum* abstinence. With the transformation of social relations and notably the evolution in women's education and employment, breast-feeding has tended to decline and sexual taboos to weaken, thus provoking significant increases in fertility in many countries, particularly in urban areas. In addition, medical progress has in general been accompanied by a reduction in infertility, especially in societies in which childlessness means a serious loss of status for a woman. In some countries particularly affected by sterilizing venereal diseases, such as Central Africa, which was for a long time a subfecund region compared with the rest of the continent (Retel-Laurentin, 1974, 1979; Larsen, 1996), this phenomenon was particularly important: fertility experienced significant growth with the arrival of antibiotics, information campaigns, and the prevention of sexually transmitted diseases.

Before the above-mentioned specific events, the countries of the Maghreb experienced a certain increase in fertility until the 1960s (for Algeria, see Breil and Biraben, 1969; Negadi *et al*., 1974). But this phenomenon mainly affected the Middle East and sub-Saharan Africa, as well as some Latin American countries to a lesser extent. Figure 69–7 traces the evolution of the TFR since 1950 in four other countries that were particularly affected by an increase in fertility that preceded the beginning of a decline. In Egypt, as well as in Kenya, this was caused by a reduction in the period of breast-feeding. In Cameroon and Gabon, the main factor was the reduction of infertility that resulted from venereal diseases. In all these cases as in many other Third World countries, fertility decline was preceded by this increase: it could even be said that the decline was announced by the increase, as the increase was itself associated with social and cultural changes that would thereafter result in fertility decline. Thus, at the first phase of the demographic transition of Third World countries, not only did mortality decline significantly more rapidly than was the case in the European countries, but also fertility was generally higher, often with more than seven children per woman, whereas in the old Europe the figure was five to six (see Chapter 68).

But as in Europe previously, once this first phase was over, with the transformation of societies, economic and cultural evolution, urbanization, increase in education, and improvement in women's status, these populations responded to the decline in mortality with increasingly improved fertility control. And just as with mortality, the decline in fertility was and will be far more rapid than it was during the European demographic transition (Fig. 69–6).

From the second half of the 1970s, all of the Third World regions, except for sub-Saharan Africa, experienced significant fertility decline, which accelerated during the last quarter of the 20th century (Table 69–2). Some investigators for a long time believed that sub-Saharan Africa would continue to be an exception (Locoh and Makdessi, 1995), but it finally fell in with the rest of the Third World at the end of the 1980s, as has been clearly confirmed by the results of each new demographic survey (Locoh and Vallin, 1998).

The factors of the fertility decline have already been discussed in Volume I. Let us therefore simply summarize the essentials here. As in the European countries in the past, the recent drop in fertility in the developing countries is affected by two key elements, the increase in age at marriage and the diffusion of contraception, which have diverse effects in different countries and cultures.

a. The Increase in Age at Marriage

The increase in age at marriage has, initially at least, had a decisive role in societies in which marriage was and still largely remains a requisite condition to any

TABLE 69–2 Evolution in Total Fertility Rates from 1950–1955 to 1995–2000

Region	Total fertility rates (children per woman)			Reduction		
	1950–1955	1975–1980	1995–2000	1950–1955/ 1975–1980	1975–1980/ 1995–2000	1950–1955/ 1995–2000
Sub-Saharan Africa[a]	6.72	6.77	5.80	0.05	–0.97	–0.92
North Africa and the Middle East	6.64	5.66	3.63	–0.98	–2.03	–3.01
North Africa	6.82	6.03	3.52	–0.79	–2.51	–3.30
Middle East[b]	6.46	5.30	3.73	–1.16	–1.57	–2.73
South central Asia	6.08	5.09	3.65	–0.99	–1.44	–2.43
India	5.97	4.83	3.45	–1.14	–1.38	–2.52
Western Asia (excluding Japan)	6.09	3.27	1.79	–2.82	–1.47	–4.29
China	6.22	3.32	1.80	–2.90	–1.52	–4.42
South East Asia	5.95	4.91	2.81	–1.04	–2.10	–3.14
Latin America	5.89	4.48	2.72	–1.41	–1.76	–3.17
Oceania, excluding Australia and New Zealand	6.35	5.43	4.27	–0.92	–1.16	–2.08
Developing countries total	6.16	4.63	3.11	–1.53	–1.52	–3.05
Total excluding sub-Saharan Africa	6.07	4.33	2.67	–1.74	–1.66	–3.40
Australia and New Zealand	3.27	2.11	1.81	–1.17	–0.30	–1.47
North America	3.47	1.78	2.01	–1.69	0.22	–1.46
Western Europe[c]	2.46	1.89	1.48	–0.57	–0.41	–0.98
Eastern Europe[d]	2.91	2.08	1.29	–0.84	–0.78	–1.62
Japan	2.75	1.81	1.39	–0.94	–0.42	–1.36
Developed countries total	2.84	1.91	1.58	–0.93	–0.34	–1.26
WORLD TOTAL	5.02	3.90	2.83	–1.11	–1.07	–2.19

[a] Consisting of four regions in the United Nations' sense: West Africa, Central Africa, East Africa, and Southern Africa.
[b] The United Nations western Asia region (including the countries of the Caucasus).
[c] Consisting of three regions in the United Nations' sense: northern Europe, western Europe, and southern Europe.
[d] Including Russian Asia.
Data from estimates adopted by the United Nations for their 2002 population projections (United Nations, 2003).

sexual relation and to procreation and in which spinsterhood was extremely rare. In Tunisia, for example, despite a very active family planning policy that has been in existence since the mid-1960s, it has been established that the decline in fertility not only was initially largely caused by the increase in women's age at marriage (Vallin and Lapham, 1969; Vallin, 1971; Jemai and Jemai, 1979) but also explained a little over half of the fertility decline observed between 1978 and 1995 (Ayad and Jemai, 2001), with contraception only becoming an essential factor at the end of the 1980s. It must be said that the increase in women's age at marriage was particularly spectacular in this country, increasing from 19.5 years in 1956 to 26.6 in 1994 (Ben Salem and Locoh, 2001). The situation is more or less the same in the majority of the Maghreb and Middle-Eastern countries such as Algeria, where the average age of women at their first marriage increased suddenly from 18.3 years in 1966 to 27.6 in 1998 (Kouaouci, 1992; Ouadah and Vallin, 2000, 2003). In other countries the evolution is less striking, for example, in Iran the average age of women at their first marriage increased from 19.7 years in 1976 to 22.0 in 1996, but this evolution nevertheless constitutes "a factor of the first order for fertility where it is nearly nonexistent outside marriage"[6] (Ladier-Fouladi, 2003). In all these countries, nuptiality has played an important role in the beginning of fertility decline, whereas contraception only complimented it. It was only at a second phase of varying duration that contraception took over.

b. The Spread of Contraception

In other contexts the evolution in age at marriage made a lesser contribution and contraception was from the start the main factor. This is obviously the case of populations in which prenuptial sexual relations and fertility outside marriage have an important place, for example, in the Caribbean and recently in sub-Saharan populations, where fertility has only very recently started to decline. In such a context, fertility decline owes little to the increase in age at marriage. But even in cultures in which marriage is traditionally a preliminary condition for reproduction, fertility decline has not necessarily been owing to the delay in nuptiality. In general, the evolution of nuptiality has only played a minor role in Latin American countries where contraception has from the beginning been in the forefront. As Lorenzo Moreno and Susheela Singh (1996) show, from the 1970s to the 1980s, women's age at first marriage increased little in this region when measured by using Bongaarts' model (1978) (see Chapter 34 of Volume I), its effect on fertility decline appears inferior to that of the development of contraception. The situation is more varied in Asia (Leete and Alam, 1993), where the increase in age at marriage has not been negligible and has at times contributed to the start of fertility decline. Nevertheless, fertility decline has often occurred without a significant increase in age at marriage (Rele and Alam, 1993). In Thailand, for example, from 1960 to 1980 women's age at their first marriage barely increased by a year (from 22.0 to 23.4), whereas fertility declined by nearly half (from 6.4 to 3.5 children per woman). On the contrary, the average age at the first marriage of Pakistani women increased by over 3 years during the same period (from 16.7 to 19.8 years), whereas fertility remained constant at 7 children per woman. It is true that marriages of very young adolescents are not always consummated immediately. The case of China (Attané, 2002; Peng, 2002), is characterized by an extremely severe policy of birth limitation associated from the start with the effects of the evolution in nuptiality (delay in marrying, separation of couples) and fertility control within marriage (contraception, induced abortion).

Whatever the role played by age at marriage in the fertility transition, the latter can only be complete with fertility control within marriage owing to contraception and induced abortion. Here again, as in Europe in the past, the paths taken vary according to country and culture. Henri Leridon and Laurent Toulemon (2003) recently drew a world panorama of the levels of contraception and the methods used. Figure 69–8 summarizes the results.

In terms of the intensity of contraceptive use, the diversity remains large. Although in East Asia and Latin America in 1997 nearly 70% of women in conjugal unions used one form of contraception or another, nearly as much as in Europe and North America, the equivalent in the rest of Asia was only 50% and a mere 20% in Africa. But there is also diversity in the contraceptive methods. Sterilization is much more widespread in Western Asia than in Europe and North America, and it is even more common in Latin America, where nearly 30% of women in conjugal union are sterilized (less than 10% in Europe and North America). This practice is marginal in Africa. The pill is used nearly as much in Latin America as in Europe or North America, and far less in western Asia, where women use the intrauterine device (IUD) much more than elsewhere. Condoms and other methods (including traditional methods) are used far less in the southern countries than in the northern countries.

[6] "Un facteur de premier ordre pour la fécondité lorque celle-ci est quasi inexistante hors marriage."

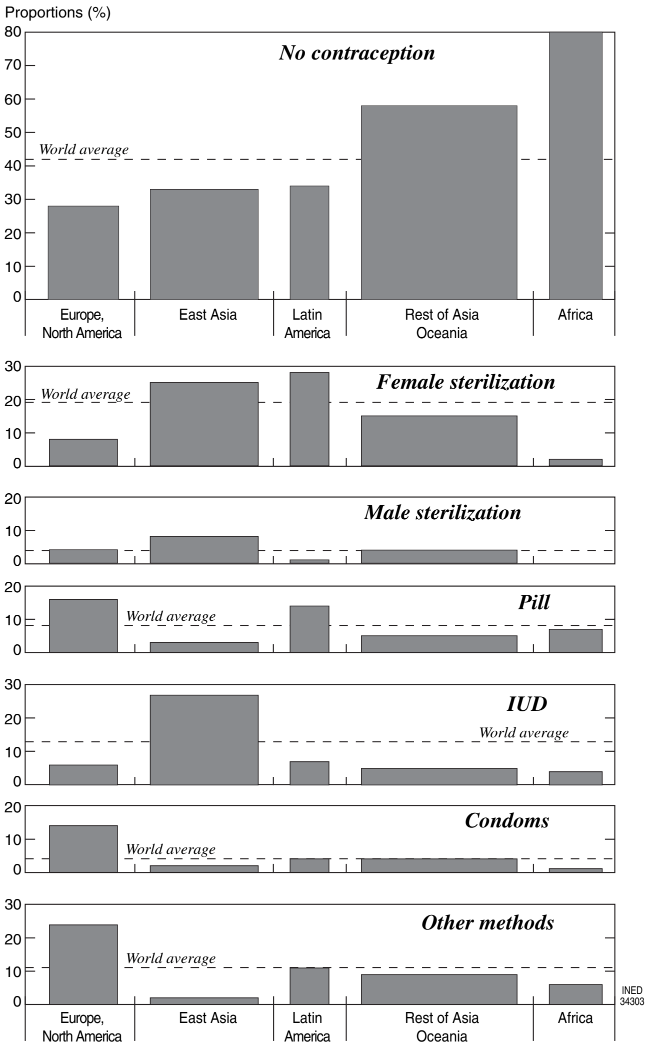


FIGURE 69–8 Distribution of 100 women from each world region in 1997, aged 15–49, married or living in conjugal union, according to contraceptive use. The width of the histograms is proportional to the population of the region. (Data from Henri Leridon and Laurent Toulemon, 2003.)

c. Recourse to Induced Abortion

Recourse to induced abortion also varies significantly from one country to another, and from one region to another. According to Henri Leridon and Laurent Toulemon, the rate of legal induced abortion per 1000 women aged 15 to 49 varied in 1995–2000, from 0 per 1000 in Africa to 79 per 1000 in eastern Europe and the rate of illegal induced abortion from 0 in East Asia to 33 in Latin America. In fact, if both legal and illegal induced abortions are cumulated, then the rate varies little between the large southern regions (36 per 1000 in east Asia, 35 in Latin America, 31 in Africa, and 30 in the rest of Asia), which all find themselves between the two extremes observed in the developed countries (25 per 1000 in Western Europe and North America, 91 in eastern Europe).

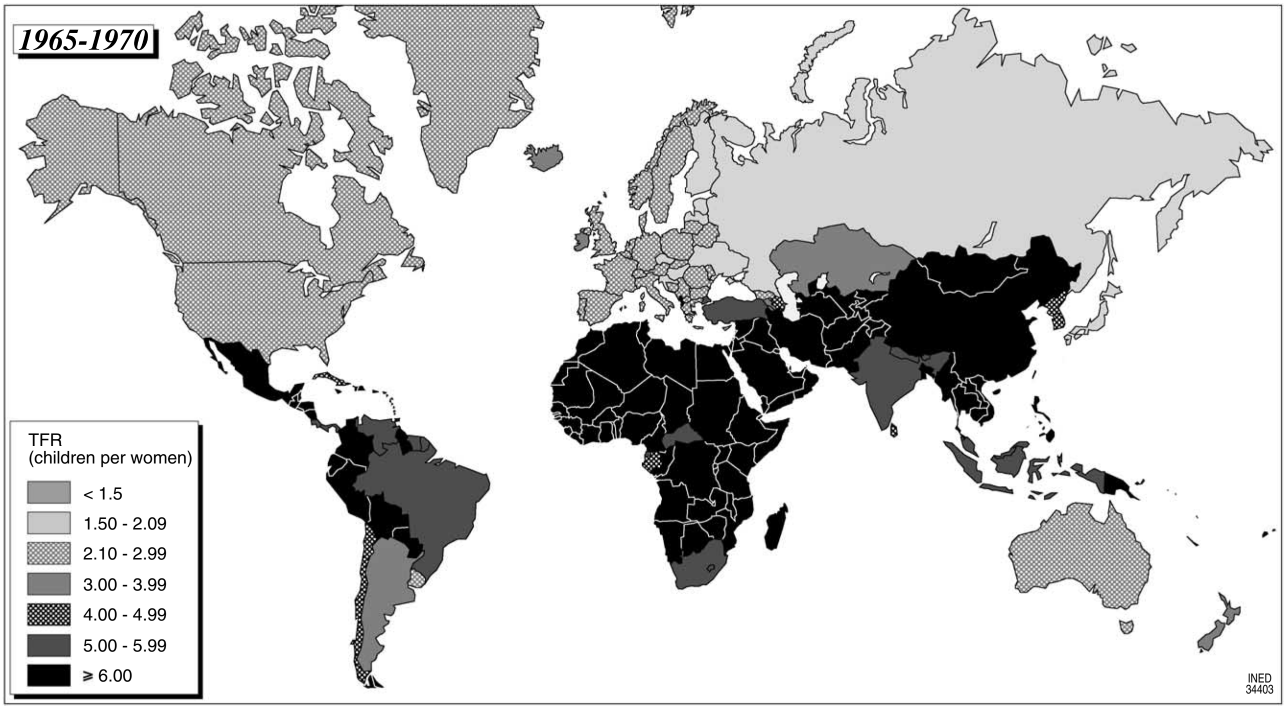
d. The Determinants of Fertility Decline

But whatever the context, age at marriage, contraception, and induced abortion are only intermediate variables. What fundamentally causes fertility decline is the evolution in couples' attitudes regarding procreation. It has certainly been affected by more or less active fertility policies carried out by various governments (see the chapters of Volume IV on this subject). However, these birth control policies, however firm and intelligent they may be, are only one element among others, and their success depends on the evolution of the society as a whole, whether it be the generalization of education, the diversification and the complexification of economic activities, the development of urban life, the evolution in women's status and their place in society and in the economy, or that of gender relationships. These questions have already been addressed in Volume I, and it is not necessary to discuss them in detail here. Let us only say that it is the diversity of experiences in all these areas that has led to the differences in timing and rates observed in the Third World countries, and it is this that explains the diversity of fertility levels observed in the world today. Fertility has declined everywhere, but the diversity of current levels remains large, as is illustrated by the two maps produced by Jean-Claude Chesnais (Fig. 69–9a,b).

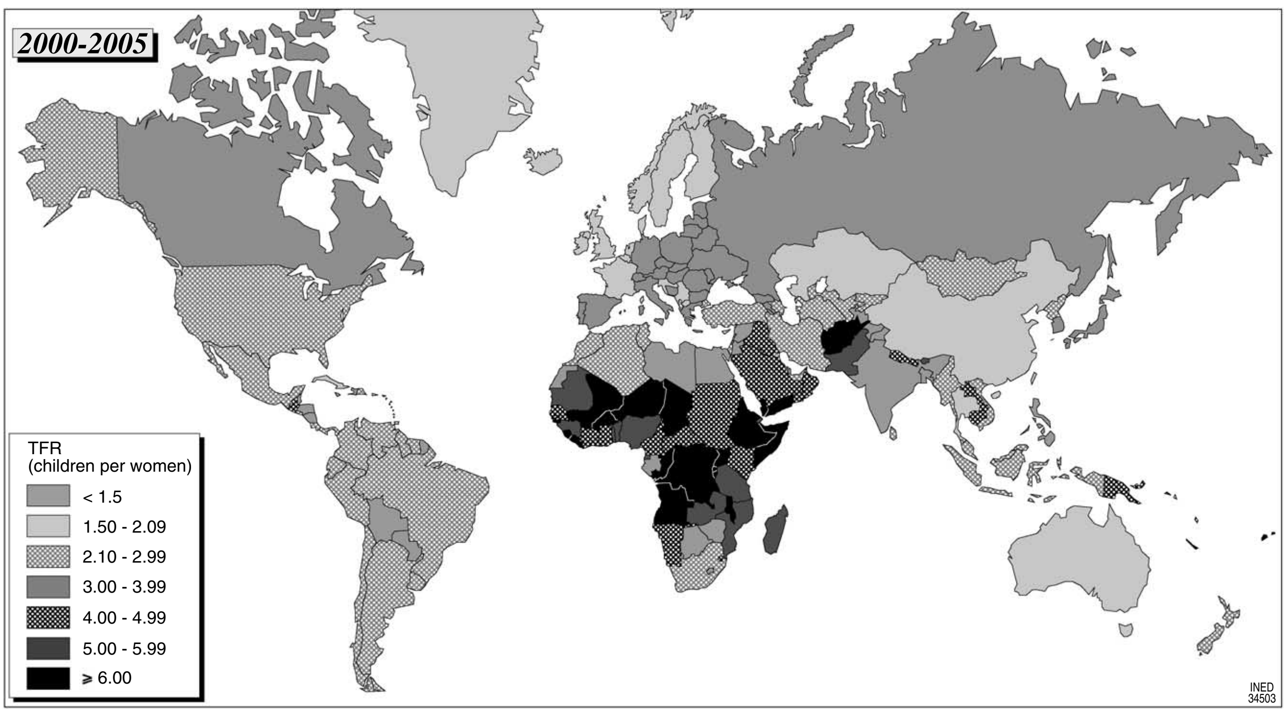
We can nevertheless begin to talk about convergence. The diversity doubtless reached its peak at the end of the 1960s, when the TFR culminated at 8 children per woman in some Third World countries such as Kenya, whereas it was already below 1.5 in some European countries. By today fertility rates of over seven children per woman have become exceptional, and rates of six are already rare.

3. Half a Century of Exceptional Growth: The Third World Explosion

The expression "Third World explosion" is most often used when the speaker is referring to the exceptional phase of world population growth that resulted from its demographic transition. Nevertheless, it could also be used in a completely different sense as we have just seen that the Third World, relatively homogenous in its demographic characteristics, literally exploded over the course of a few decades as much in the context of mortality as in that of fertility. Whereas around 1950 Third World countries were all grouped together with high levels of mortality and fertility that were very far



A



B

FIGURE 69–9 (**A**) Total fertility rate by country in 1965–1970. (**B**) Total fertility rate by country in 1995–2000. (Data from Jean-Claude Chesnais, 2003.)

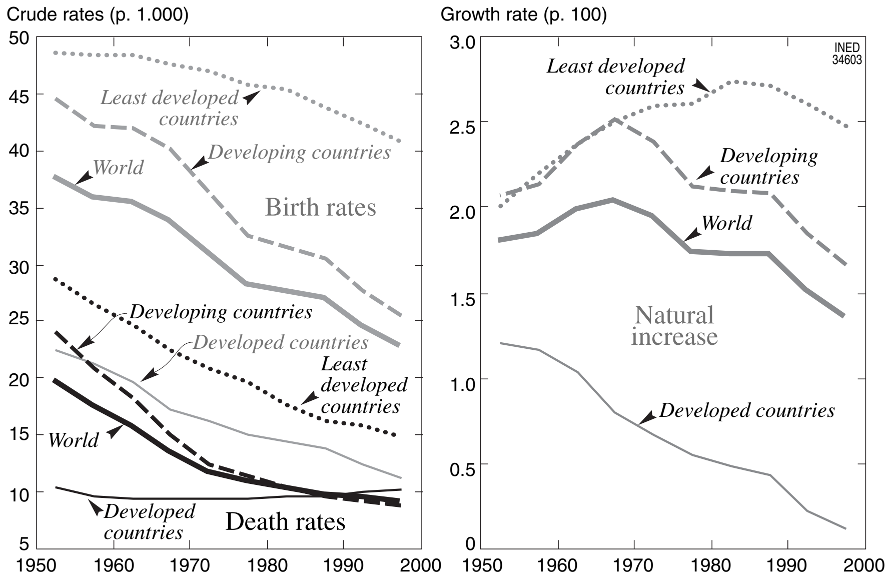


FIGURE 69–10 Evolution since 1950 of crude birth and death rates and the rate of natural increase. (Data from United Nations, 2003.)

from those of the developed countries, they are now dispersed, with some of them reaching or close to reaching European and North American standards, whereas others are still far behind on the transition path.

It is therefore in a dispersed order, with different schedules and rates, that these countries contributed to the exceptional population growth of the second half of the 20th century, generally referred to as the "population explosion."

As in Europe, this phase of exceptional growth results from the time lag between mortality decline and fertility decline. Figure 69–10, which shows the United Nations' most recent data (2003), only partially takes this phenomenon into account as it had begun before 1950 and was obviously not yet completed in 2000. In 1950, the crude death rate of the *developing countries*[7] had already significantly lessened. There is no reliable overall estimate available for this group of countries before World War II, but it is likely that it was more than 30 or even 35 per 1000. Similarly, the birth rate had also doubtless increased slightly and was probably closer to 40 per 1000 than to 45. It is quite likely that globally at the beginning of the 1950s the fertility of the Third World countries was already at its peak, even though it was not reached in a good number of the countries until later.

[7] This expression will be used constantly here according to the United Nations' definition: the countries of Africa, Latin America, and Asia (not including Japan, but including the Caucasian and Central Asian republics of the former Soviet Union).

Thus from 1950–2000, the crude birth rate of the developing countries declined regularly, falling from around 45 per 1000 to 25, whereas the crude death rate fell from 20 to less than 10. Meanwhile the decline in mortality was faster than that of birth rates during the 1950s and 1960s before becoming the opposite in the following decades. As a result the rate of natural increase of the developing countries continued to rise for another 20 years after World War II. At the beginning of the 1950s, it was around 2% per year, and it increased to 2.5% at the end of the 1960s. But at the beginning of the 1970s, the trend was reversed as the decline in birth rates was becoming greater (in absolute variation) than that of death rates. The drop in the rate of natural increase has continued to the present day. At the beginning of the 2000s, the rate was only 1.3%.

Therefore, for nearly 40 years the rate of natural increase of all the developing countries combined remained over 2% per year. This may appear modest considering what was stated about Europe in the previous chapter and all the alarmist talk concerning the Third World population explosion. However, natural growth of 2% per year results in a doubling of the population in 35 years, whereas a rate of 1% would require 70 years in order to obtain the same outcome. But above all we must not lose sight that the global evolution of the developing countries thus described is only the result of trajectories that are extremely variable in intensity and have a significant time lag.

The United Nations' current nomenclature distinguishes a subgroup of developing countries called the "*least developed*,"[8] which as we can see (Fig. 69–10) have followed a very different path to that of other developing countries. At the beginning of the 1950s, the crude death rate was still close to 30 per 1000 and the birth rate was nearly 50 per 1000. In other words, these countries only benefited from medical progress much later, and fertility significantly increased during the first phase of the demographic transition. Fertility then reached a peak and remained there for over a decade before beginning a slight decline, which only really asserted itself at the end of the century. The birth rate only dropped under the 45 per 1000 level, a culminating point for all the developing countries, at the end of the 1980s. Consequently and taking into account the significant and regular decline of the crude death rate (until the 1990s at least), the rate of natural increase, which remained the same as in the other developing countries combined until the trend reversed at the end of the 1960s, continued to increase during nearly a further 20 years until the end of the 1980s. It is only since the 1990s that this growth has started to decelerate. It thus culminated at 2.7% per year and was therefore over 2.5% per year for 35 years, from the mid-1960s until the end of the 1990s. With a growth rate of 2.5% per year, the population doubles in 28 years.

But this group of least developed countries also includes countries with different trajectories. To understand the demographic growth rates that the developing countries have experienced and are still experiencing, the level in each country must be studied individually. At the height of their growth, most of the developing countries reached growth rates of 3%, and those that exceeded 3.5% were not exceptional; some countries even exceeded 4%.

It is, however, difficult to reduce all the trajectories to Jean-Claude Chesnais' transitional multiplier (1986). On the one hand, as we saw in the preceding chapter for the European countries, it is necessary to decide on the initial point, which is often unclear because of the low quality of the data, and on the other hand, the demographic transition has not yet been completed in the majority of the southern countries. However, as a basis for the growth observed since 1950, it is also possible to use the United Nations' projections (see Chapter 77) to take into account the end of the transition as they re-estimate the trends that have occurred in each revision of their projections. By using different sources, it is possible to decide on a initial point and to include growth before 1950. The results are shown in Table 69–3 for the 25 developing countries in which the population was estimated at more than 30 million in 2000.

The first two columns of this table show the total population of each country in 1950 and 2000. The next column shows the cumulated rate of natural increase for 1950–2000. This half century is the period in which all these countries reached the peak of their growth, and this cumulated growth is obviously an important part of the transitional multiplier. Through natural growth, only the population of five of these countries more than quadrupled in 50 years: Kenya (4.7), Tanzania (4.2), and Algeria, Mexico, and the Philippines (4). For 13 other countries, the population multiplied by three to four. In only six cases was the multiplier below three (India, Indonesia, Myanmar, China, South Korea, and Argentina).

But in most cases, these coefficients are far from accounting for the totality of the transitional multiplier. The two last columns of Table 69–3 give an approximate idea by supplying the values obtained for two periods of longer duration: the first, from the presumed starting date of the transition up to 2000; the second, from the same starting date up to 2050, which includes the whole period covered by the United Nations' projections. Not only must the first be used with caution as the initial point is very arbitrary and often depends on the data available, but the second is sometimes too long or too short. Cases in which the period is too long are of little importance because beyond the date when the transition should have been completed, the countries in question are supposed to arrive at a near zero growth; however, in the cases in which the period is too short, the uncertainty of future evolution is too great for it to be useful to take it into account.

For the whole of the period up to 2050, the transitional multiplier of these 25 countries varies from 3 (China, 3.3) to nearly 15 (Ethiopia, 14.7; Pakistan, 14.2). Indeed, for China, which had more or less completed its transition by 2000, the multiplier is barely different in 2050 to what it was already in 2000 (3.0). On the other hand, adding the projection period for Ethiopia and Pakistan considerably increases the multiplier, which was only 5.7 and 5.6, respectively, when stopped at 2000. The contribution of the 2000–2050

[8] This group consists of 34 African countries (Angola, Benin, Burkina Faso, Burundi, Cape Verde, Central Africa, Comores, Congo (DR), Djibouti, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Equatorial Guinea, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Niger, Uganda, Rwanda, Sao Tome, Senegal, Sierra Leone, Somalia, Sudan, Tanzania, Chad, Togo, Zambia), 9 Asian countries (Afghanistan, Bangladesh, Bhutan, Cambodia, Laos, Maldives, Myanmar, Nepal, Yemen), 1 Latin American country (Haiti), and 5 countries in Oceania (Kiribati, Solomon Iles, Samoa, Tuvalu, Vanuatu).

TABLE 69–3 Population Growth in the 25 Developing Countries with an Estimated Population of over 30 Million Inhabitants in 2000

Country	Population		Transitional multiplier				
	in 1950	in 2000	1950–2000	1950–2050	Start	Start–2000	Start–2050
Kenya	6,265	30,549	4.7	6.9	1940	6.5	9.5
Tanzania	7,886	34,837	4.2	8.4	1940	6.7	13.4
Algeria	8,753	30,245	4.0	6.6	1922	5.6	9.2
Mexico	27,737	98,933	4.0	6.2	1920	7.5	11.9
Philippines	19,996	75,711	4.0	7.3	1920	7.6	13.8
Congo DR	12,184	48,571	3.8	11.9	1940	4.1	12.7
Iran	16,913	66,443	3.8	6.2	1920	5.5	8.9
Nigeria	29,790	114,746	3.8	8.5	1931	6.0	13.6
Colombia	12,568	42,120	3.6	6.0	1920	7.3	12
Ethiopia	18,434	65,590	3.5	9.2	1935	5.7	14.7
Pakistan	39,659	142,654	3.5	8.9	1920	5.6	14.2
Sudan	9,190	31,437	3.3	6.4	1931	5.4	10.4
Egypt	21,834	67,784	3.3	6.2	1920	4.8	9.2
Bangladesh	41,783	137,952	3.3	6.1			
Turkey	21,484	68,281	3.3	4.8	1920	5.2	7.7
Brazil	53,975	171,796	3.1	4.2	1920	5.9	8
Thailand	19,626	60,925	3.1	3.9	1920	6.1	7.7
South Africa	13,683	44,000	3.0	2.7	1930	4.7	4.3
Vietnam	27,367	78,137	2.9	4.4	1920	4.6	6.9
India	357,561	1,016,938	2.8	4.3	1921	4.1	6.2
Indonesia	79,538	211,559	2.7	3.8	1920	3.9	5.7
Myanmar	17,832	47,544	2.7	3.6	1921	3.8	5.1
China	554,760	1,275,215	2.6	2.8	1920	3.0	3.3
South Korea	18,859	46,835	2.5	2.4	1920	4.2	4.2
Argentina	17,150	37,074	2.0	2.9			

period is obviously significant in the countries in which fertility decline was late in starting, on the condition that they had not been ravaged by the AIDS epidemic in the meantime. Therefore, although fertility decline has only recently started in Kenya, which for this reason has the highest multiplying coefficient in Table 69–3 for the 1950–2000 period (4.7) and one of the highest for the period since the start of the transition up to 2000 (6.7), its coefficient only increases moderately when the projection up to 2050 is included (9.5), far below the multiplication by 14.7 of Ethiopia. Yet at the beginning of the 1980s, Kenya seemed to be destined for one of the greatest population explosions of the planet before the arrival of the AIDS epidemic. The effect of AIDS on the potential growth in south Africa is even more spectacular because its more modest multiplier coefficient before 2000 (3.0 for 1950–2000 and 4.7 for 1920–2000) decreases when the 2050 projection is included (2.7 since 1950, 4.3 since 1920).

Table 69–3 exposes the diversity of the situations in the Southern countries beyond the effect of the AIDS epidemic, which severely curtailed population growth in the African countries in which the most spectacular explosions were expected. Despite the AIDS epidemic, the populations of some African countries have increased by more than 13-fold (Tanzania, Nigeria) or 14 (Ethiopia) during their transition, as did the more dynamic Asian countries (Philippines, Pakistan) earlier and Latin American countries (Mexico). Such high transitional multipliers are a long way from the 7.5 of England and Wales, considered in the previous chapter as exceptional in the European demographic transition. The natural increase of countries such as China, South Korea, and Argentina is comparable to what was observed in Western Europe during its transition. The 600 million (and soon to be 1 billion) Chinese have been the cause of much discussion since the 1950s even though the population of China has increased far less during its transition than the population of Europe did a century earlier. The population of India will soon exceed that of China, although its transitional multiplier is lower than that of England.

In concrete terms, Table 69–4 shows the evolution of the population observed between 1950 and 2000 in the countries that today have more than 30 million inhabitants.

TABLE 69–4 1950–2000 Evolution of World Population by Large Regions and by Countries with More Than 30 Million Inhabitants in 2000

Region or country	1950	1960	1970	1980	1990	2000
Sub-Saharan Africa[a]	167,911	210,108	271,284	358,540	479,448	622,056
Nigeria	29,790	37,446	47,980	64,325	86,018	114,746
Ethiopia	18,434	22,723	29,035	35,688	48,856	65,590
Congo DR	12,184	15,438	20,603	27,909	37,370	48,571
South Africa	13,683	17,396	22,657	29,140	36,848	44,000
Tanzania	7,886	10,205	13,756	18,838	26,068	34,837
Kenya	6,265	8,285	11,370	16,368	23,585	30,549
North Africa and the Middle-East	104,225	133,873	173,138	226,104	295,635	365,837
North Africa	53,302	67,291	85,999	111,078	142,995	173,615
Egypt	21,834	27,840	35,285	43,915	55,768	67,784
Sudan	9,190	11,422	14,469	19,387	24,927	31,437
Algeria	8,753	10,800	13,746	18,740	25,017	30,245
Middle-East[b]	50,922	66,583	87,139	115,026	152,640	192,222
Turkey	21,484	28,233	36,207	46,132	57,593	68,281
South central Asia	498,508	619,722	783,331	981,313	1,225,280	1,486,049
India	357,561	442,344	554,911	688,856	846,418	1,016,938
Pakistan	39,659	48,767	61,840	80,781	110,901	142,654
Bangladesh	41,783	51,785	66,292	85,004	109,402	137,952
Iran	16,913	21,704	28,809	39,343	56,703	66,443
East Asia (except Japan)	587,360	698,132	882,446	1,061,151	1,226,424	1,354,076
China	554,760	657,492	830,675	998,877	1,155,305	1,275,215
South Korea	18,859	25,003	31,923	38,124	42,869	46,835
Southeast Asia	178,073	222,804	285,871	358,038	439,926	520,355
Indonesia	79,538	95,931	119,998	150,128	182,117	211,559
Vietnam	27,367	33,648	42,898	53,005	66,074	78,137
Philippines	19,996	27,054	36,551	48,088	61,104	75,711
Thailand	19,626	26,603	36,260	46,342	54,389	60,925
Myanmar	17,832	21,600	26,851	33,705	40,506	47,544
Latin America	167,097	218,300	284,856	361,401	441,525	520,229
Brazil	53,975	72,742	95,988	121,614	148,809	171,796
Mexico	27,737	36,945	50,596	67,569	83,225	98,933
Colombia	12,568	16,857	22,561	28,447	34,970	42,120
Argentina	17,150	20,616	23,962	28,094	32,527	37,074
Oceania, except Australia and New Zealand	2,685	3,239	4,088	5,146	6,438	8,107
Developing countries total	1,705,858	2,106,177	2,685,013	3,351,693	4,114,676	4,876,709
Least developed countries	200,380	246,145	312,714	400,164	516,907	667,757
Other developing countries	1,505,478	1,860,033	2,372,300	2,951,529	3,597,769	4,208,952
Australia and New Zealand	10,127	12,648	15,355	17,682	20,248	22,937
North America	171,616	204,152	231,937	256,068	283,549	315,915
United States	157,813	186,158	210,111	231,428	255,712	285,003
Canada	13,737	17,909	21,717	24,516	27,701	30,769
Western Europe[c]	327,204	350,926	379,436	397,455	410,806	423,448
Germany	68,376	72,815	78,169	78,289	79,433	82,282
France	41,829	45,684	50,772	53,880	56,735	59,296
United Kingdom	49,816	51,572	54,832	55,530	56,761	58,689
Italy	47,104	50,200	53,822	56,434	56,719	57,536
Spain	28,009	30,455	33,779	37,542	39,303	40,752
Eastern Europe[d]	220,199	253,475	276,419	294,976	310,776	304,538
Russia	102,702	119,906	130,392	138,660	148,292	145,612
Ukraine	37,298	42,783	47,317	50,044	51,891	49,688
Poland	24,824	29,638	32,664	35,574	38,111	38,671
Japan	83,625	94,096	104,331	116,807	123,537	127,034
Developed countries total	812,771	915,298	1,007,479	1,082,989	1,148,917	1,193,872
World total	2,518,629	3,021,475	3,692,492	4,434,682	5,263,593	6,070,581

[a] Consisting of four regions in the United Nations' sense: west Africa, central Africa, east Africa, and southern Africa.
[b] The United Nations western Asia region (including the countries of the Caucasus).
[c] Consisting of three regions in the United Nations' sense: northern Europe, western Europe, and southern Europe.
[d] Including Russian Asia.
Data from estimates adopted by the United Nations for their 2002 population projections (United Nations, 2003).
(Source: United Nations, 2003)

4. The Consequences

a. Population Growth and Economic and Social Development

Thus, at varying degrees Southern countries were or are faced with the difficulties of reconciling population growth with development. Two chapters of Volume III[9] will be dedicated to this very controversial question, and the discussion will not be anticipated here. But it can be understood that the beneficial or inhibiting effects of population growth on economic and social development are likely to be very variable. This is not only because of the diversity of the dynamics but also because of their timing. In particular, the Third World countries that entered first into transition peaked in the 1950s and 1960s, during the "*trente glorieuses*" ("the glorious 30 years") (Fourastié, 1979) of world economic prosperity before the recession caused by the oil crisis in the mid-1970s. Similar to Japan at the beginning of the 20th century, several of these countries—notably in East Asia—managed well and followed in the wake of the developed countries, such as the four "dragons" of the Far East (Singapore, Hong Kong, Taiwan, and South Korea) that were in the news during the 1980s. China itself (although the population growth of this country has never been extravagant) accomplished so many feats that the phrase "Chinese miracle" (Yifu Lin *et al.*, 2000) was born as a result. Thailand, Indonesia, and the Philippines also crossed the milestone of their peak population growth with some economic success. In Latin America the most dangerous period in population growth was also overcome. Countries such as Chile and Mexico had exceptional economic growth rates, which resulted in their rejection of the "developing country" title (which the United Nations' continues to attribute to them) and their creating a free trade zone with the United States (North American Free Trade Agreement[10]).

In contrast, the countries that entered into the demographic transition process later, notably the sub-Saharan African countries, had peak population growth in the 1980s or the 1990s, when the world economic situation had become unfavorable and, in the absence of sufficient financial flow, the International Monetary Fund and the World Bank subjected them to structural adjustment. Public budgets shrunk, the countries disengaged themselves from economic and social action, thus ruining social programs, educational systems, and health policies. And to cap it all, the AIDS epidemic arrived. The last quarter of the 20th century will no doubt be considered for a long time in the history of human population as the period of the highest economic and social disparity, not only between the richest and the poorest countries but also within what was previously called the Third World, between southern countries experiencing real development and those countries with fragile economies that were regressing in relative and sometimes absolute terms.

But beyond this global relation between population growth and economic and social development, two more specifically demographic consequences of the transition should be stressed here that have repercussions just as specific in the economic, social, and political fields: the evolution in age structure and migration (international and internal).

b. The Evolution in Age Structure

Progressively with its generalization in the developing countries, the demographic transition transforms the population pyramid of these countries (Fig. 69–11, upper part).

As was previously the case in Europe (Chapter 68), it is this transformation that will lead to the population aging that is characteristic of post-transitional populations. This change does not, however, lead directly to aging. More so than in Europe, it first of all produced a rejuvenation in the developing countries because of the high population growth caused by the reduction in infant mortality. Therefore, between 1950 and 1975 the population pyramid of the developing countries as a whole started by widening at the base. It was only during a second phase that the population of these countries began to age because of fertility decline: from 1975–2000, the base of the pyramid shrank and the central part widened.

However, in 2000 the process was far from over as can be judged by comparing the upper part of Fig. 69–11 to the central part that illustrates the evolution of the population pyramid of the developed countries over the same three periods: in 2000, the developing countries' population pyramid had not yet achieved the much older shape that has characterized the developed countries since 1950.

All in all, taking into consideration the dominating weight of the developing countries in world population, the same phenomenon of rejuvenation between 1950 and 1975 and the start of aging in the last quarter of the century is seen at the level of the population of

[9] In Chapter 85, Arnaud Dellis and Pierre Pestieau discuss the theories on the association between economic and population growth, and in Chapter 86 Jean Coussy looks more specifically at the links between population and development.

[10] The North American Free Trade Agreement (NAFTA) was signed in January 1994 between Mexico, the United States, and Canada. Shortly afterward, Chile began negotiations to join the free trade zone.

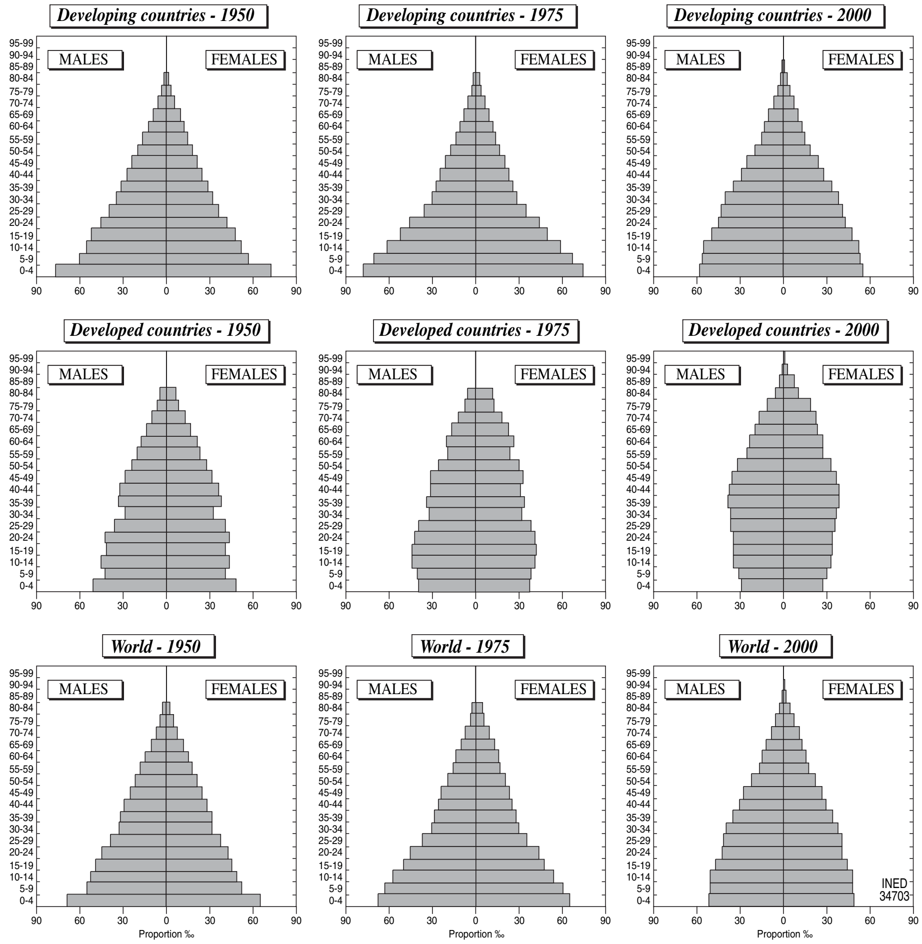


FIGURE 69–11 Evolution of the population pyramid since 1950 in the developing countries compared with that of the developed countries and the world as a whole. (Data from United Nations, 2003.)

the whole world (Fig. 69–11, lower part). The world population was barely younger in 1950 than that of the developing countries, and was a little older in 2000.

But once more, the global trajectory of the developing countries combined is only the result of the considerably different trajectories of the individual countries. Figure 69–12, which illustrates the transformation of the population pyramid for the least developing countries, exposes this clearly: already younger than that of the developing countries combined in 1950, this population continued to become even younger until 1975, and the reflux has started so slowly that the population in 2000 was as young as it was in 1950.

This transformation of the population pyramid is synthesized in Table 69–5 and Figure 69–13, which summarize the compared evolution of the three large age groups, the young (0 to 14 years), working age adults (15 to 59 years), and the elderly (60 years and over). In 1950, 38% of the population of developing countries was under the age of 15. This proportion first increased to 41% in 1975 before beginning its decline and falling to 33% in 2000. But 33% is still higher than the 27% that already prevailed in the developed countries in 1950. In addition, the singularity of the least developed countries is very clear, as 43% for the proportion of the young was still greater in 2000 than was the maximum reached in the developed countries as a whole in 1975 (Table 69–5).

At the other end of the pyramid, the proportion of those more than 60 has only just started to increase: following a slight decrease from 6.6% in 1950 to 6.2% in 1975, in 2000 it was only higher by one point than it was in 1950 (7.7%). During the whole of the second half of the 20th century, it is almost exclusively on the adult proportion that the movement of the young has been added.

This is exposed very clearly in the first graph of Fig. 69–13. Although the proportion of the elderly appears to stagnate at a level that is itself very marginal, a sig-

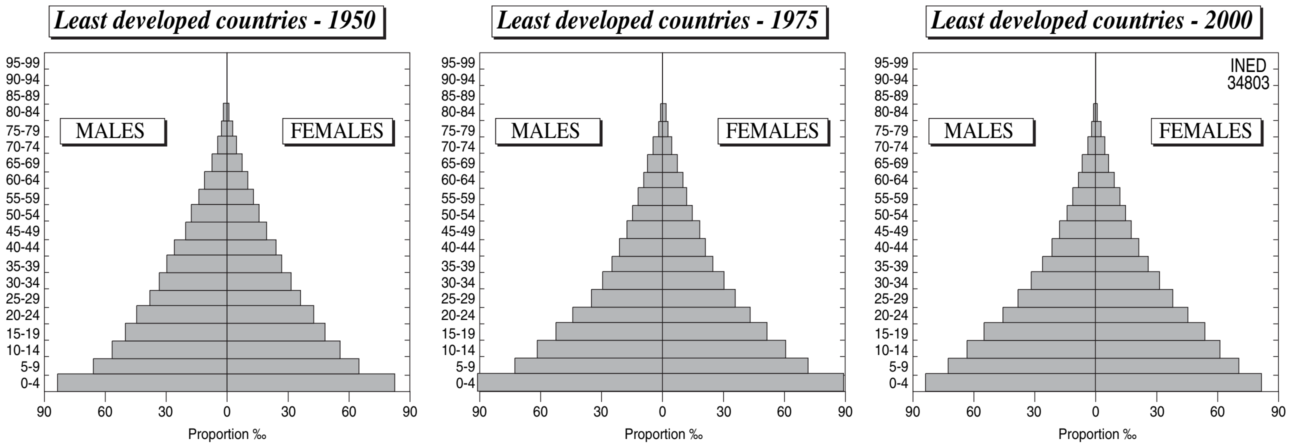


FIGURE 69–12 Evolution of the population pyramid since 1950 in the least developed countries. (Data from United Nations, 2003.)

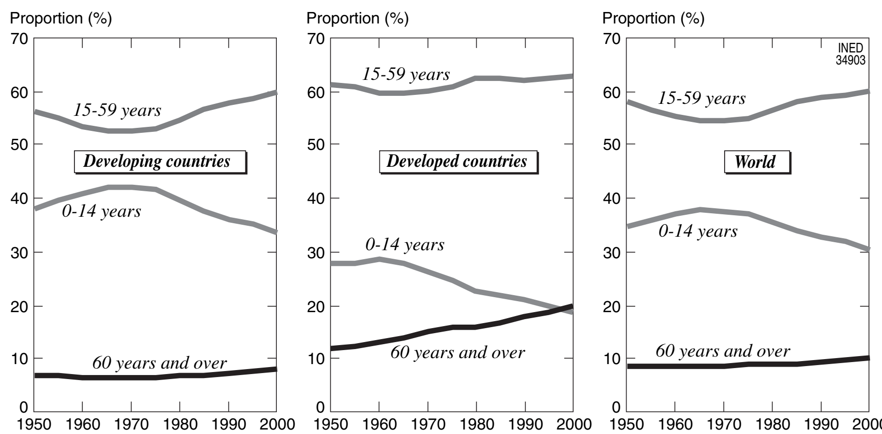


FIGURE 69–13 Evolution since 1950 in the proportions of the three large age groups in the developing countries compared with that of the developed countries and the world as a whole. (Data from United Nations, 2003.)

TABLE 69–5 Distribution of the Population in Three Large Age Groups in 1950, 1975, and 2000 in the Developing Countries Compared with the Developed Countries and the World as a Whole

Age group (years)	Developing countries			Developed countries			World			Least developed countries		
	1950	1975	2000	1950	1975	2000	1950	1975	2000	1950	1975	2000
0–14	37.6	41.1	33.0	27.4	24.2	18.3	34.3	36.8	30.1	41.1	44.7	43.2
15–59	55.9	52.7	59.3	60.9	60.4	62.3	57.5	54.7	59.9	53.5	50.3	52.0
60 and over	6.4	6.2	7.7	11.7	15.4	19.4	8.2	8.6	10.0	5.4	5.0	4.8

nificant shift has occurred between the young and the adults. The proportion of adults declined initially with the increase in the young, but since the mid-1970s, it has increased significantly as a result of the narrowing of the population pyramid's base. And although the level reached in 2000 was still considerably lower than was the level reached by the developed countries in 1950, this increase in adults will continue to grow over the next decades beyond the current level reached in the developed countries (see Chapter 77).

Thus, although during a first phase problems associated with rapid population growth in the developing

TABLE 69–6 1950–2000 Evolution of Urban and Rural Populations in the Developing Countries Compared with the Developed Countries and the World as a Whole

Regions	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000
Urban population											
World	751	873	1017	1184	1357	1543	1756	2001	2286	2565	2862
Developed countries	447	501	563	625	682	734	774	810	846	876	898
Developing countries	304	371	454	559	674	809	982	1191	1439	1690	1964
Least developed countries	15	18	23	31	40	52	69	86	107	136	171
Rural population											
World	1768	1882	2003	2150	2334	2543	2674	2824	2969	3097	3195
Developed countries	367	363	353	342	326	314	309	304	302	298	293
Developing countries	1402	1520	1650	1807	1009	2208	2365	2519	2667	2808	2901
Least developed countries	185	202	222	245	272	301	329	366	408	453	497
Total											
World	2519	2755	3020	3334	3691	4086	4430	4825	5255	5662	6057
Developed countries	814	864	916	967	1008	1048	1083	1114	1148	1174	1191
Developing countries	1706	1891	2104	2366	1683	3017	3347	3710	4106	4498	4865
Least developed countries	200	220	245	276	312	353	398	452	515	589	668
Proportion of urban population											
World	29.8	31.7	33.7	35.5	36.8	37.8	39.6	41.5	43.5	45.3	47.3
Developed countries	54.9	58.0	61.5	64.6	67.7	70.0	71.5	72.7	73.7	74.6	75.4
Developing countries	17.8	19.6	21.6	23.6	40.0	26.8	29.3	32.1	35.0	37.6	40.4
Least developed countries	7.5	8.2	9.4	11.2	12.8	14.7	17.3	19.0	20.8	23.1	25.6

Data from United Nations, 2003.

countries were combined with a specific problem of aggravation of the dependency ratio,[11] during the second phase (beginning in the mid-1970s) growth was accompanied by a distinct improvement in this ratio, which was favorable to an economic takeoff. However, the optimism of such an assessment must be moderated. On the one hand, it is once more based on a very general view that includes numerous exceptions: the least developed countries as a whole have not benefited from any advantage of this kind. But, on the other hand, even in the more numerous cases in which the dependence ratio has become favorable, this is not enough to ensure economic takeoff; it is only one factor among others as it is of little use to have a large working age population if a considerable proportion is unemployed. It is important that the developing countries take advantage of this *demographic window*, as it is open only for quite a short period of time and will be quickly followed by more rapid population aging than occurred in the developed countries.

c. Migration and Urbanization

As in Europe, the exceptional population surplus caused by the demographic transition provoked a double migratory pressure in the developing countries, with a growing proportion of the rural population migrating to the cities and causing a mass increase in candidates for expatriation.

The Urban Explosion

According to the United Nations (2002), of the 1.7 billion people living in developing countries in 1950, 1.4 lived in rural areas and only 0.3 in urban areas; by 2000 from a total of 4.9 billion, 1.9 lived in urban areas and 2.9 in rural zones (Table 69–6).[12] The urban population of developing countries has thus multiplied by six, whereas the rural population has only just doubled. The urbanization rate of developing countries has considerably increased from less than 18% in 1950 to more than 40% in 2000. This level is certainly far from the urbanization level of the developed countries as that was already 55% in 1950 and reached 75% in 2000, but there is a clear narrowing of the gap.

However, the level of urbanization varies significantly even within the group of developing countries, and this diversity can not be clearly associated with the timing of the demographic transition. Although it is

[11] Ratio of the combination of young and elderly (inactive ages) to adults (active age).

[12] These results, like all the results in this section, should be taken with precaution, insofar as the definition of urban population varies from one country to another (see Chapter 61 of Volume II). The United Nations has endeavoured to harmonize the national data as much as possible, but the data still remain influenced by the initial definitions on which they were based.

easy to understand that Africa, which entered later into the transition, is also less urbanized (the urbanization rate increased from 14.7% in 1950 to 37.2% in 2000), the same is not true of Asia and Latin America. Indeed although the urbanization rate of Latin America was already high in 1950 (41.9%) and had by 2000 caught up to the level of the developed countries (75.4%), Asia seems to have experienced a totally different process, with a very low level in 1950 (17.4%) and an identical level to Africa in 2000 (37.2%).

In any case, the second half of the 20th century saw the proliferation of the large cities of the south. In 1950, of 9 agglomerations with more than 5 million inhabitants, only 2 were located in the developing countries: Shanghai and Buenos Aires, each with barely more than 5 million inhabitants. In contrast, New York was on top of the list with over 12 million (Table 69–7), and with London, Paris, Moscow, and Tokyo, was one of the 5 most populated cities of the world. A quarter of a century later, of 22 agglomerations with more than 5 million inhabitants, 12 of them (more than half) were cities in the Third World. Although it is true that the majority of them were still far behind the leader Tokyo (19.8 million inhabitants) and New York (15.6); of the 5 cities with more than 10 million inhabitants, three were located in the Third World: Shanghai (11.4), Mexico City (10.7), and Sao Paulo (10.3). In 2000, more than three-quarters of the agglomerations with more than 5 million inhabitants were in developing countries: 29 out of 37! And although Tokyo remained in

TABLE 69–7 Agglomerations with over Five Million Inhabitants in 1950, 1975, and 2000

	1950		1975		2000
1 New York	12.3	1 Tokyo	19.8	1 Tokyo	26.4
2 London	8.7	2 New York	15.9	**2 Mexico**	**18.1**
3 Tokyo	6.9	**3 Shanghai**	**11.4**	**3 Sao Paulo**	**18.0**
4 Paris	5.4	**4 Mexico**	**10.7**	4 New York	16.7
5 Moscow	5.4	**5 Sao Paulo**	**10.3**	**5 Bombay**	16.1
6 Shanghai	**5.3**	6 Osaka	9.8	6 Los Angeles	13.2
7 Rhin-Ruhr Nord	5.3	7 Buenos Aires	9.1	**7 Calcutta**	**13.1**
8 Buenos Aires	**5.0**	8 Los Angeles	8.9	**8 Shanghai**	**12.9**
9 Chicago	5.0	9 Paris	8.9	**9 Dhaka**	**12.5**
		10 Beijing	**8.5**	**10 Delhi**	**12.4**
		11 London	8.2	**11** Buenos Aires	**12.0**
		12 Rio de Janeiro	**8.0**	**12 Djakarta**	**11.0**
		13 Calcutta	**7.9**	**13 Osaka**	**11.0**
		14 Moscow	7.6	**14 Beijing**	**10.8**
		15 Bombay	**7.3**	**15 Rio de Janeiro**	**10.7**
		16 Seoul	**6.8**	**16 Karachi**	**10.0**
		17 Chicago	6.7	**17 Manila (Metro)**	**10.0**
		18 Rhin-Ruhr Nord	6.4	**18 Seoul**	**9.9**
		19 Tianjin	**6.1**	19 Paris	9.6
		20 Cairo	**6.1**	**20 Cairo**	**9.5**
		21 Milan	5.5	**21 Tianjin**	**9.2**
		22 Manila (Metro)	**5.0**	**22 Istanbul**	**9.0**
				23 Lagos	**8.7**
				24 Moscow	8.4
				25 London	7.8
				26 Bangkok	**7.4**
				27 Chicago	7.4
				28 Tehran	**7.0**
				29 Hong Kong	**7.0**
				30 Bogota	**6.8**
				31 Rhin-Ruhr Nord	**6.5**
				32 Madras	**6.4**
				33 Bangalore	**5.6**
				34 Lahore	**5.5**
				35 Hyderabad	**5.4**
				36 Wuhan	**5.2**
				37 Kinshasa	**5.1**

Values are in millions. Cities in developing countries are in bold.
Data from United Nations, 2003.

the lead with 26.4 million, 14 of the 17 cities of over 10 million were in developing countries. Two cities in Latin America, Mexico City and Sao Paulo, have overtaken New York.

In theory, the city is the best outlet for the population surplus of the transition as it is mainly in cities that nonagricultural employment resulting from economic development is created. Nevertheless, taking into account that rural zones are incapable of absorbing the entire surplus workforce from the transition, the city has often attracted more people than it can employ. Urban growth has therefore largely consisted of extremely poor populations squeezed into unhealthy accommodation in slums and other favelas. The situation is more difficult to control when the growth is so fast. Certain authoritarian countries such as China have taken draconian measures in an attempt to stem back the flow (see Chapter 112 of Volume IV), but the suffering caused by these restrictions is hardly less.

The Barricades of the North

International migratory flows have already been discussed at length in Chapter 58 of Volume II. It is therefore not useful to go into it in detail here. Let us say that, despite the very strong demographic pressure exerted on the southern countries by a transition definitely faster and more explosive than the one Europe experienced in the 19th century, the migratory flows from the Third World countries toward the developed countries during the second half of the 20th century are, all things considered, without comparison with the European flows toward the rest of the world in the previous century. The International Organization for Migration estimates that immigrants made up only 2.9% of the world population in 2000 and that this proportion was only slightly higher than it was in 1965 (Table 69–8).

TABLE 69–8 Proportion of Immigrants in World Population

Population	1965	1975	1985	2000
Total population	3333	4066	4825	5882
Immigrants	75	84	105	175
Proportion (%) of immigrants	2.3	2.1	2.2	2.9

Data from IOM, 2003.

The proportion of immigrants in the total population certainly varies considerably from one continent to another. Although it is only 1.1% in Latin America, 1.4% in Asia, and 2.1% in Africa, it is 7.7% in Europe, 13% in North America, and as high as 19.1% in Oceania. Nevertheless, this proportion of the southern countries in the current population of the northern countries is far from that constituted in the past by the proportion of the European population in the populations of large previously poorly inhabited areas such as North America, Australia, New Zealand, and northern Asia. Migration resulting from the European transition submerged local populations and caused some of them to disappear totally, to the extent of transforming these territories into European subsidiaries. But even in Central and South America, historically more populated, the European contribution, with time, more often than not exceeded the population of the now so-called Latin America.

When Europe was exporting its population surplus overseas, no barriers could prevent the arrival of the conquering invaders endowed with military, economic, and financial means superior to those of the locals. On the contrary, when the demographic transition arrived in the southern countries, Europe with its offshoots in North America, North Asia, and Oceania, and stronger than ever before compared to the rest of the world, ensured that it could use at will the population surplus of the Third World. When after World War II there was much reconstruction to do and few hands available, the importation of labor was a widespread practice, continuing as long as economic growth made it necessary. But when the oil crisis of the 1970s and the new direction of the world economy that followed made the importation less profitable and the often inadequate treatment of immigration resulted in serious integration problems, the borders were closed. Certainly there has never been a total closure of borders and it is often bypassed by illegal immigra-

TABLE 69–9 Proportion of Immigrants in Total Population by Continent

Population	Asia	Africa	Europe	Latin America	North America	Oceania	World
Total population	3672	794	727	519	313	31	6057
Immigrants	50	16	56	6	41	6	175
Proportion (%) of immigrants	1.4	2.1	7.7	1.1	13	19.1	2.9

Data from IOM, 2003.

tion, but it was effective enough for this immigration to absorb only a tiny proportion of the potential emigration from developing countries toward developed countries.

II. INCREASING UNCERTAINTY AND THE DEATH OF THE TRANSITION PARADIGM

The evolution of the population of developing countries observed since World War II is assuredly a triumph for the demographic transition theory, but it is a limited triumph that has finally been of short duration. That which has just been described in this chapter is a generalization of the process observed in Europe a century previously. Possibly is could be called a transposition with a general difference (the acceleration of the process) and a wide diversity in application of the founding principle: demographic behavior has evolved owing to the modernization of societies, resulting in mortality decline and in fertility decline. The resulting phase of exceptional growth has only been transitory, and in this sense, the transition theory has had a major role in the explanation of a phenomenon that caused such concern during the 1960s and 1970s. It enabled scientists to propose rational forecasts and the governing powers to establish policies to complement or control the process.

However, the triumph of the theory disguised its limits. The reassuring idea of a convergence of the world population toward a modern demographic system in which low fertility would be sufficient to balance low mortality, thus leading to the absorption of the exceptional population surplus, has helped us forget that rapid population growth was not necessarily the only important issue for the future. The fact that some recent developments such as the fluctuation of fertility in Europe and the eruption of AIDS in Africa raised doubts concerning the convergence principle is not the worst. The main error is to have believed that convergence leads to stabilization. This idea of stabilization did not appear in the work of Adolphe Landry (1934), who was concerned with the risk of depopulation in Europe. It only appeared later in the work of Kingsley Davis (1945) and Frank Notestein (1945) and even later in the work of Thomas Frejka's on long-term population forecasts (1973), as a natural consequence of the fact that the Southern populations would in their turn experience a decline in fertility that would soon result in a near zero growth. It was reinforced by the success of the United Nations' population projections, which were until recently based on this convergence-stabilization principle. From the idea, soon to become reality, that all populations would approach the zero growth level, we progressed to the hypothesis that once this level was attained we would remain there.

The stabilization principle presumes that in a permanent system, fertility and mortality would become stable at a point of balance at which growth is zero and the population is *stationary* in the technical meaning of the term (see Chapter 20 of Volume I). In fact, as soon as mortality before the age of 50 becomes so low as to be nearly negligible, this balance requires a fertility rate of just below 2.1 children per woman.[13] At the highest life expectancy levels, whatever they may be, the first condition is therefore that fertility stabilizes at 2.1 children per woman; the second, that life expectancy itself stabilizes at a high level, whatever that may be. Regarding both of these aspects, the evolution observed in the most advanced countries moves increasingly further away from the demographic transition model, and if convergence is still possible, it is unlikely that it will lead to stabilization.

1. Fertility Crisis or a New Fertility System for Europe?

Although the secular fertility decline resulted in near replacement levels just before World War II in most European countries, nearly all the industrial countries later experienced an increase in fertility, the volume of which surpassed the ordinary effects of recuperation of the birth deficit of the war years. Fertility increased to levels over 2.1 children per woman everywhere, and this occurred over periods of varying duration, not only from the end of the 1940s and during the 1950s but also often in the 1960s and up to the beginning of the 1970s.

Figure 69–14 shows the United Nations' 5-year estimates (2003), which systematically retrace the trends during 1950–2000 of each country in Europe, North America plus Australia, New Zealand, and Japan. Certainly some countries such as Albania, Macedonia, Bosnia, and Malta had not as yet completed their demographic transition with fertility levels between 4 and 6 children per woman in 1950–1955. A few others are different for having not experienced the postwar baby boom and retaining (or reaching) a fertility level very near to 2.1 children per woman at the beginning of the 1950s. This is the case for Sweden and Greece and also countries of the former Soviet Union such as Latvia and Estonia. All the other countries experienced a period of relatively high fertility during the 1950s or

[13] Exactly 2.05 if mortality is zero before the age of 50 and if the sex ratio at birth is 105 boys to 100 girls.

1960s, with over 2.5 children per woman or, in some cases, over 4. However, this phenomenon did not occur in the same way or have the same signifiance everywhere.

In the countries of eastern Europe (Fig. 69–14, first two graphs) and in Japan (Fig. 69–14, seventh graph) the secular decline in fertility was only interrupted by a short baby boom or not interrupted at all, as in Russia, and the decline observed at the beginning of the 1950s is well in line with this secular decline (Chesnais, 1999).

On the contrary, in all the other countries fertility increased well over a mere recuperation of the deficit during the war. Indeed, a significant fluctuation in fertility was observed, which led to its culminating to quite high levels at the beginning of the 1960s and even later in some cases. Although this culminating level was altogether quite modest at around 2.5 children per woman in countries such as Germany, Switzerland, and Belgium, it was close to 3 in France, Austria, Norway, Italy, Spain, and Portugal; it exceeded 3.5 in Australia and the United States and reached as high as 4 in Ireland, Iceland, Canada, and New Zealand. But in an almost perfectly synchronized movement, all these countries then converged in a very pronounced decline in fertility, which resulted in levels below 2.1 children per woman from the mid-1970s.

This large fluctuation of fertility in the Western countries caused much debate. Was it the first fluctuation of the wave toward the fateful replacement level to which the demographic transition was supposed to lead all populations? Or was it an exceptional interlude in the secular fertility decline, which could then continue durably below replacement level? Different interpretations of the decline started to be made in 1965. Roland Pressat (1969a,b), for example, showed

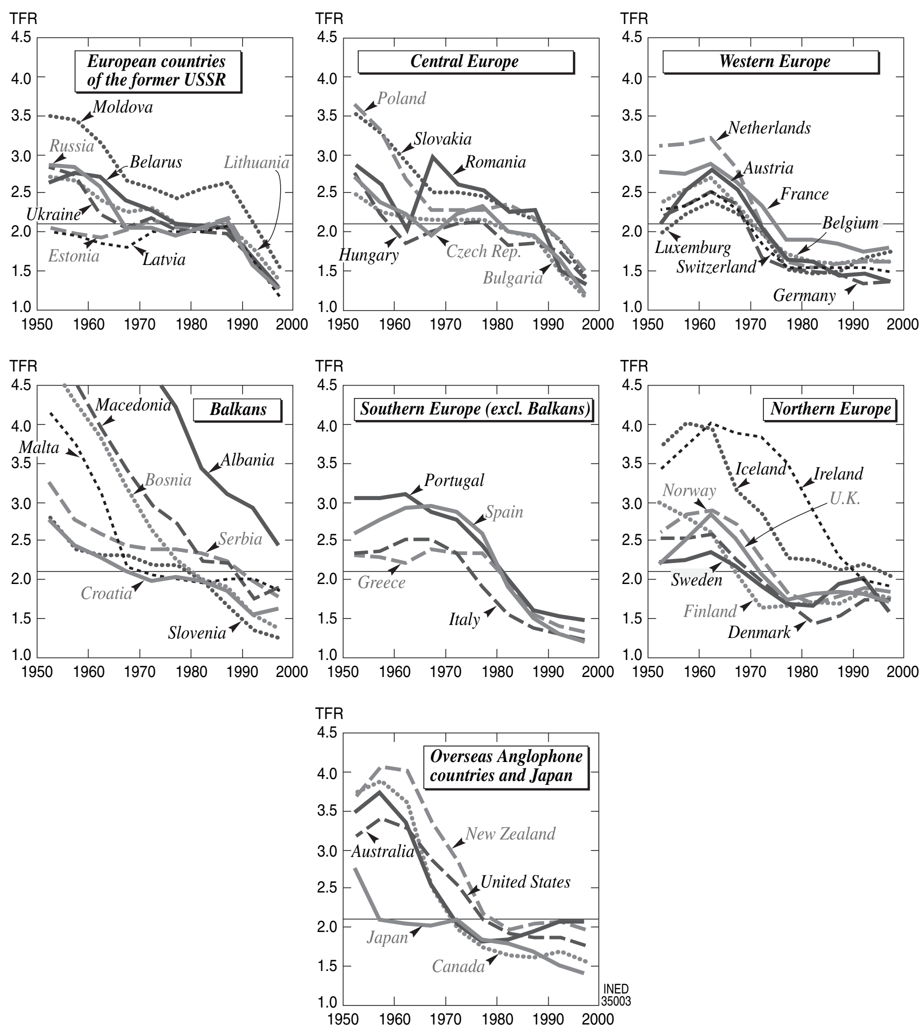


FIGURE 69–14 Evolution in total fertility rate in the developed countries since 1950. (Data from United Nations, 2003.)

that in the case of France, it could only be owing to the direct consequences of an evolution in age at first marriage: it was the reduction of age at first marriage that largely explained the increase in fertility, and the increase naturally caused the reverse to occur. Gérard Calot (1969), however, considered that it was owing more fundamentally to a reversal in fertility trends.

A little later, a demographic-economic explanation based on the ratios between generations was suggested by Richard Easterlin (1976). When the generation of reproductive age arrives on the employment market in greater numbers than did its predecessors, members of the generation experience a problems in the course of their careers that could dissuade them from having as many children as did the previous generation. If on the other hand, the new generation is less numerous, then its members integrate more easily into society and are therefore more likely to trust in the future and have more children. Thus by comparing over a period of half a century the evolution in the ratio of 35- to 64-year-olds to 15- to 34-year-olds to that of the average number of children per woman, Richard Easterlin constructed a model with results that match the reality observed in the United States. In turn, the movement of fertility affects the population pyramid, and it is then possible to imagine a series of self-sustained fluctuations associated with the interaction between age structure and fertility. Of course, this explanation could not be verified simply by observing or matching the fluctuations in a country, and the events that followed did not confirm Easterlin's theory.

More than a quarter of a century has gone by since fertility dropped below 2 children per woman in the majority of European countries without a revival preceding a new fluctuation of the same nature appearing anywhere. Some increase in fertility was observed in the 1990s, notably in the United States, but this phenomenon is rare and nowhere has fertility risen well above the 2.1 children per woman level anywhere. On the contrary, fertility has fallen below this level nearly everywhere, even in countries where it was least expected, such as Canada, Ireland, and Portugal, and even in Eastern European countries, which until the fall of communism had succeeded in maintaining quite constant levels of nearly 2 children per woman.

In 2000 no European country exceeded 1.9 children per woman except Iceland, which has fewer than 300,000 inhabitants. The TFR was between 1.8 and 1.9 only in France, Ireland, Norway, and Macedonia. It was below 1.5 in more than half of the countries and below 1.3 in nearly 20 of them (Spain, Greece, Italy, Bosnia, Slovenia, Bulgaria, Hungary, Poland, Rumania, Czech Republic, Slovakia, Russia, Estonia, Latvia, Lithuania, Byelorussia, Moldavia, Ukraine). Japan was at 1.4. Apart from the notable exception of the United States, developed countries increasingly appear to be established in a long phase of low fertility that is significantly below replacement level.

In a specific analysis of France, Henri Leridon (1987) showed that although a *second contraceptive revolution* did occur (in France from the 18th century to the 1960s, the most frequently used method was coitus interruptus), neither the wide diffusion of modern contraception (pill, IUD, sterilization) nor the liberalization of induced abortion explained the fertility decline observed since the mid-1960s. This revolution certainly resulted in a reduction in the proportion of unwanted children, but it was in no way at the root of the decline in the desire to have a child.

Beyond the benefits of technological developments that make fertility control more comfortable but do not determine its objectives, is there a *second demographic transition* (Lesthaeghe and van de Kaa, 1986; van de Kaa, 1987, 1988, 1994)? Ron Lesthaeghe and Dirk van de Kaa created this term to signify that the fertility decline observed in Europe since the 1960s did not result from the same factors as did the secular decline observed up to the World War II. On the contrary, it is caused by new changes in society that lead to new demographic behavior that could be called *postmodern.* During the past 3 or 4 decades, the family and its place in society have evolved with the reduction in official marriages, the development of cohabitation, the increase in divorces, the appearance of recomposed families, acknowledgement of homosexual couples, and the multiple facets of cohabitation outside marriage, from *couples living apart together* (Villeneuve-Gokalp, 1997) to the *PACS*.[14] At the same time, fertility outside marriage has itself changed completely both in volume and in meaning to the extent that in numerous European countries more than half of births occur outside marriage, and the evolution of these births contribute more to the TFR than does fertility within marriage. However, this does not necessarily mean that fertility decline is motivated by factors other than those that prevailed until the 1960s. Although these factors have certainly evolved since the 18th century, in particular, the evolution in women's role in society has had a growing effect, this phenomenon did not begin in the 1960s or even after the World War II. Rather than a drastic change in the determinants of the decline, one should speak of further developments in

[14] *Pacte civil de solidarité*, a status created in France, with legislation 99–944 of November 15, 1999, for unmarried couples, both heterosexual and homosexual, specifying a certain number of rights and duties pertaining to their conjugal union.

old factors. The necessity for women to have professional careers equal to that of men, one of the key factors in the most recent fertility decline, is only the outcome of feminist campaigns that began during the 19th century.

Whatever the case, whether the demographic transition theory is being replaced by a second transition theory or whether the failure of the paradigm of the stabilization of population numbers should simply be acknowledged, the general convergence toward fertility rates corresponding to replacement levels is as ineffective in explaining current trends in the developed countries as it is in making long-term forecasts for the developing countries.

2. Life Expectancy: Unexpected Divergence, Unknown Limits

If the recent history of European countries has significantly distanced us from a general system of fertility convergence toward near replacement level, the situation is hardly better from the point of view of mortality. Indeed, the unexpected evolution in mortality in several regions has distanced us from the idea of the convergence of life expectancies, but the level at which this convergence was expected has been repeatedly surpassed by reality.

a. Africa's Lateness

Although sub-Saharan Africa has just recently progressed firmly on the path of fertility decline after a long delay, the reverse has occurred in the context of health and mortality. As was the case for fertility, not only has progress been delayed but also, in many cases and notably where it was most rapid, the increase in life expectancy has been seriously affected by the AIDS epidemic.

Figure 69–15 (left) summarizes the evolution in life expectancy since 1950 in the sub-Saharan African countries least affected by the AIDS epidemic and compares these trajectories to Indonesia and Yemen, two developing countries that have followed a path more characteristic of Third World countries during the same period. This figure taken from a recent study (Vallin and Meslé, 2004) also shows, in dotted lines, the extreme values of the evolution previously observed in the developed countries. The rate of progress that all of the African countries have experienced has not been any faster than that previously experienced by the developed countries, despite the time difference. In other words, these countries have in no way benefited from the acquired knowledge that enabled most of the other countries to catch up significantly with the developed countries.

But this is not the only source of African disappointment because, as is shown in Fig. 69–15 (right), the other sub-Saharan African countries, some of which benefited during the 1950s, 1960s, and 1970s from progress nearly as rapid as in Yemen and Indonesia, have been so violently affected by the AIDS epidemic that life expectancy has abruptly dropped. In some cases the decline is such that at the end of the 1990s, life expectancy fell to the levels of the early 1950s and even lower. After gaining 20 years of life expectancy (from 44 to 64 years) between 1950–1955

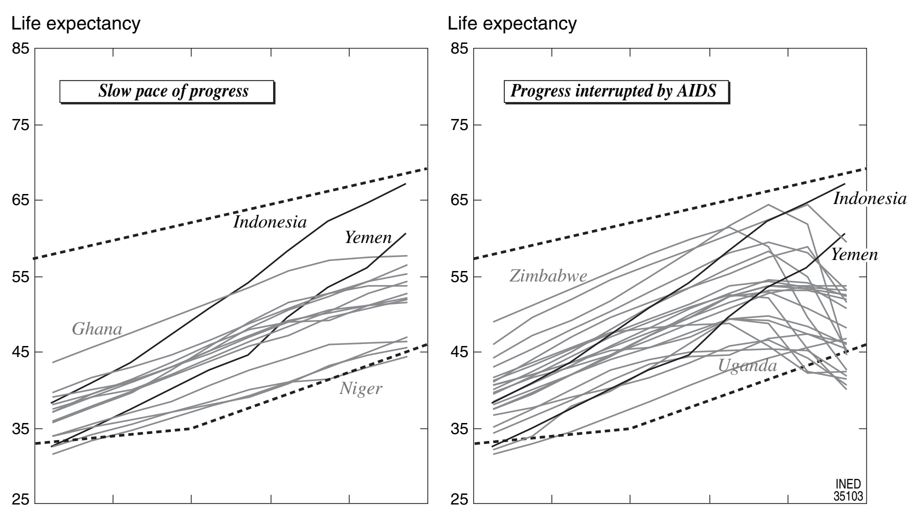


FIGURE 69–15 Evolution of life expectancy in African countries compared to two Asian countries. The dotted lines represent the extreme trajectories observed in Europe at comparable life expectancy levels. (Data from Vallin and Meslé, 2004.)

and 1980–1985, Botswana lost nearly as many years in 1995–2000, decreasing to under 45 years. The increase was less in Zimbabwe (13 years, from 49 to 62), but in 1995–2000 the decline resulted in a life expectancy of only 43 years, 6 years below its level in 1950–1955.

However, this double distortion does not fundamentally call into question the principle of epidemiological transition proposed by Abdel Omran (1971) to explain the mortality decline that is the basis of the demographic transition: if the modernization of society leads to the near eradication of infectious diseases, *a contrario,* some countries can, for reasons pertaining to their history, economy, or culture, be faced with serious obstacles in the attaining of this transition (Caselli *et al.,* 2001, 2002). Sub-Saharan African countries, more permeable than many others to a generalized spread of AIDS as a result of multiple partners and the considerable variability in age differences between partners (Brouard, 1994), were struck in full force by the epidemic at a stage when their still fragile economy was faced with the world economic crisis that had already depleted the meager health services of these countries. Prevention methods and, more recently, treatment methods can be effective, as has been seen in the northern countries, but these remain beyond the financial means of Africa.

b. Crisis in Eastern Europe

The second considerable distortion of the principle of the epidemiological transition according to Omran is that far from ceasing to progress after the victory over infectious diseases, life expectancy has been increasing since the 1970s because of successes on new fronts, such as treatment of cardiovascular diseases. However, it has been possible to reconsider the concept of epidemiological transition to the benefit of the larger concept of health transition (Frenk *et al.,* 1991), taking into consideration health effects of the modernization of society other than the control of infectious diseases in order to better understand the evolution in mortality observed during the demographic transition (Chapter 57 of Volume II). In the light of this, it is that the eastern European countries have been unable to enter into the second phase of the transition that is striking (Vallin and Meslé, 2001). In the mid-1960s, the life expectancy of the countries of central Europe, and even more so of the former Soviet Union, ceased to increase and entered into a long period of stagnation and even a decline of varying severity, notably among Russian men (Shkolnikov *et al.,* 1995a,b) and in the Ukraine (Meslé *et al.,* 2003), leading to a wide divergence between these countries and the other industrialized countries (Fig. 69–16). Communist regimes relied too exclusively on administration from the top of modern medical methods to ensure progress in public health, whereas the second phase of the health transition requires significant changes in individual behavior and thus the active participation of citizens in taking responsibility for their health. On the other hand, the economies of these countries were engaged in a ruinous competition with the West in arms and space programs and were unable to provide the necessary means for a reorientation toward an effective health policy in the area of chronic diseases (Caselli *et al.,* 2001, 2002; Vallin and Meslé, 2004).

c. The New Stages of the Health Transition

In reality, this divergence does not call into question the principle of the health transition itself as much as that of its continuity and uniformity. Like all progress, medical progress associated with economic and social modernization relies on successive inventions, the sequence and the effects of which depend considerably on the history and context of each country and each population. As a result, all major progress—in the past against infectious diseases, today against cardiovascular diseases and cancer, and tomorrow against senescence—is likely to cause an initial divergence to the benefit of the populations best prepared to benefit from it. Then a convergence occurs when the other populations finally succeed in finding ways to enable them to catch up (Vallin and Meslé, 2004). Moreover, Fig. 69–16 shows that some countries of the former communist bloc, such as the Czech Republic, Poland, and Hungary, have experienced an increase in life expectancy during the last decade of the 20th century.

In the same way, some already very advanced countries such as Japan and France continue to make rapid progress and are reaching life expectancy levels (at least its women) that were previously thought impossible, whereas others such as The Netherlands are running out of steam. This is possibly a sign of the beginning of a new phase in progress in the combat against senescence and, with it, a new divergence between pioneering countries and others (Vallin and Meslé, 2004).

In this historic succession of divergence followed by convergence, the pioneering countries are not always the same ones. Retaining for each period the life expectancies of the best-placed countries only, Jim Oeppen and James Vaupel (2002) showed that over 2 centuries the maximum human lifespan is progressing arithmetically. Not only has this progression never stopped, but it has not even decelerated. Without going so far as to follow the investigators in their hypothesis

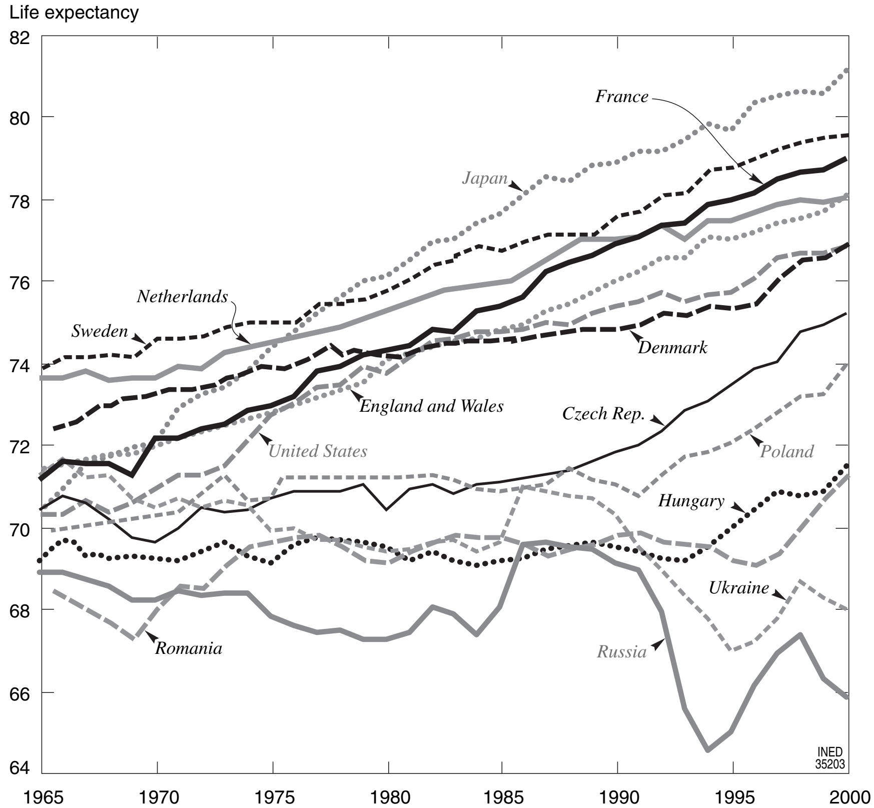


FIGURE 69–16 Evolution since 1965 of life expectancy in some eastern countries compared with some western countries. (Data from Vallin and Meslé, 2004.)

that this could continue for a long time again, the timidity of the forecasts made in the past by the investigators of population projections inspired by the demographic transition theory should be stressed. The United Nations, for example, was in the 1970s still using a life expectancy limit of 75 years as that to which each population should converge. During the 1980s this limit was increased to 85 years. It was only in 1998 that the United Nations ceased to impose an a priori life expectancy limit on its projections, although it nevertheless used mortality tables that did not go beyond a life expectancy of 92.5 years for any sex, even though we know that in a country such as France, a simple extrapolation of recent mortality trends will result in a female life expectancy of 95 years by the end of the century (Vallin and Meslé, 2001). If a limit does exist to the increase in life expectancy, no one knows its level, and the theory of a forthcoming stabilization cannot be retained for mortality any more than for fertility.

CONCLUSION

In reality, as we will see in Chapter 77, if the demographic transition theory has enabled solid world population projections to be made as long as the southern countries (which by their numbers constitute the main concern) were in mid-transition, the fact that the majority of these countries have completed most of the process and therefore have characteristics that more closely resemble those of the Northern countries gives an increasingly global scale to the considerable uncertainties that weigh on the European populations whose demographic behavior contradicts the paradigm of a general stabilization of the world population. As paradoxical as it may seem, at the beginning of the 21st century when population issues seem to be a less major concern for world leaders, the concerns that they raise in the scientific world are much greater now than when the southern countries were at the peak of their growth.

Not only are we realizing that a general and regular convergence of all these parameters is further from the reality of the 20th century than we believed (Coleman, 2002), but more so, we are discovering that the stabilization of population numbers, the *stationary population,* is far from being the most probable model for the 21st century, thereby justifying the most extravagant hypotheses for the long-term future, as we will see in Chapter 78.

Acknowledgement

This chapter was translated from the French version by Accenta, Ltd.

References

Attané Isabelle, 2002. China's family planning policy: an overview of its past and future, *Studies in Family Planning,* vol. 33(1), p. 103–113.

Ayed Mohamed and Jemai Hédi, 2001. Les déterminants de la fécondité, *in:* Jacques Vallin and Thérèse Locoh (eds.), *Population et développement en Tunisie: la métamorphose,* p. 171–202. Tunis, Cérès éditions, 801 p.

Balandier Georges (ed.), 1956. *Le "Tires Monde." Sous-développement et développement.* Paris, INED, PUF, 394 p.

Ben Salem Lilia and Locoh Thérèse, 2001. Les tranformations du mariage et de la famille, *in:* Jacques Vallin and Thérèse Locoh (eds.), *Population et développement en Tunisie. La métamorphose,* p. 143–169. Tunis, Cérès Editions, 802 p.

Bongaarts John, 1978. A framework for analyzing the proximate determinants of fertility, *Population and Development Review,* vol. 4(1), p. 105–132.

Bourgeois-Pichat Jean, 1966. *Évolution et croissance démographique.* Geneva, Dotation Carnegie pour la paix internationale, Centre européen, 68 p.

Breil Jacques and Biraben Jean-Noël, 1969. Essai d'estimation des naissances de la population algérienne depuis 1891, *Population,* vol. 24(4), p. 711–114.

Brennan-Galvin Ellen, 2003. Vers une domination des villes: une adaptation difficile, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux,* p. 581–600. Paris, INED, 768 p.

Brouard Nicolas, 1994. Aspects démographiques et conséquences de l'épidémie de sida en Afrique, *in:* Jacques Vallin (ed.), *Populations africaines et sida,* p. 120–178. Paris, La Découverte, CEPED, 224 p.

Calot Gérard, Hémery Solange, and Piro Christian, 1969. L'évolution récente de la fécondité et de la nuptialité en France, *Population,* vol. 24(2), p. 263–291.

Caselli Graziella, Meslé France, and Vallin Jacques, 1999. Le triomphe de la médecine, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998,* p. 125–181. Paris, Librairie Arthème Fayard, 792 p.

Caselli Graziella, Meslé France, and Vallin Jacques, 2001. *Les entorses au schéma de la transition épidémiologique.* Rome, Paris, DSD, INED, 43 p.

Caselli Graziella, Meslé France, and Vallin Jacques, 2002. Epidemiologic transition theory exceptions, *Genus,* vol. 58(1), p. 9–52.

Chasteland Jean-Claude and Chesnais Jean-Claude (eds.), 2003. *La population du monde: géants démographiques et défis internationaux.* Paris, INED, 768 p.

Chesnais Jean-Claude, 1986. *La transition démographique, étapes, formes, implications économiques.* Paris, INED and PUF, 580 p.

Chesnais Jean-Claude, 1999. La fécondité au XXe siècle: une baisse irrégulière mais profonde et irresistible, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe, III: Les temps incertains, 1914–1998,* p. 183–222. Paris, Librairie Arthème Fayard, 792 p.

Chesnais Jean-Claude, 2003. La transition démographique: 35 ans de bouleversements (1965–2000), *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux,* p. 455–475. Paris, INED, 768 p.

Coleman David A., 2002. Populations of the industrial world. A convergent demographic community, *International Journal of Population Geography,* vol. 8, p. 319–344.

Davis Kingsley, 1945. The world demographic transition, *Annals of the American Academy of Political and Social Science,* no. 273, p. 1–11.

Easterlin Richard A., 1976. The conflict between aspirations and resources, *Population and Development Review,* vol. 2(3), p. 417–426.

Fourastié Jean, 1979. Les Trente Glorieuses ou la révolution invisible de 1946 à 1975. Paris, Fayard, 288 p.

Frederiksen Harald, 1966. Determinants and consequences of mortality and fertility trends, *Public Health Reports,* vol. 81, p. 715–727.

Frejka Thomas, 1973. *The future of population growth: alternative paths to equilibrium.* New York, John Wiley and Sons, 268 p.

Frenk Julio, Bobadilla José Luis, Stern Claudio, Frejka Tomas, and Lozano Rafael, 1991. Elements for a theory of the health transition, *Health Transition Review,* vol. 1(1), p. 21–38.

Guzman José-Miguel, Singh Susheela, Rodriguez German, and Pantelides Edith (eds.), 1996. *The fertility transition in Latin America.* New York, Clarendon Press and IUSSP/UIESP, 450 p.

International Organization for Migration (IOM), 2003. *World migration 2003: managing migration: challenges and responses for people on the move.* Geneva, IOM, 396 p.

Jemai Yolande and Jemai Hédi, 1979. Applications of methods of measuring the impact of family planning programmes on fertility: the case of Tunisia, *in: The methodology of measuring the impact of family planning programmes on fertility: problems and issues,* p. 140–244. New York, United Nations, Department of International Economic and Social Affairs, 253 p.

Kouaouci Ali, 1992. Familles, femmes et contraception: contribution à une sociologie de la famille algérienne. Alger, CENEAP-FNUAP, 279 p.

Ladier-Fouladi Marie, 2003. *Population et politique en Iran.* Paris, INED and PUF, 356 p.

Landry Adolphe, 1934. *La révolution démographique. étude et essai sur les problèmes de la population.* Paris, Sirey, 231 p.

Larsen Ulla, 1994. Sterility in Sub-saharan Africa, *Population Studies,* vol. 48(3), p. 459–474.

Larsen Ulla, 1996. Infécondité au Cameroun et au Nigéria: tendances, *Perspectives internationales sur le planning familial,* special issue, p. 9–14.

Leete Richard and Alam Iqbal (eds.), 1993. *The revolution in Asian fertility.* New York, Clarendon Press and IUSSP/UIESP, 330 p.

Leridon Henri (ed.), 1987. *La seconde révolution contraceptive: la régulation des naissances en France de 1950 à 1985.* Paris, PUF, INED, 380 p.

Leridon Henri and Toulemon Laurent, 2003. La régulation des naissances se généralise, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux,* p. 477–495. Paris, INED, 768 p.

Lesthaeghe Ron J. and van de Kaa Dirk J., 1986. Twee demografische transities? [Two Demographic Transitions?], *in:* Dirk J. van de Kaa and Ron Lesthaeghe (eds.), *Bevolking: groei en krimp,* p. 9–24. Deventer, Van Loghum Slaterus, 115 p.

Locoh Thérèse and Makdessi Yara, 1995. Baisse de la fécondité: la fin d'une exception africaine, *La Chronique du CEPED,* no. 18(July–September), p. 1–4.

Locoh Thérèse and Vallin Jacques, 1998. Afrique noire: la baisse de la fécondité, *Population et Société,* no. 338, p. 1–4.

Meslé France and Vallin Jacques, 1996. *Mortality in the World: Trends and Prospects.* Paris, CEPED, 24 p.

Meslé France and Vallin Jacques, 2003. La mortalité dans le monde: tendances et perspectives, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux,* p. 527–546. Paris, INED, 768 p.

Meslé France, Vallin Jacques, Shkolnikov Vladimir, Pyrozhkov Serhii, and Adamets Serguei, 2003. *Mortalité et causes de décès en Ukraine au XXe siècle.* Paris, INED, 396 p.

Mitchell Brian R., 1995. *International historical statistics: Africa, Asia, and Oceania 1750–1988.* New York, Stockton Press, 1090 p.

Molineaux Louis, 1985a. The impact of parasitic diseases and their control, with special emphasis on malaria in Africa, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 13–44. Liège, Ordina Editions, 558 p.

Molineaux Louis, 1985b. La lutte contre les maladies parasitaires: le problème du paludisme, notamment en Afrique, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 11–50. Paris, INED and PUF, 542 p.

Moreno Lorenzo and Singh Susheela, 1996. Fertility decline and change in proximate determinants in the LatinAmerican and Caribbean regions, *in:* José-Miguel Guzman, Susheela Singh, German Rodriguez, and Edith Pantelides (eds.), *The fertility transition in Latin America,* p. 113–134. New York, Clarendon Press and IUSSP/UIESP, 450 p.

Negadi Gourari, Tabutin Dominique, and Vallin Jacques, 1974. Situation démographique de l'Algérie, *in:* Ahmed Bahri (ed.), *La population de l'Algérie,* p. 1–62. Paris, CICRED, 174 p.

Newman Peter, 1965. *Malaria eradication and population growth: with special reference to Ceylon and British Guiana.* Ann Arbor, The University of Michigan, School of Public Health, 259 p.

Notestein Frank, 1945. Population, the long view, *in:* T. Paul Schultz (ed.), *Food for the world,* p. 36–57. Chicago, University of Chicago Press, 367 p.

Oeppen Jim and Vaupel James W., 2002. Broken limits to life expectancy, *Science,* vol. 296(10), p. 1029–1031.

Olshansky S. Jay and Ault A. Brian, 1986. The fourth stage of the epidemiologic transition: the age of delayed degenerative diseases, *The Milbank Quarterly,* vol. 64(3), p. 355–391.

Omran Abdel R., 1971. The epidemiologic transition: a theory of the epidemiology of population change, *The Milbank Memorial Fund Quarterly,* vol. 49(4), p. 509–538.

Ouadah-Bedidi Zahia and Vallin Jacques, 2000. Maghreb: la chute irrésistible de la fécondité, *Population et Sociétés.* no. 359(June), p. 1–4.

Ouadah-Bedidi Zahia and Vallin Jacques, 2003. écarts d'âge entre conjoints en Algérie: évolution depuis 1966 et disparités régionales, *European Journal of Population/Revue démographique européenne,* vol. 19(3), p. 279–302.

Peng Xizhe, 2002. La fécondité chinoise, constat et perspectives, *in:* Isabelle Attané (ed.), *La Chine en transition, questions de population, questions de société,* p. 59–77. Paris, INED and PUF, 601 p.

Phillips James F. and Ross John A. (eds.), 1992. *Family planning programmes and fertility.* New York, Clarendon Press and IUSSP/UIESP, 340 p.

Poursin Jean-Marie, 1989. *L'Homme stable: essai démographique.* Paris, Gallimard, 380 p.

Pressat Roland, 1969a. Interprétation des variations à court terme du taux de natalité, *Population,* vol. 24(1), p. 47–56.

Pressat Roland, 1969b. Les naissances en France de 1946 à 1980, *Population,* vol. 24(3), p. 417–426.

Preston Samuel H., 1975. The changing relation between mortality and economic development, *Population Studies,* vol. 29(2), p. 231–247.

Rele Jawahar R. and Alam Iqbal, 1993. Fertility transition in Asia: the statistical evidence, *in:* Richard Leete and Iqbal Alam (eds.), *The revolution in Asian fertility,* p. 15–37. New York, Clarendon Press and IUSSP/UIESP, 330 p.

Rétel-Laurentin Anne, 1974. *L'infécondité en Afrique Noire: maladies et conséquences sociales.* Paris, Masson, 188 p.

Rétel-Laurentin Anne, 1979. *Causes de l'infécondité dans la Volta Noire.* Paris, INED, PUF, 100 p.

Sauvy Alfred, 1952. Trois mondes, une planète, *L'Observateur,* no. 118(14 August), p. 14.

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1995a. La crise sanitaire en Russie. I. Tendances récentes de l'espérance de vie et des causes de décès de 1970 à 1993, II. évolution des causes de décès: comparaison avec la France et l'Angleterre (1970–1993), *Population,* vol. 50(4–5), p. 907–982.

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1995b. Health crisis in Russia. I. Recent trends in life expectancy and causes of death from 1970 to 1993, II. Changes in causes of death: a comparison with France and England and Wales (1970 to 1993), *Population, An English Selection,* vol. 8, p. 123–154.

Tabutin Dominique, 1995. Un demi-siècle de transitions démographiques dans le Sud, *in:* Dominique Tabutin, Thomas Eggerickx, and Catherine Gourbin (ed.), *Transitions démographiques et sociétés,* p. 33–70. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 692 p.

United Nations, 1962. État et évolution récente de la mortalité dans le monde, *Bulletin démographique des Nations unies,* no. 6, p. 11–54.

United Nations, 2001. *World population prospects: the 2000 revision. Vol. 1: Comprehensive tables.* New York, United Nations, Population Division, 746 p.

United Nations, 2002. *World urbanization prospect: the 2001 Revision.* New York, United Nations, Department for Economic and Social Affairs, Population Division, 182 p.

United Nations, 2003. *World population prospect. The 2002 revision.* New York, United Nations, Department for Economic and Social Affairs, Population Division, CD-Rom.

Vallin Jacques, 1968. La mortalité dans les pays du tiers monde: évolution et perspectives, *Population,* vol. 23(5), p. 843–868.

Vallin Jacques, 1971. Limitation des naissances en Tunisie: efforts et résultats, *Population,* vol. 26(special issue "Le Maghreb"), p. 181–204.

Vallin Jacques, 1989. Théorie(s) de la baisse de la mortalité et situation africaine, *in:* Gilles Pison, Mpembele Sala Diakanda, and étienne van de Walle (eds.), *Mortalité et Société en Afrique au sud du Sahara,* p. 399–431. Paris, INED, PUF, 446 p.

Vallin Jacques, 1992. Theorie(s) of the declining of mortality and African situation, *in:* Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa,* p. 405–436. Oxford, Clarendon, 450 p.

Vallin Jacques, 1993. Life expectancy, past, present, and future possibilities, *in:* Jean-Marie Robine, Colin Mathers, Margaret Bone, and Isabelle Romieu (eds.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives,* p. 63–77. Paris, John Libbey and INSERM, 388 p.

Vallin Jacques and Lapham Robert, 1969. Place du planning familial dans l'évolution récente de la natalité en Tunisie. *Revue tunisienne de sciences sociales,* no. 17/18, p. 379–413.

Vallin Jacques and Meslé France, 2001a. Évolution de la mortalité en Europe depuis 1950: mortalité par âge, sexe et causes de décès, *in: Tendances en matière de mortalité et mortalité différentielle,* p. 33–188. Strasbourg, Council of Europe Publishing, 344 p.

Vallin Jacques and Meslé France, 2001b. Trends in mortality in Europe since 1950: age-, sex- and cause-specific mortality, *in: Trends in mortality and differential mortality*, p. 31–186. Strasbourg, Council of Europe Publishing, 334 p.

Vallin Jacques and Meslé France, 2004. Convergences and divergences in mortality. A new approach of health transition, *Demographic Research: Special Collection*, no. 2, p. 11–43.

Van de Kaa Dirk J., 1987. Europe's second demographic transition, *Population Bulletin*, vol. 42(1), p. 1–59.

Van de Kaa Dirk J., 1988. *The second demographic transition revisited: theories and expectations.* Florence, European University Institute, 55 p.

Van de Kaa Dirk J., 1994. The second demographic transition revisited, *in:* Gijs Beets, Hans van den Brekel, Robert L. Cliquet, Gilbert Dooghe, and Jenny De Jong Gierveld (eds.), *Population and family in the low countries 1993: late fertility and other current issues*, p. 81–126. The Hague, NIDI, CBGS, 246 p.

Villeneuve-Gokalp Catherine, 1997. Vivre en couple chacun chez soi, *Population*, vol. 52(5), p. 1059–1082.

Waltisperger Dominique, Vallin Jacques, and Ben Mrad Abdelhamid, 2001. La dynamique naturelle de la population depuis l'Indépendance, *in:* Jacques Vallin and Thérèse Locoh (ed.), *Population et développement en Tunisie: la métamorphose*, p. 53–88. Tunis, Cérès éditions, 801 p.

Yifu Lin Justin, Cai Fang, and Li Zhou, 2000. *Le miracle chinois. Stratégies de développement et réforme économique*. Paris, Economica, 250 p.

P A R T

II

THE GEOGRAPHY OF POPULATION AND DIVERSITY OF DEMOGRAPHIC DYNAMICS

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

From history, we move on to geography. Although short, this is a cardinally important section. The cusp of the 20th and 21st centuries is surely the one moment in human history at which demographic situations are most diverse across the world and also at which the distribution of populations across the surface of the Earth is most fluid, less because of international migration, largely halted by restrictive policies in wealthy countries, than because of the diversity of natural demographic dynamics themselves.

Whatever scale and interpretation are used, the world distribution of the population "demonstrates spectacular contrasts between densely populated and totally or nearly-unpopulated areas"[1] observes Jean-Michel Decroly in Chapter 70. After a review of the tools usable to measure this phenomenon, he seeks to explain the reasons for this and demonstrate their nature as broadly historical constants.

In Chapter 71, Bruno Schoumaker, Dominique Tabutin, and Michel Willems, stress that humankind has probably never experienced so many "demographic upheavals as in this latter half of the 20th century"[2] unprecedented growth, a break with the old numerical relationships among people, continents, high-density regions, etc., not to mention the evident radical changes in relationships between age groups and generations. The demographic transition has assumed many different forms and evolved to very different timeframes in different countries and regions of the world. The general dynamic of this has already been referred to in Chapters 68 and 69, but is reconsidered here in a structured way across its very diverse outcomes, observed mainly at the turn of the 21st century. The authors also lay some pointers for the future, and hence for Part III.

[1] "se manifeste par de spectaculaires contrastes entre zones densément occupées et zones où les hommes sont absents ou presque."

[2] "bouleversements démographiques qu'en cette deuxième moitié de XX^{e} siècle."

C H A P T E R

70

Distribution of the World Population

JEAN-MICHEL DECROLY

Laboratoire de Géographie Humaine, Université Libre de Bruxelles, Brussels, Belgium

INTRODUCTION

Whatever the scale and analytical framework used, the geographic distribution of the Earth's population displays spectacular contrasts between densely inhabited regions and areas with little or no human presence. Many images immediately come to mind to illustrate this phenomenon, such as the map of world population distribution showing the "full" worlds of Europe, India, and east Asia against the vast empty expanses in the high latitudes and the belt of arid deserts; the satellite image of Europe at night, in which the tight-knit pattern of villages and towns of inner Flanders, the Dutch Randstadt, and the Ruhr basin contrast with the more sparsely populated regions of Limbourg, the Ardennes plateau, and the Eifel; and the map of a city in an arid region showing the stark boundary between the urbanized area and the desert.

Beyond their surprising appearance, these contrasts raise several basic questions: Why are some regions of the Earth so populous? Why are there so few people elsewhere? What role have environmental and social factors played in determining these variations? All these issues are central to the analysis of the relationships between the human population and its habitat.

Addressed by scholars and philosophers since antiquity, these questions became a favorite research topic for the developing academic discipline of geography in the 19th century, at a time when it was seeking recognition. Friedrich Ratzel's *Anthropogeography* (1881–1891) and, above all, Pierre Vidal de la Blache *Les principes de géographie humaine* (1922) showcased the study of the spatial distribution of the population. They saw it as an invitation to examine the respective influence of the natural and human environments on territorial organization. This intellectual fashion would endure, as attested—in the French literature—by the lengthy discussions of settlement patterns by Max Sorre in his treatise on human geography (1948) and by Pierre George (1951), Jacqueline Beaujeu-Garnier (1956, 1958), and Wilbur Zelinsky (1966) in their volumes on population geography. Even today, the analysis of world population distribution remains a prime concern of human geography—more specifically, of population geography. The compilation of a new map of world population distribution by the International Geographic Union (IGU) Commission on Population Geography and Environmental Geography is forceful evidence of this (Fig. 70–1) (Vandermotten and Marissal, 2003). As the interest persists, the analytical approaches are evolving. The study of the impact of unequal spatial distribution is gradually replacing that of its causes (Pumain, 1992).

Rather than attempt a deceptive synthesis of the abundant literature on the topic examined, this chapter has three aims: (1) to highlight the main features of world population distribution, (2) to discuss the origins of the distribution, and (3) to supply a qualitative and historical explanatory framework. Beforehand, we discuss the concepts and tools needed for the descriptive analysis of the spatial distribution of the population. These specifications are useful because the controversies over settlement are partly generated by the ambiguity of the terms used.

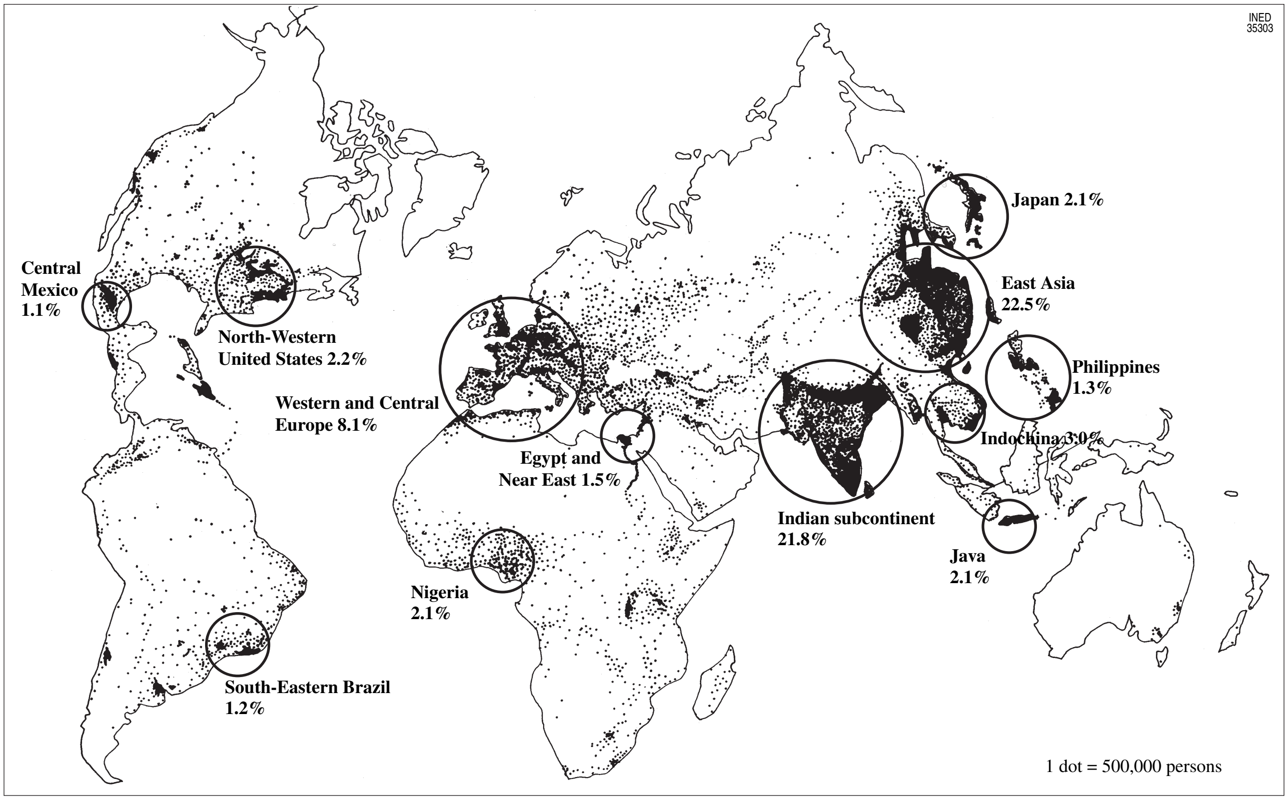


FIGURE 70–1 Map of spatial distribution of world population in around 2000. Demographic weights of population cores have been computed from national statistical yearbooks. (Data from Vandermotten and Marissal, 2003, after Baudelle, 2000 and Dolffus, 1990.)

I. DIFFERENT ASPECTS OF SETTLEMENT

In the social sciences, the term "settlement" has two different meanings. First, it describes the process by which societies establish their presence in a new territory. Examples include the conquest of the West in the United States, Russian settlements along the trans-Siberian railroad, and, more recently, the transmigration that took place in Indonesia from the island of Java to those of Sumatra, Sulawesi, and Kalimatan. Second, the term denotes the forms of land occupation by a population in a given period: presence or absence, occupation intensity, and dispersion or concentration in aggregates of varying size.

The analysis of the spatial distribution of the population specifically addresses this second aspect of settlement, as it seeks to (1) determine the Earth's inhabited area, (2) measure the variations in the intensity of human presence in that area, and (3) describe the geographic pattern of populated localities. This chapter will be restricted to the first two points, the third having been discussed, at least from a demographic standpoint, by Denise Pumain in her chapter on urbanization (Chapter 61 of Volume II).

To meet the objectives of the analysis of the spatial distribution of the population, it is important to establish the location of inhabitants. This self-evident truth conceals two questions. Who should be counted, and where?

The measurement of human presence in a given area relies on the enumeration of the population irrespective of its characteristics, be they demographic (e.g., age, sex, civil status), socioeconomic (economic activity, occupational status), or cultural (e.g., religion, language). This enumeration is usually carried out in the individuals' main place of residence. It is therefore exclusively dwelling related. People are counted in the localities where they live, hence in the structures designed to shelter them and ensure the daily reproduction of their workforce.

The analysis of population distribution at the place of residence does not provide an accurate count of populations that are continuously on the move with their dwellings (e.g., hunter-gatherers, nomadic shepherds, boatmen, gypsies) or groups engaged in seasonal or temporary migrations between several dwellings (seasonal migrants).[1] As a result, it neglects the fact that many acts of everyday life occur outside the home (Hägerstrand, 1973).

As long as transportation was slow, expensive, and difficult, people's workplaces and, more generally, their living spaces were identical to their dwellings: the vast majority of persons spent their lives within a radius of a few kilometers around the family residence. The spatial distribution of the population varied relatively little in the course of the day, the week, and the seasons. The increasing spatial division of work, the specialization of geographic areas, and the acceleration of travel have powerfully undermined the traditional pattern. Today, places of work and exchange are largely separate from places of residence. Daily or weekly commuting, as well as the temporary migration of tourists, characterizes the lives of a sizable percentage of the population, at least in the developed countries.

In consequence, there is often a wide gap between the number of persons who effectively reside in a locality and the number of persons who use that locality for their education, work, shopping, or leisure activities. For example, the resident population of the central business districts of European and North American metropolises does not exceed a few hundred people—but their daytime weekday population runs into the 10s of thousands. Likewise, in some tourist resorts, the peak-season population can be 10 times the number of all-year residents. In terms of local development and environmental impact, the significance of the two figures is, of course, totally different.

Despite its valuable lessons, the spatial distribution of the population by place of education, work, shopping, or leisure is very seldom studied. The task is admittedly complicated by the difficulty of measuring daily travel and, even more so, seasonal travel such as tourist flows (Vandermotten and Decroly, 1990). Nevertheless, further research on the subject would be useful to obtain a fuller picture of world population distribution.

Once the location of inhabitants has been established, we need to answer a third question: At what scale should we analyze their spatial distribution? This is a critical issue in the study of human settlement. The patterns of land occupation by human societies are intrinsically discontinuous: habitats are clustered into aggregates of variable size, located at different intervals from one another. Similar to a galaxy or ant colony, inhabited space therefore exhibits a granular structure, thus generating three major problems. First, such a structure does not have a single envelope but a multitude of potential ones, depending on the precision of the analytical grid adopted. Second, the contrasts inside the structure itself vary sharply with the observation scale. As Hervé Le Bras aptly notes (1993: 115): "to speak of the concentration or dispersion of the population makes sense only at a given observation scale: to say that 50% of the population is concentrated on 5% of the surface area is true only for a given segmentation. . . . The impossibility of determining the concentration of the population is the same as that of solving the famous problem of measuring the length of a country's coastline."

Third, the interpretation of the spatial distribution of the population also depends on the observation level. The processes that explain human settlement vary considerably with the scale of analysis. For example, the contrasts in population density in Belgium between urban agglomerations and their environment, on the one hand, and within the agglomerations, on the other hand, do not have the same causes. The first are largely due to the conditions of urban growth in the 19th century (economies of aggregation for industrial firms, closeness between home and work made necessary by slow transportation). The second are mainly due to the changes in urban space after World War II (shift toward service industries in urban areas, diffusion of the automobile, and promotion of the single-family home as the ideal dwelling).

II. ANALYTICAL TOOLS

The analysis of the spatial distribution of the population has long relied on three well-tested tools: the census as a source, population density as an indicator, and choropleth mapping as a method of representation. Although these tools remain very useful, they have weaknesses, and other less common tools should be mentioned.

1. Sources

A prime source for analyzing human settlement is the population census. In most countries of the world, it provides a comprehensive or nearly comprehensive enumeration of the population within the framework of often detailed administrative divisions. Analyzing the data extracted from this valuable source does,

[1] These groups currently total fewer than 50 million people worldwide. However, they intermittently occupy very large tracts of land, often lacking permanent settlements. The nomadic peoples of the old world's wide belt of arid desert and the Lapp and Siberian livestock breeders travel over territories stretching across more than 15 million km^2, or nearly twice the world's total surface area under cultivation (Debié, 1995).

however, pose two problems. The first is due to the coexistence of two different definitions of the population: (1) the de facto population, that is, the persons present at the census date in the geographic unit examined;[2] and (2) the *de jure* population, that is, the persons habitually residing in the geographic unit, regardless of whether they were present or not at the reference date. Despite international recommendations to harmonize the two concepts, countries continue to use either definition, undermining the comparison of the characteristics of their settlement patterns. In fact, the numerical difference between the de facto and *de jure* populations, which is already significant at the national level, can become substantial at the regional or local level (Decroly and Vanlaer, 1991). For example, in the Greek census of April 5, 1981, greater Athens was found to have a de facto population of 3,025,000 but a *de jure* population of 2,275,000. Barely 55% of the de facto population resided in the capital. Conversely, more than a quarter of Athens residents were enumerated in the de facto population of other geographic units.

The second problem entailed by the census data is the use of the existing administrative divisions as a basis for information collection. The divisions, which express the central government's appropriation of national territory, do not exhibit the same organizational logic or the same degree of refinement from one country to another. The contrasts are sometimes stark, even in Europe, where we find the highly disaggregated division of France (36,433 *communes*, or municipalities, with an average surface area of 15 km^2 and an average population of 1500) and the far cruder division of Belgium (589 municipalities averaging 50 km^2 and 17,000 inhabitants) and of Sweden (282 municipalities averaging 1500 km^2 and 30,000 inhabitants). This heterogeneity is a major obstacle to international comparisons, as the contrasting settlement patterns observed in the finer divisions (e.g., between rural towns and outlying villages) are wiped out in the broader divisions. Spatial smoothing methods (Le Bras, 1993; Tobler *et al*., 1997; Grasland, 1998) provide a way to overcome the obstacle but at the price of sometimes questionable assumptions.

Differences in the size of administrative units also modify the analysis of the spatial distribution of the population in very large countries with sharply contrasting settlement patterns—most notably Canada, Russia, and China. In this case, the size of the administrative units is inversely proportional to the number of persons per surface-area unit. As a result, the spatial distribution of the population is all the harder to determine accurately as the area is sparsely inhabited.

Without a census or population register, the classic way of enumerating the population is through a direct field survey. This information-collection method, however, has the disadvantage of requiring sizable human and financial resources, or it may cover only small areas. When population distribution is the focus of the analysis, the use of alternative sources—in particular, aerial photographs and satellite images—has become increasingly common. Once the dwelling locations have been enumerated and plausible occupancy ratios for the localities have been computed, both of these sources provide a good approximation of spatial distribution of the population at a moderate cost. They also offer the advantage of being suitable for counting populations that make temporary use of an area—for example, a tourist region—or that cannot be captured by standard enumerations, such as refugee camps.

2. Indicators and the Methods for Representing Them

The analysis of the spatial distribution of the population relies on two basic indicators with values that are computed in an administrative or geometrical segmentation: the absolute number of inhabitants and the population density per unit of surface area (Noin and Thumerelle, 1993). In the 19th century, the first cartographic studies of settlement used the abstract concept of density[3]—definitely an expression of that century's naturalistic approach to academic geography. In fact, the notion of density is akin to that of agricultural yields: it measures the number of persons that a given climate or soil is capable of *producing*. As Denise Pumain rightly notes (1992: 442), today "the indicator that possesses a social value would be the measure of the space available to each individual . . . in the same way as one calculates an income per capita." By this yardstick, Belgium would have 0.3 ha per inhabitant; France, 1 ha per inhabitant; and Greenland, 50,000 ha per inhabitant.

Population density should be seen as a descriptor of human presence. We must therefore refrain from giving it an explanatory value. Olivier Dollfus (1990)

[2] As a rule, the de facto population consists of people who have spent the night before the enumerator's visit in a dwelling located in the reference territory. It does not include some categories of persons despite their presence in the area, most notably visitors staying in tourist accommodation.

[3] Respectively, a map of the spatial distribution of the world population published in 1833 by the Englishman George Poulett Scrope in his *Principles of Political Economy* and a series of maps on the settlement of Ireland compiled in 1837 by Henry Drury Harness in preparation for the construction of the island's first railroads (Noin, 1979).

shows the absurdity of equating population density with a measure of the *geographic efficiency* of societies, that is, their capacity to generate yields from their land. Both the denominator (surface area) and the numerator (population) are, intrinsically, heterogeneous variables. The surface of the Earth is anything but uniform, and populations are not equal, if only because their living standards vary substantially from one place to another on the planet. Two identical densities can therefore indicate two very different realities. A comparison of Bangladesh and the Noord- and Zuid-Holland provinces (The Netherlands) is instructive in this respect (Dollfus, 1990). In terms of population density, both entities, located in flood plains and delta plains, would have roughly the same *geographic efficiency,* with 700 and 900 persons per square kilometer, respectively. Yet, on closer inspection, they differ in every way—not only in their living standards, of course, but also in their economic structure, urbanization levels, etc.

Population density is also often used as a measure of overpopulation—another serious error. Overpopulation means that the resources available in a given area are insufficient to meet the basic needs of the local population. In this sense, it is independent of population densities: overpopulated zones may have small populations, just as underpopulated ones may have many inhabitants. Whichever way one looks at it, the region of Brussels, a capital city, is not overpopulated at 5700 persons per square kilometer, despite having to cope with saturated communication infrastructure. By contrast, Mali, whose density barely exceeds six persons per square kilometer, may be described as overpopulated in that its available resources are inadequate for its resident population and its own-account production is preferred to massive imports of subsistence goods.

Even as a descriptor of human settlement, density is somewhat problematic. Its computation rests on the implicit assumption that the population is evenly distributed within each administrative subdivision considered. In other words, it assumes a consistent, homogeneous settlement pattern, which is contradicted by the facts. The discrepancy between representation and reality is not only a theoretical problem. What is the significance of the mean density of Egypt (53 persons per square kilometer) when more than 90% of the country's population is concentrated in the narrow valley of the Nile and its delta? For these different reasons, some investigators prefer to use dot maps, which locate population aggregates with precision.

Despite their being more concrete, population dot maps were not invented until the early 20th century and did not come into widespread use until after World War II. As Daniel Noin notes (1979: 38): "the principle is simple: the population is represented by dots, each of which symbolizes a given number of persons and is positioned as accurately as possible on the corresponding locality. The dots may all have the same value—and so the same surface—or different values." This system of representation offers several advantages. It gives a precise picture of population distribution while displaying certain explanatory factors such as physical features (relief, waterways, etc.) and allowing a reading on multiple levels. However, it does not permit rigorous comparisons in time and space or, more importantly, an effective display of high concentrations of population.

In the second half of the 20th century, many innovative mapping processes were developed to depict population distribution, in particular statistical relief maps, which show densities by means of *reliefs* of different heights, seen in perspective, and anamorphic maps, which display the populations of the geographic units studied by means of surfaces proportional to the number of inhabitants. In most cases, these maps are more valuable as teaching aids than as heuristic devices: they strike the imagination. An anamorphosis of the world population distribution by country shows that Russia weighs little by comparison with China, whereas Australia, which seems so large on a world map, weighs less than Taiwan (Fig. 70–2).

III. THE SETTLEMENT ENVELOPE

The first stage in the analysis of the spatial distribution of the population consists in defining the area occupied by humans. This is a crucial step, for, as Pierre Vidal de la Blache observed (1917: 81): "the area of distribution of a species, whether man or any other living species, is no less instructive for the gaps and discontinuities it reveals than for the tracts it covers." To tackle the issue, we must begin by examining the concept of *oikoumene.*

1. Oikoumene: An Ambiguous Term

The ancient Greeks coined the term *oikoumene* (from *oikos,* "household") to describe the world inhabited by the agrarian civilizations of antiquity. Today, *oikoumene* is one of the most ambiguous terms in the geographic vocabulary: its meaning varies as much as its spelling.[4]

[4] Jean-Pierre Allix (1996) notes no fewer than five different spellings in French alone: œ*koumène, écoumène,* œ*coumène, oikumène,* and œ *cumène.*

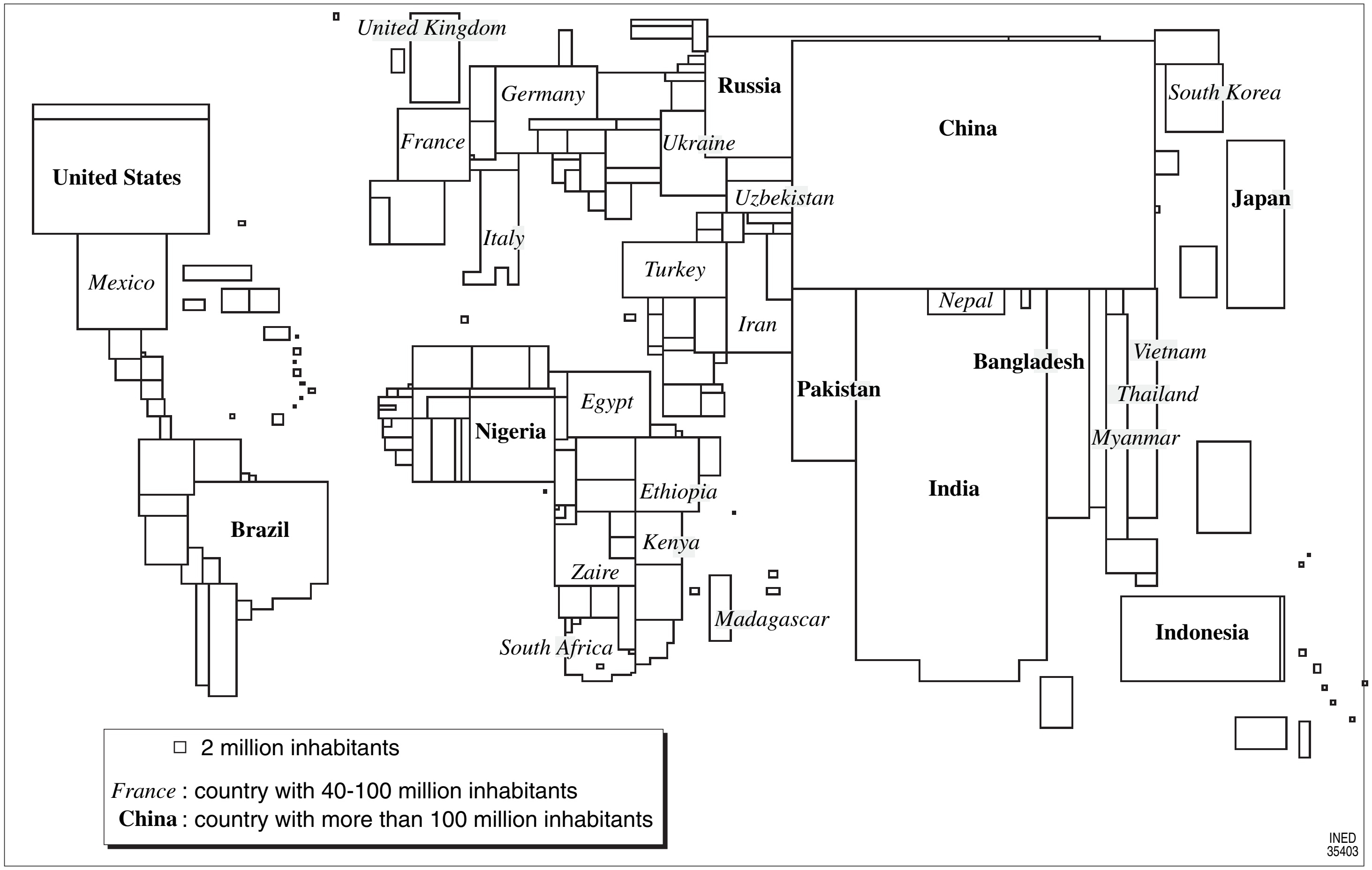


FIGURE 70–2 Anamorphosis of world population distribution. Each country is shown by a surface proportional to its population (Image from Vallin, 2000).

It is sometimes used to refer to the total landmass or, far more restrictively, to designate a market area, that is, a portion of territory in which economic consistency is provided by a network of trade routes (Wilkinson, 1993). As a rule, however, and consistently with its original meaning, *oikoumene* denotes the inhabited world. Following the method of Max Sorre (1948), Jean-Pierre Allix (1996) uses the term to describe the Earth's total inhabited landmass. Franck Debié (1995) broadens the concept somewhat by speaking of *territoire de l'humanité* (territory of humankind). But the basic idea is identical: the *oikoumene* is a concrete reality observable at a given moment. Conversely, Daniel Noin (1979) uses the term to describe the space suitable for long-term habitation by human communities or the total inhabitable areas of the planet.

Inhabited space or inhabitable space? The two notions are fundamentally different. The first is a fact observed at a given moment and on a given scale. The second is both a virtual limit—which complicates matters considerably. Inhabitability is always determined by social factors: it depends on the means of production, development approaches, and social relationships concerning production, as well as a group's representation of its environment. The limits of inhabitable space vary from one society to another, sometimes even radically. Unless we are to fall into the Eurocentric trap, it is vain to seek to define those limits for all of humanity. Consequently, in keeping with etymology, we shall regard *oikoumene* here as referring to the inhabited world as a whole.

2. The Earth as Envelope

Many investigators now regard the inhabited world as coextensive with our planet, and they reject the notion of *human deserts* in the narrow sense of the term (see Dollfus, 1990). Admittedly, the entire Earth has been mapped, explored, and appropriated. First, the successive improvements in surveying methods have made possible an exhaustive cartographic coverage of the planet. Second, the advances in methods for protecting against atmospheric and climatic constraints, combined with the development of transportation, have substantially broadened the area explored by humans. Since the 16th century, for reasons relating to strategy (marking the boundaries of a colony or country), science, or sports, innumerable expeditions have crossed the hot deserts of the old world, the dense equatorial forests, the tundra of northern Canada and northern Russia, the polar ice caps, the major mountain ranges, etc. Third, with a speed of propagation that was fitful but accelerated in the 20th century, all parts of the planet's land mass—including airspace and subsoil—were assigned to the different countries. Today, some 180 nations and 70 not-yet-independent territorial entities share the continental surfaces, forming a complex puzzle.

Although, in practice, the mapping coverage of the world is highly uneven, some areas of Antarctica and the Arctic remain untouched by human exploration, and many territories are still disputed, we have to admit that—at least at a small scale—humanity has taken possession of the entire Earth by use of its *eyes*, *feet*, and *rules*. Does this mean that humanity has extended the inhabited world to the whole planet? Without lapsing into the reductionist vision of 19th-century geography, which restricted the *oikoumene* to the sphere of European civilization and to population groups recognized as sedentary, a large part of the world is empty. In the oceans, in the cold deserts of the upper latitudes, the arid deserts of the tropics, the largest mountain massifs, and the equatorial rainforest, the rare human settlements—permanent or temporary—are separated from one another by vast uninhabited tracts, sometimes up to several 10s of thousands of square kilometers. The inescapable conclusion is that the *oikoumene* is far from covering the entire Earth.

3. The Difficulty of Drawing Boundaries

Because of the discontinuity of human settlement, the *oikoumene* does not have a single envelope, but a multitude of possible limits. On a very large scale, and without resorting to cartographic generalization, it will comprise a very large number of isolated settlements. On a very small scale, and after generalization, it will consist essentially of a few continent-sized populated tracts separated by the large expanses of human deserts. Greenland offers a good example of this. With barely 60,000 inhabitants for more than 2,150,000 km^2, the island, stretching between 59° N and 83° N, has a population density of less than 3 persons per 100 km^2. The human presence is fragmented into 19 isolated coastal settlements, including the U.S. military base of Thule at 76° N. In real life, each of these small agglomerations (the largest has 20,000 inhabitants) may be regarded as a fragment of the populated space—a crumb of the *oikoumene*. By contrast, on a planisphere, we can easily concede that Greenland as a whole is located outside the populated world, its inhabited areas being reduced to a bare minimum. Neither solution is more correct or more wrong than the other; each is correct at a given level of analysis.

The example of Greenland shows that defining the boundaries of the *oikoumene* is not self-evident to the observer but requires a set of prior choices, most notably that of the scale of analysis and the degree of generalization of reality. On a very small scale—the one being addressed here—the first approximation to be adopted may be the proposition that the *oikoumene* broadly comprises the spaces located inside an envelope curve plotted on a dot map of population distribution (Fig. 70–1). The regions outside the envelope curve form the outer limits of the *oikoumene*. They are not totally empty, but their densities rarely exceed one person per square kilometer and are very often zero over large expanses. These limits stretch over a surface area equal to 30% to 40% of the Earth's landmass (including the continent of Antarctica) and comprise less than two-hundredths of the world population.

IV. THE FRONTIERS OF THE OIKOUMENE

A wise observer of terrestrial space, Antoine de Saint-Exupéry, noted in one of his novels (*Vol de nuit*, 1931) that "the universe of fountains and houses is but a narrow ribbon on the surface of the globe; a first error—my plane went astray in the middle of the Paraguayan bush; a second error—it landed in the Saharan sands." Despite massive population growth in the past 2 centuries, the Earth still has wide expanses that are very sparsely inhabited, if at all.

1. Frontiers Shaped by Physical Constraints

The frontiers of the *oikoumene* are divided into four broad subsets shaped by the constraints on agriculture owing to certain characteristics of climate, vegetation, and altitude. To take up Guy Baudelle's (2000) classification, these consist respectively of "white deserts, dry deserts, green deserts, and altitude deserts."

a. White Deserts

White deserts combine nearly permanent cold[5] and meager precipitation. They are mostly situated in a wide band in the northern part of the Eurasian and American continents. They also comprise the Arctic islands and the Antarctic continent. In the northern hemisphere, their southern boundary varies from one continent to another. In North America, it lies consistently between 45° N and 50° N. But on the Eurasian continent, owing to the strong effects of continentality, the boundary shifts southward from west to east—from about 60° N in Scandinavia to 45° N in Siberia and the Far East. These huge territories stretch over more than 20 million km^2 (not counting Antarctica) and account for 10% to 15% of the landmass, but they are inhabited by fewer than 10 million people, or barely two-thousandths of the world population. Except for some large mining towns that are mainly a legacy of the Soviet era (e.g., Murmansk [580,000 people] at 68° N, Arkhangelsk [470,000 people] at 64° N, Norilsk [230,000 people] at 70° N, Yakutsk [190,000 people] at 62° N, Vorkuta [160,000 people] at 67° N), the human communities represent *crumbs of the oikoumene* (Racine and Villeneuve, 1992: 422)—tiny clusters of 100 to 1000 inhabitants separated by wide empty tracts.

The white deserts' low population density owes far less to the physiological effects of cold—against which humans can protect themselves—than to the minimal prospects for agriculture. The main continental glaciers and the permafrost[6] areas are totally sterile, as plants cannot find the water they need for growth. Farther south, periodic topsoil thaw promotes a meager vegetation of lichen and short grass, which can sustain large mammals such as reindeer and caribou, as well as allows hunting and fishing and even extensive livestock breeding. Cereal crops, however, are practically impossible. The mechanisms for producing organic matter are inefficient when the frost-free period does not exceed 150 to 175 days a year,[7] as is the case in white deserts. The importation of foodstuffs grown elsewhere now provides a remedy for this situation and makes it possible to sustain human communities—sometimes large—north of the cereal-growing boundary. But the cost is very high and warranted only by major economic or strategic interests.

b. Dry Deserts

Dry deserts, characterized by a precipitation below the potential evapotranspiration,[8] are dominated by

[5] One or two months with a mean monthly temperature exceeding 10°C in a temperate continental climate; none in a polar climate.

[6] Permanently frozen stratum of the soil or subsoil.

[7] Research by Canadian and Russian geographers has shown a close correlation, in their respective countries, between this climate boundary, the agriculture limit, and the main break in the settlement pattern. In Quebec, for example, the population density falls abruptly north of a line running through Lake Saint-Jean, the Abitibi, and the Témiscamingue: the line marks the boundary of 160 frost-free days.

[8] Evapotranspiration is the quantity of water absorbed by the air either through evaporation (running or stagnant surface water, soil water, and water-tables) or through transpiration by plants and other living organisms. Potential evapotranspiration measures evapotranspiration under the theoretical assumption of a consistently adequate water supply.

the great arid diagonal of Eurasia wrapping the side of the old world from the western Sahara to the Gobi Desert. Among other features, it comprises the Ténéré, the Tibesti, the Sinai, the Rubl'al-Khali, and Nefud in Saudi Arabia; the Lut Desert and Great Salt Desert of northeast Iran; the Kazakh steppe; the Karakum;[9] and the Takla-Makan.[10] In addition, there are the deserts in Latin America (Atacama, Chaco), sub-Saharan Africa (Kalahari, Namib), and Australia (Victoria, Great Sandy Desert, etc.). These arid regions are more populous than are the cold regions of the upper latitudes: covering more than 10% of the world's landmass, they are home to about 1.5% of the planet's inhabitants, with a mean density of about 5 persons per square kilometer versus the 0.5 persons per square kilometer of the white deserts. The gap narrows sharply, however, when one excludes the great irrigated valleys of the Nile, Tigris, and Euphrates, which hold approximately 70 million people. Moreover, similar to the white deserts, most dry deserts have less than one inhabitant per square kilometer.

The main reason for the sparse population of arid regions is inadequate rainfall: similar to the cold in upper latitudes, the lack of rainfall drastically curtails the production of food resources. It is now recognized that rainfall-dependent agriculture has become highly precarious (and subject to large annual variations) when rainfall is under 300 mm/year and is impossible below 150 mm. In these conditions, the formation of sedentary cores is inconceivable without supplementary water inputs—in particular, artesian water-tables and allochthonous rivers—that allow irrigated (often intensively) crops. Settlements accordingly develop as oases with distribution that matches the location of water sources: scattered concentrations around wells tapping into superficial or deep water-tables combined with linear concentrations in valleys of allochthonous rivers (Nile, Tigris, Euphrates, Amu Daria, Syr Daria, etc.) or along the source lines situated in the foothills of mountain ranges.

c. Green Deserts

The third type of *oikoumene* frontier consists of tropical and equatorial rainforests. Their population is often insignificant: 1 to 2 persons per square kilometer in the Amazon and 1 to 10 in the patches of equatorial forest in Africa (eastern Cameroon, Gabon, western Congo Brazzaville, northern Congo Kinshasa), as well as Borneo and New Guinea. Two factors play a key role here: poor sanitary conditions and the difficulty of managing the territory.

[9] Desert covering almost 80% of Turkmenistan, in central Asia.

[10] Desert covering a large portion of Xinjiang Uygur in China.

The moist heat generated by the climate and maintained by the forest cover provides a highly favorable breeding ground for pathogen complexes[11] responsible for specifically tropical diseases (e.g., malaria, leprosy, yellow fever, filariases, bilharziosis, sleeping sickness). This unhealthy environment is as much the cause as the consequence of a low settlement rate. On the one hand, it promotes high mortality and discourages the arrival of new inhabitants; on the other hand, as Pierre Gourou (1982) has clearly shown by using the example of sleeping sickness, it persists all the more easily given the lack of manpower to perform the improvements that would eliminate the vectors' places of reproduction and pathogen reservoirs. Moreover, evidence suggests that when technical resources are primitive, territorial control (e.g. penetration, clearing, extraction of natural resources) is far more difficult in closed forests than in woodlands or wooded savannas (Raison, 1988).

d. Altitude Deserts

The final category of *oikoumene* frontiers consists of the highest portions of large, steep-sloped mountain massifs. Entire sections of the Himalayas, Altai, Karakoram, Qilian, Kunlun mountains, Alps, Rockies, and southern Andes are not permanently inhabited. The same is true of some medium-altitude mountain ranges such as the Pamir,[12] Tian Shan,[13] and Japanese mountains. The upper limit of high-altitude settlements is consistently below the limit of endurance of the human body to the effects of oxygen rarefaction and the decrease in temperature. This boundary is dictated, as in the upper latitudes, by the restrictions on agriculture imposed by cold, in particular the shortening of the vegetative season. The limit therefore varies sharply with latitude: while reaching 4000 to 5000 m in the Himalayas and central Andes, it does not exceed 2000 m in the Alps and 300 m in the mountains of Iceland.

2. A Diversity of Groups Cohabiting on the Frontiers

The frontiers of the *oikoumene* are inhabited by two types of population that are very different both in the size of their settlements and in their degree of mastery and utilization of the environment—as well as the

[11] Each complex combines a micro-organism (parasite or protozoon) responsible for the disease, reservoirs in which it develops, and a vector that transmits it to humans and/or domestic animals.

[12] High mountain range between Tajikistan and Kirghizstan, in central Asia.

[13] High mountain range in Kirghizstan.

resources available to them for the purpose. The first category consists of scattered, often mobile groups that have long occupied these vast tracts; the second comprises groups that have established residence in military bases, industrial towns, or pilot farms built at great expense in recent decades.

Because of their modest resources, the first category is divided into small groups occupying extended tracts, without true territorial control. They include hunter-fishers (Inuit), hunter-gatherers (Pygmies), livestock breeders (Sames and Lapps of the European and Russian Far North), and even (but more seldom) farmers. Some are nomadic (Inuit), others are sedentary, and others combine both modes of spatial occupation (Lapps). To survive in often extreme conditions without sophisticated technologies, all have had to draw on whatever meager resources are available.

The presence of these groups in a particularly hostile environment was often owing to their exclusion (voluntary or forced) from their initial area of residence. The frontiers of the *oikoumene* have thus provided shelter for some endangered minorities (such as Druze and Christians in the Lebanese mountains). In other cases, they have served as "holding pens," where the colonizers confined the inhabitants of the region they had just conquered. In Morocco, for example, during the Arab-Muslim invasion of the eighth century, the Arab nomadic tribes drove back the Berberophones toward the Atlas. Likewise, in the United States, the European settlers steadily drove the Indian tribes out of their traditional hunting or livestock-breeding regions toward the Rockies and their foothills. In the 19th century, the settlement of the most clement regions of northernmost Norway, Sweden, and Finland (valleys and lake shores) drove the Lapps northward, as they were dispossessed of most of their hunting grounds and pastures.

The general shift toward trade-based relations challenged the traditional forms of occupation of the *oikoumene* frontiers. This trend was fueled by the expansion of the European world economy and the emergence of new countries as a result of decolonization. Very often, subsistence activities have been largely abandoned for trading activities. The Lapps, for example, have gradually converted to trading in animal hides. Likewise, in arid regions, long-distance caravan trade gradually gave way to competition from trucking. In a parallel development, political authorities have often encouraged local inhabitants to abandon their traditional areas of travel or residence. In Greenland, most Eskimo groups have been forced to settle in towns; in the Sahara, the Tuareg who moved between Algeria, Niger, and Mali were obliged to adopt a sedentary lifestyle.

These shifts have increased the dependency of the groups involved on the outside world while weakening their social structures (rise in alcohol and drug consumption, higher suicide rates). Where new activities have not generated adequate income, governments have often introduced direct or indirect subsidies. Absent such assistance, economic integration has triggered depopulation, sometimes on a massive scale as in many medium-altitude mountain regions.

For at least a century, the frontiers of the *oikoumene* have been the site of new forms of settlement driven by economic, strategic, or prestige-related motives—or a combination of all three. The extraction of energy resources (oil, natural gas, coal), rare and precious minerals (gold, silver, nickel, etc.), or timberland resources has spawned many urban developments. Some are single-purpose, short-lived, and of modest size, such as the countless small mining ghost towns in the western United States and oil-industry complexes in the Sahara; others are diversified and have become large, such as Yakutsk, a major Siberian natural-gas terminal and a wood- and paper-industry center. From 1950 onward, a drive for effective control of the Arctic led the two superpowers to build military bases and related facilities (e.g., weather stations, radar stations) in the Far North. Some of these bases gave birth to large urban centers, such as the Murmansk naval base on the Kola peninsula. And, in the past 20 years, the settlement of the frontiers of the *oikoumene* has also become an issue of prestige, as witnessed by the pilot farms in the Libyan and Arabian deserts, where cereal and vegetable crops are grown at great expense by tapping deep water tables.

The fruit of costly technological prowess, these new forms of settlement are fragile as they depend closely on changing international economic and political conditions. For example, after the break-up of the Soviet Union and the emergence of new areas of tension in the world, Arctic military bases have been gradually scaled down or simply abandoned (Radvanyi, 1996). Moreover, as Olivier Dollfus lucidly noted (1990: 448), in most of these facilities "a handful of men live in a controlled atmosphere, cooled or heated, consume products from elsewhere, and overcome the isolation and the conditions of a highly artificial and often almost all-male micro-society by frequent trips back to the *oikoumene* for restorative stays."

V. INHABITED SPACES

The spaces lying inside the envelope curve plotted on a dot map of settlements add up to 60% to 70% of the landmass and hold some 98% of the world population. They are not completely occupied, but their densities usually exceed one person per square kilometer. In their midst, where the population density exceeds 50 persons per square kilometer, are the population concentrations. They take up 10% to 20% of the landmass and contain 70% to 80% of the world's inhabitants. Between these concentrations lies the ordinary pattern of human settlement—"areas that are, in a manner of speaking, anonymous, not conspicuous for their excess or shortage of inhabitants" (Pinchemel and Pinchemel, 1988: 45)—covering approximately one-half of the landmass and holding 20% to 30% of the world population.

At the global level, the population concentrations differ in their geographic extension. First, there are three main cores—East Asia, the Indian subcontinent, and Europe—in which areas with a density greater than 50 persons per square kilometer stretch without significant breaks over several hundred thousand square kilometers. The second group consists of about 30 secondary concentrations, with cores that occupy small surface areas and are isolated from one another by large weakly-populated tracts. Both of these *oikoumene* subsets display variable patterns of territorial organization. The forms of spatial appropriation, the arrangement of cities with respect to one another, the formation of communication networks, and other characteristics differ sharply according to the intensity and degree of spatial continuity of settlement.

1. Main Concentrations

The world's three main concentrations hold 50% to 55% of the world population on about 10% of the landmass (Fig. 70–1). They are "*full worlds*" or "*human continents*," to use the expressions coined by Pierre Chaunu (1979) and Frank Debié (1995): they are massively settled, their territory is heavily developed and structured, and these characteristics have prevailed for a very long time. Far from being homogeneous, these concentrations are, in turn, dominated by cores or backbones of stronger concentrations of population, of cities, and of capital and economic activities.

In terms of population size and surface area, the largest of these *full worlds* is unquestionably southeast Asia. Stretching across the entire southeastern rim of the Asian continent, from the Mekong delta in the south up to Manchuria in the north, from the red basin of Sichuan in the west up to the Korean peninsula and Japan in the east, it contains 1.2 to 1.4 billion people on 5.5 million km^2, for a mean density of 200 to 250 persons per square kilometer. Its nucleus consists of the lowlands of northern China between Beijing and Shanghai: areas of early agricultural development and cradles of the main Chinese empires, they now combine very high rural densities—almost consistently in excess of 400 persons per square kilometer—and a uniform fabric of cities of all sizes.

The world's second-largest population concentration is the Indian subcontinent, with 1.0 to 1.2 billion people and a surface area of about 4.5 million km^2—a mean density of 200 to 250 persons per square kilometer. It is dominated by the Indus-Ganges axis and its four main components: the Punjab; the Ganges valley itself, with Uttar Pradesh and Delhi; the Brahmaputra valley; and the lowlands and deltas of both rivers. The area extends into heavy linear concentrations along the coastal plains of the peninsula: Kerala, Tamil Nadu, Maharastra, Andri Pradesh delta. In contrast, the Thar desert, on the western border between India and Pakistan, and the Chota-Nagpur plateau, west of Calcutta, are weakly populated.

The world's third-largest population concentration runs across the European continent, from the Atlantic to the Don[14] and from Sicily to the southern shores of the Baltic. This populated Europe—which excludes Scandinavia and northern Russia—is home to about 600 million people on almost 6 million km^2, which translated into a mean density of some 100 persons per square kilometer. It is dominated by a more densely populated axis that runs from Liverpool to Florence and forms the European megalopolis, a concentration of cities, manufacturing industries, and "command and control" tertiary industries. This large crescent, commonly called the *blue banana,* includes, in its center, the broadly defined German Rhineland and Benelux (apart from the Eifel and Ardennes massifs), central and southern England, the eastern and northern fringes of France, most of Switzerland, and northern Italy (excluding Trentino–Alto Adige).

The Paris basin, the Lyon conurbation, and the Scottish Lowlands to the west of the European megalopolis, as well as greater Berlin to the east, form concentrations that are fairly close to the densest area but separated from it by less populated interstices. To the east of the main axis is a secondary one, oriented west to east, running from Saxony to the Donbass and encompassing Thuringia and the northern borders of Bohemia, Moravia, and Silesia.

In southeast Asia, the Indian subcontinent, and Europe, massive settlement and high-density sprawls are associated with similar patterns of territorial organization that combine a heavy linear concentration of the highest densities; a dense scattering of

[14] River of southern Russia flowing into the Azov Sea, southwest of the Volga.

towns and cities of all sizes; a low gradient of urban density; a long-standing, robust appropriation of land by rural communities and political structures; and old, extensive communication networks with high connectivity (Debié, 1995).

Some investigators, most notably Daniel Noin (1979), link the eastern United States and southeastern Canada with the world's three main population concentrations. This statement is debatable given that, with about 100 million inhabitants spread over more than 1.2 million km^2, the area extending between the Atlantic Ocean and the Mississippi River and between the Saint Lawrence estuary and the Gulf of Mexico comprises a far smaller population than that of east Asia, the Indian subcontinent, or Europe, with lower mean densities—particularly in rural areas. A more detailed analysis also shows that the areas with a density exceeding 50 persons per square kilometer are narrowly confined to the East Coast megalopolis (Boston to Washington), the Midwest steel belt (Pittsburgh to Cincinnati), and the urban rim of the Great Lakes (Buffalo, Cleveland, Detroit, Chicago). By contrast, in east Asia, the Indian subcontinent, and Europe, areas of density stretch over hundreds of thousands of square kilometers. However, some features are comparable to the spatial organization of the *full worlds,* in particular, the linear concentration of high densities, the thick scattering of towns and cities of all sizes, and the high connectivity of communication networks.

2. Secondary Concentrations

Outside of the main concentrations, settlements are fragmented into small, high-density islands. These "*settlement archipelagos*" (Debié, 1995) hold about one-quarter of the world population on less than one-tenth of the landmass. Most are dominated by one or two large metropolises.

Among the 30 or so secondary concentrations enumerated by Frank Debié (1995), the vast majority are coastal or insular. This is true in North America (California and Vancouver–Seattle cores), Latin America (West Indies cores, western plain of Guatemala and Salvador, Venezuelan coast around Caracas, the Nordeste province of Brazil, the Porto Alegre/Sao Paulo/Belo Horizonte "backbone," Rio de la Plata around the coastal cities of Argentina and Uruguay, central Chile around Santiago-Valparaiso), sub-Saharan Africa (coastal cores of the Gulf of Guinea from Senegal to Nigeria, Cape Town and Transvaal cores in South Africa), north Africa and the Middle East (Maghreb "useful core" between Casablanca and Tunis, Levantine coast, Black Sea and Egean Sea coasts in Turkey), South Asia (Irrawaddy delta in Burma, Menam delta in Thailand, southwestern part of the Malacca peninsula, Java Island in Indonesia, Luzon archipelago in the Philippines), and Australia (coastal and urban cores of Perth, Adelaide, Melbourne, Brisbane, and Sidney).

The main exceptions to this pattern are the secondary concentrations associated with major rivers that provide the necessary water for agriculture (Nile, Tigris, and Euphrates valleys in the Middle East, Syr-Daria and Amu-Daria valleys in central Asia, Irrawaddy valley in Burma), those that have developed around Iranian oases, and a few individual concentrations located at the very center of continents, on low-altitude plateaus (Sahel core in Hausa country, northern Nigeria) and in higher-altitude or steep-sloping areas (Mexican core, Colombian Andes around Bogota, Great Lakes core centered on Rwanda and Burundi).

The spatial organization of *settlement archipelagos* differs radically from that of *human continents* (Debié, 1995). Their features are as follows: isolated population concentrations and low spatial extension; strong littoralization (increasing concentration of development on the coast) and scarcity of population enclaves at the center of continents; low density and high concentration of a scattering of towns, often dominated by an agglomeration far larger than the others; sharp contrasts in density between urban and rural areas (high urban-density gradient); recency and weakness of territorial grid; and atrophied communication networks.

VI. HISTORY OF SPATIAL DISTRIBUTION

The spatial distribution of the world population is the product of multiple, complex phenomena and has given rise to highly divergent interpretations. Consistent with the spirit of 19th-century human geography, initial studies emphasized the constraints of the physical environment on settlement. Within this framework, the contrasts in spatial distribution were explained by differences in the ease of living conditions and the disparity in the quality and quantity of available resources. In the wake of the contributions of the French school of geography—in particular, the major studies by Pierre Vidal de la Blache—environmentalist determinism was gradually discarded in favor of a possibilist explanation, largely based on the principle of proof through counter-example that stressed the diversity of production and administration methods. A more systematic approach developed in the 1960s, when population geography emerged as a separate discipline. Largely inspired by demographic methods, it sought to highlight the factors responsible for the observed differences. Accordingly, the approach underscored the history of demographic behavior, in particular, the history of migrations and their influence on settlement (see David Coleman's synthesis in Chapter 67). This type of analysis was very productive as regards the mechanisms of settlement differentiation; however, it led to a fragmentary vision of the formation of densities: the interactions between factors were often neglected, and the socioeconomic conditions in which they occurred were seldom mentioned. In the French-speaking world, it was not until 1990s and the work of researchers such as Guy Baudelle (2000), Frank Debié (1995), Olivier Dollfus (1990), and François Durand-Dastés (1995) that a resolutely cross-sectional approach emerged based on two crucial observations: the stability of the *oikoumene* borders and the persistence of settlement contrasts.

1. Stability of Oikoumene Borders

Archeological research over the past century shows that the human species had already colonized most of its current area of settlement by the Upper Paleolithic (see Chapter 67) (Fagan, 1990). Twenty thousand years ago, humans were present on the five continents, including Oceania, and occupied large tracts of land. According to the evidence available today, the only parts of the world that still eluded humanity's grasp were the high-latitude regions (northeast Canada, Greenland, Iceland, Scandinavia, northern Russia, southern tip of the American continent, eastern Australia) and some islands far removed from the continents (Tasmania, New Zealand, Polynesia, Hawaii, Easter Island, etc.).

These observations corroborate the hypothesis formulated in the early 20th century by Pierre Vidal de la Blache (1917: 241), namely, that the human population *has not broadened its geographic reach so much as it has deepened its localized presence.* In other words, the successive increases in world population since the Upper Paleolithic have resulted in a densification of already inhabited places rather than a spatial extension of the area of settlement through large-scale conquest or clearing: "a crystallization of sorts has agglomerated batches of human population in certain points. These groups have, through their intelligence, increased the natural resources and value of the sites, so that others have come . . . and successive layers have accumulated on the chosen lands" (Vidal de la Blache, 1922). In consequence, the settlement envelope is highly stable, particularly in the main concentrations. In Japan, as in China or Europe, the pattern of territorial occupation has remained basically the same for several centuries. This stability is linked to the spatial logic of intensive agriculture. Improved yields enabled farmers there to reduce working time—a saving that was not used to enlarge the holding but was reinvested in capital to enhance crop performance. This logic, in turn, promoted the accumulation of human inhabitants in the same limited spaces (Baudelle, 2000).

Although the long-term stability of the borders of the *oikoumene* is undeniable on a small scale, the same is not true on a medium or large scale. The successive waves of demographic growth have generally caused a spatial extension of settlement, most notably after the clearing of woodlands or the drainage of wetlands. But this extension usually occurs near already inhabited and developed localities, in already frequented or exploited spaces.

The only significant exceptions to the logic of on-site accumulation of settlement are pioneering frontiers. As a vehicle for the agricultural development of previously low-populated or empty lands, they can change the settlement envelope. Unlike other forms of settlement on the boundaries of the *oikoumene*, they are often charted across vast territories and in a particular direction, mobilizing large populations, and they tend to be more enduring. Pioneering episodes occur in waves, in response to phases of accelerated population growth or political plans. The most significant examples date from the past 2 centuries and took place in the interior or on the peripheral islands of large countries under construction. They include the settlement of the regions west of the Mississippi River in the second half of the 19th century, the pioneering frontiers of the Patagonian pampa in the late 19th century, the colonization of western Siberia along the axis of the Trans-Siberian railroad in the first half of the 19th century, the colonization of Mindanao[15] starting in the interwar years, and the Indonesian transmigration to Sumatra, Sulawesi, and Kalimatan in the last quarter of the 20th century (Pain, 1998).

2. Enduring Contrasts in Settlement Patterns

In the past 20 years, specialists have compiled fairly comprehensive estimates of the population changes since the start of our era by macroregion (Biraben, 1979) or by country (MacEvedy and Jones, 1978). By using these figures, several investigators have emphasized the long-term persistence of major population concentrations. In the first volume of *Civilisation matérielle, économie et capitalisme* (*Civilization and Capi-*

[15] The largest island of the Philippines.

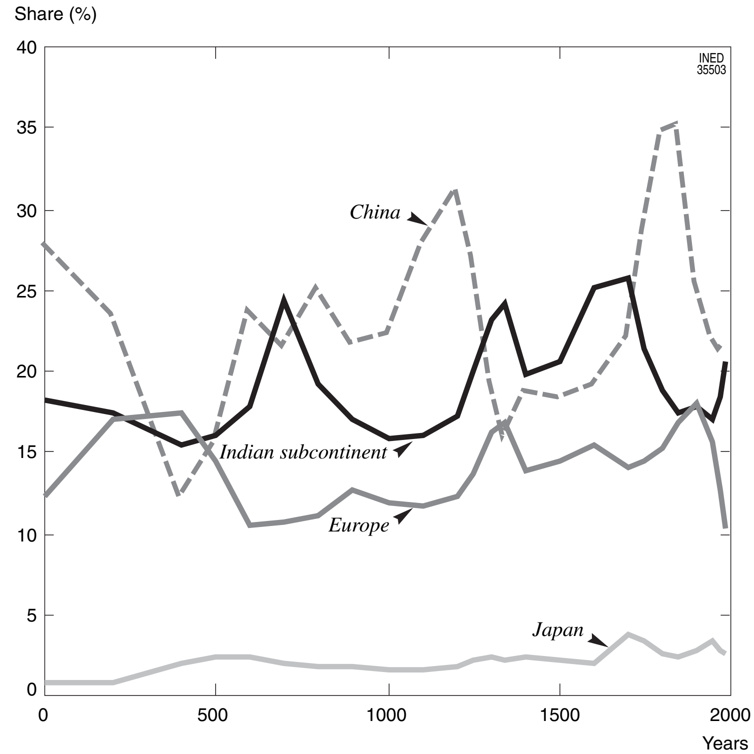


FIGURE 70–3 Changes in relative shares (percentage) of China, Japan, Indian subcontinent, and Europe in the world population since the start of our era (Data from Biraben, 1979).

talism), Fernand Braudel (1979) noted, for example, that the narrow band of high population densities observed in Europe and Asia in the late 15th century has persisted—all other things being equal—in the present geography of the world. Jean-Noël Biraben (1979) points out that, between the start of our era and today, China has consistently held 15% to 35% of the world population; the Indian subcontinent, 15% to 25%; and Europe, 10% to 15% (Fig. 70–3). In other words, for the past 2000 years, between 50% and 70% of the world population appears to have lived continuously on one of the three *human continents,* as defined by Franck Debié (1995) (Fig. 70–4).

We can refine these observations by a systematic comparison of the population distribution by macroregion at the start of our era and at later dates using a temporal autocorrelation method[16] (Fig. 70–5). This shows a slow change in the distribution until the late 15th century. At that date, the contrasts in settlement patterns between macroregions were very similar to those observed more than a 1000 years earlier (r ≈ 0.8). The population distribution then stabilized, roughly until 1750. From that point on, its change accelerated, with the correlation coefficient falling steadily to 0.6 in 1990.

These results show that, inside a rather stable envelope, the spatial distribution of the human population has undergone little change in the past 2 millennia. Until the start of the Industrial Revolution, the proportional relationships between the populations of the world's macroregions remained nearly identical. Afterward, as Olivier Dollfus has already suggested (1990: 457), bursts of sharp demographic growth "altered the existing balances and introduced new distortions between continents" (see Chapters 68, 69). Despite these changes, the broad features of the initial distribution endured, in particular the location of population concentrations. These findings are often valid on a large scale, but for shorter periods of time, as in Europe, many factors in the distribution of high population densities were active by the Middle Ages—for example, the high densities in Flanders and the Po valley in the 13th century. Nineteenth-century industrialization therefore accentuated the density contrasts of the preindustrial period. The centers of industrial development were located in places that had regis-

[16] The principle consists in calculating the correlation coefficient between the initial distribution of the population and later situations. To neutralize the changes in the relative weights of the most populous regions, the value has been computed on the logarithms of the population rather than on the populations themselves.

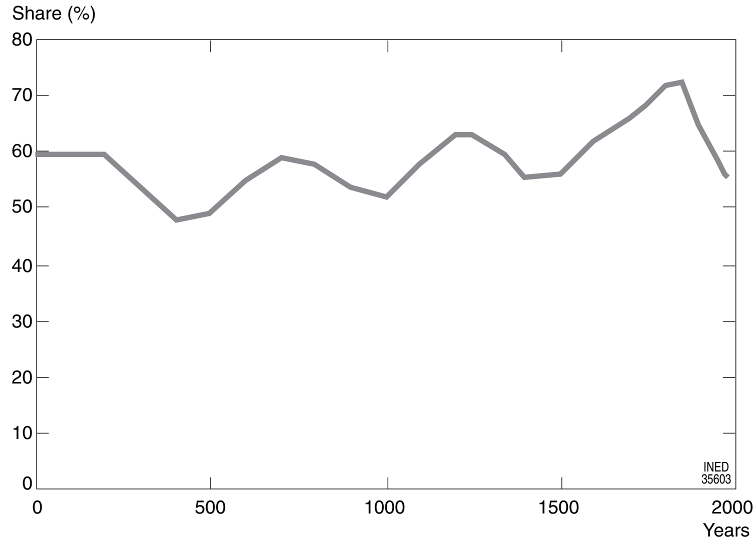


FIGURE 70–4 Changes in relative share (percentage) of the combined population of the three human continents in the world population since the start of our era (Data from Biraben, 1979).

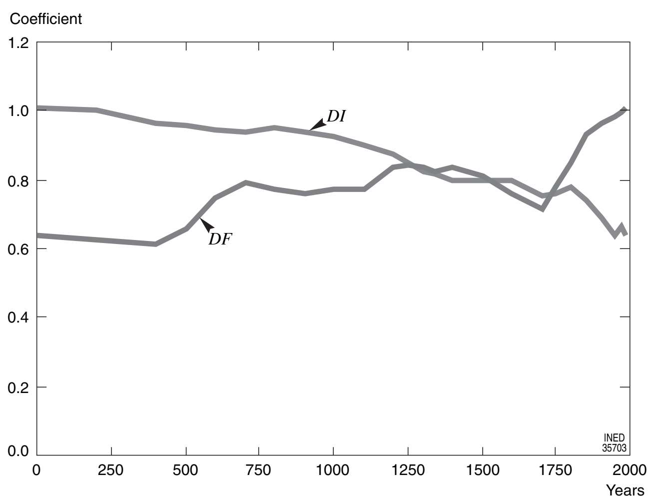


FIGURE 70–5 Change in correlation coefficient for world population distributions (12 macroregions) since the start of our era. DI, correlation coefficient between initial and later distributions; DF, correlation coefficient between final and earlier distributions. (Data from Biraben, 1979)

tered the earliest and most vigorous capital accumulation, where a prosperous merchant bourgeoisie had emerged and where agriculture was already the most advanced and the most closely integrated into the town-centered trading networks (Decroly and Vanlaer, 1991). In short, the areas of population concentration were already potentially present in the shaping of the European merchant economy of the 17th and 18th centuries.

The dot maps of the spatial distribution of world population over 2 millennia, such as the ones prepared by Daniel Noin (1979: 92–93) and Olivier Dollfus (1990: 447), provide a useful complement to this analysis. Admittedly, they rely on highly uncertain and fragmentary data, especially for periods before the 18th century, at least for the areas outside the great empires of the Mediterranean world, India, and China. Nevertheless, they offer a broad picture of the changes in global population distribution (for a detailed analysis, see Noin, 1979 and Debié, 1995).

Of the Earth's 250 million or so inhabitants at the start of our era, almost 60% lived in one of the three human continents. The north China plain, between the Huang-he and the Yangzijiang, is believed to have held alone nearly 50 million people, that is, one-fifth of the world population and four-fifths of the Chinese population. It was thus already the planet's largest population concentration. The rest of China presumably comprised some 20 million people living in small, scattered settlements. The Guangdong region, similar to the middle Chang Jiang valley and Sichuan, was not yet an area of relatively concentrated population.

The Indian subcontinent is estimated to have had nearly 50 million inhabitants, of whom about 30 million were in the Indus-Ganges plain and roughly 10 million were in the Dekkan coastal plain. Europe's approximately 30 million people were concentrated within the borders of the Roman Empire, particularly in the plains and basins near the western Mediterranean (10 million people) and on the Egean coasts (about 3 million people on the European rim).

The most novel aspect of world population distribution at the start of our era was the relative importance of the Middle East. The region, bounded by the

Red Sea and Persia at one end and the Black Sea and Indian Ocean at the other, is thought to have had some 50 million inhabitants, most of them concentrated in four areas: lower Mesopotamia, the Turkish coast, the Levant coast, and the western coast of the Arabian peninsula. If we add north Africa's estimated 15 million or so inhabitants, we find over 60 million people—nearly one-quarter of the world population in that period—living on the confines of the arid zones of the old world and in the oases at the heart of those zones (Nile, Tigris, and Euphrates valleys). In later periods, even at the apogee of the Arab-Muslim Empire, the region's share of the world population never reached these initial levels.

The rest of the world population of the time (about 50 million) was scattered across the vast expanses of sub-Saharan Africa and the American continent, which held about 12 million people each.

As mentioned earlier, the spatial distribution of the world population underwent minimal change between the start of our era and the end of the 15th century. By the latter date, the three *human continents* of the old world contained slightly more than 55% of the planet's 460 million or so inhabitants: east Asia (China, Japan, Korea) and the Indian subcontinent each held 90 to 100 million people, Europe nearly 70 million. Within these human masses, the Huanghe and Yangzijiang plains in northern China and the Indus-Ganges and Dekkan plains in India remained by far the largest population concentrations. However, the areas of high Asian density expanded substantially, both in China (Manchuria, Korea, Chang Jiang valley, Guangdong, and Indochina) and in the Indian subcontinent (central Dekkan). With almost 10 million inhabitants in its southern lowlands, Japan formed a northeastern extension of the east Asian area of high relative density. Although Europe preserved its relative share of world population (about 15%), the geography of its settlement changed significantly. After the center of gravity of high densities shifted from the Mediterranean coast to the North Sea coast, the space between southeast England and northern Italy gradually emerged as the main axis of the Old Continent.

The most important changes in world population distribution occurred in the areas outside the human continents. Between the start of the Christian era and the Age of Discovery, North Africa and the Middle East lost nearly half their population. By 1500, they held just 7% to 8% of the world population. The only dense, continuous demographic core in their midst remained the Nile Valley. Conversely, sub-Saharan Africa and the American continent registered fairly brisk population growth. Sub-Saharan Africa added about 60 million inhabitants to reach 15% to 20% of the world population by the end of the 15th century. The American population grew by some 30 million to nearly 10% of the world total by around 1500. The spatial patterns of growth were very different in the two continents. In sub-Saharan Africa, most of the new inhabitants scattered across the landmass, so that only one population concentration emerged: that of the Hausa, Mossi, and Yoruba countries in west Africa. In America, by contrast, population growth was more narrowly confined. It gave rise to a series of concentrations, in particular the Aztec core at the center of Mexico and the Inca core along the Andes cordillera.

Between 1500 and 1800, during the expansionary phase of Europe's world-economy, the world population doubled to about 950 million inhabitants. This growth mainly benefited the three human continents: in 3 centuries, their population grew 250%, versus only 30% for the rest of the world. As a result, by 1800, east Asia, the Indian subcontinent, and Europe comprised no less than 70% of the world population. Densification was significant in all the concentrations already established by about 1500, in China (northern plain, Chang Jiang plain, Sichuan, Guandong), Japan, Indochina, India (Indus-Ganges plain, Dekkan), northwest Europe, and northern Italy. Meanwhile, new high-density zones were forming, most notably on the Korean peninsula in east Asia, in the Punjab in the Indian subcontinent, and along the Danube corridor and northern Hercynian massifs, between the Ruhr and Silesia, in Europe.

The weak total growth outside the three main concentrations masks highly contrasting regional patterns. On the Asian continent, the demographic momentum was brisk, as evidenced by the massive densification of the island of Java and the formation of secondary settlement cores along the Burmese and Thai coasts. Similarly, Muscovy and the southern Ukraine gradually developed into distinct high-density centers, and the first pioneering frontiers were created in Siberia. By contrast, the population of Africa, deeply affected by the various slave trades, registered modest gains from its 1500 level (15 million or more inhabitants) and actually fell by about 15 million from its 1600 peak of 105 million. The high-density core in Nigeria remained stable and even weakened slightly; no new significant concentration emerged. Meanwhile, the American continent lost almost half of its population during the period. In the wake of brutal European colonization, the settlement cores clearly identifiable in about 1500 were either almost totally destroyed (Maya core in Yucatan) or sharply reduced (Aztec core in central Mexico and Inca core in the Andes).

A comparison of the world population distribution in 1800 and today shows that the massive 2-century growth has caused major changes in the geography of settlement but has not fundamentally altered its pattern. At the global level, the key phenomenon is the decline in the human continents' share, which has fallen from 70% of the world population in 1800 to 55% today. Conversely, the other regions of the Earth, sub-Saharan Africa excepted, have seen their shares of world population rise, some very sharply (North America, Latin America, southeast Asia), others more marginally (north Africa and Middle East). These contrasting trends are owing, of course, to the different paces of population growth over the past 2 centuries and to the resulting changes in the tempo and amplitude of the demographic transition. Since 1800, the population of the human continents has quadrupled, but that of the rest of the world has increased more than eightfold.

Two of the changes during the period deserve special emphasis: the settlement of North America and the formation of several high-density cores, often coastal, in Latin America and Africa.

In about 1800, North America, including Canada, had a combined population of barely 5 million. Two centuries later, the total stands at 280 million, one-third of whom are concentrated in the "*outlined continent*" (Debié, 1995) consisting of the eastern part of North America. The speed of development of this extended population concentration is spectacular. Whereas it took millennia to produce the *full worlds* of Eurasia, a few decades sufficed for the United States and Canada to build a territory with many similarities but of far smaller size. This speed was the result of massive immigration spread out over barely 2 centuries. It also bears powerful witness to the territorial efficiency of American society as regards the appropriation of space (balanced administrative division into *townships*), land use, and establishment of links between localities.

European immigration and the demographic transition combined to multiply the population of Latin America by 20 between 1800 and today. This prodigious growth—the steepest outside the United States and Canada—was achieved in a territory deeply shaped by an extraverted colonial economy and polarized around a few port towns, Mexico City excepted. This explains the formation of a series of concentrations, most of them coastal, that together constitute the backbone of the Latin American settlement archipelago.

Because of the lateness of its demographic transition and the absence of actual settlement colonies, sub-Saharan Africa experienced far weaker growth than did Latin America. From the early 19th century to the present, its population has risen fivefold from 90 million to nearly 500 million. As in South America, demographic growth led to the formation of a string of coastal settlements. But it also contributed to the strengthening of the settlement core in northern Nigeria and the emergence of other high-density cores inside the continent—most notably the Great Lakes region and, to a lesser extent, the Ethiopian mountains.

VII. TOWARD AN INTERPRETATION

1. Environmental Prerequisites for Settlement

The broadening of the ecological bases of inhabited territories in the past 50 years has led to the gradual abandonment of the analysis of the environmental prerequisites for settlement (Pumain, 1992). We believe, however, that these preconditions must be taken into account, as the formation of population densities is ultimately rooted in the capacity of societies throughout history to tap the resources at their disposal.

To begin with, the settlement of a portion of the landmass requires the ability to extract, produce, or introduce sufficient quantities of two commodities essential to survival: food and water (Claval, 1974). The notion has long prevailed that this possibility depended exclusively on the characteristics of the natural environment, with climate and soil determining the quantity and quality of available resources, and terrain and hydrography determining the capability to bring foodstuffs into the area. Today, it is admitted that availability of food and water are jointly determined by biophysical and socioeconomic characteristics. The productive capacity of a cultivated ecosystem depends on the surface area and fertility of land actually under cultivation. These two variables are determined, in turn, by the characteristics of the cultivated ecosystem and by the methods used in the existing agrarian system to restore soil fertility (Mazoyer and Roudart, 1997).

Population volume in a given area therefore depends on the existing local ecosystem and on the performance of the agrarian systems that have been developed there. The characteristics of these biophysical and socioeconomic components of productive capacity are not immutable: they evolve at the slow pace of climate change or the far quicker pace of social, economic, and political change. Moreover, their state at any given moment is influenced by their previous state: an ecosystem's fertility and the method of replenishing it depend, in turn, on the legacy of earlier agrarian systems. Recent research has also emphasized the role of collective perceptions in the relationship that develops between environmental conditions and settlement (for a summary, see Baudelle, 2000). In this interpretation, what matters is not so much the quality or weaknesses of a particular environment but the value assigned to it—by comparison with neighboring environments—by the human communities that have settled there or would like to. This value depends not only on objective social conditions (production methods and relationships) but also on beliefs and habits. These enable us to understand some original features of settlements. In Japan, for example, the sacred value attributed to mountains has shaped a striking contrast between the relative emptiness of high-altitude regions and the dense occupancy of the plains (Gourou, 1984).

Within the framework outlined above, it is not surprising to observe that the relationships between environmental conditions and population densities vary considerably from one part of the inhabited world to another, and that no generalization is possible on this issue. The relationships between settlement and climate and between settlement and terrain provide convincing illustrations of this.

In the early 1960s, Jozef Staszewski (1961) showed that the highest demographic densities (on average, 60 persons per square kilometer) were found in humid temperate climates and humid subtropical climates (monsoon climate); excluding polar, desert, and equatorial climates, the lowest densities were observed in dry tropical climates (8 persons per square kilometer) and humid tropical climates (14.4 persons per square kilometer). Most important, he demonstrated that the population density for a given climate varied widely from one continent to another, and that the climate prevalence ranking changed from one continent to another. He thus demonstrated at the global scale what many monographs had already underlined or would later underline at the local level, namely, that, in comparable climate conditions, human densities can vary across a very broad range of values (Gourou, 1973): the Mississippi plains have 10 times fewer inhabitants than do the Yangzijiang or Huang-He plains, the Congo basin is nearly empty whereas in Java rural densities reach nearly 2000 persons per square kilometer, and so on.

The analysis of population distribution by altitude also leads to nuanced conclusions (Staszewski, 1957; for a summary, see Bähr *et al.*, 1992). Admittedly, at the global scale, 60% of the world population lives below 200-m altitude and 80% below 500 m. Likewise, again at the global scale, density falls continuously with altitude, consistently with the reduction of farming opportunities as the temperature decreases. As with climates, however, there are sharp differences between continents: less than 1% of the population lives above 1000 m in Australia and Europe, but more than 20% does in Africa and South America. On the latter continent, the proportion of the population living above 2000 m averages 11%, but exceeds 50% in Peru and Ecuador and 70% in Bolivia. These variations show that the appeal of mountains depends on their environment: although the contrast between plain settlements and altitude settlements is sharp at high and medium latitudes, it is more modest, or even reversed, at Mediterranean and tropical latitudes (Belbeoch *et al.*, 1986). For the latter, Olivier Dolffus (1990: 454) notes that, "the plains may seem less safe, less salubrious, less fertile or harder to clear. The Andes were less challenging to occupy than the Amazon forest or the deserts below. This difference is partly due to the societies' technological level. Slopes become an ever more costly obstacle to deal with when motors and axles are used in transportation and agriculture. By contrast, in rural societies that rely on basic tools, the work yield on a 15° slope is barely lower than on a flat terrain."

2. A Distribution Whose Origins Go Back a Long Way

The long-term inertia of the spatial distribution of the world population prompts us to look to Antiquity for an explanation of the contrasts observed today. From this standpoint, the uneven diffusion of agricultural techniques in the Neolithic revolution seems to play a key role—far more decisive than the variations in the absolute age of settlements.

As long as transportation remained slow, human societies had to procure the food and water needed for their survival in a narrow radius around their settlements. The intensity of their presence in an area therefore depended on their capacity to find resources in their immediate surroundings. Throughout the Upper Paleolithic, this capacity remained limited, as the prevailing survival techniques (gathering, hunting, fishing) were simple collection methods without restitution. They required human groups to be mobile over large areas and prevented the formation of significant concentrations. As a result, the density contrasts in inhabited regions were presumably weak.

Ten thousand years before our era, the advent of agriculture and livestock raising during the Neolithic revolution brought slow but deep changes in the characteristics of the earlier settlement. By allowing a sharp increase in food production per unit of surface area, it made possible—in places where it took hold—the establishment of the first significant population concentrations. In the Middle East, for example, densities rose from no more than one person per 20 to 40 km^2 among the hunter-gatherers of the Upper Paleolithic to 2 to 5 persons per square kilometer with the advent of the earliest forms of rain-dependent agriculture. This created the first settlement contrasts between preagricultural economies, in which densities stayed below one person per square kilometer, and agricultural economies, in which densities substantially exceeded that limit. Consistently with Ester Boserup's model (1965), the progressive differentiation of farming techniques later broadened the range of possible densities—to the point at which the following patterns were found across the world at the start of the 16th century (Debié, 1995; Grigg, 1996; Mazoyer and Roudart, 1997): a few people per square kilometer for the itinerant agriculture practiced, for example, in the west African savannas and the lowlands of Central and South America; about 10 persons per square kilometer when this form of agriculture, with shorter rotations, shifted toward more intensive forms as on the highlands of west Africa, in the Andes, and in Mexico; from 10 to 50 persons per square kilometer for agriculture with short fallow periods (biennial or triennial rotation), and animal-traction farming in the Mediterranean basin and most of Europe; nearly 100 persons per square kilometer for permanent rainfall-dependent agriculture (without fallowing) in Flanders, Tuscany, and Lombardy; and 200 to 300 persons per square kilometer, or locally more, for irrigation farming in the Indus-Ganges plain, the Gulf of Bengal, the Malabar and Coromandel coast, southern China (Chang Jiang valley, Sichuan, Guandong lowlands), Japan, Java, and the deltas of southeast Asia.

In sum, a long-standing spatial concordance exists between areas of intensive agriculture and dense settlement. This link still harbors many mysteries, most notably because the functional relationship between agrarian structures and settlement is so complex and varies according to the political framework in which it operates. Research on the issue has yielded two important findings. First, the studies by Ester Boserup have shown that the relationship between farming techniques and population densities constituted a system: progress in farming techniques allowed population growth, but, reciprocally, increased densities required or stimulated advances in agrarian structures. Second, archeological studies indicate that, as early as the Neolithic, agricultural progress generated surplus production that allowed work specialization, fostered trade, encouraged the emergence of political control structures, and promoted the birth and rise of towns. Reciprocally, these transformations stimulated technical progress, most notably in agriculture.

François Durand-Dastés (1995) has drawn on these findings to propose a model for interpreting the high overall densities observed in India. He begins by pointing out that rice farming is an agricultural production system characterized by very high yields (quantity of food produced per surface-area unit) but very low productivity (quantity of food produced per labor unit). Rice farming therefore sustains high population densities, which, in return, help to perpetuate and even expand it. Durand-Dastés also identifies three links that have developed between rice farming and control structures. First, high rice yields generated large surpluses, allowing urbanization and the formation of fairly centralized governments. Second, the mobilization of the labor force and the management of large-scale infrastructure development—which only a central government was capable of achieving—played a likely role in the growth of rice farming. Third, organized states have the capability of setting conflicts inside their borders and can thus ensure the existence of *peaceful areas* conducive to population growth. In combination, these three feedback *loops* form a system with an operation that—according to Durand-Dastés—is responsible for the close correlation between rice farming in floodable areas and high population densities observed in India.

The historical developments between the early Neolithic to the Age of Discovery were therefore largely responsible for the differential pattern of human settlement. For the most part, developments in later centuries "strengthened the continents, the high-density axes, and the settlement cores already sketched out, multiplying the densities of the early modern era by five, ten, and sometimes a hundred" (Debié, 1995: 327). They also brought some important adjustments linked to the rise of the European world-economy (Braudel, 1979; Wallerstein, 1980a,b). Examples include the demographic debacles in America and Oceania, population transfers owing to the growth of the slave trade in Africa, but also, more indirectly, to the concentration of the population in the coastal areas of colonized territories. The latter phenomenon was owing to the centrality of ports in the colonial economy and the expansion of "funnel-shaped" transportation networks. Likewise, the economic and demographic disruptions of the 19th century, by triggering trans-Atlantic migrations and fostering the expansion of pioneer frontiers, enlarged the settlement envelope to some extent. But the outcome of the process had already been determined long before—indeed, very long ago.

CONCLUSION

A panel at the main entrance of the Lisbon World Exhibition (1998) dedicated to oceans stated that, at the close of the 20th century, almost 60% of the world population lived on sea coasts. This claim is highly fanciful, as the recent computations by Daniel Noin (1999) found that barely 16% of the Earth's inhabitants live within 15 km of a coast. Nevertheless, the panel—widely quoted in the media—points to a key aspect of human settlement: its pericontinental distribution. Today, more than one-quarter of the world population lives within 50 km of a coast, almost one-half at 200 km, and more than two-thirds at 500 km. Most of the main population centers are located in the immediate vicinity of the sea or communicate with a seaboard; many metropolises with populations in the millions are coastal. As many other aspects of settlement, this concentration seems to be an enduring pattern. It was already observed at the start of our era and was consolidated by the expansion of the European colonial economy from the 16th century onward. However, it may be called into question in the near future. Experts predict that climate change in the 21st century could raise the average sea level and threaten the coastal plains, some of which—such as northwest Europe and the Bangladesh lowlands—are heavily populated. Will the potentially displaced populations find shelter along the new emerging coastal settlement bands, thereby increasing concentration in the lowlands or, on the contrary, will they be forced into new locations, perhaps outside the settlement envelope? Such questions belong to the realm of *geography fiction* for now but could become topical at fairly short notice.

Acknowledgment

This chapter was translated from the French version by Jonathan Mandelbaum.

References

Allix Jean-Pierre, 1996. *L'Espace humain. Une invitation à la géographie.* Paris, éditions du Seuil, 419 p.

Bähr Jürgen, Jentsch Christoph, and Kuls Wolfgang, 1992. *Bevölkerungsgeographie.* Berlin, Walter de Gruyter, 1158 p.

Baudelle Guy, 2000. *Géographie du peuplement.* Paris, Armand Colin, 192 p.

Beaujeu-Garnier Jacqueline, 1956, 1958. *Géographie de la population,* 2 vols. Paris, Génin, 893 p.

Belbeoch Olivier, Charbit Yves, Faron Olivier, Magescas Jean-Bernard, and Mata Carmen, 1986. *La population mondiale. Vers une stabilisation au XXI^{e} siècle?* Paris, La Documentation Française, 143 p.

Biraben Jean-Noël, 1979. Essai sur l'évolution du nombre des hommes, *Population,* vol. 34(1), p. 13–26.

Boserup Ester, 1965. *The conditions of agricultural growth: the economics of agrarian change under population pressure.* London, Allen and Unwin, 124 p.

Braudel Fernand, 1979. *Civilisation matérielle, Economie et Capitalisme: XV^{e}–XVII^{e} siècle,* Vol. 1: *Les stuctures du quotidien.* Paris, Armand Colin, 544 p.

Chaunu Pierre, 1979. *Un futur sans avenir. Histoire et population.* Paris, éditions Calmann-Lévy, 320 p.

Claval Paul, 1974. *Éléments de géographie humaine.* Paris, éditions M.T. Génin, 412 p.

Debié Franck, 1995. *Géographie économique et humaine.* Paris, PUF, 750 p.

Decroly Jean-Michel and Vanlaer Jean, 1991. *Atlas de la population européenne.* Brussels, Éditions de l'Université de Bruxelles, 172 p.

De Saint-Exupery Antoine, 1931. *Vol de nuit.* Paris, Gallimard, 182 p.

Dollfus Olivier, 1990. Le système monde, *in:* Roger Brunet (ed.), *Géographie universelle,* Vol. 1: *Mondes nouveaux,* p. 274–529. Paris, Hachette/Reclus, 551 p.

Durand-Dastes François, 1995. Monde indien, *in:* Roger Brunet (ed.), *Géographie universelle,* Vol. 8: *Afrique du nord, Moyen-Orient, monde indien,* p. 246–463. Paris, Hachette/Reclus, 352 p.

Fagan Brian M., 1990. *The journey from Eden: the peopling of our world.* London, Thames and Hudson, 256 p.

George Pierre, 1951. *Introduction à l'étude géographique de la population du monde.* Paris, PUF, 284 p.

Gourou Pierre, 1973. *Pour une géographie humaine.* Paris, Flammarion, 299 p.

Gourou Pierre, 1982. *Terres de bonne espérance. Le monde tropical.* Paris, Plon, 391 p.

Gourou Pierre, 1984. *Riz et civilisation.* Paris, Fayard, 456 p.

Grasland Claude, 1998. La composante d'échelle dans l'analyse des distributions spatiales, *Revue belge de géographie,* vol. 122(4), p. 435–457.

Grigg David, 1996. *The agricultural systems of the world: an evolutionary approach.* Cambridge, Cambridge University Press, 358 p.

Hëgerstrand Torsten, 1973. The domain of human geography, *in:* Richard Chorley (ed.), *Directions in geography,* p. 67–87. London, Methuen, 331 p.

Le Bras Hervé, 1993. *La planète au village. Migrations et peuplement en France.* Paris, DATAR and Éditions de l'Aube, 222 p.

Mazoyer Marcel and Roudart Laurence, 1997. *Histoire des agricultures du monde du néolithique à la crise contemporaine.* Paris, éditions du Seuil, 533 p.

Mcevedy Colin and Jones Richard, 1978. *Atlas of world population history.* Harmondsworth, Penguin, 368 p.

Noin Daniel, 1979. *Géographie de la population.* Paris, Masson, 325 p.

Noin Daniel, 1997. *People on earth: world population map.* Paris, UNESCO Publishing, 43 p.

Noin Daniel, 1999. La population des littoraux du monde, *L'information géographique,* vol. 2, p. 65–73.

Noin Daniel and Thumerelle Pierre-Jean, 1993. *L'étude géographique des populations.* Paris, Masson, 124 p.

Pain Marc, 1998. Politique de peuplement en Indonésie. Transmigration et migrations spontanées au centre des débats, *Hérodote,* no. 88, p. 26–46.

Pinchemel Philippe and Pinchemel Geneviève, 1988. *La face de la Terre. Eléments de géographie.* Paris, Armand Colin, 520 p.

Pumain Denise, 1992. Le peuplement, *in:* Antoine Bailly, Robert Ferras, and Denise Pumain (eds.), *Encyclopédie de géographie,* p. 439–462. Paris, Economica, 1132 p.

Racine Jean-Bernard and Villeneuve Paul, 1992. Canada, *in:* Roger Brunet (ed.), *Géographie universelle,* Vol. 4: *états-Unis, Canada.* p. 231–458. Paris, Hachette/Reclus, 480 p.

Radvanyi Jean, 1996. *La nouvelle Russie. L'après 1991: un nouveau "temps des troubles."* Paris, Masson / Armand Colin, 406 p.

Raison Jean-Pierre, 1988. La colonisation des espaces vides forestiers tropicaux, *L'information géographique,* vol. 52(5), p. 206–208.

Ratzel Friedrich, 1881, 1991. *Anthropogeographie, Oder Grundzüge der Anwendung der Erdkunde auf die Geschichte,* 2 vols. Stuttgart, Engleborn, 1287 p.

Sorre Max, 1948. *Les fondements de la géographie humaine*. Paris, Armand Colin, 608 p.

Staszewski Jozef, 1957. *Vertical distribution of world population.* Warsaw, State Scientific Publishing House, 115 p.

Staszewski Jozef, 1961. Bevölkerungsverteilung nach Klimagebieten von W. Köppen, *Petermanns Geographische Mitteilungen,* vol. 105(2), p. 133–138.

Tobler Waldo, Deichmann Uwe, Gottsegen John, and Maloy Kelly, 1997. World Population in a Grid of Spherical Quadrilaterals, *International Journal of Population Geography,* no. 3, p. 203–225.

Vallin Jacques, 2000. La fin de la transition démographique: soulagement ou inquiétude? *in:* Yves Michaud (ed.), *Qu'est-ce que l'humain?* p. 202–217. Paris, Éditions Odile Jacob, 602 p.

Vandermotten Christian and Decroly Jean-Michel, 1990. La grande périphérie, *in:* Christian Vandermotten (ed.), *L'Europe et ses régions: diagnostic interrégional*, p. 69–79. Charleroi, Brussels, CIFoP, SRBG, and DULBEA, 223 p.

Vandermotten Christian and Marissal Pierre, 2003. *La production des espaces économiques,* Vol. 2. Brussels, Éditions de l'Université de Bruxelles, 295 p.

Vidal de La Blache Pierre, 1917. La répartition des hommes sur le globe, *Annales de géographie,* nos. 140 and 142, p. 81–93, p. 241–254.

Vidal de La Blache Pierre, 1922. *Principes de géographie humaine.* Paris, Armand Colin, 327 p.

Wallerstein Immanuel, 1980a. *Le système-monde du XV^e siècle à nos jours,* Vol. 1: *Capitalisme et économie-monde, 1450–1640*. Paris, Flammarion, 331 p.

Wallerstein Immanuel, 1980b. *Le système-monde du XV^e siècle à nos jours,* Vol. 2: *Le mercantilisme et la consolidation de l'économie-monde européenne. 1600–1750.* Paris, Flammarion, 503 p.

Wilkinson David, 1993. Civilizations, cores, world economies, and oikumenes, *in:* Andre G. Frank and Barry K. Gills (eds.), *The World-System: five hundred years or five thousand?* p. 221–277. London, Routledge, 320 p.

Zelinsky Wilbur T., 1966. *A prologue to population geography.* Englewood Cliffs (NJ), Prentice-Hall, 150 p.

CHAPTER

71

Demographic Dynamics and Diversity in the World (1950–2000)

BRUNO SCHOUMAKER, DOMINIQUE TABUTIN, AND MICHEL WILLEMS

Institut de Démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

INTRODUCTION

Humankind has never experienced such demographic upheavals as in this second half of the 20th century: a huge growth of the population (which increased from 2.5 billion to more than 6 billion between 1950 and 2000), a new distribution of the population on the planet (with the regions of the south now representing 80% of the total versus 67% in 1950), a concentration of the population in the towns or on the coasts, and consequent changes in the demographic dynamics of most societies. As mentioned in Chapter 69, after and largely in response to the transformations observed in Europe (Chapter 68), a genuine demographic revolution began that is irreversible, although it occurs at different rates in the different cultures or economic systems. Although the transition was quite advanced by the 1950s in the countries in the north, it was just beginning in those in the south, gaining momentum in the 1970s and 1980s. Heterogeneous situations, diverse rhythms of changes, spatial and social inequalities are the dominant features of contemporary demography that will be described in this chapter. The focus is on the 1960–2000 period, in which major changes occurred. First, the patterns of demographic transition and population growths will be explained, fertility, mortality and age structures will be described, and finally a few questions concerning the future will be introduced, including the demographic impact of AIDS.[1]

[1] Migration will not be discussed in this chapter.

I. SOURCES, INDICATORS, AND GEOGRAPHIC DIVISIONS

In an overview of the demography of various regions and countries in the world over 40 years, statistic works of historic compilation and perspectives conducted by specialized organizations such as the United Nations (UN) or the World Bank should be consulted. We will use the database of the UN Population Division (UN, 1996a,b, 2001a,b), which is probably the best known and most detailed. Although it is revised every 2 years, we shall mostly use the 2000 revision, as well as results on fertility and child mortality from recent surveys (such as the Demographic and Health Surveys [DHSs] conducted in the 1990s).

The indicators selected are the usual ones—birth, death, and natural growth rates to examine the transition patterns, the total fertility rate (TFR), life expectancy, and infant mortality,[2] as well as the proportions of population aged under 15 and over 65 years as indicators of the age composition.

We have gathered the above-mentioned indicators on 187 countries[3] and over three periods (1960–1965,

[2] We wish we could have used the 0- to 5-year-old probability of dying, which is a better indicator of mortality at early ages, but it is not systematically published by the United Nations nor is it available from other sources in all the countries.

[3] The United Nations publish them only on countries and territories with a population of more than 140,000. The data on Melanesia, Micronesia, and Polynesia sometimes are not included in the analyses, making a total of 179 countries in certain tables and charts.

1975–1980, 1995–2000).[4] In Appendix Table 71–1, the average numbers are presented in table form according to the 6 major regions and 21 subregions as defined by the UN with a view to synthesis and minimal diversification. UN terminology divides the countries among three categories: developed (44 countries, 20% of the world population), least developed (which we will refer to as the poorest; 45 countries, 32 of which are in sub-Saharan Africa, 10% of the world population), and intermediate (98 countries, 70% of the total).[5] Although these three major categories offer different demographic systems, some heterogeneity can be observed within each of them. Another form of classification is a division of the world into nine major regions: Europe; North America, South America, Central America, and the Caribbean; North Africa; sub-Saharan Africa; the Near East; Central Asia; South Asia; Southeast Asia; and the Far East. These countries are listed in table form according to those geographic classification in Appendix Table 71–1.

All averages, variances and other indices are weighted by population size of the countries under study.

II. THE TRANSITION PATTERNS FROM 1950–2010

The chapter moves from a world perspective, as outlined in Chapter 69, to examine regional patterns, going back to 1950 and using the projection (mean hypothesis, medium variant) suggested as far ahead as 2010, in the framework of the longer-term projections that will be discussed in Chapter 77.

1. The World

The population of the planet has never increased as quickly as it did in the second half of the 20th century. With birth and death rates of about 38% and 20% in 1950 (Fig. 71–1), the population increased at an annual rate of 1.8%, or about 48 million individuals per year. With decreasing mortality in the countries of the south, the growth gained momentum to about 1970 (2.1% a year) before slowing down as a consequence of declining fertility, particularly in China (from the early 1970s). For 15 years (from 1975–1990), growth remained at about 1.7%. In the late 1990s, growth was estimated at 1.3%, with a birth rate of 22% and a death rate of 9%, which meant an annual increase of approximately 80 million people. According to UN projections, the growth rate will still probably be 1.1% in about 2010. The transition takes place by stages, and on the whole, it has been slower than was expected in the 1960s or the 1970s, leading to several revisions of the mid- and long-term projections in the past 30 years.

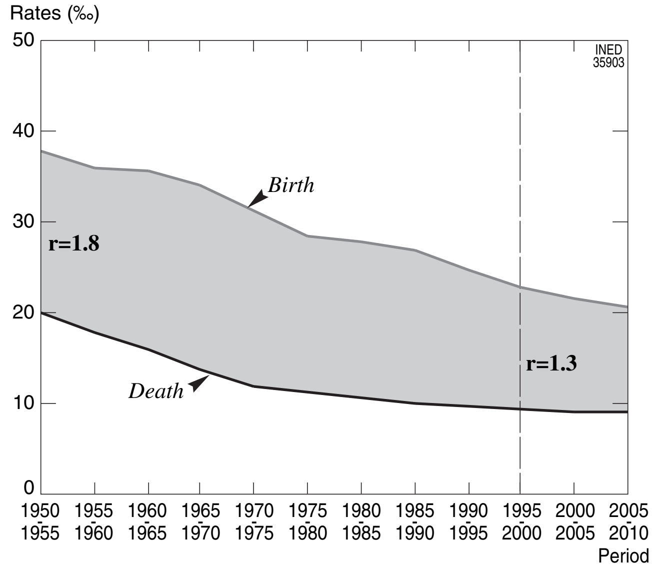


FIGURE 71–1 The transition pattern of the world population from 1950–2010. (Data from United Nations, 2001a,b.)

2. The Three Major Groups of Countries

World demography is an abstraction; in fact, some people doubt that the concept is of any interest. From the distinction between developed, intermediate, and poor countries emerge three major categories of transition histories, which as early as 1950 have been very different from one another (Fig. 71–2; Table 71–1).

The developed, rich countries—In these countries the decline of fertility and mortality dates back to the second half or the end of the 19th century.[6] By the beginning of the 1950s, the birth (22%), death (10%), and natural growth (12%) rates were comparatively low among the countries of this group. Nevertheless, after the baby boom of the 1940s and 1950s, fertility declined again at the end of the 1960s, falling drastically in the 1970s and 1980s and causing the

[4] Sometimes we go as far back as 1950 or 2010.

[5] Those numbers are 55, 48, and 125, respectively, when countries with a population of less than 140,000 are included.

[6] For most of these countries at least, as some had begun earlier (e.g., France, where fertility started to decline as early as the late 18th century) and others later (Southern and Eastern Europe, where the trend really started only in the early 20th century). Those historical discrepancies between the Western countries have been previously discussed, in Chapter 68. Readers may also see the works by Jean-Claude Chesnais (1986) or Patrick Festy (1979).

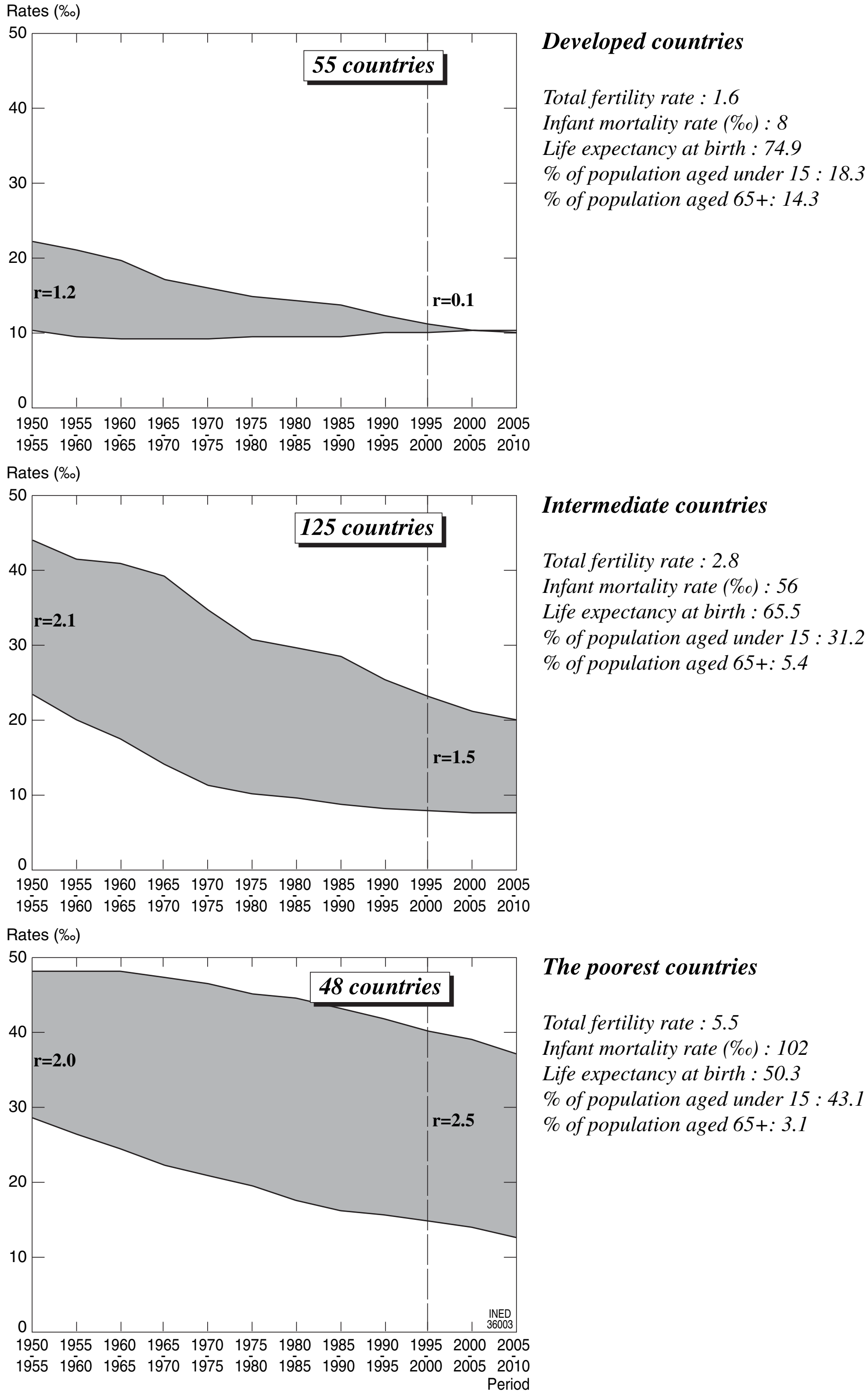


FIGURE 71–2 Patterns of transition from 1950–2010 in the three major groups of countries. (Data from United Nations, 2001a,b.)

growth to cease (or even become negative), as well as accelerating the aging of the population.

The intermediate countries—In this group, which is the largest and most heterogeneous, birth rates have been decreasing by stages (Fig. 71–2), sometimes since the 1950s but mostly since the mid-1960s. Death rates also declined, enough to counterbalance the decline of fertility, so that the growth rates went down only in 1950 (1.5% versus 2.1%). With an average fertility rate of 2.8 children per woman, a life expectancy of 65

TABLE 71–1 A Few Indicators of Structures and Demographic Change in the Three Major Country Groups and in the Countries in Extreme Situations

Indicators	Period	World average (n = 228)	Poorest countries (n = 48)	Intermediate countries (n = 125)	Developed countries (n = 55)	Countries and extreme values: Minima	Countries and extreme values: Maxima
Crude birth rates (%)	1960–1965	35.4	48.4	41.1	19.7	13.4 Hungary	56.3 Niger
	1975–1980	28.0	45.4	30.7	14.9	10.3 Germany	57.2 Malawi
	1995–2000	22.5	40.4	23.1	11.2	7.7 Latvia	55.4 Niger
Crude death rates (%)	1960–1965	15.6	24.6	17.4	9.4	5.8 Byelorussia	32.3 Sierra Leone
	1975–1980	10.9	19.7	10.3	9.5	4.1 Guam	45.0 East Timor
	1995–2000	9.0	14.9	7.8	10.2	2.2 Kuwait	26.4 Sierra Leone
Natural growth rates (%)	1960–1965	2.0	2.4	2.4	1.0	0.3 Hungary	3.6 Costa Rica
	1975–1980	1.7	2.6	2.0	0.5	–2.0 East Timor	3.8 Kenya. Solomon
	1995–2000	1.4	2.6	1.5	0.1	–0.6 Bulgaria	4.1 Yemen
Total fertility rates	1960–1965	5.0	6.7	5.9	2.7	1.8 Hungary	8.1 Kenya
	1975–1980	3.9	6.4	4.4	1.9	1.5 Germany, Luxembourg	8.5 Rwanda
	1990–1995	2.8	5.5	2.8	1.6	1.1 Latvia	8.0 Niger
Life expectancy at birth	1960–1965	52.4	39.8	48.9	69.7	32.0 Sierra Leone	73.5 Sweden
	1975–1980	59.8	45.3	58.7	72.3	31.2 Cambodia, East Timor	76.3 Iceland
	1995–2000	65.0	50.3	65.5	74.9	37.3 Sierra Leone	80.5 Japan
Infant mortality rates (‰)	1960–1965	119	171	131	33	15.0 Sweden	221 East Timor
	1975–1980	88	140	91	18	8.0 Sweden	263 Cambodia
	1995–2000	60	102	56	8	4.0 Japan	165 Sierra Leone
Proportion of people aged 15 years or less (%)	1965	37.6	43.7	41.4	27.5	21.0 Sweden	50.8 Bahrain
	1980	35.1	44.6	38.4	22.5	18.5 Germany	50.2 Yemen
	2000	30.0	43.1	31.2	18.3	14.3 Italy	50.1 Yemen
Proportion of people aged 65 years and older (%)	1965	5.3	3.1	3.9	9.0	1.6 Samoa	13.2 Austria
	1980	5.9	3.1	4.2	11.6	1.1 Qatar	16.3 Sweden
	2000	6.9	3.1	5.4	14.3	1.5 Qatar	18.1 Italy

n is the number of countries.
Data from UN, 2001.

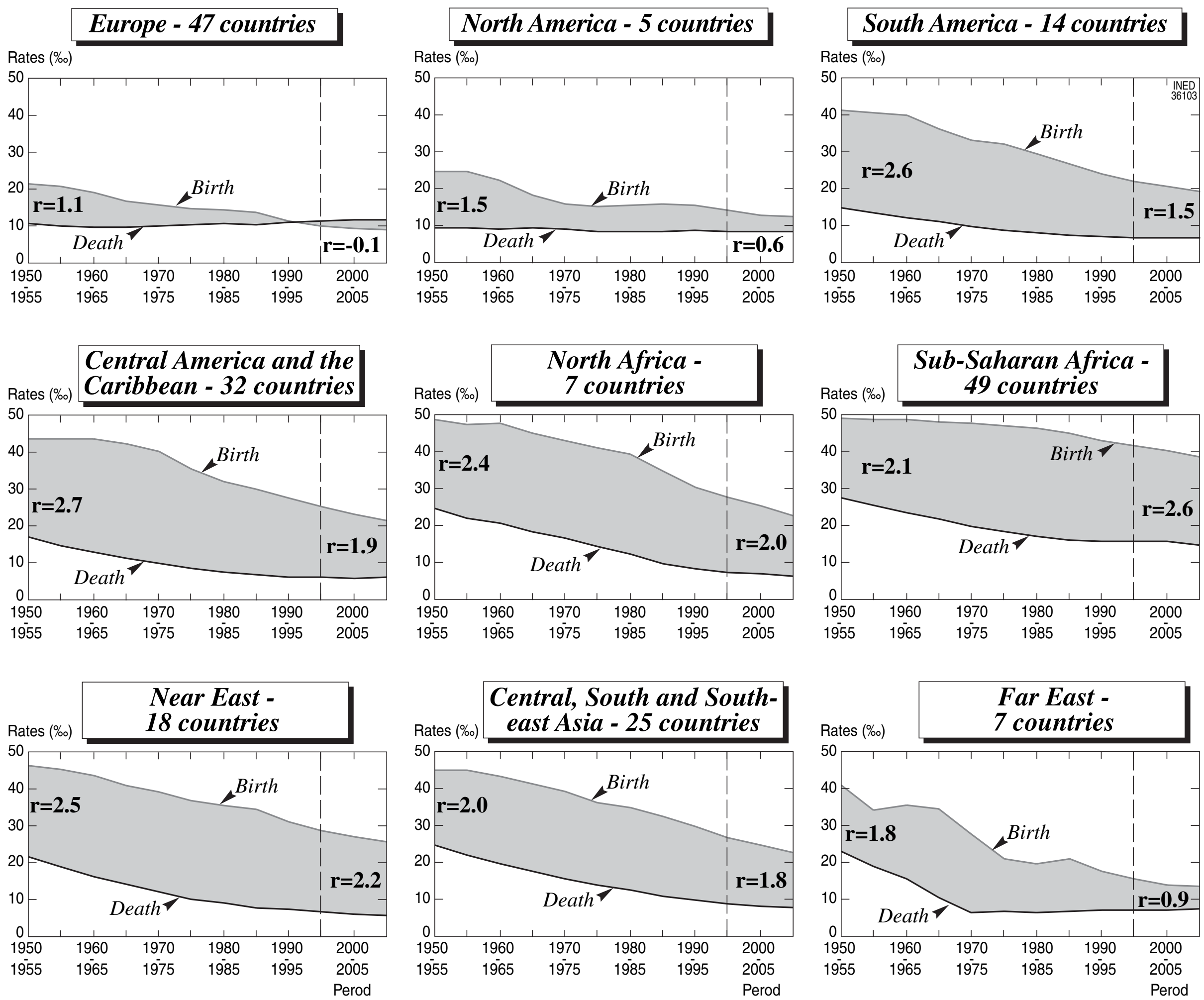


FIGURE 71–3 The transition patterns of nine major regions from 1950–2010. (Data from United Nations, 2001a,b.)

years, and the fast aging of the age structure (the proportion of the population aged under 15 dropped from 45% down to 31%), this group is well advanced on the path of demographic transition.

The poorest countries—Some 50 countries, the poorest ones in the world and those who are excluded from economic and social development, represent over 10% of the world population and 13% of all the developing countries. On the whole, they are still at an early stage. Birth (41%) and fertility (5.5 children per woman) rates have changed little, but mortality has decreased (the crude death rate dropped from 25% to 15% between 1960 and 2000). As a consequence, in the second half of the 20th century, growth increased, going from 2.0% in 1950–1955 to 2.5% in 1995–2000; according to UN projections, growth will reach even higher levels in 2010. However, the age structure will remain as young as it is now (about 43% of the population are aged under 15).

3. The Regional Diversification

In addition to the slowing down of world demographic population growth from 1950–2000, which was rather slow to occur, there is a great diversity in the history and present situations of the 21 subregions selected in Appendix 1, or even of the 9 major regions given in Fig. 71–3.

In the north, two groups of developed countries should be distinguished:[7] (1) Europe, in which birth

[7] On the figure we have not represented the Australia—New Zealand subregion (which appears as such in the table in Appendix 1), nor have we isolated Japan, integrating it into the Far East.

and death rates are more or less equal (about 10%) and growth is virtually nil; and (2) North America, in which growth is still slightly positive and is expected to remain so in 2010.

In the south, seven regions appear to be at different, more or less advanced stages of transition.[8] Within some 40 years, this group, the situation of which used to be comparatively homogeneous, has acquired unprecedented diversity. Although in 1955, all the subregions had pretransition birth rates greater than 40%,[9] the birth rates now range from 16% in the Far East to 43% in sub-Saharan Africa. Death rates, which already were more varied, now range from 5.4% (in Central America) to 15% (in sub-Saharan Africa). Last, growth rates, which in the early 1960s were between 2% and 2.7%, ranged from 0.9% (Far East) to about 3% (sub-Saharan Africa) in 2000. When the 1990s are compared with the 1950s, it appears that everywhere the rates went down, but the decrease occurred at diverse speeds: fast in the Far East and in South and Central America, and more slowly in the Near East, in north Africa, and Central and South Asia, as well as sub-Saharan Africa, where growth is still 2.6% per year.

Although in the 1950s and 1960s it was still possible to refer to the Third World population and quote some average numbers to situate it in contrast with the population of the Western world, that has changed. For some 30 years now, the various regions, countries, and the many diversified societies of which they are made, started their own transitions at more or less early dates and then continued at more or less rapid, more or less steady paces. Except for the Far East, the new balance that was promised or forecast, based on a zero growth with a fertility rate of about 2 children per woman and a life expectancy of 75 or 80 years, is still far from being achieved. Regional diversification and inequalities between the different countries have never been so great, a situation that is likely to persist for a large part of the 21st century.

III. FERTILITY: DECLINE IS WIDESPREAD, THOUGH AT UNEQUAL PACES

Fertility at the world scale has declined since the 1950s, dropping from 5.0 to 2.8 children per woman, particularly in the past 20 years. This decline is gradually becoming widespread, although at paces that vary from one region and one country to another.

As previously mentioned, the developed countries completed their fertility transition after a long process that usually started in the second half of the 19th century (see Chapter 68). The Far East has recently achieved the same rate (1.8 children per woman in 1998), after a decrease of fertility that will probably remain as the most rapid and dramatic of the whole history of humankind. On the whole, the decline was significant in all of the intermediate countries: the TFR dropped from 5.9 to 2.8 children per woman within 40 years (Table 71–1). For about 10 years now, the group of the 48 poorest countries has also been affected by the decline, although at a much slower pace (in these countries, the average TFR is still 5.5 children per woman).

The diversity is still greater when the trends *by region* (Fig. 71–4; Table 71–2) are examined, leaving aside the Western countries and the Far East. The starting point, in the 1950s and 1960s, is high levels of birth and fertility rates, which everywhere are far higher than those in pretransition Europe: about 45% (as against 35%) for the birth rate, and 6 to 7 children per woman (as against under 5) for the TFR. These fertility rates of the 1940–1960 period, at such a level of aggregation, were probably the highest of the history of humankind, as each region later experienced a decline of fertility, each at its own time and pace: first Latin America, as early as the beginning of the 1960s; then Central America and the Caribbean in the late 1960s; South and Southeast Asia in the early 1970s; and, last, the Near East (at a rather slow pace) and North Africa, which started from a higher level (7 children per woman) but decreased comparatively quickly, particularly from the 1980s (3.6 children by 1998). As for sub-Saharan Africa, a region that was long considered a pocket of resistance (some people even called it the African exception), the decline of fertility started in the 1990s (Fig. 71–4).

In fact, regional reproductive behaviors have been experiencing a rapid, deep change for the past 25 or 30 years. This change includes both significant changes in the area of family and marriage (particularly as the age at which women marry has increased) and the extension of contraception. The decline has been substantial on the whole, and it has even been faster than certain observers had predicted in the 1960s or 1970s.

At the *national level*, both the intensity and variability of changes were confirmed by studying the evolution of fertility between 1963 and 1998 in 179 countries (Fig. 71–5). Fertility remained virtually the same, or even slightly increased, in some 30 countries, mostly in sub-Saharan Africa (in addition to Bhutan, Laos, and

[8] For further details concerning those regions, see Tabutin (1995).

[9] It should be recalled that they were higher than in pretransition Europe in the 18th and 19th centuries (30% to 35%) as a result of marriage models that were traditionally earlier and more intense in most of the other societies and cultures.

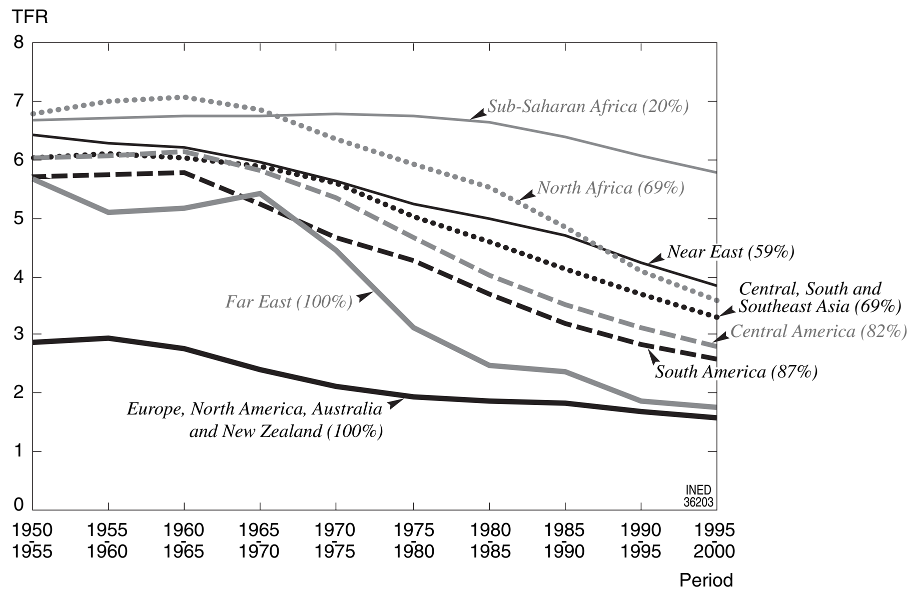


FIGURE 71–4 The evolution of the total fertility rate between 1950–1955 and 1995–2000 by major region. The figures between brackets are the proportions showing how much has been achieved with regard to replacement level (2.1) from 1950–1955 to 1995–2000. (Data from United Nations, 2001a,b.)

TABLE 71–2 Averages, Standard Deviations, and Coefficients of Variation of Total Fertility Rates in 1963, 1978, and 1998 in the Eight Major Regions

Regions	Average			Standard deviation			Coefficient of variation (%)		
	1963	1978	1998	1963	1978	1998	1963	1978	1998
Sub-Saharan Africa (46)	6.8	6.8	5.9	0.51	0.74	1.11	8	11	19
North Africa (7)	7.1	6.0	3.6	0.21	0.72	0.66	3	12	18
Near East (18)	6.3	5.3	4.0	1.24	1.59	1.80	20	30	45
Central, south, and southeast Asia (25)	6.0	5.0	3.4	0.58	0.66	0.84	10	13	25
Far East (7)	5.3	3.1	1.8	1.19	0.48	0.14	23	15	8
Central America and Caribbean (20)	6.5	5.0	3.0	0.84	1.15	0.82	13	23	28
South America (13)	5.8	4.3	2.6	1.21	0.62	0.45	21	14	17
Europe, North America, Australia, New Zealand (43)	2.8	1.9	1.6	0.45	0.30	0.34	16	16	21
Total, World (179)	5.0	4.0	2.9	1.65	1.63	1.47	33	41	51

The figures in parentheses are the number of countries in each region. The averages and standard deviations are weighted.
Data from UN, 2001.

Afghanistan), during that period. In all other countries, fertility has declined, but in different ways.

When the standard deviations and the coefficients of variation of TFRs among the countries in the eight regions in 1963, 1978, and 1998 (Table 71–2) are compared, or when the dispersions are examined graphically (Fig. 71–6), it is confirmed that the situations within regions are growingly heterogeneous. The only coefficients of variation that remain more or less stable are those concerning South America and Europe, and the only one that really decreases is that of the Far East (understandably, as it includes only six countries). At world level (Table 71–2), the standard deviation gets slightly smaller, but the coefficient of variation increased from 1963 to 1998.

Between regions, as well as countries, there is great diversity as to when the decline begins and to its pace, and it will probably remain so for some time yet. “All situations are possible: an early start and slow pace (India); an early start and fast pace (Mauritius); a late start and fast pace (Mexico); a late start and slow pace. Though the process, once irreversibly launched, on the

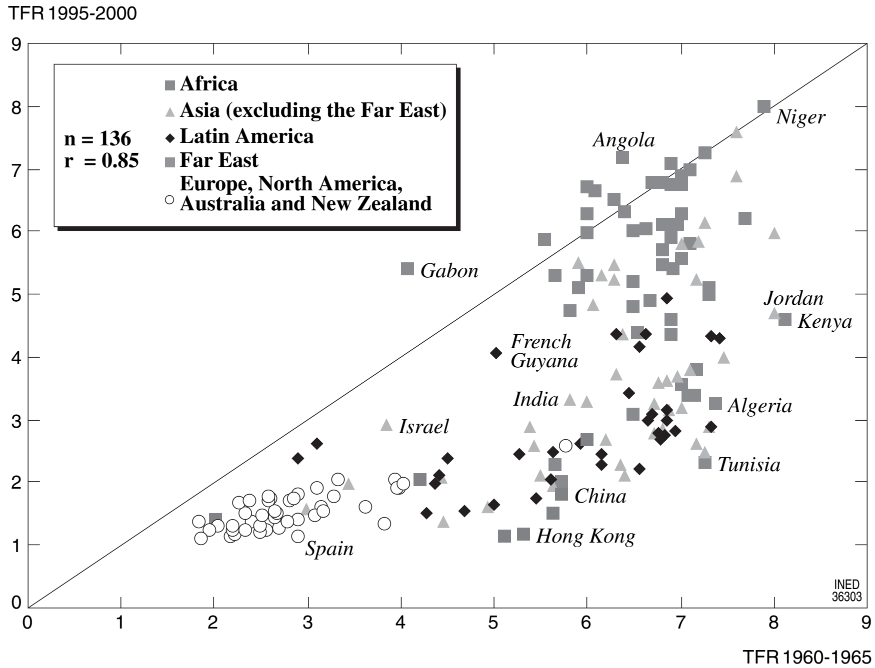


FIGURE 71–5 The evolution of total fertility rates between 1960–1965 and 1995–2000 in 179 countries. (Data from United Nations, 2001a,b.)

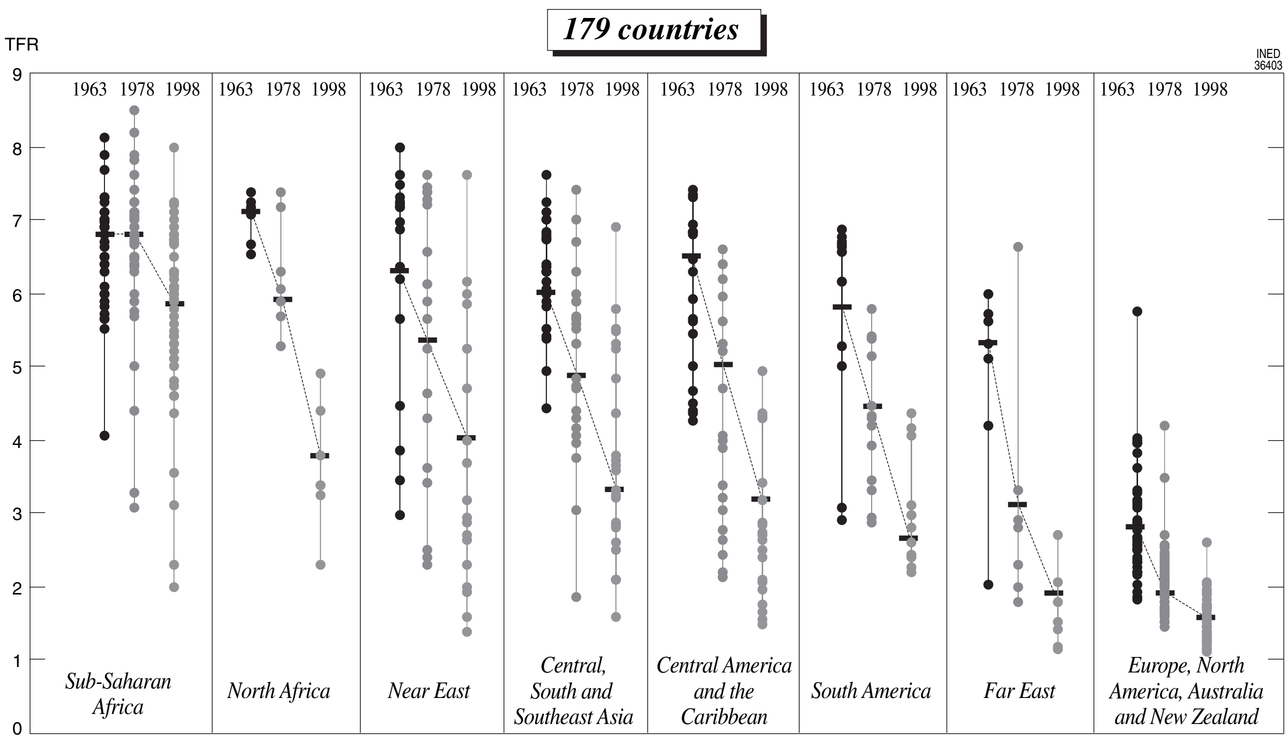


FIGURE 71–6 Dispersion between countries and average of total fertility rates in 1960–1965, 1975–1980, and 1995–2000 by major region. The dots represent the countries, and the dotted lines link the weighted regional averages between them. (Data from United Nations, 2001a,b.)

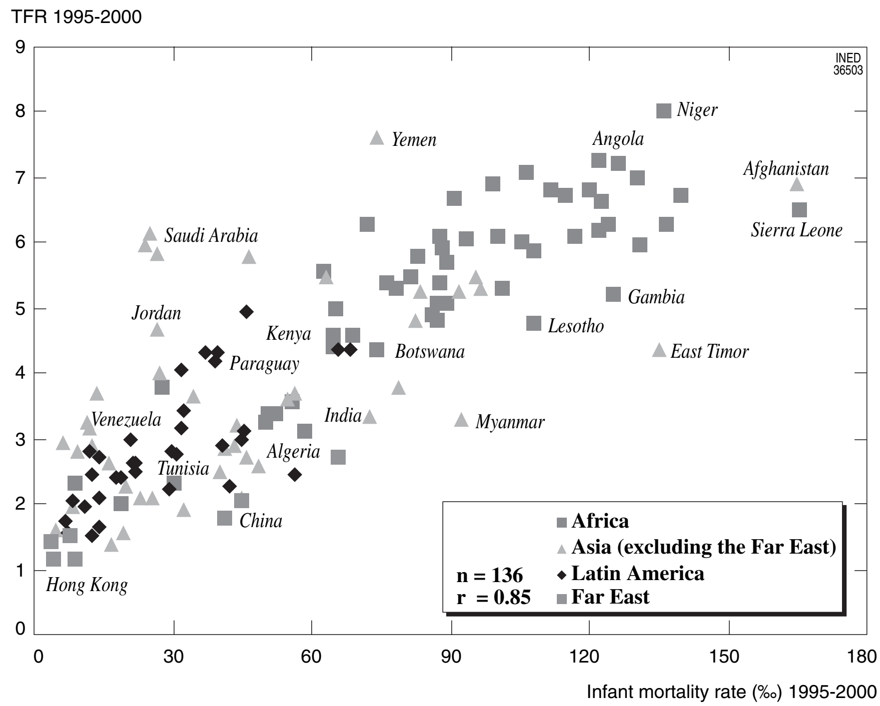


FIGURE 71–7 Relations between fertility and infant mortality in 136 countries in the south in 1995–2000. (Data from United Nations, 2001a,b.)

whole is faster than it was in Western countries in the past, there is no rule, nor general relations between starting levels, period and decline paces. A society's reproduction model is derived from complex dynamics that depend on historical, cultural, social, economic and political factors" (Tabutin, 1995: 46). In other words, every country has is own history, so that within one culturally homogeneous region (e.g., North Africa), the fertility decline processes, which are only one element of the evolution of societies among others, may occur in rather different ways, although they eventually converge.

In the theory of transition, *child mortality* always appears as one of the major variables that account for the variations and evolution of fertility: it is impossible to have very low fertility rates in contexts of high mortality levels or the population would be threatened with extinction. The decline of mortality is often presented as a precondition for the decline of fertility. What is the situation then with these two variables on the international level? Figure 71–7 presents the relation, in about 1998, between the TFR and the infant mortality rate in 136 countries in the south. The relation is clearly a positive one, with a correlation coefficient of 0.85, although some countries differ from the general pattern–for example, a few of the Emirates (which have a high fertility and a very low mortality rate), the India—Bangladesh—Myanmar group (with intermediate fertility but still high levels of mortality), or some countries in which mortality levels are clearly exceedingly high (East Timor, Liberia, Sierra Leone). However, it should be acknowledged that the latter group of countries have very specific contexts of political and military unrest.

Apart from a few exceptions, all countries in which the infant mortality rate is greater than 90% have a fertility rate of more than 5 children per woman; all countries in which mortality rate is under 30% have a fertility level of less than 3.5 children per woman. None of the countries with a fertility rate of 2 children per woman has an infant mortality rate over 25%, except for China (44%). All considered, no society in the south has achieved its fertility transition without its infant mortality first decreasing in a significant and often rapid way.

IV. MORTALITY: IMPROVEMENTS BUT GROWING INEQUALITIES

The health transition started in the late 18th century in Northern and Western Europe,[10] spreading on to North America, Japan, and southern and eastern

[10] For further details, see Chesnais (1986) or the 1989 issue of the *Annales de démographie historique*; also see Chapters 57 (in Volume II) and 68 (in Volume III).

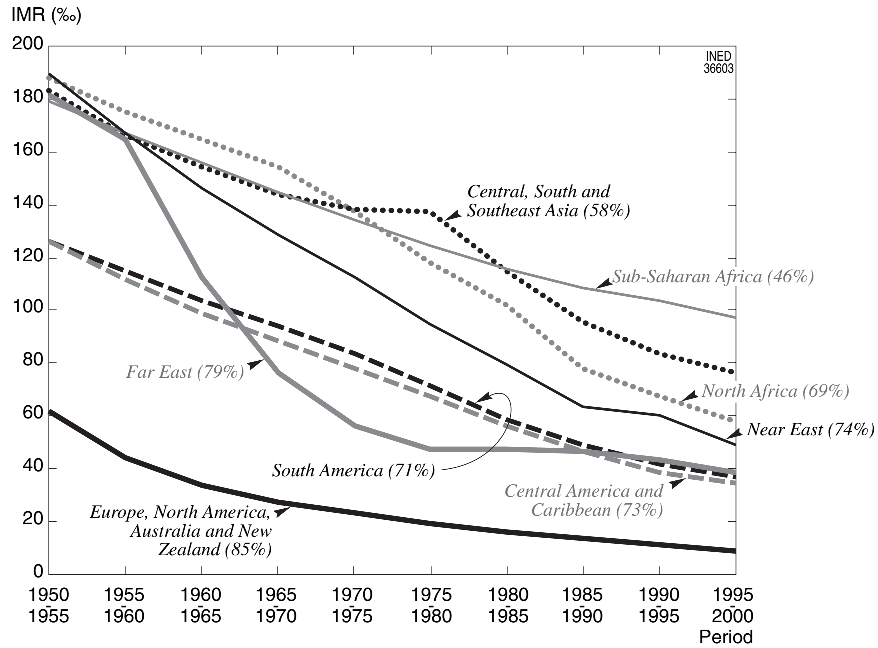


FIGURE 71–8 The evolution of infant mortality rates in 1950–1955 and 1995–2000 by major region. The figures between brackets are the proportions of decline from 1950–1955 to 1995–2000. (Data from United Nations, 2001a,b.)

Europe in the second half of the 19th century. Then it reached South America and a few countries in Asia between the two World Wars, before spreading to the whole of the planet in the 1950s. Just as for fertility, there have been significant differences in the timing and the pace of mortality changes.

As a consequence, just after the end of World War II, the situation was extremely diversified. In the early 1960s (see table in Appendix 1), life expectancy was 70 years in the developed countries, 49 years in the intermediate group, and barely 40 years among the poorest countries; the figures nowadays are 75, 66, and 50, respectively. Infant mortality rates then ranged between 33‰ and 171‰, in 1998 they ranged from 8‰ to 102‰. In the 21 subregions (see table in Appendix 1), life expectancy ranged from 39 to 71 years in the early 1960s; in 1998 the range is from 46 (East Africa) to 78 years (Australia and New Zealand).

It thus appears that the social and health progress that caused mortality to decline has affected every region and every country. Inequality, however, is now as great as it used to be, if not greater: there is a difference of 32 years between the extreme life expectancy figures by region, and infant mortality rates range from 1 to 20 (whereas they ranged from 1 to 8 in the 1960s). The poorest countries have hardly caught up, and in those countries most seriously affected, AIDS now tends to slow down or destroy the progress achieved in the area of mortality. Last, the diversity of situations has never been so great within the regions in the south.

Let us more closely examine this diversity in the area of infant mortality, the best known and probably most reliable at international level. It still represents 30% to 40% of death cases in many countries in the south, its relation to life expectancy[11] remains strong, and the pace at which it changes (on which the pace of life expectancy largely depends), is considered a good indicator of social and health progress.

In the eight regions that we have drawn (Fig. 71–8; Table 71–3), infant mortality has much declined in some 40 years, but at different paces: from 1963–1998, the decline was 76% in the developed countries (starting from a much lower level), 66% in the Far East, 64% in Latin America and North Africa, 54% in Asia, but only 38% in sub-Saharan Africa (where the gap with the other regions in the world has been widening since the 1980s). Just as in the area for fertility, the Far East has experienced an unprecedented health revolution as early as the beginning of the 1960s, although its effects on infant mortality have considerably slowed down in the past 15 years.

[11] Calculated on 172 countries (of all categories), the coefficients of linear correlation between life expectancy at birth and infant mortality are –0.96 in 1960–1965, –0.96 in 1975–1980, and –0.94 in 1990–1995. Calculated on 130 countries in the south, they are –0.94, –0.94, and –0.93, respectively.

TABLE 71–3 Averages, Standard Deviations, and Coefficients of Variation of Infant Mortality Rates in 1963, 1978, and 1998 in the Eight Major Regions

Regions	Average			Standard deviation			Coefficient of variation (%)		
	1963	1978	1998	1963	1978	1998	1963	1978	1998
Sub-Saharan Africa (46)	156	124	97	30	26	23	19	21	23
North Africa (7)	165	118	58	9	15	16	5	13	29
Near East (18)	146	95	49	52	37	25	35	39	52
Central, south, and southeast Asia (25)	151	119	69	27	27	23	18	23	33
Far East (7)	112	48	38	25	12	10	22	26	27
Central America and the Caribbean (20)	95	65	34	29	25	12	30	39	36
South America (13)	104	71	37	22	21	11	21	29	31
Europe, North America, Australia, New Zealand (43)	33	19	9	14	8	5	42	41	52
Total, world (179)	119	88	60	49	45	32	41	51	54

The figures in parentheses represent the number of countries in each region. The averages and standard deviations are weighted.
Data from UN, 2001.

Even at that macroregional level, there is great diversification. The classification in three major groups of regions prevalent in 1955 (Fig. 71–8) has been changed into a four-group classification: Europe and North America (infant mortality rate, 9‰); Latin America and the Far East (38‰); south Asia and the Near East, including North Africa (about 60‰); and sub-Saharan Africa, which is lagging behind (97‰).

What exactly is happening within each region? Figure 71–9, in which the infant mortality rates in 1960–1965 and 1995–2000 for 179 countries are compared, illustrates the diversity of the paces of change at national level. We have drawn the diagonal (no changes) and the straight lines corresponding to relative declines of 25%, 50%, and 75% compared with the 1960–1965 rates.[12] Although the quality of some data must be considered, a few major trends and groups of countries can be distinguished.

The countries in which virtually no progress was achieved—Some 10 countries (situated close to the diagonal line or with a decline of mortality of less than 25%) have mortality rates that are not much different in 1998 from what they were in the 1960s, sometimes after certain periods of decline. These are mostly very poor countries (Sierra Leone, Uganda, Zambia, Mali) or countries that have been through serious armed conflicts (Rwanda, Burundi, Liberia or Iraq, Cambodia, and Afghanistan).

The countries with comparatively slow progress (25 to 50% decline)—There are many of these countries, and they are mostly in sub-Saharan Africa, where mortality rates are highest; in central south Asia, although at lower levels; and in Paraguay and Haiti.

The countries with appreciable progress—These countries form a more heterogeneous group, including north Africa; many countries in Latin America, Iran, and Turkey; and a few African countries such as Senegal, Madagascar, and Namibia.

The countries with exceptional progress—These countries have a far higher than average in the world or in the region to which they belong (more than 75% decline). These are the United Arab Emirates, Reunion island in Africa, and Macedonia in Europe.

All considered, on the whole those countries in which mortality was highest 30 years ago (mostly in sub-Saharan Africa) have also progressed the least. Conversely, those countries in the south that were at the best starting points have made the most remarkable progress in their fight against infant mortality.

Nevertheless situations are diversified within the regions. When the graphic dispersions (Fig. 71–10) or the standard deviations and coefficients of variation by region (Table 71–3) are considered, it seems no reduction of the inequalities between countries within the regions is observed. From 1963 to 1998, all the standard deviations have narrowed down,[13] but most of the

[12] The relative decline of infant mortality in the world was 50% from 1963 to 1998 (table 3).

[13] But at various paces, North Africa's has even slightly increased.

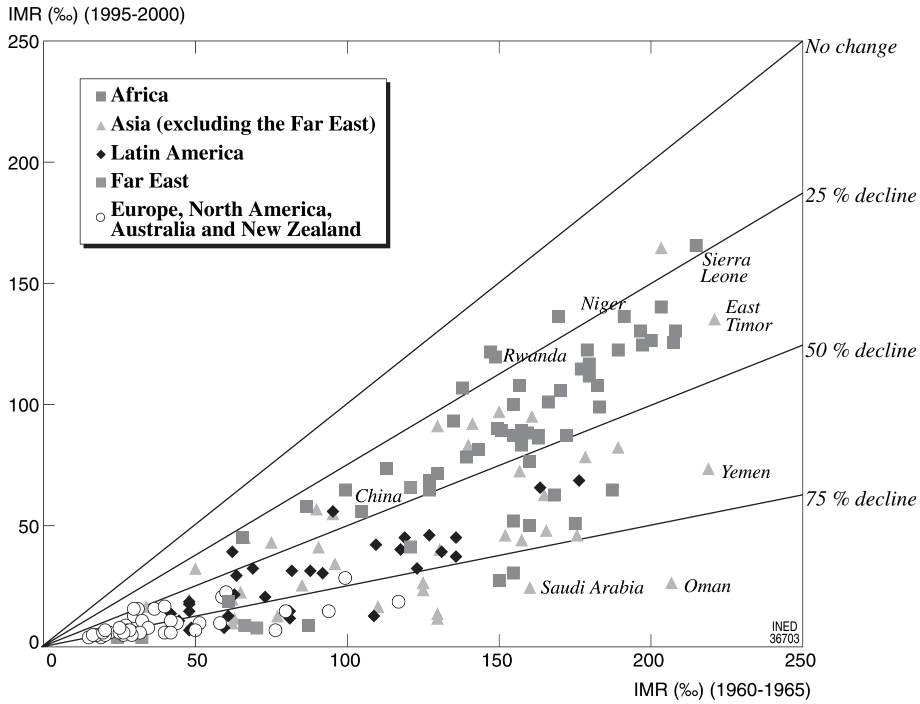


FIGURE 71–9 The evolution of infant mortality between 1960–1965 and 1995–2000 in 179 countries by pace of decline. (Data from United Nations, 2001a,b.)

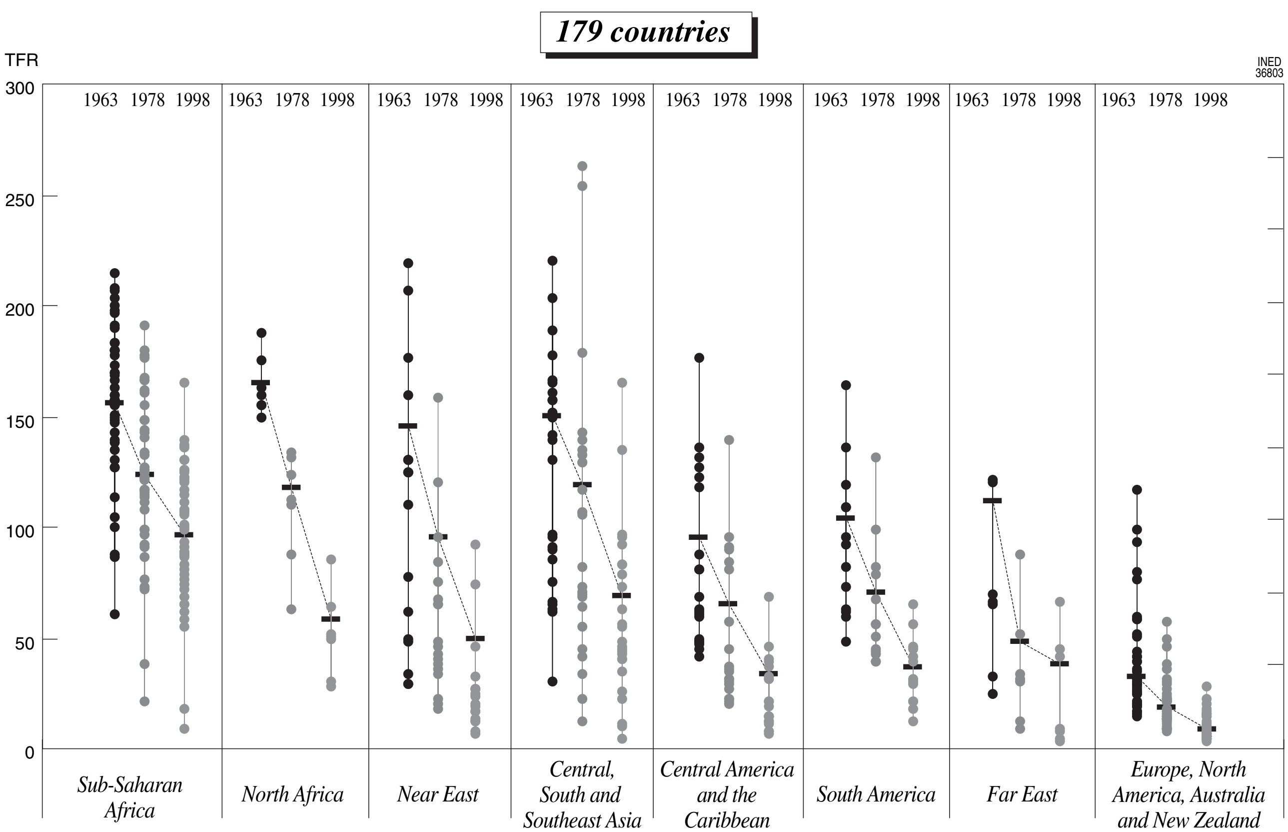


FIGURE 71–10 Dispersion between countries and average infant mortality rates in 1960–1965, 1975–1980, and 1995–2000 by major region. The dots represent countries, and the dotted lines link the weighted regional averages between them. (Data from United Nations, 2001a,b.)

coefficients of variation have clearly increased. They are particularly high in the Near East, Central America, and Europe/North America.

When the persistently high mortality rates, the widening inequalities, and the serious public health problems in many countries in the south are considered, the "health for all in 2000," as promoted by the World Health Organization, is far from a reality. Now AIDS also has a major effect on the evolution of mortality in Africa, which will be discussed later in this chapter.

V. FROM NATIONAL AVERAGES TO SOCIAL OR REGIONAL INEQUALITIES IN THE COUNTRIES OF THE SOUTH

Behind national averages usually lie regional or social inequalities in areas such as fertility and its intermediate variables (Chapters 34, 37 of Volume I), marriage (Chapter 28 of Volume I), or mortality (Chapters 53, 54, 55 of Volume II). There are—or were—differences in pretransition societies, in which demographic systems have never been homogeneous. However, just as on the international level, it is often during the demographic transition that internal disparities widened. The decline of mortality and fertility do not affect all social groups at the same period or pace. There are always groups that are ahead of others, innovative groups, and groups that are kept away from progress or are more averse to change (Tabutin, 1995). It seems that after a phase of rather high homogeneity, a phase of growing heterogeneity usually comes, which is followed by a phase of convergence. Such diffusion, such gaps between social and cultural groups or between living environments, are fairly well documented for Europe but are less so in the countries in the south. Cross-sectional analyses are still often used, in which countries are compared at various stages of transition, although the diversity of political, cultural, economic, and health contexts in those various countries is likely to cause confounding.

However, what do such analyses show? Fertility is now lower in urban than in rural regions and is lower among the more educated groups than among the least educated groups. However, differences, both absolute and relative, vary from one country to another. On the whole, the differences according to the living environment corroborate a dynamic pattern in which a growing heterogeneity, in a first transition phase, is followed by a certain convergence. According to fertility surveys conducted in the 1970s, 1980s, and 1990s (UN, 1987; Muhuri and Rutstein, 1994), the smallest relative gaps are to be observed in high and low fertility countries, whereas the largest gaps are in intermediate fertility countries. Such inequality is also greater in Latin America than it is in Africa or Asia.[14] Differences caused by educational level, which usually creates the largest disparities, are also greatest in intermediate fertility countries; again, the gaps are wider in Latin America than anywhere else (UN, 1995).

Child *mortality* is also lower in urban areas, and the gaps vary from one country to another. In a few countries (e.g., Trinidad and Tobago or the Dominican Republic), mortality levels are much the same in urban and in rural areas, but in most cases, there is a great difference. Relative differences, however, do not depend much on the mortality level. The relation between infant-child mortality and the mothers' educational level is stronger: everywhere, the mortality of children whose mothers have been through secondary education is much lower than that among children whose mothers are uneducated.[15] Absolute differences usually decrease along with the mortality level, whereas relative inequalities either increase or remain unchanged. Again, inequalities are much greater in Latin America than elsewhere.

Table 71–4 illustrates and confirms some of the relations described above for countries at various stages of transition. It also shows the inequalities that may exist for two important intermediate variables of fertility, that is, age at first marriage and prevalence of modern contraception.

In Mexico, for example, the absolute gaps in contraceptive prevalence rates are 30 points between place of residence and about 45 between educational levels, whereas in Indonesia, Sri Lanka, or Trinidad and Tobago, although the prevalence at national level is also about 40%, the gaps between place of residence are very small, a few points only. The gaps between educational levels are larger. It is also between educational levels that differences between median ages at first marriage are greatest—ranging from 2.3 years in Tunisia to 8 years in Senegal.

Even in high fertility countries (where apparently no change is occurring), there may be significant differences between social classes. For example, in Burkina Faso (Table 71–5) the fertility rate is 6.9 children per woman at national level. Three major reproductive models can currently be distinguished in this country (Tabutin, 1997): (1) that of executives, who form a small minority and are ahead of other groups

[14] Except for Trinidad and Tobago, where the urban-rural gaps are some of the lowest in the world.

[15] The ratios are very often one to three, or even four. For a comprehensive view, see Sullivan *et al.* (1994). For Africa in the 1970s and 1980s, see Akoto and Tabutin (1989).

TABLE 71–4 Inequalities in about 1990 in Eight Countries in the Areas of Marriage, Fertility, and Child Mortality

Indicators	Countries	Place of residence			Educational level			National average
		urban (1)	rural (2)	≠ % (1)–(2)/(2)	illiterate (3)	secondary education and over (4)	≠ % (4)–(3)/(3)	
Median age at first marriage (women aged 30 to 34 years)	Senegal	18.3	15.7	+17	16.0	24.0	+50	16.2
	Bolivia	20.5	20.1	+2	19.6	21.8	+11	20.3
	Ecuador	20.7	19.5	+6	18.2	24.1	+32	20.2
	Tunisia	22.0	20.5	+7	20.6	22.9	+12	21.3
	Mexico	20.4	18.1	+13	16.7	22.0	+32	19.4
	Indonesia	19.1	17.0	+11	15.9	21.8	+37	17.6
	Trinidad	19.7	19.7	0	18.9	22.5	+19	19.7
	Sri Lanka	25.4	23.1	+10	—	—	—	24.4
Total fertility rate	Senegal	5.1	6.7	–24	6.5	3.7	–43	6.0
	Bolivia	3.8	6.3	–40	6.5	2.7	–58	4.8
	Ecuador	3.5	5.5	–37	6.4	2.3	–64	4.4
	Tunisia	3.4	5.7	–40	5.1	2.4	–53	4.3
	Mexico	3.1	5.9	–47	6.1	2.5	–59	3.8
	Indonesia	2.8	3.6	–22	3.7	2.4	–35	3.3
	Trinidad	3.0	3.1	–3	3.6	2.3	–36	3.1
	Sri Lanka	2.1	3.3	–36	—	—	—	2.7
Rate of contraceptive prevalence (modern methods)	Senegal	11.8	1.4	(a)	2.2	29.3	(a)	4.8
	Bolivia	25.3	6.9	+266	(a)	35.9	—	17.8
	Ecuador	43.2	26.5	+63	15.0	45.8	+205	36.0
	Tunisia	47.6	30.1	+37	37.4	44.5	+19	40.4
	Mexico	63.3	32.5	+95	23.7	69.9	+195	44.8
	Indonesia	48.1	42.3	+12	31.3	56.0	+79	44.0
	Trinidad	45.7	43.4	z+5	36.9	55.1	+49	44.4
	Sri Lanka	42.9	40.3	+6	43.9	37.0	–16	40.5
Child Mortality Probability of dying between 0 and 5 years of age (‰)	Senegal	102	184	–45	171	52	–70	157
	Bolivia	104	162	–36	187	49	–74	132
	Ecuador	65	111	–41	160	43	–73	90
	Tunisia	62	88	–30	86	39	–55	74
	Mexico	37	104	–64	114	29	–75	71
	Indonesia	78	124	–37	142	43	–70	111
	Trinidad	(a)	(a)	(a)	(a)	(a)	(a)	34
	Sri Lanka	39	73	–47	71	26	–64	42

Countries are classified by decreasing level of fertility. —, data not available.
[a] Number of cases too low.
Data from the Demographic and Health Surveys national reports. These surveys were conducted between 1987 (Ecuador, Trinidad, and Tobago) and 1994 (Bolivia).

in every respect, with a fertility of 3.1, a postpartum abstinence period extending over 7 months, and a contraceptive prevalence of 33%; (2) that of clerks/traders/industrial workers, who are in the early stages of transition, with contraceptive prevalence ranging from 11% to 17% and a slightly lower fertility; and (3) that of farmers, who have kept their traditional behaviors virtually unchanged (7.6 children, 27-month breastfeeding, a long period of abstinence, and very low contraception prevalence).

Now what is the situation when the evolution of the mortality and fertility inequalities is examined in the countries for which adequate chronological data are available? In Honduras, Paraguay, or Indonesia, from the 1960s to the 1980s, fertility first declined in the more privileged regions or in the more educated urban families, causing disparities to widen. Between 1970 and 1980 in Cuba, Chile, or Thailand, where fertility was already very low, the fertility decline affected all classes of the population before a process of convergence started. In contrast, John Cleland et al. (1992) showed that, over the 1975–1985 period, child mortality declined in the 12 countries in the south they studied (and in all social classes), without giving any

TABLE 71–5 Indicators and Intermediate Fertility Variables by Social Groups and (Extreme) Educational Levels of the Spouses in Burkina Faso (1993)

Social groups	TFR (1)	No. of children at age 45	Desired no. of children (2)	Breast-feeding period (3)	Period of abstinence	Median age at first marriage (4)	% women in polygamous situations (5)	Current modern contraception (6)
Executives (5)	3.1	5.4	3.6	20	7	20.4	14	33
Clerks (6)	5.9	7.5	4.9	24	16	17.3	45	12
Traders (11)	6.3	6.8	4.9	24	15	17.8	58	13
Artisans (7)	6.8	8.2	4.9	24	14	17.8	35	11
Industrial/unqualified workers (5)	5.9	7.8	4.2	22	12	19.0	25	17
Well-off farmers (15)	7.7	7.6	6.1	26	17	17.2	67	3
Poor farmers (46)	7.5	7.4	6.8	28	17	16.9	50	1
Illiterate parents (72)	7.4	7.4	6.1	27	17	17.1	61	3
Parents with secondary education (or more) (7)	2.9	4.8	3.4	19	6	22.0	8	34
Total (100)	7.1	7.4	5.7	24	16	17.5	53	7

The figures in parentheses represent the weight (in %) of each group in the survey.

(1) Ever-married women; (2) ever-married women aged 30 to 39 years; (3) average length of breastfeeding and abstinence periods (ever-married women aged 15 to 49 years), expressed in months; (4) ever-married women aged 30 to 39 years; (5) at age 30 to 39 years; (6) proportion (%) of women who were using a modern contraceptive method at the time of the survey.

Indicators calculated from the 1993 Demographic and Health Survey by Youssouf Langani (Institut de Démographie, Louvain-la-Neuve).

obvious sign of converging. The gaps between the place of residence or the mothers' educational levels even seemed to have widened.

Social, regional, or even cultural disparities are then great, virtually universal, and persistent. Though convergence, or the narrowing of inequalities, has started in many countries in the south where fertility is concerned, it is not a fact yet where mortality and health are concerned. On the contrary, inequality may even worsen concerning the latter.

But before the impact of AIDS and future uncertainties are addressed, we should deal briefly with the structures and the aging of populations, the consequences of the transitions in the past few decades.

VI. AGE STRUCTURES AND DEMOGRAPHIC AGING

The age structures of populations vary considerably from one country or one region to another. At one extreme, in 2000, the population of Sweden, one of the oldest in the world, had virtually as many old people (those aged 65 years and older were 17.4% of the population) as young people (18.2% were aged under 15 years). At the opposite extreme, in Senegal there were almost 15 times as many young (44.5%) as there were old (2.5%). Between the two extremes are countries whose population pyramids have a more or less large base: in Mexico, for example, young people are about 33%, or in China, where people aged under 15 years were only about one-quarter of the total population in 2000 (24.9%). On the scale of the three major groups of countries, the group of the poorest has a very young population: 50% of the population were aged under 18 in 2000, and about 3% were 65 or older. In the developed countries, the structure is clearly older, with a median age of 37, and young people form less than 20% of the population; people aged 65 years and older, 14%. Last, in the intermediate countries, young people are about 31% of the population; people aged 65 and older, 5%.

Those differences in the structures are linked to the evolution of fertility and to that of mortality and sometimes migration (like in such cases as Israel or Burkina Faso). Examining the pyramids (Fig. 71–11), it appears that the structures have remained unchanged in the poorest countries in the past 30 years. For example, in Senegal, where mortality has declined between 1965 and 2000 (life expectancy has climbed from 39 up to 52 years), no aging of the population is observed. In contrast, the population even got slightly younger, owing to the decline of child mortality. The main cause of population aging is the decline of fertility, at least in a first phase, in which fertility and mortality are still comparatively high (Grinblat, 1986). This accounts for the stability of the population pyramids in the poorest countries, most of which have not expe-

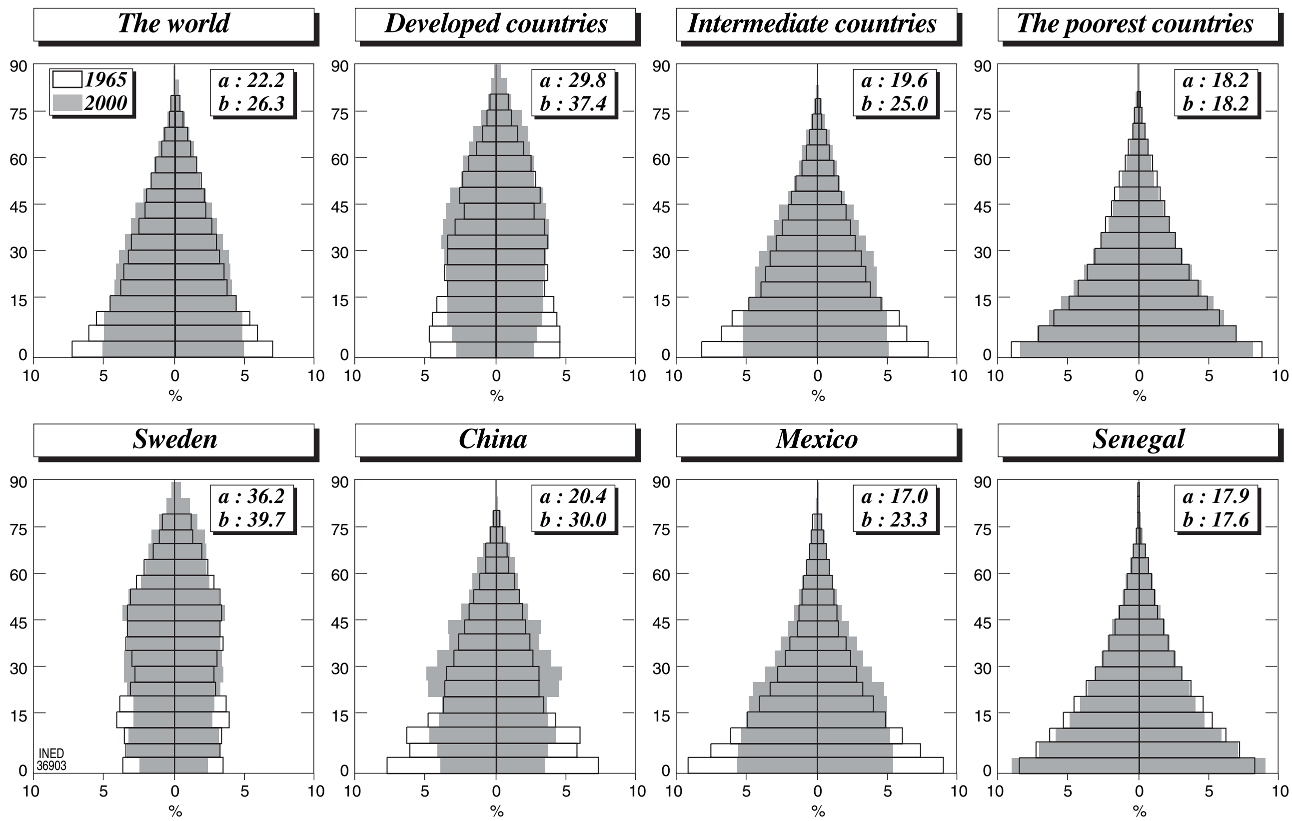


FIGURE 71–11 Relative structure by sex and age of the populations of the three major groups of countries, and in Sweden, China, Mexico, and Senegal in 1965 and 2000. **(a)** Median age in 1965. **(b)** Median age in 2000. (Data from United Nations, 2001a,b.)

rienced any significant decline of their fertility. As for the intermediate countries, they have started the aging process: between 1965 and 2000, the base of the pyramid in the countries taken as a whole has considerably narrowed (41% people aged under 15 in 1965 versus 31% in 2000), owing to the decline of fertility. Although the proportion of people over 65 has not increased in any appreciable way, it is likely to reach 6% in 2010 and 9% in 2025, which means more than 500 million people aged over 65. The developed countries have been experiencing an increase of the proportion of the elderly for a much longer period, starting in 1950 (Fig. 71–12).[16] As for the proportion of youth, it remained the same and then slightly increased in the 1950s before it went on a decline, starting in the 1960s.

Aging, although it has widely affected the developed countries and, to a lesser extent, the intermediate ones, has not really affected the poorest countries. However, one important fact should be kept in sight despite this: the population of those countries will be aging in the coming century. In a first phase, the proportions of young people will go down along with the decline of fertility, as is currently happening in the intermediate countries (Fig. 71–1). The proportion of the elderly will increase only later, as a consequence of the decline of the proportion of youth and through a decline of mortality at older ages, as is now the case in developed countries. However, beyond proportions, it is the total population of people aged over 65 that is increasing already and will keep on increasing in an impressive way. Aging is a problem not only of structures but also of absolute numbers (Tabutin, 1986). Only 20 million people aged over 65 currently live in the 48 poorest countries. There will be 27 million of them in 2010 and some 46 million in 2025. At that date, there will also be 520 million of them in the intermediate countries and 260 million in the developed countries. There are currently about as many elderly people in the rest of the world as in the developed countries. There will be twice as many of them in 30 years. The countries in the south will all be faced with new social and health constraints. And the aging process will be all the more dramatic as fertility has declined faster. In this respect the case of China is exemplary (Fig. 71–11).

[16] Except for a stagnation in the early 1980s.

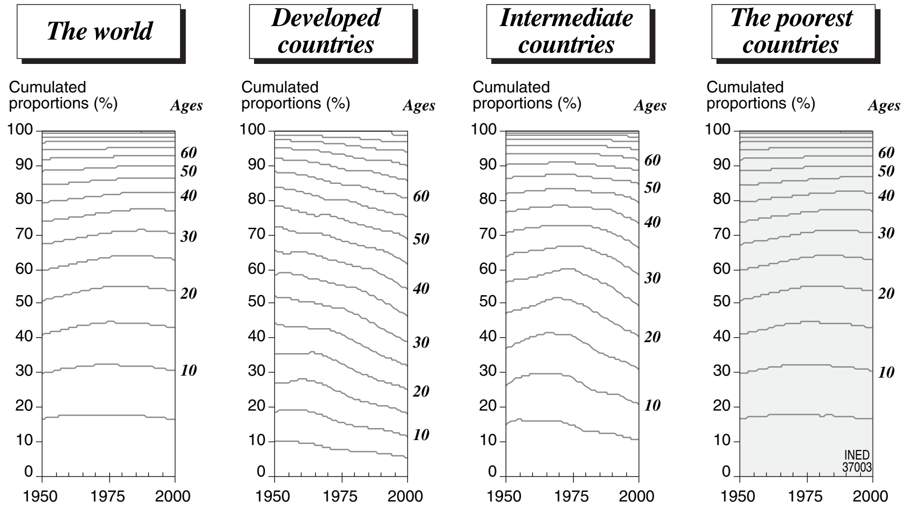


FIGURE 71–12 The evolution of the cumulated proportions (percentage) of the population by age from 1950–2000 in the world and in the three major groups of countries. (Data from United Nations, 2001a,b.)

We should probably worry not so much about any loss of dynamism in society or any aging of mentalities but about the transformations made necessary by such a revolution of age structures, as well as about the lot of those masses of elderly people (Tabutin, 1986). Social security systems are still budding or are nonexistent in many countries, and family solidarity, which was traditionally important, is changing.

VII. AIDS AND ITS DEMOGRAPHIC CONSEQUENCES

As we have seen above, mortality has noticeably declined in all regions of the world since the 1960s. Even in sub-Saharan Africa, where progress has been slower than elsewhere, infant mortality has dropped from 156% in the early 1960s to 100% in the late 1990s, and in the same period, life expectancy has improved from 40 to 48 years. However, since the early 1980s, the evolution of life expectancy has noticeably slowed down or stagnated. It was estimated as 48.3 years in the 1980–1985 period and was still 48.4 years in the late 1990s. This halt in the improvement of life expectancy reflects the tragedy of the AIDS epidemic that many African countries are experiencing.

Sub-Saharan Africa is the region most seriously affected by the epidemic: 70% of the 42 million HIV-infected individuals in the world now live in sub-Saharan Africa (UNAIDS/WHO, 2002), although only about 10% of the world's population live in the region. The prevalence of HIV for the whole of sub-Saharan Africa was estimated at 9% at the end of 2001, far ahead of the Caribbean, the second most affected region, with a little more than 2% of its adult population infected by HIV (UNAIDS, 2002).[17] Last, the 20 countries most seriously affected in the world are all in sub-Saharan Africa, with a prevalence of more than 20% among the adult population in seven countries of southern and east Africa (Botswana, Zimbabwe, Swaziland, Lesotho, Namibia, Zambia, South Africa).

What are and will be the demographic consequences of the epidemic? Although there are still many uncertainties, clearly the AIDS epidemic has and will continue to have devastating effects on mortality in years to come in many countries, particularly in sub-Saharan Africa, where AIDS is now the first cause of mortality (Murray and Lopez, 1996), but also in the Caribbean (e.g., Haiti, Guyana) and a few countries in Asia (e.g., Cambodia, Thailand).[18] In those countries that are most seriously affected, life expectancy has already deteriorated since the early 1980s, falling below its 1960s level in Botswana and Zimbabwe (Table 71–6). Child mortality also seems to be increasing in several African countries owing to the

[17] In all the other regions prevalence is less than 1%, and the prevalence among the 15- to 49-year-olds at world level was estimated as 1.2% in late 2002 (UNAIDS, 2002).

[18] The consequences of AIDS on fertility are not as well known, but it seems that the fertility of HIV-positive women is about 25% to 40% lower than that of noninfected women; such differences might largely be accounted for by the effects of other sexually transmitted diseases associated with HIV infection (United Nations, 2002). The indirect influences of the AIDS epidemics on fertility, particularly through the change in the behaviors of noninfected individuals, are potentially important but have not been extensively studied (United Nations, 2002).

TABLE 71–6 Populations, Life Expectancies, and HIV Prevalence in the 10 Countries Most Affected by AIDS in Late 2001

The 10 countries most affected by AIDS	HIV prevalence (%) in late 2001 (among 15–49 year old)	Life expectancy						Population (in thousand)			
		Past estimates		Estimates with AIDS		Estimates without AIDS		Estimates with AIDS		Estimates without AIDS	
		1963	1983	1998	2008	1998	2008	2000	2010	2000	2010
Botswana	38.8	47.5	59.6	44.4	38.2	67.6	71.6	1,541	1,628	1,654	2,129
Zimbabwe	33.7	51.7	59.4	42.9	46.3	66.5	70.1	12,627	15,028	14,083	18,659
Swaziland	33.4	41.6	52.6	50.8	35.3	60.2	65.0	925	987	950	1,211
Lesotho	31.0	44.5	53.7	51.2	36.3	61.4	66.1	2,035	2,127	2,097	2,610
Namibia	22.5	43.7	54.5	45.2	49.0	62.1	66.7	1,757	2,097	1,869	2,454
Zambia	21.5	42.8	51.4	40.5	47.3	57.6	61.6	10,421	12,989	11,369	15,283
South Africa	20.1	50.0	56.5	56.7	42.4	63.3	67.8	43,309	45,140	44,091	52,451
Kenya	15.0	46.0	55.8	52.2	49.5	63.6	67.9	30,669	36,941	31,863	41,469
Malawi	15.0	38.4	45.0	40.7	40.5	51.2	55.3	11,308	14,024	11,914	16,042
Mozambique	13.0	38.5	43.6	40.6	38.4	47.0	51.0	18,292	21,649	18,942	24,358

Sources for HIV prevalence: UNAIDS (2002).
Source for data on population and life expectancy: United Nations (2001).

AIDS epidemic (U.S. Bureau of the Census, 1999). Consequently, the countries most affected, which used to be some of the African countries with lowest death rates, now have some of the lowest life expectancies. According to the UN Population Division estimates (UN, 2001a,b), life expectancy is lower by more than 15 years than it would have been if not for AIDS in several countries, and the gap might be as much as 30 years in the next few years (Table 71–6). As a result, crude death rates are already three times as high as those that would have been observed without AIDS in those countries most affected,[19] and the gaps are likely to widen in years to come.

The impact of AIDS on mortality is already reflected in population growth clearly slowing down in several African countries. The natural growth rate in Botswana between 2000 and 2010 would have been about 2.5% without AIDS; it is now probably about 0.5%. The populations of Botswana and Zimbabwe in 2000 are already fewer by 7% and 10%, respectively, than the populations that would have been observed without AIDS, and these differences are likely to reach almost 20%, if not more, in the next few years. Population growth is thus perceptibly slowing down; however, it is likely to remain positive in virtually all countries. According to UN projections, only South Africa and Guyana (two countries with comparatively low birth rates) may have negative growth rates (UN, 2001a,b).[20]

On the other hand, AIDS affects a majority of young adults and, to a lesser extent, very young children, which causes deep changes in the age distribution of deaths among the populations seriously affected by the epidemic (Murray and Lopez, 1996). This has a particular impact on the age structures of those populations (Stanecki, 2000). By disproportionately affecting young adults, the epidemic also has major economic consequences in the countries most affected. One more consequence is the significant increase of the number of orphans. The proportion of children under 15 years of age who have lost their father and/or mother was estimated 12% in sub-Saharan Africa in 2001, whereas it would have been about 8% without AIDS (USAID/UNICEF/UNAIDS, 2002). At that date in Zimbabwe orphans were about 18%, three-quarters of which were owing to AIDS. Among the main effects of these trends is that children are more vulnerable, and the elderly have a greater part to play in taking care of them.

The demographic, health, social, and economic consequences of the AIDS epidemic in Africa and some countries in the Caribbean and Asia thus appear to be considerable. At the world scale, however, the effect of AIDS over mortality and demographic growth is com-

[19] In the 1995–2000 period, the crude death rate in Botswana was estimated as 17.0%, whereas without AIDS, it would have been about 6%. These indicators are 18% (with AIDS) and 6.4% (without AIDS) in Zimbabwe, and 10.8% and 7.6%, respectively, in South Africa.

[20] Of course, those results depend on the hypotheses on the evolution of the epidemics in the countries involved. Stanecki (2001), for example, shows that, according to the projections of the U.S. Bureau of the Census, Botswana and Zimbabwe are also likely to have negative population growth.

paratively moderate in relative terms. Nevertheless, an estimated 20 million individuals have died of AIDS since the epidemic started, about 3 million of whom died in 2002 alone (UNAIDS/WHO, 2002), which represents more than 5% of deaths in the world during that year.

CONCLUSION

Everywhere in the world, mortality has been declining in the past 50 years. Even in the group of the poorest countries, life expectancy went up from 40 to 50 years between 1960–1965 and 1995–2000, and infant mortality went down from 171% to 102%. But the starting levels in the 1950s, as well as the decline paces, have been variable, and inequality between regions and countries is currently as great as or greater than it was. Mortality remains high and health precarious in many countries, particularly in sub-Saharan Africa, where "an increasing number of countries appear to be kept out of the phase of fast progress which Third World countries experienced before them on the way to epidemiological transition" (Vallin and Meslé, 1995: 18). "Health for all in 2000" unfortunately remained a slogan in many poor regions. Complete control of infectious diseases is far from being achieved, and the dissemination of AIDS contributes to stagnation of life expectancy in sub-Saharan Africa. The epidemic has devastating consequences in some countries, where life expectancy is noticeably decreasing. The economic crisis and its consequences on health care or education systems are also likely to slow down the decline of mortality in the poorest countries.

At diverse periods, most of the regions in the world have also entered a phase of transition to low fertility. The latter was over 5.5 children per woman in the 1950s in 6 of the 8 major regions. That level currently prevails only in sub-Saharan Africa, where the decline began only very recently. There has been—and will be for time to come—diversity in the starting periods and speeds of the decline , between regions and countries, and between social groups within each country. Unlike mortality, the decline of fertility seems hardly reversible; countries, once they have launched in the transition process, continue and will probably continue to experience declining fertility. Some countries, such as China, Thailand, or Sri Lanka, have been through such decline and have currently reached levels close to those of developed countries. Many other countries have started on the same path.

Although it is comparatively clear that all countries will have their fertility rates go down, it remains unknown when the decline will begin and at what pace in those countries that have not yet entered the second phase of demographic transition (some 50 countries in the world, about one-sixth of the world population). The improvement of the populations' living standards, which was long considered a precondition for fertility decline, is now superseded by concepts of crisis-led transition and poverty Malthusianism. It seems that deterioration of living standards would not be incompatible with fertility decline. On the contrary, it tends to give it momentum. However, it is not clear whether the economic, social, and health difficulties in some countries, particularly in sub-Saharan Africa, can be integrated into such patterns in the short term and contribute to the decline of fertility. By limiting girls' education and slowing down the decline of mortality or access to health, they may possibly contribute to delay or slow down the decline rather than accelerate it.

At the world scale, however, sub-Saharan Africa currently represents only a comparatively small part of the world population (some 10% in 2000). Consequently, its transition lag will hardly affect the evolution of the world or of the undeveloped countries. The fertility transition is now largely under way, with an average of 2.9 children per woman in 2000 at world level. The world life expectancy has increased by 12 years in the past 4 decades, and infant mortality has gone down from 119‰ to 60‰.

Although, because of such data, we might think that the demographic transition at world level will soon be completed, Africa and some other countries and regions in the world remind us that huge differences and inequalities persist, that the process will be a long one in some places, and that many things remain uncertain for the future. Among these are the direct and indirect and more or less immediate effects of economic recession, AIDS, civil wars and local conflicts, and resurgent or emergent religious or political fundamentalism.

Acknowledgment

This chapter was translated from the French version by Mireille Rabenoro.

References

Akoto Eliwo and Tabutin Dominique, 1989. Les inégalités socio-économiques et culturelles devant la mort, *in:* Gilles Pison, étienne van de Walle, and Mpembélé Sala Diakanda (eds.), *Mortalité et société en Afrique*, p. 35–63. Paris, INED/PUF, 446 p.

Akoto Eliwo and Tabutin Dominique, 1992. Socio-economic and cultural differentials in the mortality of sub-Saharan Africa, *in:* étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 32–64. Oxford, Clarendon Press, 450 p.

Brouard Nicolas, 1994. Aspects démographiques et conséquences de l'épidémie de sida en Afrique, *in:* Jacques Vallin (ed.), *Populations africaines et sida*, p. 119–178. Paris, La Découverte, CEPED, 223 p.

Chesnais Jean-Claude, 1986. *La transition démographique, étapes, formes, implications économiques.* Paris, INED, PUF, 580 p.

Chesnais Jean-Claude, 1995. *La transition démographique, trente ans de bouleversements (1965–1995).* Paris, CEPED, 25 p.

Cleland John, Bicego George, and Fegan Greg, 1992. Socioeconomic inequalities in childhood mortality: the 1970s to the 1980s, *Health Transition Review*, vol. 2(1), p. 1–18.

Douglas Robert, Jones Gavin, and D'Souza Rennie (ed.), 1996. *The shaping of fertility and mortality declines: the contemporary demographic transition*. Canberra, ANU, 396 p.

Festy Patrick, 1979. *La fécondité des pays occidentaux de 1870–1970*. Paris, INED, PUF, 322 p.

Grinblat Joseph-Alfred, 1990. Le vieillissement des populations mondiales: tendances démographiques récentes et futures, *in:* Michel Loriaux, Dominique Remy, and éric Vilquin (eds.), *Populations âgées et révolution grise*, p. 53–76. Louvain-la-Neuve, Ciaco, 1118 p.

Joint United Nations Programme on HIV/AIDS [UNAIDS], 2002. *Report on the global HIV/AIDS epidemic.* Geneva, UNAIDS, 231 p.

Joint United Nations Programme on HIV/AIDS [UNAIDS]/World Health Organization [WHO], 2002. *Le point sur l'épidémie de SIDA.* Geneva, UNAIDS/WHO, 42 p.

Leridon Henri and Toulemon Laurent, 1996. *La régulation des naissances se généralise.* Paris, CEPED, 19 p.

Meslé France and Vallin Jacques, 1995. *La mortalité dans le monde: tendances et perspectives.* Paris, CEPED, 25 p.

Meslé France and Vallin Jacques, 1996. *Mortality in the world: trends and prospects.* Paris, CEPED, 24 p.

Muhuri Pradip and Rutstein Shea, 1994. *Socioeconomic, demographic and health indicators for subnational areas.* Columbia (MD), Macro International, 213 p.

Murray Christopher and Lopez Alan (eds.), 1996. *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge (MS), Harvard University Press, 990 p.

Société de Démographie Historique, 1989. *Le déclin de la mortalité.* Paris, Annales de Démographie Historique, 493 p.

Stanecki Karen, 2000. *The AIDS pandemic in the 21st century: the demographic impact in developing countries* (Paper presented at the 13th International AIDS Conference, Durban, South Africa, July 9–14, 2000).

Sullivan Jeremiah, Rutstein Shea, and Bicego George, 1994. *Infant and child mortality.* Columbia (MD), Macro International, 58 p.

Tabutin Dominique, 1990. L'âge vermeil du tiers monde: perspectives des populations âgées dans les pays jeunes, *in:* Michel Loriaux, Dominique Remy, and Eric Vilquin (eds.), *Populations âgées et révolution grise*, p. 1087–1103. Louvain-la-Neuve, CIACO, 1118 p.

Tabutin Dominique, 1995. Un demi-siècle de transitions démographiques dans le Sud, *in:* Dominique Tabutin, Thierry Eggerickx, and Catherine Gourbin (eds.), *Transitions démographiques et sociétés*, p. 33–70. Louvain-la-Neuve, Academia, Paris, L'Harmattan, 692 p.

Tabutin Dominique, 1997. Les transitions démographiques en Afrique sub-saharienne: spécificités, changements et incertitudes, *in: International Population Conference, Beijing, 1997*, vol. 1, p. 219–247. Liège, IUSSP, 980 p.

Tabutin Dominique, Eggerickx Thierry, and Gourbin Catherine (ed.), 1995. *Transitions démographiques et sociétés.* Louvain-la-Neuve, Academia, Paris, L'Harmattan, 692 p.

United Nations [UN], 1987. *Fertility behaviour in the context of development.* New York, United Nations, Population Division, 383 p.

United Nations [UN], 1995a. *Women's education and fertility behaviour.* New York, United Nations, Population Division, 113 p.

United Nations [UN], 1995b. *World population prospects, the 1994 revision.* New York, United Nations, Population Division, 886 p.

United Nations [UN], 1996a. *World population prospects, the 1996 revision,*: Annex I: *Demographic indicators.* New York, United Nations, 253 p. (provisional documents).

United Nations [UN], 1996b. *World population prospects, the 1996 revision*, Annexes II and III: *Demographic indicators by major area, region and country.* New York, United Nations, Population Division, 441 p. (provisional documents).

United Nations [UN], 2001a. *World population prospects, the 2000 revision, Vol. I: comprehensive tables.* New York, United Nations, Population Division, 745 p.

United Nations [UN], 2001b. *World population prospects, the 2000 revision, Vol II: sex and age*. New York, United Nations, Population Division, 919 p.

United Nations [UN], 2002. *HIV/AIDS and fertility in sub-Saharan Africa: a review of the research literature.* New York, United Nations, Population Division, 10 p.

U.S. Agency for International Development [USAID]/United Nations' Childrens' Fund [UNICEF]/Joint United Nations Programme on HIV/AIDS [UNAIDS], 2002. *Children on the brink 2002. A joint report on orphan estimates and program strategies.* Washington DC, USAID, 36 p.

U.S. Bureau of the Census, 1999. *World Population Profile: 1998*, Washington DC, U.S. Government Printing Office, 172 p.

ANNEX TABLE 71–1 Numbers, Structures, Rates, and Demographic Indicators by Major Region and Subregion in 1965, 1980, and 2000

Region	Number of countries	Population (in thousand)			Relative importance (in %)			Mean annual growth rate (%)			Proportion of people under 15 years old (%)			Proportion of people 65 years old and older (%)		
		1965	1980	2000	1965	1980	2000	1965–1980	1980–2000	1965–2000	1965	1980	2000	1965	1980	2000
Africa	56	313,125	466,871	793,627	9.4	10.5	13.1	2.7	2.7	2.7	44.4	44.8	42.6	3.1	3.1	3.3
Asia	50	1,898,404	2,630,746	3,672,342	56.9	59.4	60.6	2.2	1.7	1.9	40.3	37.4	30.2	4.0	4.5	5.9
Europe	47	634,827	693,232	727,304	19.0	15.6	12.0	0.6	0.2	0.4	26.5	22.2	17.5	9.4	12.4	14.7
Latin America	46	250,351	361,328	518,809	7.5	8.2	8.6	2.4	1.8	2.1	42.9	39.6	31.5	4.1	4.5	5.4
North America	5	219,566	255,034	314,113	6.6	5.8	5.2	1.0	1.0	1.0	30.7	22.5	21.5	9.3	11.0	12.3
Oceania	24	17,442	22,536	30,521	0.5	0.5	0.5	1.7	1.5	1.6	33.0	29.2	25.4	7.2	8.1	9.9
East Africa	18	94,097	143,271	250,318	2.8	3.2	4.1	2.8	2.8	2.8	45.4	46.5	45.3	2.7	2.7	2.9
Central Africa	9	35,584	52,184	95,404	1.1	1.2	1.6	2.6	3.0	2.8	42.5	45.0	47.2	3.1	3.1	3.1
North Africa	7	75,885	110,828	174,150	2.3	2.5	2.9	2.5	2.3	2.4	44.6	42.5	35.6	3.6	3.7	4.1
Southern Africa	5	22,439	32,882	49,567	0.7	0.7	0.8	2.5	2.1	2.3	41.9	42.1	35.0	3.8	3.1	3.6
West Africa	17	85,121	127,706	224,189	2.6	2.9	3.7	2.7	2.8	2.8	44.6	45.6	44.8	3.0	2.9	3.0
Far East	7	874,284	1,177,958	1,481,075	26.2	26.6	24.5	2.0	1.1	1.5	38.6	34.2	23.9	4.6	5.1	7.7
Central and south Asia	14	696,559	981,616	1,480,868	20.9	22.2	24.5	2.3	2.1	2.2	41.2	39.6	35.2	3.6	3.9	4.6
Southeast Asia	11	252,107	357,873	522,121	7.6	8.1	8.6	2.3	1.9	2.1	43.0	40.7	32.4	3.3	3.7	4.7
Near East	18	75,453	113,298	188,277	2.3	2.6	3.1	2.7	2.5	2.6	42.5	40.8	35.9	4.2	4.3	4.7
East Europe	10	266,945	294,976	304,172	8.0	6.7	5.0	0.7	0.2	0.4	28.7	22.6	18.1	7.3	10.8	12.9
North Europe	13	84,928	89,892	95,076	2.5	2.0	1.6	0.4	0.3	0.3	23.7	21.3	18.9	11.6	14.6	15.5
South Europe	15	122,974	137,902	144,935	3.7	3.1	2.4	0.8	0.2	0.5	26.7	24.3	15.8	8.8	11.5	16.4
West Europe	9	159,980	170,462	183,121	4.8	3.8	3.0	0.4	0.4	0.4	24.2	20.2	17.0	12.1	14.6	16.0
Caribbean	24	22,664	29,247	37,941	0.7	0.7	0.6	1.7	1.3	1.5	40.4	36.6	29.5	4.8	6.0	6.9
Central America	8	57,667	89,833	135,129	1.7	2.0	2.2	3.0	2.0	2.4	46.2	45.1	34.9	4.1	3.6	4.5
South America	14	170,021	242,248	345,738	5.1	5.5	5.7	2.4	1.8	2.0	42.1	37.9	30.4	4.0	4.6	5.6
North America	5	219,566	255,034	314,113	6.6	5.8	5.2	1.0	1.0	1.0	30.7	22.5	21.5	9.3	11.0	12.3
Australia–New Zealand	2	14,021	17,682	22,916	0.4	0.4	0.4	1.5	1.3	1.4	30.2	25.5	20.9	8.4	9.7	12.2
Melanesia	5	2,845	4,074	6,482	0.1	0.1	0.1	2.4	2.3	2.4	44.1	43.0	39.3	2.3	2.4	2.7
Micronesia	7	219	308	516	0.0	0.0	0.0	2.3	2.6	2.4	44.4	40.1	38.7	2.0	2.8	4.5
Polynesia	10	357	472	606	0.0	0.0	0.0	1.9	1.3	1.5	48.1	40.9	35.0	2.0	2.5	4.4
Poorest countries	48	272,135	392,707	658,192	8.2	8.9	10.9	2.4	2.6	2.5	43.7	44.6	43.1	3.1	3.1	3.1
Intermediate countries	125	2,094,285	2,954,285	4,207,094	62.8	66.7	69.5	2.3	1.8	2.0	41.4	38.4	31.2	3.9	4.2	5.4
Developed countries	55	967,295	1,082,755	1,191,429	29.0	24.4	19.7	0.8	0.5	0.6	27.5	22.5	18.3	9.0	11.6	14.3
World	228	3,333,716	4,429,747	6,056,715	100.0	100.0	100.0	1.9	1.6	1.7	37.6	35.1	30.0	5.3	5.9	6.9

Data from UN, 2001.

(Continues)

ANNEX TABLE 71–1 (*Continued*)

Region	Crude birth rates (%)			Crude death rates (%)			Natural growth rates (%)			Total fertility rates (TFR)			Life expectancy at birth			Infant mortality rates (‰)		
	1960–1965	1975–1980	1995–2000	1960–1965	1975–1980	1995–2000	1960–1965	1975–1980	1995–2000	1960–1965	1975–1980	1995–2000	1960–1965	1975–1980	1995–2000	1960–1965	1975–1980	1995–2000
Africa	48.3	45.6	38.7	22.8	17.5	14.1	2.5	2.8	2.5	6.8	6.6	5.3	42.1	48.3	51.4	158	123	91
Asia	39.7	29.3	22.3	17.7	10.4	7.9	2.2	1.9	1.4	5.6	4.2	2.7	48.4	58.4	65.8	135	95	59
Europe	19.1	14.8	10.1	9.7	10.4	11.5	0.9	0.4	–0.1	2.6	2.0	1.4	69.6	71.5	73.2	37	22	10
Latin America	40.9	33.2	23.1	12.4	8.8	6.5	2.8	2.4	1.7	6.0	4.5	2.7	56.8	63.0	69.3	101	69	36
North America	22.4	15.1	14.2	9.3	8.5	8.4	1.3	0.7	0.6	3.3	1.8	2.0	70.2	73.4	76.7	25	14	7
Oceania	26.5	20.9	18.2	10.6	8.8	7.5	1.6	1.2	1.1	4.0	2.8	2.4	64.0	67.7	73.5	49	40	26
East Africa	49.8	48.1	43.0	23.3	18.5	17.5	2.7	3.0	2.5	7.0	7.0	6.1	41.0	46.7	45.7	155	125	103
Central Africa	46.5	47.2	46.0	24.5	19.2	16.2	2.2	2.8	3.0	6.0	6.5	6.4	39.6	46.0	48.9	163	126	98
North Africa	47.6	40.9	27.6	20.6	14.2	7.5	2.7	2.7	2.0	7.1	5.9	3.6	46.3	53.9	64.6	165	118	58
Southern Africa	42.1	36.1	27.8	17.1	12.1	11.5	2.5	2.4	1.6	6.5	5.1	3.3	49.2	54.7	55.4	94	77	63
West Africa	49.6	48.8	42.3	25.2	19.9	15.1	2.4	2.9	2.7	7.0	7.0	5.9	39.3	45.5	50.0	167	131	96
Far East	35.6	20.9	15.6	15.7	6.6	7.0	2.0	1.4	0.9	5.2	3.1	1.8	51.4	66.4	70.9	112	48	38
Central and south Asia	43.3	36.5	27.9	20.0	14.3	9.2	2.3	2.2	1.9	6.0	5.1	3.6	45.5	52.5	61.5	157	127	76
Southeast Asia	43.2	35.1	23.8	18.5	12.7	7.4	2.5	2.2	1.6	6.1	4.9	2.8	46.7	54.6	65.3	134	95	47
Near East	43.5	36.7	28.9	16.1	10.1	6.7	2.7	2.7	2.2	6.2	5.2	3.9	52.1	60.5	67.9	146	95	49
East Europe	19.4	16.6	9.2	8.5	10.4	13.4	1.1	0.6	–0.4	2.4	2.1	1.3	68.7	69.4	68.2	40	27	15
North Europe	18.0	12.9	11.7	11.1	11.4	10.8	0.7	0.2	0.1	2.7	1.8	1.7	71.2	73.0	76.7	22	13	6
South Europe	20.6	15.9	10.0	9.4	9.1	9.8	1.1	0.7	0.0	2.7	2.2	1.3	68.6	73.0	77.0	52	23	8
West Europe	18.1	11.9	10.7	11.3	11.1	10.0	0.7	0.1	0.1	2.7	1.6	1.5	70.8	73.1	77.7	26	13	5
Caribbean	38.8	27.4	20.8	12.1	8.7	8.2	2.7	1.9	1.3	5.5	3.6	2.5	58.5	64.5	67.5	94	72	38
Central America	45.3	38.1	26.5	13.0	8.3	5.4	3.2	3.0	2.1	6.8	5.4	3.0	56.4	63.5	71.0	96	63	33
South America	39.7	32.1	22.1	12.3	8.9	6.7	2.7	2.3	1.5	5.8	4.3	2.6	56.7	62.6	68.9	104	71	37
North America	22.4	15.1	14.2	9.3	8.5	8.4	1.3	0.7	0.6	3.3	1.8	2.0	70.2	73.4	76.7	25	14	7
Australia–New Zealand	22.6	16.2	13.7	8.7	7.8	7.2	1.4	0.8	0.7	3.4	2.1	1.8	70.9	73.3	78.4	20	13	6
Melanesia	42.3	39.5	32.9	19.8	13.6	9.3	2.2	2.6	2.4	6.2	5.5	4.4	43.0	50.6	58.7	120	88	58
Micronesia	42.0	34.8	30.9	11.1	6.9	5.4	3.1	2.8	2.6	6.4	4.2	4.3	57.3	64.3	71.8	91	56	21
Polynesia	46.0	33.5	24.6	12.3	8.0	5.3	3.4	2.5	1.9	7.0	4.6	3.2	53.5	60.4	70.3	86	58	19
The poorest countries	48.4	45.4	40.4	24.6	19.7	14.9	2.4	2.6	2.5	6.7	6.4	5.5	39.8	45.3	50.3	171	140	102
Intermediate countries	41.1	30.7	23.1	17.4	10.3	7.8	2.4	2.0	1.5	5.9	4.4	2.8	48.9	58.7	65.5	131	91	56
Developed countries	19.7	14.9	11.2	9.4	9.5	10.2	1.0	0.5	0.1	2.7	1.9	1.6	69.7	72.3	74.9	33	18	8
World	35.4	28.0	22.5	15.6	10.9	9.0	2.0	1.7	1.3	5.0	3.9	2.8	52.4	59.8	65.0	119	88	60

Data from UN, 2001.

ANNEX TABLE 71–2 The List of the 228 Countries and Territories Featured in the Database of the United Nations Population Division, Classified by Groups and Regions

The poorest countries (48 countries, 45 of which have a population of more than 140,000 people)

North Africa
Sudan

Sub-Saharan Africa
Angola
Benin
Burkina Faso
Burundi
Cape Verde Islands
Comoros
Djibouti
Eritrea
Ethiopia
Gambia
Guinea
Guinea Bissau
Equatorial Guinea
Lesotho
Liberia

Sub-Saharan Africa
Madagascar
Malawi
Mali
Mauritania
Mozambique
Niger
Central African Republic
Congo Democratic Rep.
Rwanda
Sao Tome and Principe*
Sierra Leone
Somalia
Chad
Togo
Uganda
Tanzania
Zambia

Central America and Caribbean
Haiti

Central, south, and southeast Asia
Afghanistan
Bangladesh
Cambodia
Laos
Maldives
Myanmar
Nepal

Near East
Yemen

Oceania
Solomon Islands
Kiribati*
Samoa
Tuvalu*
VanuatuBhutan

Intermediate countries (125 countries, 98 of which have a population of more than 140,000 people)

North Africa
Algeria
Egypt
Libya
Morocco
West Sahara
Tunisia

Sub-Saharan Africa
South Africa
Botswana
Cameroon
Congo
Côte d'Ivoire
Gabon
Ghana
Kenya
Mauritius
Namibia
Nigeria
Reunion
Seychelles*
St-Helen*
Swaziland
Senegal
Zimbabwe

Central America and Caribbean
Anguilla*
Antigua and Barbuda*
Netherlands Antilles
Aruba*
Bahamas
Barbados
Belize
Costa Rica
Cuba
Dominica*

Central America and Caribbean
Grenada*
Guadeloupe
Guatemala
Honduras
Cayman Islands*
Turks and Caïque Islands*
American Virgin Islands
British Virgin Islands
Jamaica
Martinique
Mexico
Montserrat*
Nicaragua
Panama
Porto Rico
Dominican Republic
St-Kitts and Nevis*
St-Lucia
St-Vincent and Grenadines*
Salvador
Trinidad and Tobago

South America
Argentina
Bolivia
Brazil
Chile
Colombia
Ecuador
Guyana
French Guyana
Falkland Islands*
Paraguay
Peru
Surinam
Uruguay
Venezuela

Central, south, and southeast Asia
Brunei
India
Indonesia
Iran
Kazakhstan
Kirghizstan
Malaysia
Myanmar
Uzbekistan
Pakistan
Philippines
Singapore
Sri Lanka
Tajikistan
Thailand
East Timor
Turkmenistan

Far East
China
North Korea
South Korea
Hong Kong
Macao
Mongolia

Near East
Saudi Arabia
Armenia
Azerbaijan
Bahrain
Cyprus
United Arab Emirates
Georgia
Israel
Iraq
Jordan
Kuwait
Lebanon
Oman
Qatar
Syria
Occupied Palestinian territories
Turkey

Oceania
Fiji
Guam
Cook*
North Mariana Islands*
Marshall*
Pitcairn*
Micronesia*
Nauru*
Niue*
New Caledonia
Palau*
Papua New Guinea
French Polynesia
American Samoa*
Tokelau*
Tonga*
Wallis and Futuna*

(Continues)

ANNEX TABLE 71–2 (*Continued*)

Developed countries (55 countries, 44 of which have a population of more than 140,000 people)			
Europe	*Europe*	*Europe*	*Far East*
Albania	Greece	Norway	Japan
Germany	Hungary	Netherlands	
Andorra*	Isle of Man*	Poland	*North America*
Austria	The Channel islands	Portugal	Bermudas*
Belgium	The Faroe islands*	Czech Republic	Canada
Byelorussia	Ireland	Rumania	United States
Bosnia Herzegovina	Iceland	United Kingdom	Greenland*
Bulgaria	Italy	Saint-Marin*	Saint-Pierre and Miquelon*
Croatia	Latvia	Slovaquia	
Denmark	Liechtenstein*	Slovenia	
Spain	Lithuania	Ukraine	
Estonia	Luxembourg	Holy See*	*Oceania*
Russian Federation	Macedonia	Yugoslavia	Australia
Finland	Malta	Sweden	New-Zealand
France	Moldavia	Switzerland	
Gibraltar*	Monaco*		

Those countries and territories with a population of less than 140,000 in 2000 are asterisked (*). Data from UN, 2001.

PART

III

THE FUTURE OF POPULATIONS: DEMOGRAPHIC ISSUES AND PROJECTIONS FOR THE FUTURE

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Having retraced the history of the human population, especially the recent and still ongoing transition in the developing countries from an old to a modern demographic regime reflected–as it still is in the least developed countries–in exceptional population growth, and having remarked on the wide range of current situations–especially the contrast between southern regions of the world, where population growth remains strong, and northern regions, where the decline is increasing in intensity–is it possible to foretell the future?

If we are to produce projections and population forecasts from which to conjecture probable future trends, we must first have reliable estimates of actual current situations and recent trends, but also a set of cogent methods that draw extensively on the range of principles and tools of demographic analysis. But for those projections to be reliable, they must be underpinned by hypotheses of demographic relevance on trends in the main factors of change–fertility, mortality, and migration.

In the first chapter of Part III (Chapter 72), therefore, Christine Wattelar takes a historical look at the methods and basic principles behind the reliable projections made over the past 50 years by various national and international statistical and demographic bodies, chiefly the cohort-component method but also the useful and promising companion methods of simulations and probabilistic models.

But it is not enough simply to forecast population size and age-sex structure trends; one often needs projections for much more specific target groups than the general population: school populations, working population, pension/retirement forecasts, health outlooks, etc. This is where purely demographic studies must be combined with forward economic, social, or medical studies to evolve new hypotheses with which to produce what are known as *derived* forecasts–so called because they are generally derived from existing general population findings made by a set of key projections. But not in all cases. In Chapter 73, Marlène Lamy gives a concise overview of this, exemplified by educational forecasting.

As stated at the outset, however, the most challenging stage of a population projection is formulating plausible hypotheses for fertility, mortality, and migration trends. For that reason, specialists have been enlisted to discuss in three separate chapters how future trajectories in these three components of population change can be conjectured. Luciano Ciucci and Piero Giorgi will consider hypotheses construction techniques for fertility forecasting in Chapter 74, and Graziella Caselli, Odo Barsotti, and Alberto Bonaguidi will do likewise for mortality and migration forecasts in Chapters 75 and 76.

With the tools and hypotheses to hand, it remains to make the projections and assess the outcomes. This section connects up with the chapters in Part I on the history of the human population by focusing on the future of the world population. In Chapter 77, Jacques Vallin and Graziella Caselli take critical stock of the outcomes of nearly half a century's world population projections by United Nations experts compared with actual outcomes and other attempts by other rival authors and organizations. This covers short- and medium-term projections (over about 50 years) and longer-term projections by which to conjecture not only the final outcomes of a general stabilization of all parameters at strict replacement level but also those of selected alternative scenarios and textbook hypotheses by which to assess how material a role behavioral changes now in progress will play.

In the concluding chapter (Chapter 78), the same two authors conjecture what the possible *post-transition* development towards new changes in mortality may be, such as access to a life expectancy of 150 years with the *rectangularization of the survival curve* or conversely the *re-expansion of ages at death* or fertility, with the adoption of new patterns such as the *one-child family, delayed fertility*, or *two-mode fertility*, let alone the impact of changes in the sex ratio at birth. The findings offer much food for thought. Although they set out more to invite reflection than prognosticate the future, what can be said of findings is that they are no less improbable than the absolute all-round stabilization accepted until very recently as the most likely outcome of the demographic transition.

C H A P T E R

72

Demographic Projections: History of Methods and Current Methodology

CHRISTINE WATTELAR
Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

INTRODUCTION

As Louis Henry (1973) wrote, "demographic projections enable us to calculate a population's future size and structure, according to any given set of hypotheses." This definition highlights the conditional nature of this type of exercise (see de Jouvenel, 1964: chapter 5); it implies that the hypotheses are as important as the method chosen, if not more so, as the latter is essentially quantitative (Pollard, 1975; Wattelar, 1980).

The term "demographic projection" actually encompasses two very different types of analysis. On the one hand, it denotes population forecasts, the goal of which is the short-term prediction of a population's size and structure based on plausible assumptions about the future; on the other hand, it also refers to longer-term demographic projections that, as simulations, are not constrained by the need for immediate plausibility. For instance, a population projection that extrapolates from tendencies observed during the recent past into the near future is often termed a prediction; in contrast, normative models that push a given dynamic process to its extreme, in order to assess the consequences (e.g., of certain government policies), are more strictly-speaking "projections." This type of model experiments with more speculative scenarios.

Since the late 1970s or early 1980s, however, the term "prospective analysis" is often preferred to "prediction" (Godet, 1977),[1] in the sense that "a prospective analysis aims not at predicting the future . . . but at helping us create it" (de Jouvenel, 1993: 51). Prospective analysis requires a more active involvement on the part of the researcher than does the purely intellectual exercise of prediction, however sophisticated (Romaniuc, 1990). Finally, during the same period and spurred by systemic models, the expression "sectoral forecasts" replaced that of "derived forecasts," thereby stressing the interaction of demographic variables with economic and social sectors (e.g., employment, education, health) rather than assuming that population estimates necessarily precede other predictive models.

This chapter aims to trace the evolution of these various approaches,[2] from the works of earliest pioneers to those developed in the early 1980s, before examining more recent developments, still highly influenced by these earlier methods. New techniques have appeared, but works of pure "futurology" remain rare in demography.

[1] Gaston Berger had already created this new word, prospective, in 1957, in the *Revue des deux mondes* (Cazes, 1986: 337), emphasizing the close link between action and future.

[2] Computer software for projections or population simulation will not be discussed here (Pinelli and Birindelli, 1996), as they will be examined in the specialized chapter of Volume IV.

I. GLOBAL MATHEMATICAL PERSPECTIVES

The history of demographic perspectives[3] is not a long one. The first scientific attempts to extrapolate populations go back barely 2 centuries. Influenced by the physiocrats and the newly developed *mathématique sociale*, the first authors who ventured into the prediction of population dynamics were convinced that, whatever the internal composition of the population, human evolution was governed by mathematical and physical laws.

Thus, by the mid-17th century in England, writers such as Thomas Browne and James Harrington had already attempted a mathematical demonstration of population growth. Nonetheless, the three true forefathers of demographic projection, all writing at the end of the 17th century, were John Graunt, William Petty, and Sébastien le Prestre de Vauban (Hecht, 1990: 290).

Eighteenth-century demographers remained within the tradition of these three pioneers, and Europe became the center of a heated controversy regarding the real or imagined risk of depopulation. Authors such as Montesquieu,[4] Jean-Jacques Rousseau, Condorcet,[5] and Jean-Baptiste Moheau sided with the pessimists or with the optimists, sometimes alternating between the two camps, in their common search for the "optimum" population.

Not before the early 19th century, with Thomas Malthus and his "Essay on the Principles of Population," was population growth expressed in terms of a fixed rate geometric progression, leading to population doubling every 25 years.[6] After Malthus, others introduced a more refined variant of the exponential law: the logistic law. Thus, according to the Belgians, Adolphe Quetelet (1835) and Pierre Verhulst (1845), and, later on (1923), to the Americans, Raymond Pearl and Lowell Reed, human populations do not grow at a constant rate but go through several phases, according to an interplay between the environment and the resources available. According to this law, a population's exponential increase slows in proportion to the square of its volume and cannot exceed a given threshold.

Raymond Pearl and Lowell Reed adjusted this logistic curve to the American population, registered in the 1870 and 1910 censuses. Their method generated satisfactory population predictions for the years 1920 and 1930 (Myers, 1954; Orcutt *et al.*, 1961). Others also used third- and fourth-degree functions to extrapolate the growth of the American population (Myers, 1954; Orcutt *et al.*, 1961).

These first attempts all produced acceptable results because of the long-term stability in population growth in the United States during the 19th and early 20th centuries. The 1930s crisis and, to an even greater degree, World War II, radically disrupted these trends. Since then, no fixed and aggregative law has ever successfully adjusted to population movement taken as a whole. All global mathematical projections failed and were discarded, at least as far as national forecasting[7] was concerned.

II. POPULATION PROJECTION AS FORECASTING: THE COMPONENT METHOD

As early as 1924, demographers such as Arthur Bowley abandoned global methods and drew attention instead to the importance of a fundamental element: population age structure. In 1925, the Americans Louis Dublin and Alfred Lotka demonstrated that the real growth rate of the American population was lower than the rate actually recorded, because of immigration's role in driving up numbers of young people and adults. Louis Dublin even came to envisage a decline in the American population before the end of the 20th century.[8]

1. Origins and Development of the Component Method

In 1926, the Swedish economist Sven Wicksell was among the earliest users of the component method, estimating the evolution of the Swedish population on the basis of hypotheses relating to each

[3] For more details the reader can consult the comprehensive works and articles of Thomas Frejka (1981, 1996), Bernard Cazes (1986), Marcel Lachiver (1987), Jacqueline Hecth (1990), and Éric Vilquin (1996).

[4] Charles de Segondat, baronet of Brède and Montesquieu.

[5] Marie Jean Antoine Nicolas, Marquis of Condorcet.

[6] Even before Malthus, Vauban had already predicted that the population would double every 30 years. He had, in fact, elaborated population or "populating" projections for the period 1730–2000, using 50,000 households from the Canadian colony as his starting point, and assuming four children per household: "Moyen de rétablir nos colonies de l'Amérique et de les accroître en peu de temps," cf. *Les Oisivetés de M. de Vauban* (quoted by Vilquin, 1972).

[7] At the local level, nonetheless, and more particularly for projections of urban growth, mathematical methods, including the logistic one, are still recommended (see United Nations, 1975).

[8] An earlier English author, Edwin Cannan, had proceeded in a similar manner in 1895. Before him, John Graunt had already stressed the importance of immigration in the growth of the city of London (Vilquin, 1977: 84, 87, 110).

component of demographic growth: fertility, mortality, and migration.[9]

In 1928, the American Pascal Whelpton began his long series of American population forecasts. He was the first to take a longitudinal approach to projections, using the cohort-survival method to estimate the future population, cohort by cohort; he was also the first to apply the component method to regional predictions (Thompson and Whelpton, 1934; Whelpton, 1936).[10] In this same year, 1928, Alfred Sauvy wrote his first essay on demographic projection. His later projections (Sauvy, 1932, 1937) demonstrated the importance of this new method through their success in alerting the French public to the risk of declining fertility rates faced by France at that time.[11]

The radical upheavals wrought by World War II[12] and the early postwar years left many highly skeptical about the validity of demographic predictions (Dorn, 1950). Demographic phenomena no longer displayed any regular pattern of development but were constantly fluctuating: deficient cohorts, baby boom, marriage recovery. The World Population Conference, held in Rome in 1954, expressed deep pessimism in this respect (Henry, 1954; Hajnal, 1954; Myers, 1954); at the 1965 Belgrade conference, the ups and downs of demographic projection were still on the agenda (Grauman, 1965: 10–14; Shryock, 1965).

It is since the war, nonetheless, that national statistical institutes have started to carry out regular population forecasts, based on results of 5- or 10-year censuses, depending on the country. Although the methodology is evolving only gradually, the tools used to make the calculations are developing rapidly, making it possible to incorporate multiple hypotheses and alternative scenarios (Calot *et al.*, 1970: 12).

International agencies also regularly conduct vast comparative projections, by continent, region, or country. The first world population projections carried out by the United Nations (UN) go back to 1951; since 1958, the UN has updated these projections on a regular basis, first every 5 years and[13] then every 2 years (see Chapter 77). In 1953, Jean Bourgeois-Pichat published his first European population projections at INED (*Institut national d'études démographiques*). In 1956, the Organization for European Economic and Development (OEEC), since renamed the Organization for Economic Cooperation and Development (OECD), started conducting demographic forecasts for its member states (OEEC, 1956), revising them periodically. Also worthy of mention is CELADO's (*Centro Latinoamericano de Demografia*) work in the field of projection, initiated in the 1960s for several Latin American countries (CELADE, 1960–1968). Finally, the World Bank published its own world demographic projections for the first time in 1978 (Zachariah and Mi Thi Vu, 1978).

2. Refining Component Analysis and Microsimulation

During the same period, demographers sought to achieve increasing precision in their analyses. Within period measures, they strove to identify more fundamental tendencies, contrasting a longitudinal or cohort vision of events to a cross-sectional or period one (Ryder, 1956; Akers, 1965; Hency, 1996) and making use of so-called translation methods (see Chapter 17 of Volume I) (Ryder, 1963; Frinking, 1969) to reconcile the two points of view.

a. *Fertility Perspectives*

The principal objective of the 1955 and 1960 Growth of American Families (GAF) studies (Freedman *et al.*, 1959; Whelpton, 1963; Whelpton *et al.*, 1966) was to find an efficient means of predicting marital fertility by asking couples directly about their projected family size. Results proved interesting in the study of couples' fertility intentions, as long as the necessary qualifications were taken into account when interpreting the data. In the responses given, it was clearly necessary to dissociate the now familiar distinctions between ideal, expected, and completed family size (Ryder and Westoff, 1967; Siegel and Akers, 1969). However, such surveys cannot provide information on couples not yet constituted and are also strongly influenced by period effects.

Meanwhile, knowledge of the biological process underlying human reproduction progressed greatly, thanks to microsimulation models. The REPSIM model, elaborated by Mindel Sheps and her team, is particularly worthy of mention (Perrin and Sheps, 1964). Mindel Sheps' (1971) historical sketch of the evolution of these microsimulation models remains a useful reference. Thus emerged stochastic types of projection models that follow the individual trajectories of a selected sample, no longer using the mean and aggregate indicators of larger populations (Orcutt

[9] See United Nations (1975).

[10] Jacob Siegel (1972: 67) re-examined the series of American projections.

[11] The history of French projections can be found in Nizard (1974), and INED (1987: 63–81).

[12] In 1943, the League of Nations had already conducted international projections in collaboration with the Princeton team (Notestein *et al.*, 1944).

[13] Jacques Vallin (1976) studied the results of successive UN projections in depth (see Frejka, 1981).

et al., 1961). The University of Göteborg developed such models (Hyrenius *et al*., 1967; Holmberg, 1968; Widen, 1969).

New problems emerged, however. From 1965 onward, fertility plummeted in most industrialized nations, with no means of assessing whether this trend was temporary or permanent. It appeared that couples' fertility intentions were themselves influenced by conditions at the time and could lead to serious mistakes when it came to assessing their own reproductive future (Westoff and Ryder, 1977; Monnier, 1978).

b. Regional Forecasts

Regional forecasting, characterized by internal migration, followed a path that closely paralleled that of fertility projection; as the latter, it benefited from developments in fundamental demographic research. Studies of spatial mobility now dissociate the longitudinal from the cross-sectional perspective (Eldridge, 1964). The demographic concepts of frequency, duration, and rank (Goldstein, 1954; Pourcher, 1966; Morrison, 1971; Courgeau, 1980) are increasingly integrated into the very notion of migration and are starting to be used to elaborate projections (Hobcraft, 1977).

Regional perspectives henceforth emphasize gross migratory flows rather than net migration (Siegel, 1954; Schwartz, 1963; Schryock, 1965; Lowry, 1966; Zitter, 1969): as a species, the *net migrant* is virtually extinct! The individual motivations behind aggregate measures are also drawing attention (Wolpert, 1965; Morrison, 1971; Chevallier, 1973). More recently, some have advocated applying event history analysis techniques to the study of migration (Courgeau and Lelièvre, 1989).

After a certain obsession with deterministic, even mechanistic,[14] methods, based on essentially gravitational models within which the notion of distance is pivotal,[15] regional projections are now struggling to incorporate the socioeconomic context within which the migratory movement necessarily takes place; multiple regression methods have come to the forefront (Tarver, 1961; Lowry, 1966). In addition to the more traditional concepts of zones of attraction (or repulsion), often defined by the opportunities available to migrants (Stouffer, 1940), demographers have added those of social distance (Somermeyer, 1961) and life space (Fremont, 1972; Brunet, 1973; Courgeau, 1984).

Here also, regional forecasting has had periods of hope and optimism followed by extended periods of despair (UN, 1954; Hamilton, 1961). Researchers soon realized that the smaller the area studied and the faster the population's growth rate, the greater the inaccuracies at the regional level (Siegel, 1954: 125). According to the most pessimistic, no method was better than any other, and regional demography had progressed no further than the stage of trial and error (Siegel, 1954; Giard, 1973).

If regional forecasts, whatever the methodology, are less reliable than are national ones, it is not only because they deal with smaller populations, more sensitive to the effect of small numbers but also because they are more affected by the evolution of migration than of fertility or mortality. Unlike migration, the latter contain a double inertia, which make them more predictable—a definite advantage for those involved in forecasting. Not only do fertility and mortality generally evolve in a steadier and less extreme fashion than does migration, but their influence on population dynamics and structure is also slower and more diffuse. By contrast, migration fluctuates more acutely in response to political and economic conditions, and the repercussions for population dynamics, and especially structure, can be very swift. It is precisely the weight of this kind of inertia that makes demographic forecasts so much more reliable than are economic ones. The larger the population and the weaker the role of migration, the stronger this inertia becomes. Consequently, world population forecasts are the most reliable of all, not only because errors made about one country can cancel out inverse errors made about another one but also because migration plays no role whatsoever.

Finally, coming back to regional migration, the debate is not over between the advocates of a methodology that could be described as "distributive"—one that distributes national projection estimates between regions in proportion to their population (the ratio, or quota, method)—and the proponents of a disaggregated approach, in which each territorial subdivision has a dynamic of its own, and one that could ultimately affect the national average.[16] From this point on, it became clear that regional perspectives increas-

[14] The study of migrations has long remained under the influence of the laws established by Edward Ravenstein (1885) at the end of the 19th century. Thus, migration between two areas can be determined, among others, by the distance separating them and by the size of their respective populations. This is what George Zipf (1946) called the $P_1.P_2/D$ hypothesis. This type of formulations is often called "Pareto formula."

[15] For a retrospective study of migration models and their importance for demographic projections, see Ter Heide (1963), Termote (1967), and Giard (1973).

[16] Papers presented at the seminar "The improvement of disaggregated methods and techniques of demographic projection" at the General Congress of the IUSSP, in Mexico City, in 1977, are a useful reference.

ingly require the collaboration of various disciplines. However, given the growing complexity of each variable taken into consideration, integrating and reconciling all points of view sometimes raises insuperable problems for projection.

III. PROJECTIONS AND DYNAMIC POPULATION MODELS: QUESTIONING STABILITY

The first mathematical models of population growth were created a little before that of Malthus.[17] Nonetheless, it is really to Alfred Lotka, who provided the most fundamental demonstration of the intrinsic properties of demographic movement, that we owe their definitive formulation (see Chapter 20 of Volume I). His theory has remained the reference for any study connected, however remotely, to dynamic population models; the concepts of stable or stationary populations are alluded to daily in the specialized demographic literature.

All the properties that Lotka established in terms of infinitesimal analysis were integrated, in a different format, into a new projection approach based on matrix calculus (Leslie, 1945; Keyfitz, 1964). The population matrix is often referred to as the *Leslie matrix*, in honor of its author.[18]

New computer technology made it possible to construct gigantic matrices that are particularly well adapted to the analysis and projection of interregional population movement (Muhsam, 1961; Rogers, 1968; Feeney, 1973; Rees and Wilson, 1977). These large matrices are often termed the *Rogers matrix.*

These methodological developments spurred the elaboration of large-scale projection models spanning long periods of time; unlike the microsimulation models described above, these projections received enthusiastic acclaim during this period, and the large systemic world models developed at the Massachusetts Institute of Technology were much talked about (Forrester, 1971; Meadows, 1971; Meadows *et al.*, 1974; Mesarovic and Pestel, 1974). In these dynamic models, populations interact with all the other variables of the surrounding system. The Bachue simulation models, of the International Labour Office, is another example (Blandy and Wery, 1973).[19] Initially, the notion of zero-population growth spread like wild fire and appeared as a miraculous solution to development; it nonetheless soon provoked a heated controversy (Bourgeois-Pichat, 1970; Frejka, 1973; Cole *et al.*, 1974).

From a purely theoretical point of view, Lotka's dynamic model is a process closed to migration. It became imperative, therefore, to find a way of integrating the migration process into macromodels (Tabah and Cataldi, 1963; Le Bras, 1971a). The initial hypothesis of convergence towards an upper limit can even be waived, replaced by notions such as weak convergence (or weak ergodicity) (Lopez, 1961; Coale, 1972), or quasi-stable (Coale, 1963) or even unstable populations (Le Bras, 1971b, 1974; Bourgeois-Pichat, 1994).

Dynamic population models were thus mostly elaborated at the macroanalytical level. Some studies on projection, nonetheless, make more explicit use of probabilities (Pollard, 1966; Sykes, 1969) and endeavor to be less deterministic than are classical approaches. Markovian migration processes (Goodman, 1961; Tarver and Gurley, 1965) belong to this category of models, known as stochastic models. In more recent work, transition probabilities are no longer held constant over time[20] in these models and can even become a function of migrant's personal characteristics; such processes are then labeled semi-Markovian (Ginsberg, 1971). In this way, the new probabilistic population models seek to give more of a place to extrademographic motives, and attempt to get closer to a "predictive process" in the strict sense of the term.

IV. CONTEMPORARY METHODS[21]

1. The Component Method Yet Again

Contemporary textbooks always specify that, "In the present state of projection methodology, population forecasting essentially constitutes an exercise in extrapolation based on the demographic transition model. Hypotheses about the evolution of fertility and mortality from a reference year to the final year of the projection are applied to a given population, divided by sex and age. The computer does the rest" (Tapinos, 1996: 116–117). At first glance, nothing has changed.

[17] It would be even more precise to say that they go back to Leonhard Euler (1760).

[18] For a more complete retrospective study of the various projection matrices and their creators, see Wattelar and Gillet de Stefano (1971).

[19] See René Wéry's retrospective analysis (1996).

[20] The Cornell University team postulated the axiom of cumulative inertia, which expressed the probability not to migrate as an function increasing with the duration of residence (Myers *et al.*, 1967). Seymour Spilerman's research should also be mentioned.

[21] Nico Keilman's (1985b) and Joël Cohen's syntheses constitute useful references concerning recent innovations in mathematical demography applied to models and population projections (see also Hatem, 1993; Lutz *et al.*, 1996).

Its universality is undoubtedly a great advantage of the component method. It can be applied at every level of observation (individual, local, regional, or national). Furthermore, countries with incomplete statistical data can use it as easily as can countries with a permanent population register, compensating for the lack of complete and up-to-date data through the use of indirect methods adjusting age-specific fertility or migration curves mathematically, and model life tables.

Numerous recent studies have nonetheless improved on earlier work on the mathematical adjustment and parameterization of demographic functions developed previously (Heligman and Pollard, 1980; Zaba, 1985; Kostaki, 1992; Wunsch, 1966; Romaniuc, 1973; Brass, 1974; Duchêne and Gillet de Stefano, 1974; Rogers *et al.*, 1978). Others focus on the three dimensions of demographic phenomena—age, period, and cohort (APC models) (Hobcraft *et al.*, 1982; Rodgers, 1982); they apply this three-dimensional approach to projections of fertility (Willekens and Baydar, 1984) and mortality (Osmond, 1985). "Time-series"[22] are back in fashion and have been improved by incorporating the variance that "surrounds" these statistical trends.[23]

Moreover, projections tended to be conducted separately for each sex, giving priority[24] to the female population, especially in fertility estimates; the new projection models, on the contrary, aim at simulating both populations simultaneously (Pollard, 1977; Wijewickrema, 1980; Preston, 1983; Keilman, 1985a; Pollak, 1986).

2. "Multistates" and "Multilevels"

During the 1980s, projections techniques that used large multiregional matrices continued to develop (Willekens and Drewe, 1984).[25] The multistate approach moved beyond the field of regional studies and started to be applied to other dynamic processes, such as marital status projections (Lavoie, 1984; Espenshade, 1985), educational forecasts (Guyot, 1991), and even an attempt at social accounting (Cohen and Tuyl, 1991). Finally, Claude Dionne (1995) used the Leslie matrix to create an intergenerational model of projection and retrojection.

Moreover, some felt compelled to apply their analyses to different levels of observation (multilevel analysis), making it necessary to integrate individual and aggregate approaches simultaneously (Courgeau, 1994; Goldstein, 1995; Pinelli and De Rose, 1997). During the same period, semi-Markovian models remained fashionable (Rajulton and Lee, 1988), and event-history analyses proliferated (Ravanera *et al.*, 1993).

3. Household Projections

Although household projections began as early as the 1930s (Siegel, 1972; Pressat, 1959), especially in the United States with the work of Pascal Whelpton mentioned above, this type of model blossomed during the 1980s (Bongaarts *et al.*, 1987; Keilman *et al.*, 1988). As with classical projections, there are two types of methods: aggregate household projection methods and microsimulation models (van Imhoff *et al.*, 1995; Duchêne, 1996).

Macromodels (UN, 1974) are essentially based on the computation and extrapolation of household headship (or reference person) rates, or household size (Kono, 1981). Micromodels rather draw on individual data in an effort to simulate the family trajectory of each individual (Bongaarts, 1981; Nelissem and Vossem, 1989).

One method, which can be viewed as midway between micro- and macrosimulations, consists in dividing a population by subcategories or multistates (Ledent, 1992), on the basis of geographic locality or household status. This type of approach led to the LIPRO model (van Imhoff and Keilman, 1991; Boulanger *et al.*, 1994). In addition, family reconstitution models (Rallu, 1986) extrapolate to the entire population the family dynamics observed in a reference sample. Certain demographers (Mason and Racelis, 1992) attempted to integrate both sexes within household projections, and in the final analysis, fuzzy logic has entered the realm of household projections (Murphy, 1996): reality is no longer perceived in a dichotomous manner but rather in terms of networks linking people belonging to neighboring configurations, with blurred boundaries.

4. Small Populations

"The projection of subpopulations is strangely lacking in the current debate on regional development" (Banens, 1994). The methodological problems are numerous (Termote, 1980; Ledent, 1995). Several new methods have been proposed; some are deliberately hybrid, such as the "quasi-component method"

[22] For a history of the times series method, and its use in population projections, see Cohen (1985: 182).

[23] Consult Burgio and Frova (1995) and Caselli (1996) for a critical analysis and a comparative application of these methods, particularly to mortality.

[24] In the past, some authors had already dealt with this problem (Goodman, 1953; Keyfitz, 1968; Feeney, 1972).

[25] Following the example of the Netherlands, Canada, and Belgium, national statistics offices increasingly use this approach, particularly for regional forecasts.

(Hakkert, 1985; Smith and Shahidullath, 1995); others make use of new spatial representations (geographical or social), such as fractal scales (Le Bras, 1993; Toulemon, 1991; Frankhauser, 1996).

Here also, the component method remains the rule (Rees, 1994) but introduces other variables of local interest, such as housing (Dittgen, 1995). Methods inspired by thermodynamics (Prigogine's team: Allen and Sanglier, 1981) have been applied to the mechanisms of urban growth (Pumain, 1991). The geodemographic dimension (Tayman, 1996) needs no further proof, and the inclusion of a policy variable (local policy) is an essential element of the predictive process (Thumerelle, 1987).

CONCLUSION: UNCERTAINTY

This short historical review of demographic research in the field of projections shows that, in demography as in other sciences, any progress in methodology relies largely on the results of fundamental research. The time has certainly not yet arrived for population forecasting to find its particular specificity. Its line of reasoning, which could only be autonomous at the risk of becoming sterile, is perhaps taking a new direction.

Finally, it must be acknowledged that researchers who indulge in the perilous practice of projection are well aware of the random nature of what they are doing and regularly and willingly subject their work to a retrospective evaluation by confronting the results of their models with reality. The maxim, to predict is not to explain, remains to the point.

Acknowledgment

This chapter was translated from the French version by Heather Juby.

References

Akers Donald S., 1965. Cohort fertility versus parity progression as methods of projecting births, *Demography*, vol. 2, p. 414–428.

Allen Peter, 1991. Spatial models of evolutionary systems: subjectivity, learning and ignorance, *in:* Denise Pumain (ed.), *Spatial analysis and population dynamics*, p. 147–160. Paris, INED, 458 p.

Allen Peter M. and Sanglier, M., 1981. Urban evolution, self-organization, and decisionmaking. Environment and Planning, vol. 13, p. 167–183.

Banens Maks, 1994. La prévision de sous-populations par la méthode des tendances partielles, *Population*, vol. 49(4–5), p. 1130–1138.

Berger Gaston, 1967. *Les étapes de la prospective*. Paris, PUF, 344 p.

Blandy Richard and Wery Rémy, 1973. Bachue-1: the dynamic economic-demographic model of the population and employment project of the world employment programme, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population, Liège, 1973*, vol. 1, p. 17–51. Liège, IUSSP, 494 p.

Bongaarts John, 1981. Simulation of the life cycle, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population, Manilla, 1981*, vol. 3, p. 399–415. Liège, Ordina éditions, 697 p.

Bongaarts John, Burch Thomas, and Wachter Kenneth, 1987. *Family demography: methods and their applications*. Oxford, Clarendon Press, 365 p.

Boulanger Paul-Marie, Lambert André, Deboosere Patriek, and Lesthaeghe Ron, 1994. *La formation des familles: étude prospective.* Brussels, Service de programmation de la politique scientifique, Inbel, 102 p.

Bourgeois-Pichat Jean, 1953. Les problèmes de population européenne, II: perspectives sur les populations, *Population*, vol. 8(1), p. 3–56.

Bourgeois-Pichat Jean, 1970. L'accroissement nul pour les pays en voie de développement. Rêve ou réalité, *Population*, vol. 25(5), p. 957–974.

Bourgeois-Pichat Jean, 1994. *La dynamique des populations. Populations stables, semi-stables et quasi-stables.* Paris, INED and PUF, 296 p.

Bowley Arthur L., 1924. Birth and population of Great Britain, *Journal of the Royal Economic Society*, vol. 34, p. 188–192.

Brass William, 1969. A generation method for projecting death rates, *in:* Frank Bechhofer (ed.), *Population growth and the brain drain*, p. 75–91. Edinburgh, Edinburgh University Press, 236 p.

Brass William, 1974. Perspectives in population prediction. Illustrated by the statistics of England and Wales, *Journal of the Royal Statistical Society A*, vol. 137(4), p. 532–583.

Brunet Pierre, 1975. Pour une nouvelle définition de la migration, *in: IVe Colloque national de démographie, Caen, 1973*, p. 527–536. Paris, CNRS, 564 p.

Burgio Alessandra and Frova Luisa, 1995. Projections de mortalité par cause de décès: extrapolation tendantielle ou modèle âge-période-cohorte, *Population*, vol. 50(4–5), p. 1031–1052.

Calot Gérard, Bodin Jean-Louis, Salais Robert, and Hemery Solange, 1970. *Projections démographiques pour la France.* Paris, INSEE, 123 p.

Cannan Edwin, 1895. The probability of cessation of the growth of population in England and Wales during the next century, *The Economic Journal*, vol. 5, p. 505–515.

Cazes Bernard, 1986. *Histoire des futurs. Les figures de l'avenir de saint Augustin au XXIe siècle.* Paris, Seghers, 475 p.

Celade, 1960 and subsequent years. *Proyeccion de la poblacion por sexo y grupos de edades.* Santiago du Chili, CELADE.

Chevallier Maurice, 1975. Méthodes de recherche des causes de migrations: approches par le comportement des migrants, *in:* CNRS (ed.), *IVe Colloque national de démographie, Caen, 1973*, p. 321–357. Paris, CNRS, 564 p.

Clark Jeanne C. and Sheps Mindel C., 1966. An analytic simulation model of human reproduction with demographic and biological components, *Population Studies*, vol. 20(1), p. 297–310.

Coale Ansley J., 1963. Estimates of various demographic measures through the quasi-stable age distribution, *in: Emerging techniques in population research: proceedings of a round table at the Thirty-Ninth Annual Conference*, p. 175–193. New York, Milbank Memorial Fund, 307 p.

Coale Ansley J., 1972. *The growth and structure of human populations: a mathematical investigation.* Princeton (NJ), Princeton University Press, 227 p.

Cohen Joel E., 1985. Mathematical demography: recent developments in population projections, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Florence*, *1985*, vol. 4, p. 179–189. Liège, Ordina Editions, 556 p.

Cohen Solomon I. and Tuyl J.M.C., 1991. Growth and equity effects of changing demographic structures in the Netherlands: simulation within a social accounting matrix, *Economic Modelling*, vol. 8(1), p. 3–15.

Cole Hugh S., Freeman Christopher, Jahoda Marie, and Pavitt Keith, 1974. *L'Anti-Malthus: une critique de "Halte à la croissance."* Paris, Le Seuil, 244 p.

Courgeau Daniel, 1980. *L'analyse quantitative des migrations humaines.* Paris, Masson, 225 p.

Courgeau Daniel, 1984. Relations entre cycle de vie et migration, *Population*, vol. 39(3), p. 483–513.

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, vol. 49(1), p. 7–26.

Courgeau Daniel and Lelievre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p.

Dionne Claude, 1995. Les relations intergénérationnelles et quelques modèles qui en découlent, *European Journal of Population*, vol. 11, p. 85–101.

Dittgen Alfred, 1999. Le démographe au défi des projections de populations locales, *in:* Christine Wattelar and Josianne Duchêne (eds.), *Le défi de l'incertitude: nouvelles approches en perspective et prospective démographiques*, p. 299–312. Louvain-la-Neuve, Université catholique de Louvain, Institut de démographie, 436 p.

Dorn Harold F., 1950. Pitfalls in population forecasts and projections, *Journal of the American Statistical Association*, vol. 45, p. 311–334.

Dublin Louis I. and Lotka Alfred J., 1925. On the true rate of natural increase as exemplified by the population of the United States, 1920, *Journal of the American Statistical Association*, vol. 20(105), p. 305–339.

Duchêne Josianne, 1996. *Modèles de projection des ménages et des familles.* Louvain-la-Neuve, UCL, Institut de démographie, 5 p.

Duchêne Josianne, 1999. Le traitement de l'incertitude dans les perspectives démographiques, *Qüestiió*, vol. 23(1), p. 113–128.

Duchêne Josianne and Gillet de Stefano Stefania, 1974. Ajustement analytique des courbes de fécondité générale, *Population et Famille*, no. 32, p. 53–93.

Eldridge Hope T., 1964. A cohort approach to the analysis of migration differentials, *Demography*, vol. 1(1), p. 212–219.

Espenshade Thomas J., 1985. Multistate projections of population by age and marital status, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Florence*, *1985*, vol. 4, p. 151–162. Liège, Ordina éditions, 556 p.

Euler Leonhard, 1760. Recherches générales sur la mortalité et la multiplication du genre humain, *Mémoires de 1'Académie Royale des Sciences et Belles Lettres*, no. 16, p. 144–164.

Feeney Griffith M., 1972. Marriage rates and population growth: the two-sex problem in demography. Berkeley, University of California, 230 p. (Ph.D. thesis).

Feeney Griffith, 1973. Two models for multiregional population dynamics, *Environment and Planning*, vol. 5(1), p. 31–43.

Forrester Jay W., 1971. *World Dynamics.* Cambridge (MA), Wright Allen Press, 142 p.

Frankhauser Pierre, 1996. L'analyse fractale, un nouvel outil pour l'analyse spatiale des tissus urbains, *in:* Jean-Pierre Bocquet-Appel, Daniel Courgeau, and Denise Pumain (eds.), *Spatial analysis of biodemographic data*, p. 311–340. Paris, INED, 367 p.

Freedman Ronald, Whelpton Pascal K., and Campbell Arthur A., 1959. *Family planning*, *sterility and population growth.* New York, McGraw Hill, 515 p.

Frejka Thomas, 1973. *The future of population growth: alternative paths to equilibrium.* New York, John Wiley and Sons, 268 p.

Frejka Thomas, 1981. World population projections: a concise history, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Manilla*, *1981*, vol. 3, p. 505–527. Liège, Derouaux, 696 p.

Frejka Thomas, 1996. Long-range global population projections: lessons learned, *in:* Wolfgang Lutz (ed.), *The future population of the world: what can we assume today?* p. 1–13. London, Earthscan Publications and IIASA, 500 p.

Fremont Armand, 1972. La région: essai sur 1'espace vécu, *in: La pensée géographique française contemporaine. Mélanges offerts à André Meynier*, p. 663–678. Saint-Brieuc, Presses universitaires de Bretagne, 716 p.

Frinking Gérard, 1969. L'utilisation des formules de translation dans les modèles de prévision, *in:* IUSSP (ed.), *International Population Conference*, *London*, *1964*, vol. 1, p. 65–72. Liège, IUSSP, 844 p.

Giard Vincent, 1975. Modèles descriptifs et prévisions à moyen et long terme des phénomènes migratoires, *in:* CNRS (ed.), *IVe Colloque national de démographie*, *Caen*, *1973*, p. 375–410. Paris, CNRS, 564 p.

Ginsberg Ralph B., 1971. Semi-Markov processes and mobility, *Journal of Mathematical Sociology*, no. 1, p. 233–262.

Godet Michel, 1977. *Crise de la prévision. Essor de la prospective: exemples et méthodes.* Paris, PUF, 188 p.

Goldstein Harvey, 1995. *Multilevel statistical models.* London, Edward Arnold, 178 p.

Goldstein Sidney, 1954. Repeated migration as a factor in high mobility rates, *American Sociological Review*, vol. 19, p. 536–541.

Goodman Leo A., 1953. Population growth of the sexes, *Biometrics*, no. 9, p. 212–225.

Goodman Leo A., 1961. Statistical methods for the mover-stayer model, *Journal of the American Statistical Association*, vol. 56, p. 841–868.

Grauman John V., 1965. Success and failure in population forecasts of the 1950's. A general appraisal, *in: Proceedings of the Second World Population Conference*, *Belgrade*, *1965*, vol. 3, p. 10–14. New York, United Nations, 460 p.

Guyot Jean-Luc, 1991. *Micro-populations et perspectives: présentation de la méthodologie développé dans l'étude de la population étudiante de l'UCL.* Louvain-la-Neuve, UCL, Institut de démographie, 46 p.

Hajnal John, 1954. The prospects for population forecasts, *in: Proceedings of the World Population Conference*, *Rome*, *31st August-10th September 1954*, vol. 3, p. 43–53. New York, United Nations, 906 p.

Hakkert Ralph, 1985. Quasi-component models for small area population projections: examples from the State of Sao Paulo, Brazil, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Florence*, *1985*, vol. 4, p. 163–174. Liège, Ordina éditions, 556 p.

Hamilton Horace C., 1961. Some problems of method in internal migration research, *Population Index*, vol. 27(4), p. 297–307.

Hatem Fabrice *et al.*, 1993. *La prospective. Pratiques et méthodes.* Paris, Economica, 385 p.

Hecht Jacqueline, 1990. L'avenir était leur affaire: de quelques essais de prévision démographique au XVIII[e] siècle, *Revue européenne de démographie / European Journal of Population*, vol. 6(3), p. 285–322.

Heligman Larry and Pollard John H., 1980. The age pattern of mortality, *Journal of the Institute of Actuaries*, vol. 107(434), p. 49–80.

Henry Louis, 1954. Perspectives de naissances après une perturbation de la natalité, *in: Proceedings of World Population Conference*, *Rome*, *1954*, vol. 3. New York, United Nations, p. 55–56.

Henry Louis, 1966. Analyse et mesure des phénomènes démographiques par cohortes, *Population*, vol. 21(3), p. 465–482.

Henry Louis, 1972. Passé, présent et avenir en démographie, *Population*, vol. 27(3), p. 383–395.

Henry Louis, 1973. *Perspectives démographiques*, 2nd ed. Paris, INED, 115 p.

Hobcraft John, 1977. Forecasting the components of population change, with special reference to migration: data, problems, suggestions, *in: Joint Meeting of the British Society for Population Studies and the Population Geographv Study Group*, *Liverpool*, *1977*, p. 99–126.

Hobcraft John, Menken Jane, and Preston Samuel H., 1982. Age, period, and cohort effects in demography: a review, *Population Index*, vol. 48(1), p. 4–43.

Hoem Jan M., 1973. *Levels of error in population forecasts*. Oslo, Statistisk Sentralbyra, 46 p.

Holmberg Ingvar, 1968. *Demographic models. DM4.* Göteberg, University of Göteborg, Demographic Institute, 55 p.

Hyrenius Hannes, Holmberg Ingvar, and Carlsson Margareta, 1967. *Demographic models. DM3.* Göteberg, University of Göteborg, Demographic Institute, 39 p.

Institut national d'études démographiques [INED] (ed.), 1987. *Les projections démographiques*, *Actes du VIIIe colloque national de démographie*, *Grenoble 1987*, vol. 1 Paris, PUF, 275 p.

Institut national d'études démographiques [INED] (ed.), 1988. *Les projections démographiques*, *Actes du VIIIe colloque national de démographie*, *Grenoble mai 1987*, vol. 2. Paris, PUF, 122 p.

International Union for the Scientific Study in Population [IUSSP], 1977. L'amélioration des méthodes et techniques de projections démographiques désagrégées: Papers presented at the Round Table no. 5, *in: IUSSP*, *International Union for the Scientific Study of Population de Mexico*, *1977*. *Les Actes*, p. 482–490. Liège, IUSSP, 678 p.

Jouvenel Bertrand de, 1972, 1964. *L'art de la conjecture*. Paris, SEDEIS, 385 p.

Jouvenel Hugue de, 1993. Sur la démarche prospective: un bref guide méthodologique, *Futuribles*, no. 179(September), p. 51–71.

Keilman Nico, 1985a. Nuptiality models and the two-sex problem in national population forecasts, *European Journal of Population*, vol. 1(2–3), p. 207–235.

Keilman Nico, 1985b. Application of recent developments in mathematical demography to major socio-demographic projects, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Florence*, *1985*, vol. 4, p. 91–203. Liège, Ordina éditions, 556 p.

Keilman Nico, 1990. *Uncertainty in national population forecasting: issues*, *backgrounds*, *analyses*, *recommendations*. Amsterdam, Swets & Zeitlinger, 211 p.

Keilman Nico, Kuijsten Anton, and Vossen Add, 1988. *Modelling household formation and dissolution.* Oxford, Clarendon Press, 298 p.

Keyfitz Nathan, 1964. The population projection as a matrix operator, *Demography*, vol. 1(1), p. 56–73.

Keyfitz Nathan, 1968. *Introduction to the mathematics of population.* Reading (MA), Addison-Wesley, 450 p.

Keyfitz Nathan, 1972. On future population, *Journal of the American Statistical Association*, vol. 67(338), p. 347–363.

Kono Shigemi, 1981. Further contrivances on methods of household projections with special attention to household size and to social development planning, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Manilla*, *1981*, vol. 3, p. 485–502. Liège, Ordina éditions, 696 p.

Kostaki Anastasia 1992. A nine-parameter version of the Heligman-Pollard formula, *Mathematical Population Studies*, vol. 3(4), p. 277–288.

Lachiver Marcel, 1987. Les projections avant l'époque contemporaine, *in:* INED (ed.), *Les projections démographiques; Actes du VIII*[e] *colloque national de démographie*, *Grenoble*, *mai 1987*, vol. 1, p. 29–39. Paris, PUF, 275 p.

Land Kenneth C. and Rogers Andrei (eds.), 1982. *Multidimensional mathematical demography: an overview for applied systems analysis.* New York, Academic Press, 41 p.

Lavoie Claudette, 1984. Description et application au Québec d'un modèle de mobilité matrimoniale, *Cahiers québecois de démographie*, vol. 13(2), p. 215–245.

Le Bras Hervé, 1971a. Équilibre et croissance de populations soumises à des migrations, *Theoretical Population Biology*, vol. 2(1), p. 100–121.

Le Bras Hervé, 1971b. Éléments pour une théorie des populations instables, *Population*, vol. 26(3), p. 525–572.

Le Bras Hervé, 1974. Populations stables aléatoires, *Population*, vol. 29(3), p. 435–464.

Le Bras Hervé, 1987. Nature et limites des prévisions de population, *in:* INED (ed.), *Les projections démographiques; Actes du VIII*[e] *colloque national de démographie*, *Grenoble mai 1987*, vol. 1, p. 13–39. Parus, PUF, 275 p.

Le Bras Hervé, 1993. *La planète au village*. *Migrations et peuplement en France.* Paris, Datar, éditions de l'Aube, 220 p.

Ledent Jacques, 1992. *Vers des perspectives de familles/ménages sur la base d'un modèle de type multidimentionnel.* Montréal, INRS-Urbanisation, 48 p.

Ledent Jacques, 1995. Les perspectives démographiques à micro-échelle. Une évaluation constructive, *in:* Michel Loriaux and Dominique Rémy (eds.), *Entre l'utile et le futile: pistes pour une démographie du XXI*[e] *siècle*. Louvain-la-Neuve, UCL, Institut de démographie.

Leslie Paul H., 1945. On the use of matrices in certain population mathematics. *Biometrika*, vol. 33 p. 183–212.

Lopez Alvaro, 1961. *Some problems in stable population theory.* Princeton (NJ), Office of Population Research, 107 p.

Lotka Alfred J., 1939. *Théorie analytique des associations biologiques. Deuxième partie: analyse démographique avec application particulière à l'espèce humaine*. Paris, Herman & Cie, 139 p.

Lotka Alfred J., 1970. Relation between birth rates and death rates, *Science*, vol. 26(653), p. 21–22.

Lowry Ira S., 1966. *Migration and metropolitan growth: two analytical models.* Los Angeles, University of California, Chandler Pubishing, Institute of Government and Public Affairs, 118 p.

Lutz Wolfgang (ed.), 1996. *The future population of the world: what can we assume today?* London, Earthscan Publications, IIASA, 500 p.

Lutz Wolfgang, Goldstein Joshua R., and Prinz Christopher, 1996. Alternative approaches to population projection, *in:* Wolfgang Lutz (ed.), *The future population of the world: what can we assume today?* p. 14–44. London, Earthscan Publications, IIASA, 500 p.

Malthus Thomas R. 1798. *An essay on the principle of population.* London, J. Johnson, 396 p.

Mason Andrew and Racelis Rachel, 1992. A comparison of four methods for projecting households, *International Journal of Forecasting*, vol. 8(3), p. 509–525.

Meadows Dennis L., Behrens III William W., Meadows Donella, Nail Roger F., Randers Jorgen, and Zahn Erich K.O., 1974. *Dynamics of growth in a finite world.* Cambridge (MA), Wright-Allen Press, 637 p.

Meadows Dennis L., Behrens III William W., Meadows Donella, Nail Roger F., Randers Jorgen, and Zahn Erich K.O., 1977.

Dynamique de la croissance dans un monde fini. Paris, Economica, 610 p.

Meadows Donella H. and Club of Rome, 1972. *The limits to growth: a report for the Club of Rome's project on the predicament of mankind.* New York, Universe Books, 205 p.

Meadows Donella H., Meadows Dennis L., Randers Jørgens J., and Behrens III, William W., 1972. *Halte à la croissance: rapport sur les limites de la croissance.* Paris, Fayard, 314 p.

Meadows Donnella H., Meadows Dennis L., and Randers Jorgen, 1992. *Beyond the limits: confronting global collapse. Envisioning a sustainable future.* Post Mills (VT), Chelsea Green Publishing, 300 p.

Mesarovic Mihajlo and Pestel Eduard, 1974. *Stratégie pour demain.* Paris, Seuil, 208 p.

Monnier Alain, 1978. Projets de maternité et comportements réels: une enquête longitudinale, 1974–1976, *Population*, vol. 33(4–5), p. 813–853.

Morrison Peter A., 1971. Chronic movers and the future redistribution of population: a longitudinal analysis, *Demography*, vol. 8(2), p. 171–184.

Muhsam Helmut V., 1961. Toward a formal theory of internal migration, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *New York*, *1961*, vol. 1, p. 333–340. Liège, IUSSP, 797 p.

Murphy Mike, 1996. The dynamic household as a logical concept and its use in demography, *European Journal of Population*, vol. 12(4), p. 363–381.

Myers George C., McGinnis Robert and Masnick George, 1967. The duration of residence approach to a dynamic stochastic model of internal migration: a test of the axiom of cumulative inertia, *Eugenics Quarterly*, vol. 14(2), p. 121–126.

Myers Robert J., 1954. Comparison of population projections with actual data, *in: World Population Conference Report*, *Rome*, *1954*, vol. 3, p. 101–112. New York, United Nations, 905 p.

Nelissem Jan H.M. and Vossen Add P., 1989. Projecting household dynamics: a scenario-based microsimulation approach, *European Journal of Population*, vol. 5(3), p. 253–279.

Nizard Alfred, 1974. Les perspectives de population, *Population*, vol. 29(special number), p. 239–284.

Notestein Frank W., Taeuber Irene B., Coale Ansley J., Kirk Dudley, and Kiser Louise K., 1944. *La population future de l'Europe et de l'Union Soviétique. Perspectives démographiques 1940–1970.* Geneva, League of Nations, 332 p.

Organisation for European Economic Cooperation [OEEC], 1956. *Demographic trends in Western Europe*, *1951–1971: a report by the manpower committee.* Paris, OEEC, 155 p.

Organisation for European Economic Cooperation [OEEC], 1961. *Demographic trends*, *1956–1976 in Western Europe and the United States.* Paris, OEEC, 149 p.

Orcutt Guy H., Greenberger Martin, Korbel John, and Rivlin Alice M., 1961. *Microanalysis of socioeconomic systems: a simulation study.* New York, Harper and Bros, 425 p.

Osmond Clive, 1985. Using age, period, and cohort models to estimate future mortality rates, *International Journal of Epidemiology*, vol. 14(1), p. 124–129.

Pearl Raymond and Reed Lowell J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation, *Proceedings of the National Academy of Sciences*, vol. 6(6), p. 275–288.

Perrin Edward B. and Sheps Mindel C., 1964. Human reproduction: a stochastic process, *Bulletin of the International Statistical Institute*, vol. 40(2), p. 1067–1085.

Petty William, 1686. *An essay concerning the multiplication of mankind*, 2nd ed. London, Printed for M. Pardoe, 50 p.

Pinelli Antonella and Birindelli Anna-Maria, 1996. Interest and limits of computer programs and demographic simulation, *in:* Graziella Caselli (ed.), *Demography: analysis and synthesis: proceedings of the International Seminar*, *Sienna April 1996*, vol. 2, p. 389–407. Rome, University of Rome, DSD, 251 p.

Pinelli Antonella and De Rose Alessandra, 1997. *Micro and Macro Determinants of Family Formation and Dissolution.* Rome, DSD, 11 p. + annex.

Pittenger Donald B., 1976. *Projecting state and local populations.* Cambridge (MA), Ballinger Publishing, 247 p.

Pollak Robert A., 1986. A reformulation of the two-sex problem, *Demography*, vol. 23(2), p. 247–259.

Pollard John H., 1966. On the use of the direct matrix product in analysing certain stochastic population models, *Biometrika*, vol. 53, p. 397–416.

Pollard John H., 1975. Modelling human populations for projection purposes: some of the problems and challenges, *Australian Journal of Statistics*, vol. 17(2), p. 63–76.

Pollard John H., 1977. The continuing attempt to incorporate both sexes into marriage analysis, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Mexico*, *1977*, vol. 1, p. 291–309. Liège, Derouaux, 522 p.

Pourcher Guy, 1966. Un essai d'analyse par cohorte de la mobilité géographique et professionnelle, *Population*, vol. 21(2), p. 357–378.

Pressat Roland, 1959. Un essai de perspectives des ménages, *in: Proceedings of the International Union for the Scientific Study of Population*, *Vienna*, p. 112–121. Vienna, IM Selbestverlag, 735 p.

Preston Samuel H., 1983. An integrated system for demographic estimation from two age distributions, *Demography*, vol. 20(2), p. 213–228.

Pumain Denise (ed.), 1991. *Spatial analysis and population dynamics.* Paris, INED, 458 p.

Quetelet Lambert Adolphe J., 1835. *Sur l'homme et le dévelopement de ses facultés ou Essai de physique sociale*, *Tome 1.* Paris, Bachelier, 329 p.

Rajulton Fernando and Lee Hawa Y., 1988. A semi-markovian approach to using event history data in multiregional demography, *Mathematical Population Studies*, vol. 1(3), p. 289–315.

Rallu Jean-Louis, 1986. Projection des familles aux 1.01.1990 et 1.01.1995, *Population*, vol. 41(3), p. 511–532.

Ravanera Zenaida R., Rajulton Fernando, and Burch Thomas K., 1993. From home-leaving to nest-emptying: a cohort analysis of life courses of Canadian men and women, 1910–1970, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population*, *Montreal 1993*, vol. 2, p. 207–218. Liège, Ordina édition, 493 p.

Ravenstein Edward G., 1885. The laws of migration, *Journal of the Royal Statistical Society*, vol. 48, p. 167–227.

Rees Philip H., 1994. Estimating and projecting the population of urban communities, *Environment and Planning*, vol. 26(11), p. 1671–1697.

Rees Phillip H. and Wilson Alan G., 1977. *Spatial population analysis.* London, Edward Arnold, 356 p.

Ridley Jeanne C. and Sheps Mindel C., 1966. An analytic simulation model of human reproduction with demographic and biological components, *Population Studies*, vol. 19(3), p. 297–310.

Rodgers Willard L., 1982. Estimable functions of age, period, and cohort effects, *American Sociological Review*, vol. 47(6), p. 774–787.

Rogers Andrei, 1968. *Matrix analysis of interregional population growth and distribution.* Berkeley, Edward Arnold, 115 p.

Rogers Andrei, Raquillet Richard, and Castro Luis J., 1977. *Model migration schedules.* Laxenburg, IIASA, 61 p.

Romaniuc Anatole, 1973. A three parameter model for birth projections, *Population Studies*, vol. 27(3), p. 467–478.

Romaniuc Anatole, 1990. Les projections démographiques en tant qu'instruments de prédiction, de simulation et d'analyse prospective, *United Nations Demographic Bulletin*, no. 29, p. 21–41.

Ryder Norman B., 1956. La mesure des variations de la fécondité au cours du temps, *Population*, vol. 11(1), p. 29–46.

Ryder Norman B., 1963. The translation model of demographic change, *in: Emerging techniques in population research: proceedings of a round table at the thirty-ninth annual conference of the Milbank Memorial Fund, September 18–19, 1962*, p. 65–81. New York, Milbank Memorial Fund, 307 p.

Ryder Norman B., 1973. A critique of the national fertility study, *Demography*, vol. 10(4), p. 495–506.

Ryder Norman B. and Westoff Charles F., 1967. The trend of expected parity in the United States: 1955, 1960, 1965, *Population Index*, vol. 33(2), p. 153–168.

Sauvy Alfred, 1928. La population française jusqu'en 1956, essai de prévision démographique, I, *Journal de la Société statistique de Paris*, no. 12, p. 321–327.

Sauvy Alfred, 1929. La population française jusqu'en 1956, essai de prévision démographique, II, *Journal de la Société statistique de Paris*, no. 1, p. 8–14.

Sauvy Alfred, 1932. Calculs démographiques sur la population française jusqu'en 1980, *Journal de la Société statistique de Paris*, no. 7–9, p. 319–338 .

Sauvy Alfred, 1937. Perspectives statistiques sur la population, l' enseignement et le chômage, *Journal de la Société statistique de Paris*, no. 6, p. 227–248.

Schwarz Karl, 1963. Vorausschatzung der Bevölkerung des Bundesgebietes bis zum Jahr 2000, *Wirtschaft und Statistik*, no. 12, p. 729–735.

Sheps Mindel C., 1971. A review of models for population change, *Review of the International Statistical Institute*, vol. 39,(2), p. 185–196. OK

Shryock Henri S., 1965. *Projections of total population and of age-sex structure.* New York, UN, 44 p.

Siegel Jacob S., 1953. Forecasting of population of small areas, *Land Economics*, vol. 29, p. 72–88.

Siegel Jacob S., 1954. Some aspects of the methodology of population forecasts for geographic subdivision of countries, *in: Proceedings of the World Population Conference, Rome, 1954*, vol. 3, p. 113–133. New York, UN, 906 p.

Siegel Jacob S., 1972. Development and accuracy of projections of population and households in the United States, *Demography*, vol. 9(1), p. 51–68.

Siegel Jacob S. and Akers Donald S., 1969. Some aspects of the use of birth expectations data from sample surveys for population projections, *Demography*, vol. 6(2), p. 101–115.

Somermeyer W.H. 1961. Een analyse van de binnenlandse migratie in Nederland tot 1947 en van 1947–1957, *Statistische en Econometrische Onderzoekingen*, vol. 16(1), p. 115–174.

Spilerman Seymour, 1972. The analysis of mobility processes by the introduction of independent variables into a markov chain, *American Sociological Review*, vol. 37(3), p. 277–294.

Stouffer Samuel A., 1940. Intervening opportunities: a theory relating mobility and distance, *American Sociological Review*, vol. 5(6), p. 845–867.

Sykes Zenas M., 1969. Some stochastic version of the matrix models for population dynamics, *Journal of the American Statistical Association*, vol. 64(325), p. 111–130.

Tabah Léon and Cataldi Alberto, 1963. Effets d'une immigration dans quelques populations modèles, *Population*, vol. 18(4), p. 683–696.

Tapinos Georges, 1980. De l'utilité des perspectives démographiques, *in:* Éric Vilquin (ed.), *Perspectives de population, d'emploi et de croissance urbaine*, p. 15–53. Louvain-la-Neuve, UCL, Département de Démographie, Ordina éditions, 430 p.

Tapinos Georges, 1996. *La démographie.* Paris, Le Livre de Poche, 255 p.

Tarver James D., 1961. Predicting migration, *Social Forces*, vol. 39, p. 207–213.

Tarver James D. and Gurley William R., 1965. A stochastic analysis of geographic mobility and population projections of the census divisions in the United States, *Demography*, vol. 2(3), p. 134–139.

Tayman Jeff, 1996. The accuracy of small area population forecasts based on a spatial interaction land use modeling system, *Journal of the American Planning Association*, vol. 62(1), p. 85–98.

Ter Heide Henk, 1963. Migration models and their significance for population forecasts, *Milbank Memorial Fund Quarterly*, vol. 41(1), p. 56–76.

Termote Marc, 1967. Les modèles de migration. Une perspective d'ensemble, *Recherches économiques de Louvain*, vol. 33(4), p. 413–444.

Termote Marc, 1980. Problèmes méthodologiques des perspectives de population urbaine, *in:* Éric Vilquin (ed.), *Perspectives de population, d'emploi et de croissance urbaine*, p. 347–378. Louvain-la-Neuve, UCL, Département de démographie, Ordina éditions, 430 p.

Thompson Warren S. and Whelpton Pascal K., 1934. *Estimates of future population by states.* Washington, National Resources Board, 521 p.

Thumerelle Pierre-Jean, 1987. Les perspectives démographiques à l'échelle locale: pour qui? pour quoi?, *in:* INED (ed.), *Les projections démographiques, Actes du VIIIe colloque national de démographie, Grenoble, May 1987*, vol. 1, p. 169–188. Paris, INED, 275 p.

Toulemon Laurent, 1991. Une représentation fractale de l'espace social: à quoi bon?, *in:* Denise Pumain (ed.), *Spatial analysis and population dynamics*, p. 187–198. Paris, INED, 458 p.

United Nations [UN], 1951. The past and future growth of world population: a long range view, *Population Bulletin*, no. 1, p. 1–14.

United Nations [UN], 1954. 13th session: Méthodes de projection démographique, *in: World Population Conference Report, Rome, 1954*, vol. 3, p. 1–146. New York, United Nations, 906 p.

United Nations [UN], 1956. *Manual III: methods for population projections by sex and age.* New York, United Nations, 81 p.

United Nations [UN], 1974. *Manuel VII: Méthodes de projection des ménages et des familles.* New York, United Nations, 110 p.

United Nations [UN], 1975. *Manuel VIII: méthodes de projection des populations urbaine et rurale.* New York, United Nations, 131 p.

Vallin Jacques, 1976. Les perspectives de population mondiale des Nations unies, *Population*, vol. 31(3), p. 561–592.

van Imhoff Evert and Keilman Nico, 1991. *LIPRO 2.0: an application of a dynamic demographic projection model to household structure in the Netherlands.* Amsterdam, Swets & Zeitliger, 246 p.

van Imhoff Evert, Kuijsten Anton, Hooimeijer Pietre, and van Wissen Leo, 1995. *Household demography and household modeling.* New York, Plenum Press, 369 p.

Verhulst Pierre-François, 1845. Recherches mathématiques sur la loi d'accroissement de la population, *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles Lettres de Belgique*, vol. 18(1), p. 1–38.

Vilquin Éric, 1972. *Vauban et les méthodes de statistique démographique au siècle de Louis XIV.* Paris, Institut de démographie de l'Université de Paris, 154 p.

Vilquin Éric, 1977. *John Graunt: Observations naturelles et politiques . . . sur les bulletins de mortalité, 1662, édition critique et traduction.* Paris, INED, 174 p.

Vilquin Éric, 1996. Histoire de la pensée démographique, *in:* Graziella Caselli (ed.), *Démographie: analyse et synthèse, actes du Séminaire de Sienne April 1996*, vol. 1, p. 213–239. Rome, DSD, 251 p.

Wattelar Christine, 1980. Les perspectives démographiques par âge: simplisme ou sophistication ?, *in:* Éric Vilquin (ed.), *Perspectives de population, d'emploi et de croissance urbaine*, p. 57–88.

Louvain-la-Neuve, UCL, Département de démographie, Ordina éditions, 430 p.

Wattelar Christine and Gillet de Stefano Stefania, 1971. Représentations matricielles du mouvement naturel et du mouvement migratoire d'une population, *Recherches Economiques de Louvain*, vol. 37(4), p. 379–432.

Wery Rémy (ed.), 1996. *Modélisation démo-économique: l'expérience des modèles Bachue*. Louvain-la-Neuve, Paris, Academia-l'Harmattan, 215 p.

Westoff Charles F. and Ryder Norman B., 1977. The predictive validity of reproductive intentions, *Demography*, vol. 14(4), p. 431–453.

Whelpton Pascal K., 1928. Population of the United States, 1925 to 1975, *American Journal of Sociology*, vol. 34(2), p. 253–270.

Whelpton Pascal K., 1936. An empirical method of calculating future population, *Journal of the American Statistical Association*, vol. 31(195), p. 457–473.

Whelpton Pascal K., 1963. Cohort analysis and fertility projections, *in: Emering Techniques in Population Research*, p. 39–64. New York, Milbank Memorial Fund, 307 p.

Whelpton Pascal K., Campbell Arthur A., and Patterson John E., 1966. *Fertility and family planning in the United States*. Princeton (NJ), Princeton University Press, 443 p.

Wicksell Sven D., 1926. Sveriges framtida befolking under olika fôrutsattingar, *Ekonomisk Tidskrift*, vol. 28(1), p. 91–123.

Widen Lars, 1969. *Methodology in population projection: a method study applied to conditions in Sweden*. Göteborg, University of Göteborg, Demographic Institute, 53 p.

Wijewickrema Stanley M., 1980. *Weak ergodicity and the two-sex problem in demography*. Brussels, Vrije Universiteit Brussel, 209 p. (PhD thesis).

Willekens Frans and Baydar Nazli 1984. *Age-period-cohort models for forecasting fertility*. Voorburg, NIDI, 79 p.

Willekens Frans and Drewe Paul, 1984. A multiregional model for regional demographic projection, *in:* Henk Ter Heide and Frans Willekens (eds.), *Demographic research and spatial policy: the Dutch experience*, p. 309–336. London and Orlando, Academic Press, 412 p.

Wolpert Julian, 1965. Behavioral aspects of the decision of migrate, *Papers of the Regional Science Association*, vol. 15, p. 159–169.

Wunsch Guillaume, 1966. Courbes de Gompertz et perspectives de fécondité, *Recherches économiques de Louvain*, vol. 32(4), p. 457–468.

Zaba Basia 1985. A parameterized procedure for projecting population, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population, Florence, 1985*, vol. 4, p. 137–150. Liège, Ordina éditions, 556 p.

Zachariah Kunniparampil C. and Mi Thi Vu, 1978. *Population projections: 1975–2000*. Washington, World Bank, 440 p.

Zipf George Kingsley, 1946. The P_1P_2/D Hypothesis: on the intercity movement of persons, *American Sociological Review*, vol. 11(6), p. 677–686.

Zitter Meyer, 1969. Gross migration and enrollment projections, *in:* IUSSP (ed.), *International Union for the Scientific Study of Population, London, 1969*, vol. 3, p. 1990–2004. Liège, IUSSP, 2296.

C H A P T E R

73

Educational Forecasts

From Derived Forecasts to the Application of the Flow Method

MARLÈNE LAMY

Institut de démographie de l'université de Paris (IDUP), Paris, France

INTRODUCTION

In the majority of countries, the proportion of public spending on education, often already considerable, is increasing. In Third World countries, the phenomenon is associated with the establishment of compulsory education, which affects particularly large school-age populations; in countries where the education system is older, public demand for pre- and postcompulsory education continues to increase and, taking into account the economic and social issues involved, governments are striving to respond positively.

For example, in France—where education is compulsory between the ages of 6 and 15–in the beginning of the 1990s nearly all 3- to 5-year-olds attended school, and since 1987, approximately 35% of all 2-year-olds attended nursery school. Moreover, few adolescents left school at the age of 16. The rate of school attendance among 20 year olds more than doubled between the beginning of the 1980s and the beginning of the 1990s. In 1995–1996, 57% of young people between the ages of 6 and 15; in 2001–2002, 65.9% were currently pupils or students. Thus, in 1995–1996 and in 2001–2002, nearly 15 million youngsters in France were attending school or university.

Educating such large numbers requires considerable budgets and planning for student needs. Planning is mainly based on forecasts of the pupil population, which in addition to demographic variables, take into account the future demand for different types of education, the rules that give access to it, and political environment that determines the availability of the education.

There are basically two types of educational forecasts. The first is based on population projections (which is why it is called "derived") and the enrollment rate. The others come from analyses of pupils' progress within the education system and apply flow rates to the number of pupils in attendance. These two methods are not interchangeable as they often have different objective and use different data, but they can complement each other.

The derived forecast method is usually used to shed light on general long-term trends and to make international comparisons. The flow method is used for more short-term forecasts, such as ascertaining the grade at a local level or within an educational establishment.

I. DERIVED FORECASTS

Derived forecasts can be used to evaluate the school-age population and to distribute them by grade. The principle consists of estimating the school-age population by applying the enrollment rate to the projected population by sex and age. Therefore, for the year n,

$$E_{x,n} = P_{x,n} \times T_{x,n}$$

where $E_{x,n}$ represents the total number of pupils aged x;

$P_{x,n}$, the projected population aged x; and

$T_{x,n}$, the enrollment rate at the age of x.

Technically, the calculation is simple, but it implies the respect of certain hypotheses, and the required data collection is not always easy.

On one hand, the formula assumes that there is independence between the rate and the size of the population; on the other hand, it assumes that there is concordance in the definition of ages relating to the population forecasts and to the enrollment rate.

For population size not to affect the enrollment rate, the supply must meet the demand for education without delay. In countries where the educational system is old and where the objective of forecasts is to fulfill the spontaneous demand for education, this hypothesis for the future seems obvious. However, as future rates are determined from past values, it must be verified that they have not been influenced by a lack of elasticity in the supply. For example, in France, the enrollment rate of 2-year-olds in is greater when the number of 3- to 5-year-olds is less, and vice versa. This association is obviously caused by the number of places available in nursery schools. In such circumstances, it would be risky to base future rates on the most recent ones; it is therefore recommended to use series that are sufficiently long to reduce the impact of these period variations.

This kind of competition between several demands that, in the example in question, relate to spontaneous education exists on the level of compulsory education in countries where the educational system is recent. Unless the rules of access to education are applied in a draconian manner and all the children who are not eligible are excluded, the school enrollment rate at the legal age is not likely to immediately match the unit value. Indeed, the desire of older children to take advantage of the new system will compete with the obligation of those who have reached the required age; if the supply is not increased to match the demand, the rate will remain lower than the unit value until the overall demand is satisfied.

To avoid the influence of numbers on the determination of the rates, it is also preferable to use rates by age in years rather than rates by age groups that vary with changes in population structure.

The second verification to make before applying the above formula applies to declared age. Apart from the appropriate censuses and surveys, demographic statistics and educational statistics do not originate from the same sources and are not collected at the same dates. The former are generally collected at the end or in the middle of the year, whereas the latter are collected at the beginning of the school year. Unless the data is sorted by cohort and the ages under consideration are ages reached during the year (see Chapter 6), it is not easy to establish rates that differentiate children enrolled in school from within the total population at a given time. Data comparisons are made easier if information is collected according to the age the pupils reached during the year or by birth cohort; completed age at the beginning of the school year, which depends on when the school year begins, should not be used. If total population and school population are estimated annually, then the gap between the two dates of collection will be 6 months at the most. At a national scale, if mortality and migration levels are low, the corrections may not be needed to compare both statistical series; however, corrections are required if both levels are high.

The reliability of derived forecasts is related to the reliability of the population forecasts, on the one hand, and the quality of the evaluation of the enrollment rates on the other hand. At ages at which compulsory education is effective, the uncertainty of education forecasts depends only on the uncertainty of population projections. At a national level, in the absence of significant variations in mortality and migration, the forecast is accurate if the cohorts concerned have already been born. If they have not yet been born, the degree of uncertainty will depend on future fertility.

At the ages of pre- and postcompulsory education, demographic uncertainty is compounded by spontaneous school attendance, which is more or less variable according to the level of education.

The accuracy of the projections is therefore not the same at these various levels. At university level, it depends almost entirely on changes that occur in the desire of young people to continue their education or on the wish of governments to steer the active population toward certain sectors. At the pre-elementary level, parents' wish to have more or fewer children and to place them earlier or later in school must be taken into consideration.

Long-term forecasts by level of education are essential for planning the recruitment of teachers and the creation or reorientation of educational establishments. Here again, educational forecasts will be based on population projections. The distribution of the population by level of education is obtained by applying an indicator of school enrollment to a specified projected population figure. This indicator, which is also called a *crude school enrollment rate*, is calculated by applying the number of pupils at a certain educational grade to the number who, according to legislation, are of the age to be in that grade. Thus, in a country where children start school at the age of 6 and where primary education lasts 5 years, we would have for the primary level the following crude rate Tp:

$$\mathrm{Tp} = \frac{\text{Number in primary education}}{\text{Population aged 6–10 years}}$$

In France, for example, the number of pupils enrolled at elementary level are compared to children aged 6 to 10 years, those in the first cycle (junior secondary school) to children aged 11 to 15 years, those in the first short cycle to the 14- to 17-year-olds, and the first long cycle (junior and senor secondary school) to the 15- to 19-year-olds.

Such rates can surpass the unit value if the number of pupils repeating a grade is significant; it decreases if measures are implemented to reduce academic failure. These measures should be integrated into the underlying hypotheses of the projections.

In developing countries, crude rates are likely to be even higher as pupils entering the system late are added to higher numbers of pupils repeating a grade. The resorption of the demand for education from older pupils reduces the crude rates by level of education, whereas the reduction in the dropout rate over the course of a cycle causes an increase. Moreover, any change in the policy relating to repeating a grade (regression or even suppression) during a cycle is likely to have repercussions on the distribution of pupils of other cycles; any innovation at a given grade should be integrated into the forecasts for all the grades.

Grade repeating, dropping out, and progression are terms found in the majority of flow model variants on which shorter-term educational forecasts are based. This type of model is preferred in evaluating the number of pupils per year, or the grade of study at a local level and within individual schools or universities.

II. THE FLOW METHOD

At the end of a school year, pupils can either proceed to the next grade, repeat the grade, or drop out. If there is no external contribution, the numbers present during a given year at a certain grade are equivalent to what has become of the pupils enrolled at the preceding grade during the previous year. It is therefore through the comparison of two successive stocks that the flow is estimated. This is how real, imaginary, or apparent educational progress is evaluated.

If only the numbers by year of study are available for two consecutive school years, then the *retention rate* (*rg*) is calculated as the number of pupils in a grade (g) at the beginning of the school year compared with those of grade $g - 1$ at the beginning of the previous school year. These rates are then applied to the number present in one grade at the beginning of the school year t, ($E_{g,t}$), and by iteration, the number of pupils at the beginning of the following school years is obtained:

$$E_{g+1,t+1} = rg \times E_{g,t}$$

The *rg* rates vary under the influence of school results and demographic events such as mortality and migration. In the case of external contributions being greater than departures, their value can be greater than one. These are also called *apparent flow rates* and are used in schools to evaluate the number of pupils per class in the very short term, such as the beginning of the next school year or the year following. In developing countries, they are used at the national level as little data collection is required.

When possible, it is nevertheless preferable to differentiate, at each grade, those pupils that are repeating a grade from new entrants. Thus, by difference, the net balance between entries and departures of one grade are obtained, and the succeeding flow rates calculated are less global than the are previous ones:

Grade progression rate (P_g)—proportion of pupils of grade g during the school year t who will be in the next class in the following year

Grade repetition rate (R_g)—proportion of pupils of grade g during the school year t who will still be in g in the year $t + 1$

Grade dropout rate (A_g)—proportion of pupils of grade g who will drop out during or at the end of the school year t

The outcome for pupils at grade g is calculated by the following:

$$E_{g,t} = P_g \times E_{g,t} + R_g \times E_{g,t} + A_g \times E_{g,t}$$

with

$$P_g + R_g + A_g = 1.$$

And in the absence of external contributions, the following equation makes it possible to obtain the number at grade $g + 1$ at the beginning of the school year $t + 1$:

$$E_{g+1,t+1} = P_g \times E_{g,t} + R_{g+1} \times E_{g+1,t}$$

However, it is not possible with this flow method to forecast the number of entries into the educational system, which must be estimated by using population forecasts.

The analysis can be even more precise. For a given grade (if the data are available), those passing to the next grade and those repeating the grade can be esti-

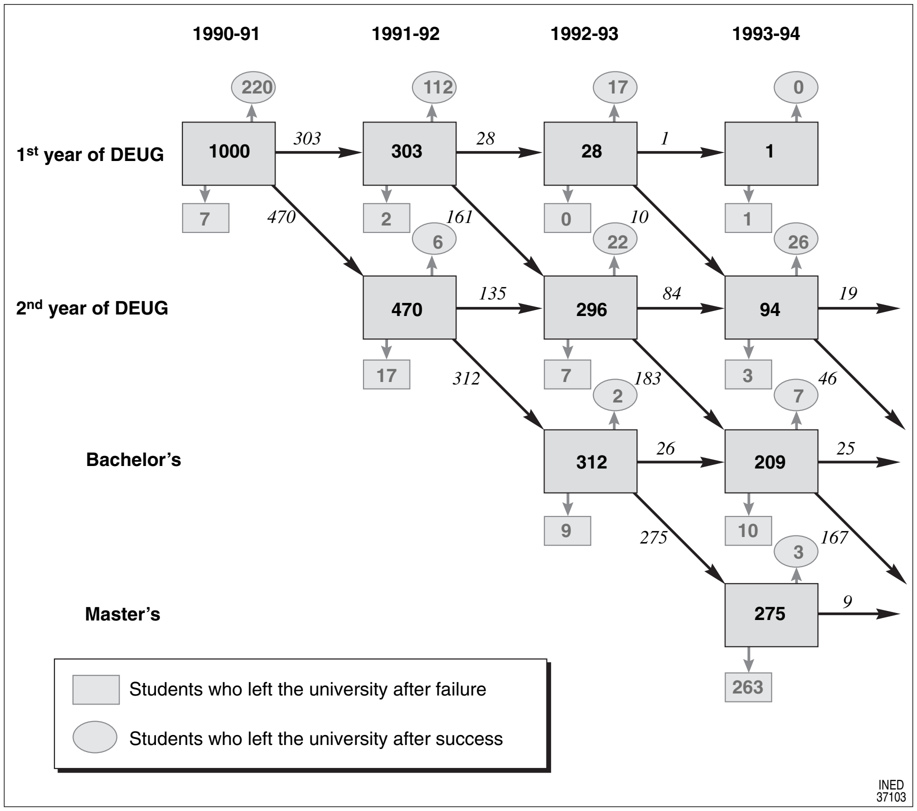


FIGURE 73–1 Cohort study at the *Université de Paris I Panthéon-Sorbonne.*

mated according to whether the pupils were enrolled for the first, the second, or even the third time. Cycle and grade repeating can be matched with dropout rates; transfer rates between regions or educational institutions can also be introduced.

The possible influence of one level or cycle on another must also be taken into account. A restriction on entries at one level can provoke a surplus of repeating pupils waiting to enter the cycle, and the preceding level as well as the adjacent levels will be affected by its repercussions.

Although they are more precise than are the previous analysis, these variants remain based on period data analysis; information relating to two consecutive years is sufficient to establish the required indicators.

An analysis of behavior, based on real case studies is by far preferable. But collecting the required data is an onerous task. At the national or local level, it requires the existence of panels of pupils. These pupils are studied on their arrival at primary level in order to evaluate the rate of passage to the next grade and the repetition rate, or they are studied in the French system on their arrival at *collège*[1] in order to estimate the distribution of pupils between the short and long cycle. Progression can be traced from secondary school to show the proportion of holders of the *baccalauréat* and of pupils dropping out during the cycle.

The issues vary according to education level and become more complex when schooling is no longer compulsory. These issues can often be treated in a differential study according to social and family environment. This latter variable, which is important for the future of pupils, is also important in the design of the hypotheses on which the forecasts are based. However, it loses impact as pupils progress through the educational system and can be replaced by the age reached and the diplomas received.

Educational institutions, because of their size, prefer studying entire cohorts to samples. They attract a more homogenous population compared to the one at the national level because of their geographic and social localization or the selection process that they operate. As a result, they can often dispense with an internal sociological differentiation in establishing their projections. Moreover, the smaller population size would probably make this impossible.

Cohort studies within an establishment can be very informative. In the following example based on the French educational system, for readability the distri-

[1] Junior secondary school that French children attend between the ages of 11 and 15.

bution of those who passed one level at a given grade is shown, as well as their characteristics: age reached, *baccalauréat* type, social and family environment, and place of birth and residence. One can take these factors into account by retracing the educational history according to them, thereby refining the forecast. Such detailed data collection even makes projections by simulation of individual trajectories possible.

The degree of detail contained in forecasts depends on their objectives. A mayor who wishes to know the number of meals to provide in the school dining hall is only interested in the total number of pupils. Academic inspection departments need to know the distribution of pupils per class in order to create new classes or close down some according to predefined criteria of numbers per class. The Education Department needs to be able to plan the recruitment of teachers. These different forecasts require educational projections.

In the long-term, it is preferable to use the derived projection method, which requires the design of population projections beforehand, for which the risks of error at each stage are independent. In the more short-term, the flow method is better, but it has the disadvantage of increasing the size of any potential initial error when the projection period is long. The choice of objectives is at least as important as the choice of methods for projections, including educational forecasts.

References

Guillois Pascale, 1983. Les projections nationales d'effectifs dans le premier et second degré: méthodologie, résultats à long terme, *éducation et formations*, no.2, p. 85–92.

Ministere de l'éducation nationale (France), 1996. 1996–2005: Projection du système éducatif, ralentissement? *éducation et formations*, vol. 47(special issue) 95 p.

Thonstad Tore, 1983. *Analyse et projection des effectifs scolaires dans les pays en développement: manuel de méthodologie.* Paris, Nations unies/United Nations, 131 p.

CHAPTER

74

Hypotheses for Fertility Forecasts

LUCIANO CIUCCI AND PIERO GIORGI

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION

The advances made in the description and interpretation of reproductive behavior have often provided the hope of increasing the capacity to forecast the future of fertility. However, reflection on forecasting tools (Lee, 1978; de Beer, 1992) as well as the ex post evaluation of earlier results have periodically led researchers to increasing pessimism about the possibility of being able to anticipate future fertility trends with any certainty at all.

It is true that forecasters rarely aim to predict what will really occur. Instead, they usually suggest possible scenarios for fertility trends without taking the risk of predicting which scenario is most likely to occur; however, the results are hardly satisfactory as the anticipation of trends is rarely based on clearly stated theoretical presuppositions.

The reason for this uncertainty is owing to insufficient understanding of the individual and collective determinants of reproductive behavior, which makes it impossible to predict whether the couples of the future will make different reproductive choices compared with those of the couples of today.

A review of currently available forecasting tools and possible methodological developments cannot, however, be based on this single premise. To understand what can reasonably be imagined and what can reasonably be accomplished, it is not possible in forecasting fertility, even when methods are based on the most recent expertise, to disregard objectives or targets assigned to *knowledge of the future* of fertility.

This knowledge of the future can only be understood as an intellectual construction or conjecture on the probable future, or even of several probable futures because the directions in which the trend can develop are multiple and what will actually occur will depend as much on spontaneous factors as on political and social decisions and choices, without mentioning possible innovations that are difficult to predict. The modern mind cannot accept the hypothesis of a determinist future, and a future that is not controlled by human decisions and that depends only on spontaneous evolution is unimaginable. The political, economic, and social implications associated with the progressive aging of populations of the developed countries and the rapid demographic growth of the developing countries should also be taken into consideration insofar as they can lead to interventions or population policies likely to modify current trends.

Yet, in the short and medium term, the essential operational needs of administration and management require knowledge of the future, which is equivalent to an *anticipation of what will occur.* In this case, indeed, such a requirement can be satisfied by supposing that the near future is, at least partly, already written in the present and past trends and that possible changes are, barring major upheavals, limited. In addition, it is assumed that retroactive effects likely to modify spontaneous trends are improbable.

I. APPROACHES AND METHODS OF FORECASTING FERTILITY

Traditionally, among fertility forecast methods, it is possible to make a distinction between those based on extrapolation, which consider the future as a continuation of the past and can only be evaluated through an accurate analysis of past experience; those based on desired fertility expressed by couples; those based on explanatory fertility models; those based on the opinion of experts; and those that are based on the construction of scenarios.

Various authors have already evaluated the problems of forecasting fertility and also drawn up exhaustive assessments of the different approaches and methods used (Lee, 1978; Willekens, 1990; de Beer, 1992). In this chapter, we will restrict ourselves to a rapid examination, essentially critical, of the different approaches before discussing the potential of these various methods according to the aims of the forecast.

1. Extrapolation

The extrapolation of past fertility trends is based on more or less detailed measurements of the phenomenon: intensity and tempo, age-specific rates, order-specific rates, parity-specific rates, etc. (see Chapters 12, 14 of Volume I).

The independence of measures, the number of variables, and the related logical and operating choices are the difficulties in the application of these various methods. Moreover, empirical evidence exists that supports more or less disaggregated methods, and the choice of various operators is not univocal.

The complexity of the issue is owing to the interdependence of these two aspects and their relation to the choice of a longitudinal (Chapter 12 of Volume I) or period approach (Chapter 14 of Volume I). In other words, the choices made on one of the dimensions of the issue necessarily engender repercussions on the other dimensions. For example, using a parametric curve, the tool most often used to make a dimensional reduction, transfers the problem of the independence of specific rates to the parameters of the model used to represent them. Attempts have been made to overcome the problem of variable interdependence, but these have not yet resulted in projection methods (Thompson *et al.*, 1989).

In addition to the problem of independence that rises nearly exclusively from logical coherence, the problem of the consistency of projected variables must also be dealt with, if only in terms of the coherence of the profiles of the extrapolated fertility rates.

Various authors have studied the consistency problem. Some have tried to solve specific stages such as that of the dimensional data reduction. This is the case for the procedure based on the factor analysis of the fertility rate matrix (Bozik and Bell, 1987). In other cases, the consistency problem has been automatically solved by using approaches in which the main objective is not consistency itself or dimensional reduction as a basis, as is the case in approaches based on the multivariate *Arima* models adapted to cohorts (de Beer, 1985). The main objective of these models is to extrapolate period indicators up to a level already attained in terms of cohorts or, as in the application of age/period/cohort (APC) models—be it at their simplest (Lee, 1974) or more complex (Willekens and Baydar, 1984; Giorgi and Viola, 2003)—to take into account the analysis of the effects of age, cohort, and period. More recently, the consistency problem was tackled by the representation of the fertility process through a model that, by associating the parity progression ratios, the time elapsed since the preceding event, and the age of the mother at the time of the preceding event, directly produces a consistent set of age-specific rates (Martinelle, 1989, 1993).

In addition to the above-mentioned methods that produce age-specific fertility rates or other measures from which they are directly deductible, there are also other methods based on the extrapolation of fertility indicators that are nonreducible to age-specific rates, such as methods based on fertility measures by age and parity (Al-Osh, 1986), by parity and the time elapsed since the preceding event (Feeney, 1985), or by parity only. In all cases, however, these methods require the construction of appropriate population projection models capable of producing the female population distribution necessary for the application of the indicators used. This construction can introduce problems in incorporating mortality and migration, which produce the number of births to be inserted in the classic forecast of population by sex and age. These methods are only useful if birth order is one of the objectives of the forecast.

2. Methods Based on Desired Fertility

For short- and medium-term forecasts, some authors rely on models based on the observed desired fertility of women (Lee, 1980; de Beer, 1991; Van de Giessen, 1992). Although some official organizations use this type of application, the real usefulness of such data for fertility forecasts remains largely controversial.

The particularity of this strictly microapproach is the use of intrinsically predictive variables that are based on women's replies to questions regarding their future fertility. However, although fertility intentions are highly predictive from a theoretical point of view, in reality they cannot be directly used, as actual births do not generally match the intentions expressed (van Hoorn and Keilman, 1995).

In actual fact, fertility intentions only represent potential reproductive behavior if the individual and structural aspects of the socioeconomic situation remain unchanged. Some authors see only purely conjectural indicators in these fertility intentions and believe forecasts based on desired fertility depend too heavily on current conditions to be of any use (Westoff, 1981).

On the other hand, other authors attribute a real predictive value to this approach, even if they do not directly use fertility intentions to estimate future fertility, as fertility plans will not be carried out in the same way or in the same terms implicitly or explicitly predefined by the women. This is why the transformation of intentions into actual forecasts requires the development of adjustment methods (van de Giessen, 1992; van Hoorn and Keilman, 1995) that are based on the use of two conceptual approaches that can, moreover, be combined: the *limiting factors* model and the *partial adjustment of the projection*.

The limiting factors model consists of weighing the declared intentions with individual variables, identified beforehand, directly or indirectly associated with factors that obstruct the achievement of these goals. The partial adjustment of the projection is based on the inverse relation between age and desired fertility empirically observed within the cohorts, and evaluated from successive surveys.

Data on desired fertility are mainly used to complete the fertility of female cohorts who have not yet completed their fertility, or to develop fertility hypotheses to use in other projection methods. A significant example of the use of data on desired fertility can be seen in the fertility forecasts of The Netherlands' Central Statistical Office (Statistics Netherlands) (de Beer, 1992).

3. Methods Based on Explanatory Models

Models based on an explanatory approach to fertility use the results of regression analyses of the time trends in variations of fertility and other variables, particularly economic ones. In general, they do not rely on a very solid theoretical basis, so much so that often variables that appear correlated to fertility levels are accepted, even if there is no logical association between them. Moreover, these models do not simplify the issue of forecasting, insofar as they only substitute the problem of fertility forecast with that of predicting the phenomena, premonitory or explanatory, which are associated with it. This is the case of two models proposed by Ermisch (1992) that are based on the relation between certain economic variables (particularly female employment) and (1) the trend in age-specific fertility and (2) and the parity progression ratios by age and parity. This type of approach obviously comes up against the possibilities of actually predicting the trend of explanatory variables (Long, 1985), and of validating the explanatory models adopted, particularly with reference to their invariance with time, either in terms of relative weight of the different variables or in terms of the explanatory variables taken into consideration in the model. Social and economic variables are indeed subject to rapid changes, and their development is often more difficult to predict than is that of the demographic variables with which they are associated (Keyfitz, 1982). However, in developing countries where the trends of certain variables are sometimes easier to predict, it can be useful to base fertility forecasts on those of their explanatory variables. Thus, the development of female education has become a solid base for fertility forecasts in some countries.

In Richard Easterlin's model (1976), which directly produces series of future births, explanatory mechanisms of a highly economic nature are in actual fact represented by variables that are exclusively demographic (Ahlburg, 1986). In substance, this model's subjacent theory is that individuals belonging to a birth cohort, or to a group of cohorts, that is larger (for whatever reason) than the preceding one suffer greater stress owing to the imbalance between their aspirations and the resources to achieve them. This stress, which is expressed, for example, in terms of entry into the employment market or economic and social progress and which is therefore associated with an objectively disadvantageous situation, encourages individuals to reduce their number of children because they are unable to cope with the costs related to the arrival of children. On the contrary, the cohort of their children, fewer in number, will experience an objectively more advantageous situation with less imbalance—or even none at all—between their actual resources and their aspirations, an advantage that, according to the theory, should lead them to have more children.

Such a cyclicity, although resulting from essentially economic mechanisms, can be seen through the periodic oscillations of the population structure observed at the level of birth cohorts, with an amplitude of fluctuation double the interval between two generations, and, from there, can easily be transformed into a predictive instrument.

4. Methods Based on Objectives/Hypotheses and Expert Opinion

Population forecasts by official bodies are often based not on actual fertility forecasts (which, on the contrary, appears to be the case for mortality) but according to objectives promoted to the rank of hypotheses, even if they are not always presented as such by their authors. A critical analysis of the fertility hypotheses employed by the various bodies responsible for official forecasts show to what extent, in reality, such objectives have been frequently adopted and how they have often conditioned, albeit only implicitly, forecasts theoretically based on extrapolation (Lutz, 1991; Keilman and Cruijsen, 1992). The most frequent focal point was initially the *replacement* level under which period fertility should not fall; followed by the level of 1.5 children per woman that no generation was supposed to cross; together with alternative trends toward fertility decline, a hypothesis set aside a priori for a long time, even when it appeared clearly in the extrapolation of current trends; and, finally, the return to the replacement level, which continues imperturbably, despite all odds, to fascinate.

5. The Scenarios Method

The word *scenario* has quite varied meanings, and to understand the content of fertility forecasts based on the scenario method, a precise reference frame must be specified. These forecasts are in no way, unlike population forecasts, simulations of the future resulting from the presumed trend of fertility dynamics without reference to a clearly defined theoretical framework.

The method is based on the construction of pertinent alternatives for the future, none of which are considered to be more likely but which are seen as "a coherent group of realistic trends for the future" (Cruijsen, 1991). It thus aims to build an instrument for analyzing the future and is particularly useful in evaluating possible consequences of political decisions, leaving the field clear for creativity while not reducing it to a purely academic exercise. That is, it supplies images of the future based on a meticulous quantitative analysis of processes that, by hypothesis, could lead from today's situation to that of the future.

The scenario method is therefore meant to be very different and innovative compared with the extrapolation of the past. It seems to be closer to the view of forecast not as an anticipation of the future but rather as prospect or conjecture, that is, as an intellectual construction of the future or as an exploration of the various possible, although sometimes contrasting, futures. It is a long-term projection method aimed at showing the effect of potential structural changes.

Careful examination of the past is once again an essential factor of validity, but extrapolation of past trends is not in itself proof of a good definition of future trends and it is not justified by the application of analogous methods. It aims rather to interpret trends in terms of the internal structural relations within the system under observation and to respect major changes in order to avoid discontinuity between the current situation and future scenarios. In this sense, as well as the synchronic approach, used to construct representations of the future, a diachronic approach is also used to define the direction to follow in order to link the present with the future.

The definition of reliable logical rules and clear procedures is needed to limit the arbitrary values in this method that can produce sufficiently operational projections, insofar as it succeeds in translating qualitative evaluations of future trends into quantitative terms. Using appropriate quantitative methods is therefore also important, even if it is likely to reduce the scope of creativity available with this method.

In an interesting application of this method (de Jong, 1995), scenarios of future fertility were obtained by imposing deviations on an observed trend to extrapolate the fertility of cohorts of women. The direction and scale of these deviations were derived from the analysis and theorization of the relation between developments in socioeconomic and cultural systems and fertility trends.

II. FORESEEABILITY OF FERTILITY AND CRITICAL ANALYSIS OF THE METHODS

Table 74–1 presents a classification of available methods, some of which are simple tools, according to whether they are based on the observation of past trends of the reproductive process or, on the contrary, on theoretical or explanatory models of reproductive behavior. A residual category assembles the approaches that do not fit exactly or completely into the two main categories.

We have singled out some important characteristics in order to better distinguish between the different methods:

Use of measures of past fertility trends
Ability to anticipate trend reversals
Degree of subjectivity
Longitudinal or period approach
Quality of the estimate
Significance of the ex-post evaluation
Content of the projection

Short-term methods based on neutral projections, and therefore on extrapolation, are characterized by a complete and systematic use of the past, a minimal degree of subjectivity (reference period, trend model),

TABLE 74–1 Characteristics of Fertility Forecast Methods According To Whether They Are Based on Past Observations or on Theoretical or Explanatory Aspects

Classification elements	Approaches based on:					
	Past observations	**Theories and paradigms**		**Others**		
Basis	observed results of the reproductive process in terms of event and measure	interpretative theories of reproductive behavior		partial or mixed		
Product	"projection" of the past	"anticipation" of the future		"projections" partially associated with the past and dependent on other criteria		
Methods	extrapolation of (1) measures, (2) model parameters (historical, age/period/cohort)	use of expressed desired fertility	construction of scenarios	reference to analogical schemes	realization of objective or reference hypothesis	expert evaluation
Use of past trends of fertility indicators	total	nil	nil	partial	partial	partial
Ability to anticipate trend reversals	nil in cases of direct extrapolation but some models (age/period/cohort) can produce reversals	yes	yes	yes	yes	yes
Degree of subjectivity	weak and limited to (1) the reference period, (2) the trend model	weak (choice of impeding factors)	high	high	high	high
Longitudinal/period	promotes the evaluation of trends (disregarding the logical contents of the longitudinal approach)	longitudinal	both, but mainly longitudinal	longitudinal	longitudinal (the objectives are expressed in longitudinal terms)	longitudinal (but often the criterion is period related)
Quality of the estimate	depends on (1) identification of the trend, (2) its variability; method fails if there is a break in the trend (need to work on disaggregation to increase inertia and flexibility)	ability to recognize the associations between the wishes (goals) and their achievement	depends on the ability to process the theory and its ability in anticipating the future	not assessable		
Significance of the expost evaluation	reduced or nil	important as a knowledge base of relations in expectations and achievements	important in verifying the ability of a theory in anticipating the future	useful for revisions		
Predictive content						
Short-term	good	high	nil	limited (difficulty in evaluating the gap)	nil (by construction)	limited (no logical frame of reference)
Long-term	nil	nil	limited			

the production of estimates whose quality depends (because of the significant short-term inertia of disaggregated indicators) on the identification of trends and the parameters to be projected and their variability. For an improved identification of trends, the longitudinal perspective is often adopted. Such methods cannot generally anticipate trend reversals unless they result from cohort effects that are detectable in complex models such as APC models (Chapter 18 of Volume I) or translation models (Chapter 17 of Volume I).

Using data on desired fertility in short-term projections is characterized by its more or less complete lack of recourse to the past, a strictly longitudinal perspective, and a low degree of subjectivity in the choice of factors limiting the achievement of desired fertility. When used for short-term forecasts, these data help to complete the fertility of cohorts currently of reproductive age, and their main asset is the theoretical capacity to predict trend reversals absent in other forecast methods.

The advantage of this type of approach depends largely on the stability of desired fertility and the extent to which it can be achieved, taking into account the obstacles that couples could encounter as well as current events that could have affected the estimates. Indeed, some authors maintain that desired fertility depends largely on the socioeconomic and cultural climate of the time and, as a result, can only be conjunctural. The concept of cohorts' fertility plans is being challenged today with the return of the idea that cohort fertility, its intensity, and tempo are caused by a dynamic process of interaction among desired fertility and the obstacles, limitations, and constraints imposed by the circumstances at the time. Despite theoretical uncertainties, these methods are still used to a great extent by official statistical institutes in their forecasts.

The construction of long-term scenarios is characterized by the absence of recourse to the past, integrated use of longitudinal and period dimensions, and a large dose of subjectivity. There does not as yet seem to be sufficient knowledge on the complex mechanisms that associate, on one hand, the collective and structural factors and individual factors with, on the other hand, fertility trends in order to produce the certainties necessary for evaluating the degree of likelihood of the constructed scenarios. Numerous obstacles that prevent the theories and the interpretation frameworks from being transformed into efficient forecasting tools (Keyfitz, 1982) actually occur, even when we simply wish to extract likely future scenarios from the theories.

If imagining future changes in behavior is beyond the actual cognitive possibilities that exist today, the static, comparative, and conditional nature of current theories impede the simple translation into trend scenarios, as does the fragmentation and partiality of the various theories. Often two theories are applicable or better, their mode of action on behavior occurs jointly but without it being possible to determine their relative weight. The qualitative nature of a number of theories is not helpful in constructing scenarios. One cannot say that the possible availability of a complete theory of reproductive behavior (a possibility that some authors challenge) can be transformed into a tool both useable and manageable for normal human intelligence.

Furthermore, even by presupposing the credibility of conjectures on future trends of reproductive behavior according to the dynamics of the socioeconomic context, the technical tool for constructing quantitative scenarios does not seem to respond to precise rules and thus imposes a large dose of the arbitrary in the actual construction of scenarios for the future.

III. PROSPECTIVE

To evaluate what might occur in the short-term, in other words to make a real forecast, it is possible to use different tools that are, whatever happens, capable of making short-term forecasts. It is possible to use all those methods based on the results of the reproductive process, expressed in terms of event/measure that base the forecast entirely on the analysis of past trends in fertility indicators and their projection into the future with the aid of a neutral extrapolation method. Still in the short-term, variables on desired fertility—which by their very nature are intrinsically predictive and can, theoretically at least, anticipate manifestations of reproductive behavior—can also be used.

For long-term forecasts, on the contrary, the tools rely mainly on the development of theoretical knowledge, both qualitative and quantitative, to be used not for making forecasts but for constructing scenarios that can prefigure various possible futures, none of which, however, might actually occur. As previously stated, these scenarios are reasoned conjectures that, although based on theoretical presuppositions, do not claim to be forecasts, as it is impossible to control the complexity of the theoretical relations and difficult to control all the dimensions of the phenomenon's variability. In the abstract, scenarios could also anticipate change processes that lead to new forms of behavior and are therefore not foreseeable, a prospect that should certainly not be excluded in the long term. In concrete terms, however, giving theoretical justifications to possible but unforeseeable trend reversals, such as a rise in fertility in the European countries, does not seem easy if we consider that in the very rare cases in which it has been verified (e.g., Norway), either it cannot be explained or explanatory factors of contrary effects have been adopted (a favorable economic development that results in fewer relative costs of children and more women undertaking an economic activity).

Disregarding, in any case, the desirable development of the theoretical knowledge and techniques necessary for increasingly more effective applications of the scenario method, progress is still needed particularly in optimizing the use of the more traditional methods and their capacities, always in the context of their effective possibilities of forecast.

According to these considerations, the only real forecast would be to explore the short term only. And for projection methods, methods of demographic analysis aimed at separating the demographic mechanisms of reproductive behavior should be promoted in order to better determine the trend of their elementary components by eliminating the effects of structural factors from the measure. Recent or future progress in the field of fertility analysis by birth order or by parity (Martinelle, 1989; Rallu and Toulemon, 1993) could therefore be used to study the effects of age, period, and cohort (Willekens and Baydar, 1984) and to examine the translation mechanism beyond the traditional model of event frequencies (Keilman, 1994; Keilman and van Imhoff, 1995).

A more detailed description of reproductive mechanisms does not necessarily improve the quality of short-term forecasts and is of no value for long-term forecasts, but it reinforces our knowledge and assumes a fundamental importance when it comes to associating reproductive behavior with its determinants, as fertility scenarios attempt to do.

Evidently, the trend is to abandon or at least to go beyond the constraints that constitute, on one hand, the use of probabilities by age that integrate measures of another nature into their forecast (Martinelle, 1989) and, on the other hand, the use of completely different approaches aimed at directly predicting births that thus become an exogenous component of the system of population projection by the component method. Recourse to multistate models would make it possible to directly insert fertility indicators by parity, accompanied, however, by a significant increase in the complexity of the whole system.

In terms of fertility projections, progress in the analysis and the identification of the most basic components of reproductive behavior can be favorable to a new method of constructing *conditional forecasts*, understood not as the results and consequences on the population of hypotheses concerning the trends in fertility (and mortality) but as the trends of aggregate fertility indicators within a set of possibilities identified by analyzing what is considered as improbable or incompatible at the level of basic components. This presupposes that the disaggregated methods can distinguish a group of associations that help to define this set of possibilities.

The possibility of relating the fertility measures adopted in the projection or forecast model to the reproductive processes attributed to cohorts becomes more important to give greater demographic significance to the input of the model by attaching it directly to hypotheses of reproductive behavior, to make rational predictions on the reproductive behavior itself, to associate the fertility forecasts with the desired fertility observed at an individual level (de Beer, 1991). In other words, we refer to the diversity of the possibilities provided by these models to translate the hypotheses on the *input* into underlying hypotheses or objectives on reproductive behavior, or (in light of what has been stated) refer to the capacity that such measures have of representing the evolution of the reproductive process, if they are analyzed on a temporal basis, and enabling the forecast to be based on actual knowledge of the phenomenon.

The association of the variables of the model to the subjacent reproductive process can also have repercussions on the possibility of correctly using statistical methods capable of identifying and quantifying dependency between the variables. This is only possible if the measures under consideration are not ambiguous as a result of a lack of control over certain variables. Thus only the relationship with the subjacent reproductive process allows effective explanatory models to be produced.

Moreover, the forecast procedure, be it mathematical or statistical, would produce the best results if the associations between the variables of a purely demographic nature have already been eliminated and if techniques do not require an internal method of control (Giorgi and Viola, 2003). This last case is equivalent to the hypothesis that even with measures that do not control the population's age structure, the forecast method could grasp and distinguish, in a way that we cannot be aware of, between the evolution of components relating to the measure itself, in the sense of reproductive behavior, and trends of component, which are owing to structural elements. We can understand why such hypotheses have never been attempted and that, even in a case in which the method had this possibility, it would work better particularly in the dynamic model if the series of measures were not influenced by structural elements but could be directly related to changes in reproductive behavior.

The possibility of translating the hypotheses of the model's parameters in a bi-univocal manner into hypotheses on the reproductive process and on cohort behavior makes it possible to interpret the *input* of the models and to clearly formulate forecast hypotheses; once again, models with variables that are essentially reducible to the reproductive process are the most promising.

In conclusion, it would appear that, in purely demographic terms, the most frequently applicable methods, disregarding the availability of raw data, are the Martinelle model and the probability model (Giorgi, 1993).

In the case of desired fertility, the logical frame of interpretation that associates it to individual reproductive behavior can be improved, particularly by identifying the obstacles and constraints (collective and individual) in a theoretical-behavioral, and not exclusively mechanical, context that can stop desired fertility from being fully realized.

It is more difficult to give indications on the direction to follow in order to improve the capacity to produce probable scenarios. The transition model would seem to be a useful structural reference, within which we can place theoretical contributions that explain each partial aspect in the context of a more general and comprehensive explanation of fertility trends. This would help to reduce the present mechanical character of the transition model. However, it seems impossible to obliterate the difficulty of explaining the changes in fertility by the action of the social and economic factors to which fertility determinants are attributed.

For interpretative theories to be useful in producing probable scenarios, they must be intrinsically dynamic; that is, they must be capable of taking into account the association between the past and the present in order to project into the future. These theories, particularly economic theories, are essentially static, with the exception of those that use the concept of life cycle in the wake of Easterlin's macro theory. The possibility of creating probable scenarios depends largely on the ability to produce explanatory theories with a high dynamic content.

In any case, it is indispensable to avoid the intellectual isolation of forecasters who limit themselves to the statistical extrapolation of the past to formulate long-term hypotheses. One can also wish for a greater sensitivity among those who deal with the interpretation of reproductive behavior in their comparisons for the aims of forecast, as these aims require dynamic models and possibly univocal management.

Acknowledgment

This chapter was translated from the French version by Accenta, Ltd.

References

Ahlburg Dennis A., 1986. A relative cohort size forecasting model of Canadian total live births, *Social Biology*, vol. 33(1–2), p. 51–56.

Al-Osh Mohamed, 1986. Birth forecasting based on birth order probabilities with application to U.S. data, *Journal of the American Statistical Association*, vol. 81(395), p. 645–656.

Bozik J. James E., and Bell William R., 1987. Forecasting age specific fertility using principal components, *Proceedings of the Social Statistics Section, American Statistical Association*, p. 396–401.

Ciucci Luciano, Gambini Antonio, and Keyfitz Nathan, 1984. *Expectations and auxiliary variables in fertility forecasts*. London, WFS, 58 p.

de Beer Joop, 1985. A time series model for cohort data, *Journal of the American Statistical Association*, vol. 80(391), p. 525–530.

de Beer Joop, 1991. From birth expectations to birth forecasts: a partial-adjustment approach, *Mathematical Population Studies*, vol. 3(2), p. 127–144.

de Beer Joop, 1992. Methods of fertility projections and forecasts, *in:* Nico Keilman and Harri Cruijsen (eds.), *National population forecasting in industrialized countries*, p. 27–48. Amsterdam, Swets & Zeitlinser, 365 p.

de Jong Andries, 1995. *Long-term fertility scenarios for the countries of European economic area*. Voorburg, CBS, 43 p. (Paper presented at the International Seminar on "New long term population scenarios for the European economic area," Luxemburg, November 8–10, 1995).

Easterlin Richard A., 1976. The conflict between aspirations and resources, *Population and Development Review*, vol. 2(3), p. 417–426.

Ermisch John F., 1992. Explanatory models for fertility projections and forecasts, *in:* Nico Keilman and Harri Cruijsen (eds.), *National population forecasting in industrialized countries*, p. 201–222. Amsterdam, Swets & Zeitlinser, 365 p.

Feeney Griffith, 1985. Parity progression projection, *in:* IUSSP (ed.), *International Population Conference, Florence, 1985*, vol. 4, p. 125–136. Liège, Ordina éditions, 556 p.

Giorgi Piero, 1993a. *Previsioni di fecondità che considerano l'ordine di nascita: aspetti metodologici ed operativi. Analisi comparativa con altri metodi di previsione della fecondità*. Rome, DSD, 141 p. (Doctoral thesis).

Giorgi Piero, 1993b. Una rilettura della fecondità del momento per ordine di nascita in Italia nel periodo 1950–1990, *Genus*, vol. 49(3–4), p. 177–204.

Giorgi Piero and Viola Giulia, 2003, Il completamento della fecondità delle coorti: un modello basato sull'analisi delle serie storiche, *Quaderni di Statistica*, vol. 5, p. 189–214.

Hendershot Gerry E. and Placek Paul J. (ed.), 1981. *Predicting fertility*. Lexington (MA), Lexington Books, 335 p.

Keilman Nico, 1994. Translation formulae for non-repeatable events, *Population Studies*, vol. 48(2), p. 341–358.

Keilman Nico and Cruijsen Harri (eds.), 1992. *National population forecasting in industrialized countries*. The Hague, Swets en Zeitlinger, 365 p.

Keilman Nico and van Imhoff Evert, 1995. Cohort quantum as a function of time-dependent period quantum for non-repeatable events, *Population Studies*, vol. 49(2), p. 347–352.

Keyfitz Nathan, 1972. On future population, *Journal of the American Statistical Association*, vol. 67(338), p. 347–363.

Keyfitz Nathan, 1981. The limits of population forecasting, *Population and development review* vol. 7(4), p. 579–593.

Keyfitz Nathan, 1982. Can knowledge improve forecasts? *Population and development review,* vol. 8(4), p. 729–751.

Keyfitz Nathan, 1985. Demography in the twenty-first century: the uses of forecasting, *in:* IUSSP (ed.), *International Population Conference, Florence, 1985,* vol. 1, p. 59–82. Liège, Ordina éditions, 230 p.

Land Kenneth C., 1986. Method for national population forecasts: a review, *Journal of the American Statistical Association,* vol. 81(347), p. 888–901.

Lee Ronald D., 1974. Forecasting births in post-transition populations: stochastic renewal with serially correlated fertility, *Journal of the American Statistical Association,* vol. 69(347), p. 607–617.

Lee Ronald D., 1979. New methods for forecasting fertility: an overview, *in:* United Nations (ed.), *Prospects of population: methodology and assumptions,* p. 215–225. New York, United Nations, 300 p.

Lee Ronald D., 1980. Aiming at a moving target: period fertility and changing reproductive goals, *Population Studies,* vol. 34(2), p. 205–226.

Long John F., 1985. *The utility of demographic-economic models in projecting state and major metropolitan area populations in the United States.* Liège, IUSSP (Paper presented at the International Population Conference, Florence, 1985).

Lutz Wolfgang, 1991. *Future demographic trends in Europe and North America.* New York, Academic Press, 585 p.

Martinelle Sten, 1989. A Cohort model for analysing and projecting fertility by birth order, *in:* IUSSP (ed.), *International Population Conference, New Delhi, 1989,* vol. 1, p. 315–329. Liège, Ordina éditions, 462 p.

Martinelle Sten, 1993. The timing of first birth. Analysis and prediction of Swedish birth rates. *European Journal of Population/Revue européenne de démographie,* vol. 9(3), p. 265–286.

Rallu Jean-Louis and Toulemon Laurent, 1993a. La mesure de la fécondité transversale, I: construction des différents indices, *Population,* vol. 48(1), p. 7–26.

Sardon Jean-Paul, 1995. Les prévision de fécondité, *in:* Christine Wattelar and Josianne Duchêne (ed.), *Le défi de l'incertitude. Nouvelles approches en perspectives et prospectives démographiques,* p. 173–191. Louvain-la-Neuve, Academia-Bruylant; Paris, L'Harmattan, 436 p.

Thompson Patrick A., Bell William, Long John F., and Miller Robert B., 1989. Multivariate time series projection of parameterized age-specific fertility rates, *Journal of the American Statistical Association,* vol. 84(407), p. 689–699.

van de Giessen Hans, 1992. Using birth expectations information in national population forecasts, *in:* Nico Keilman and Harri Cruijsen (eds.), *National population forecasting in industrialized countries,* p. 223–241. Amsterdam, Swets & Zeitlinser, 365 p.

van Hoorn Willem and Keilman Nico, 1995. *Birth expectations and their use in fertility forecasting.* Luxemburg, Eurostat, 39 p.

Westoff Charles F., 1981. The validity of birth intentions: evidence from US longitudinal studies, *in:* Gerry Hendershot and Paul Placek (eds.), Predicting fertility : demographic studies of birth expectations, p. 51–59. Lexington (MA), Lexington Books, 335 p.

Willekens Frans J., 1990. Demographic forecasting; state-of-the art and research needs, *in:* Cornelius A. Hazeu and Gerard A.B. Frinking (eds.), *Emerging issues in demographic research,* p. 9–66. Amsterdam, Elsevier Science Publishers, 314 p.

Willekens Frans and Baydar Nazli, 1984. *Age-period-cohort models for forecasting fertility.* Voorburg, NIDI, 79 p.

C H A P T E R

75

Mortality Forecasts

Hypotheses and Methods

GRAZIELLA CASELLI

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION

Because demographic phenomena respond to and reflect long-term changes in society and the economy, forecasts of future demographic behavior can usually be based on past behavior. This particularly applies to mortality forecasts, even more so for countries for which long series of data disaggregated by sex and age are available.

Mortality forecasts often hypothesize gradual convergence of rates toward a specified target that will be attained by the end of the forecast period. The target usually coincides with the mortality rate observed for another population, or a subpopulation, taken to be further along with regard to health achievements (Olshansky, 1988). The target could also be mortality rates by age with level and structure that describe ideal mortality schedules that the population could expect to achieve. For example, 1994 United Nations (UN, 1995) forecasts were based on the assumption that countries would achieve the same life expectancy levels worldwide by 2020–2025: the rate of improvement for the most advanced countries would slow while those farthest behind would rapidly catch up (for UN forecasts, see Chapter 77). As opposed to other approaches, targeting is often applied to populations for which the data are incomplete or not very reliable, and recourse is often made to standard life tables or to those for more developed countries (van Poppel and de Beer, 1996).

It could be said that targeting leaves more scope for the imagination whereas other approaches are primarily tied to the past. The approach chosen largely depends on the statistical data available.

This chapter looks at the various approaches, indicating the hypotheses and methods most widely used in demography. Ex post data comparisons will be used to assess the ability and potential of each method to affect forecast results.

I. FROM SIMPLE TO MORE COMPLEX MODELS

This section covers methods based on extrapolations of the past. When a forecast is based on the hypothesis that past trends will continue into the future, it is crucial to define not only the method used to estimate trends but also the length of the data series involved and the year selected to start the forecast. These choices greatly influence the final outcome (Beaumont and Isserman, 1987), so much so that they overshadow even the efficacy of more sophisticated methods. Particularly in recent times when elderly mortality has declined both considerably and unexpectedly, the choice of whether data for the past decade or the past 2 or 3 decades is used, or of whether the starting year is the current year are not arbitrary. Although at a theoretical level, these observations may appear obvious, empirical verification is always needed to assess their effect. For this reason, after a brief introduction of the most widely used forecast methods, later sections focus on applications that clarify the various methods and hypotheses used in estimating future mortality trends.

There are numerous examples of mortality forecasts based on extrapolation methods, from simple graphic extrapolations to analytical extrapolations of an indicator or a mathematical equation modeling the survivorship of a population. What is important is that the selected indicators have an easily recognized underlying trend; that is, they are not subject to strong perturbations and at the same time are sensitive to variations in observed mortality levels. The choice of indicator is usually either the probability of death or age- and sex-specific mortality rates.

At its simplest, a *graphic extrapolation* plots the data by time with the aim of identifying trends. This type of extrapolation often suffices because, although it may appear to be subjective and nonscientific, the results often do not differ very much from those obtained with more objective, complex mathematical approaches.

Analytical extrapolations project trends for a demographic indicator by using a mathematical equation that describes changes over time (McNown *et al.*, 1995). The equation can incorporate the dynamics of one or more dimensions of the chosen indicator: that is, age (A), period (P), or cohort (C).

Among the equations used, John Pollard's (1988) is the most frequently applied and models an exponential drop in mortality in time for each age (with varying P):

$$q_{n,x} = b_x(c_x)^n, \tag{1}$$

where

$q_{x,n}$ is the probability of death at age x in year n,

b_x is mortality at age x at a particular moment in time, and

c_x (ranging between zero and one) expresses the decline over time of mortality at age x.

By applying the logarithmic transformation of Eq. 1 we have

$$\ln(q_{n,x}) = \ln(b_x) + n\ln(c_x). \tag{2}$$

If we wish to fix a lower limit a_x for the mortality decline at age x, we use

$$q_{n,x} = a_x + b_x(c_x)^n. \tag{3}$$

Once representative ages of the mortality model by age are chosen, the corresponding values c_x for different calendar years is estimated with the ratio $(q_{n,x}/q_{m,x})^{1/(n-m)}$, (where $n > m$); then the values of c_x are calculated to obtain those forecasts. The results, applied to one of the three formulas above, yield the probability of death at the selected age. By interpolation, q_x for ages in between can be obtained.

During the late 1970s and 1980s, *parametric models* were popular in making mortality forecasts. Over the past century, mortality experts have tried to infer mathematical laws through a combination of a certain number of parameters (where A varies, fixed P) from regularities in mortality by age. There is, for example, Gompertz's law (1825), named after the mathematician and later modified by William Makeham (1867) and Wilfrid Perks (1932). The main defect of these functions is that they only model mortality beyond 30 to 35 years of age, that is, to a mortality curve that increases with age and thus not to the mortality curve for infancy and youth. Larry Heligman and John Pollard overcame this drawback in 1980 with a three-component model. Their equations, with eight parameters, are

$$q_x/(1-q_x) = A^{(x+B)c} + D\exp\{-E[\ln(x/F)]^2\} + GH^x \tag{4}$$

or, alternatively,

$$q_x = A^{(x+B)c} + D\exp[-E(\ln x - \ln F)^2] + GH^x/(1+GH)^x] \tag{5}$$

The first component of each function models infant mortality, the second models youth mortality from violent causes, and the third follows mortality trends among the elderly.

Thus, parametric models gradually adapt to the age-specific mortality curve as the number of parameters to be estimated increases. Although ensuring greater similarity between theoretical and observed values, this has generated such complex analytical functions that their use has been limited. This helps explain the success enjoyed by the two-parameter Brass model (1974)—particularly in countries where data are lacking—which uses the logit of the survival functions of a life table:

$$L_{x,n} = 0{,}5\ln[(l_{0,n} - l_{x,n})/l_{x,n}]^s. \tag{6}$$

This models the logits of a given life table as a linear function of the logits of a standard or reference life table, or

$$L_x = a + bL_x^{(s)}, \tag{7}$$

whose parameters a and b are independent of age x (where P varies, fixed A). Its application thus depends of the availability of sufficiently lengthy series of the probability of death, so that the survivorship of a cohort may be reconstructed. If this is possible, survivors for certain selected ages x of the cohort born in the calendar year n can then be calculated and the corresponding logit can be linearly extrapolated. The values obtained, applied to the inverse of the logit transformation—that is, $l_{x,n} = l_{0,n}/[l + \exp(2L_{x,n})]$—yields the estimated survivorship ($l_{x,n}$). Where cohort data are missing, a frequent occurrence, period life tables are used. In their absence, as happens in developing countries, standard UN or Ansley Coale and Paul Demeny tables (1966, 1983) are employed, using subsequent series through a longitudinal reading of mortality at different ages.

In summary, the application of parametric models for forecasts comprises three steps: an estimation of the parameters, an extrapolation of the relative trend over time, and the subsequent integration of estimated values of the parameters into the original mathematical equation. Parameters are estimated by fitting to the data, often by least-squares methods, by chi-squared methods, or by examining the maximum likelihood function (Wilmoth, 1983). The corresponding values forecast for a given year are attained by substituting these values into the model so that mortality at each age can be calculated.

Among the many drawbacks arising from the application of parametric models, we focus on two. First, there is a possible inconsistency of the results obtained by extrapolating the parameters independently of each other. For example, when fitted over age for a fixed period, it should be verified that for every cohort that the number of survivors at age x is greater than at age $x + 1$. Second, it may be difficult to identify the trend of parameters characterized by large variability, so that the choice between the many possible solutions becomes subjective.

Since the 1980s, demographic forecasts have turned with increasing interest to methods based on the analysis of economic time series.[1] The simplest of these is a deterministic linear model, where the dependent variable is mortality in age group x (expressed by rates or logarithms of rates), and the independent variable is time t:

$$X_t = a + bt. \tag{8}$$

A simple stochastic model is

$$DX_t = DX_{t-1} + e_t \tag{9}$$

where the change in X_t (i.e., DX_t) depends both on variations in previous time periods (DX_{t-1}) and on a random term (e_t). This formulation is included within a wide class of models called ARIMA (autoregressive integrated moving average) models. McNown and Rogers (1988) experimented a method combining parametric models with time series analysis. In practice, an ARIMA model is often used to extrapolate from "pure" George Box and Gwilym Jenkins (1970) time series methods, which are theoretic and make the introduction of additional substantive information difficult. Robert McNown and Andrei Rogers applied their model to mortality data for the United States by performing an ex post forecast: based on the historical series for the period 1941–1970, they checked whether forecasts matched observed data for 1971–1983. Although the results obtained were not completely satisfactory, this study still represented a stepping stone in the potential offered by ARIMA models in interpreting and making mortality forecasts.

However, ARIMA models are frequently unable to capture changes in the latter years of the observation period, so during the extrapolation step they have less influence than do overall longer-term trends in the time series. This is a major drawback in demographic research and particularly in the analysis of mortality in the near past, which has been distinguished by a sudden and marked decline among the elderly. However, as all variations in mortality must be accounted for—and, in particular, those in the last years of the reference period—a mathematical function is needed that can interpret these and provide the best indications of possible future change. Numerous attempts have been tried, particularly in the economic literature, where the principal time series techniques have been developed. The most recent of these are based on linear regressions. Simple linear relationships do model mortality well as long as there are no underlying variations in the trend. If not, which often occurs when studying mortality for a given cause of death, more complex mathematical functions are needed. Thus, although still dealing with time series, the probabilistic component has been replaced by a deterministic model. The data are interpreted by models like Eq. 8, where the dependent variable is mortality at age x (expressed by logarithms of mortality rates) and the independent variable is time t. The linear relationship is denoted by

$$\ln(m_{x,t}) = a_x + b_{x,}t \tag{10}$$

or by the quadratic relationship:

$$\ln(m_{x,}t) = c_x + b_{x,t} + c_{x,t}{}^2 \tag{11}$$

By using least-squares methods, coefficients can be estimated and the extrapolation begun. As far as forecasts are concerned, however, the disadvantage of these functions is that they fail to impose limits on the time trend, so these are only valid in short- or medium-term forecasts.

Among *stochastic models*, Ronald Lee and Lawrence Carter (1992; Lee, 1998) proposed what is today the

[1] In the standard definition for an historical time series (X_t, $t = 1, 2, \ldots, N$), what is intended is the finite realization of a stochastic process X_t, where t is time, that describes the trend.

most well-known and frequently used method. Their analysis focused on stochastic variations within a single model of mortality change. The procedure "combines a rich yet parsimonious demographic model with statistical time series methods" and relies on a simple model of past mortality trends. Their model is

$$\ln(m_{x,t}) = a_x + b_x k_t + (e_{x,t}), \tag{12}$$

where

$m_{x,t}$ is the mortality rate at age x for time t;

a_x, b_x, and k_t are the parameters of the model; and

$e_{x,t}$ is an error term, with expected value $E(e_{x,t}) = 0$ and variance $V(e_{x,t})$. Although Lee and Carter fit their model by ordinary least squares, other methods can be used (Lett *et al*., 1995, 1998). The interpretation of the parameters is quite simple: the fitted values of a_x are exactly equal to the average of $\ln(m_{x,t})$ at age x over time, b_x represents the age-specific pattern of mortality change, and k_t represents the time trend (Wilmoth, 1996). The advantage of mortality trends obtained by using this model is that it provides a confidence interval range within which the results can vary.

The Lee-Carter method can be expanded to include age and period, by adding additional parameters to be estimated (Wilmoth, 1993). Gomez de Leon (1990) tackled these aspects in an exploratory analysis of Norwegian mortality data for 1846–1988.

The Lee-Carter method presumes that mortality variations are a function of the components of age and period (where A and P vary); however, particularly when forecasts are made for cause of death, it may be of interest to consider the third component of mortality, cohort (C). In such a case the model becomes increasingly complex but formally more accurate in analyzing cause-specific mortality trends in which variations are strictly linked to changes in generalized behavior patterns.

Age-period-cohort (APC) models have been widely applied to past mortality time series (Osmond and Gardner, 1982; Hobcraft *et al*., 1982; Caselli and Capocaccia, 1989; Wilmoth *et al*., 1989), but they are not frequently used in forecasts (Caselli, 1993, 1996; Burgio and Frova, 1995), and even in those cases, they are usually used only to forecast cause-specific mortality (Osmond, 1985). In a recent article, Alessandra Burgio and Luisa Frova (1995) hypothesized that the logarithms of mortality rates can be interpolated by using a polynomial function of the age, period, and cohort covariates, based on the fact that the mortality rate can generally be expressed as a smooth function of factors and parameters. Their model is

$$\ln(y^*_{t,x}) = a + a(x) + p(t) + c(t - x) \tag{13}$$

or

$$\ln\left(y^*_{t,x}\right) = a + \sum_i b_i x^i + \sum_j c_j t^j + \sum_k d_k (t-x)^k, \tag{14}$$

with $i = 1, \ldots, h_1$; $j = 1, \ldots, h_2$; and $k = 1, \ldots, h_3$,

where $y^*_{x,t}$ is the theoretical mortality rate at age x and time t, and $b_1, \ldots, b_{h1}, c_1, \ldots, c_{h2}, d_1, \ldots, d_{h3}$ are the parameters estimated with the least-squares method.

This model may be acceptable if used to describe past mortality trends. For projections, however, the use of a period effect curve may be inadequate owing to the volatility of short-term factors. Thus, the period curve is subdivided into two additional contributions: the trend underlying the period component is given by the straight line uniting the first and last year of the observation, and deviations from this average trend. An additional hypothesis supposes that the underlying trend holds for the future with no fluctuations from this trend. This is possible by transforming the period function $p(t) = c_0 + c_1 t + \ldots + c_{h2} t^{h2}$ so that the new function $p^*(t)$ results as $p^*(t_1) = p^*(t_n)$, where t_1 and t_n are, respectively, the first and last year of observation. Subsequently, as long as the mortality rate remains unchanged, age and cohort functions all undergo a comparable transformation. By these transformations, it is possible both to identify the model (Caselli and Capocaccia, 1989) and to control the period effects.

II. TARGET FORECASTS

Unlike an extrapolation approach, targeting allows for a larger number of hypotheses about future mortality. As a method it is more flexible; however, it tends to favor the influence of subjective elements, which could compromise the results. Moreover, once the end target has been set, an additional problem arises: how to get there from here (Keyfitz, 1981; Alho and Spencer, 1985). This will depend in part on the completeness of mortality data. Where data are missing, model tables are used or reference is made to those for other countries. When sufficiently detailed data are available, an approximation to the final value can be obtained by means of ad hoc functions that have been constructed based on the information at hand.

An approximation to targets using model tables carries the implication that the system of tables can be adapted to fit the mortality structure. A table having the value of the desired parameter (e.g., infant mortality or life expectancy at birth) is chosen that is suited to representing the new mortality model and is thus able to provide the targeted probabilities of death. Additional hypotheses are then made, or interpolation methods are employed, which determine the other values of the parameter so that the remaining model tables will be representative of different phases in future mortality trends.

A similar approach may be used for mortality forecasts for developed countries. To illustrate this point, let us recall the health transition in Europe: overall mortality patterns, especially for causes of death, were similar for the majority of countries, although with individually varying time schedules and some variations in pattern (Caselli, 1993, 1994): some countries forge ahead while others lag behind, experiencing the same phenomena but with a certain phase displacement.

A good example is provided by mortality trends for lung cancer in adult men (although the same applies to other causes of death). Incidence has tended to increase with time, and until quite recently, it was thought impossible to bring this trend to a halt. Following major antismoking campaigns, dating from the 1950s in England and Wales, male mortality from lung cancer began to decline. At the same time, lung cancer mortality in other Northern and Central European countries continued to rise, a trend that was not reversed or halted until the late 1980s. Naïve trend extrapolation for these areas might not take into account more recent trends elsewhere, although it might be reasonable to expect the England/Wales experience to be replicated elsewhere but with a time displacement.

A crucial aspect of such an approach is assessing the potential for phase displacement, such that it would be possible to link mortality trends across countries. Naturally, both the theoretical and operational problems are numerous and would necessitate solutions specific to each country, for each cause and for each age.

Less sophisticated approaches would include those in which hypothesized mortality trends are extreme, the aim being to achieve ideal values expressed by an "optimal" mortality table. In this case one must establish how and when an approximation to the ideal table is to be achieved. For this type of forecast, Eq. 3 is generally used, where q_x represents the probability of death in the ideal table.

Targeting models also include *cause-delay models*, which foresee the elimination, either partial or total, of one or more causes of death in order to assess their impact on the population's survival levels (Olshansky, 1987, 1988). When formulating the hypotheses, the researcher attempts to identify intervenable risk factors so as to gradually abolish the cause or causes of death directly linked to these risks. It is then possible to express a particular cause of death in terms of cost in life expectancy. Cause-delay models are particularly desirable in long-term forecasts if it is possible to intervene on one or more risk factors.

III. DIFFERENT METHODS COMPARED

Figure 75–1 compares mortality rates for English males aged 60 to 84 years for the period 1971–1988; the forecasts are for the same period and are made by applying three different models to the series of mortality rates by age observed for the years 1950–1970. The

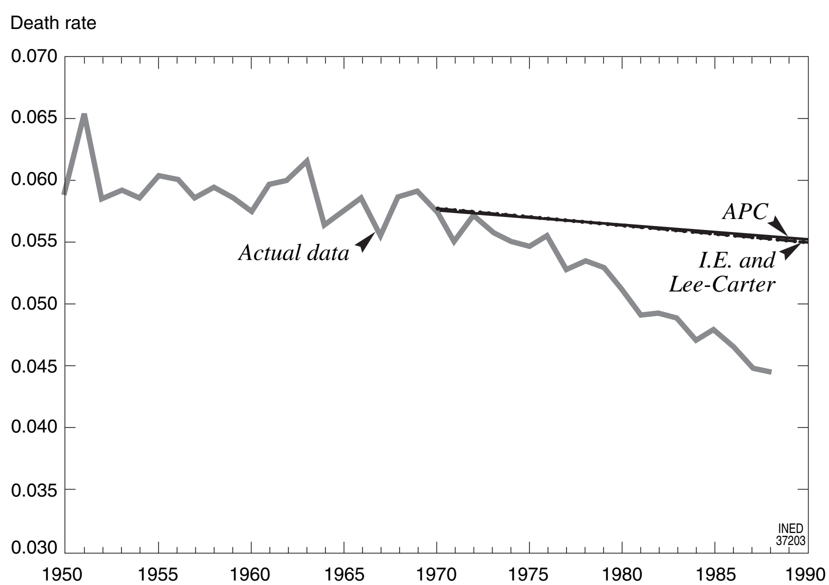


FIGURE 75–1 Extrapolations for 1971–1988 of trends for 1950–1970, according to three models, compared with actual trends. Direct extrapolation of total mortality rates (data on England and Wales, males).

three models are (1) interpolation and extrapolation as expressed by the Eq. 10 or 11, hereafter referred to as model I.E.; (2) "Lee and Carter"; and (3) APC. First, observe that when the reference data present an easily identified underlying trend, the use of any of these models, or even a simple graphical extrapolation, largely generates the same result. Second, note that none of the forecasts was able to capture the sizeable decline in mortality among these age groups from the 1970s onward. Could a mortality forecast, obtained from single forecasts of mortality for the major causes of death, be capable of it? We know that changes in trends after the 1970s, especially for the ages considered, are the outcome of modifications in mortality levels for certain causes of death, particularly cardiovascular diseases and lung cancers. As can be seen from Fig. 75–2, regardless of the model used, the outcome of forecasts for the five major groups of causes, although quite different from the levels observed in recent years, are without doubt better than anything else previously obtained (Caselli and Vallin, 1999).

In this case, the reason why the fit between the forecast and the actual was so poor must be traced back to certain cause-specific mortality trends, particularly car-

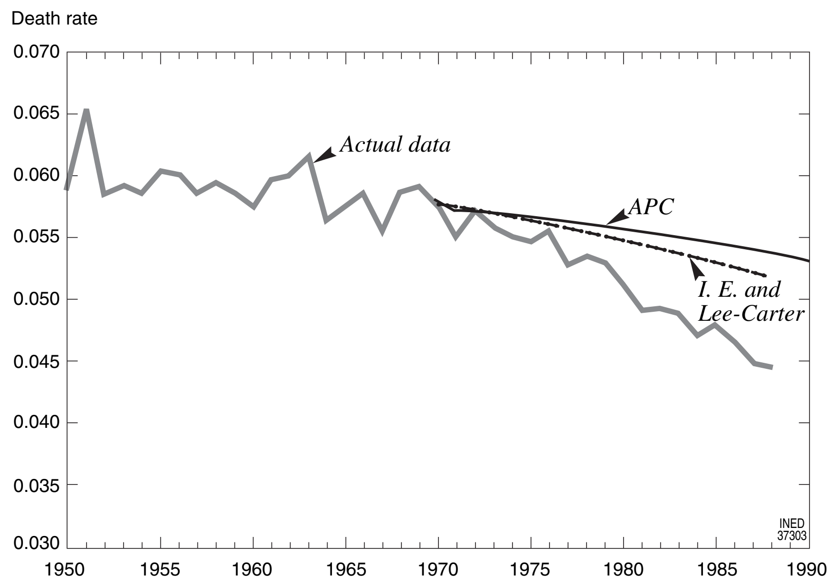


FIGURE 75–2 Extrapolations for 1971–1988 of trends for 1950–1970, according to three models, compared with actual trends. Sum of extrapolations of rates by cause (data on England and Wales, males).

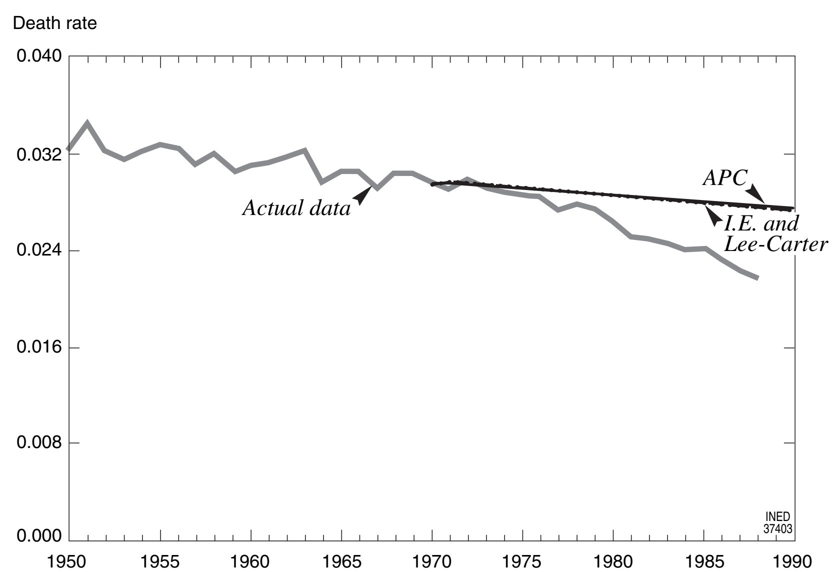


FIGURE 75–3 Extrapolations for 1971–1988 of trends for 1950–1970, for mortality by cardiovascular diseases according to three models, compared with actual trends (data on England and Wales, males).

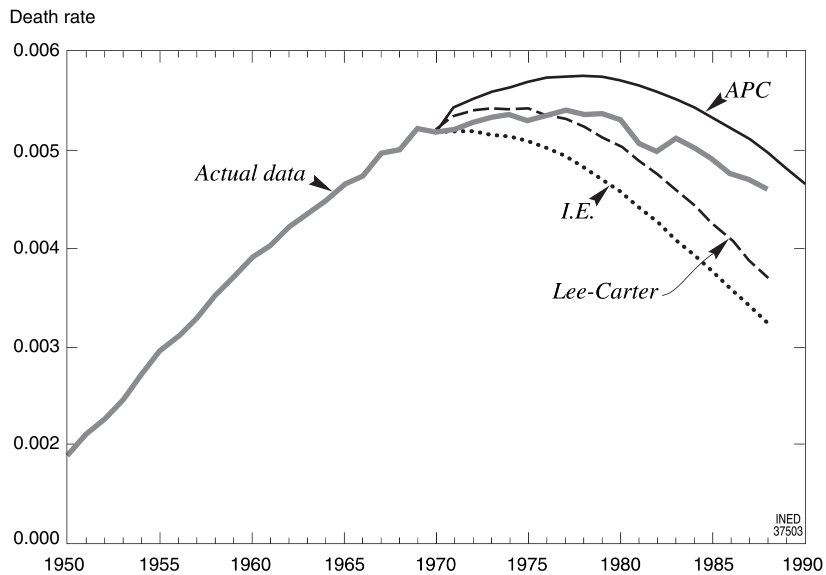


FIGURE 75–4 Extrapolations for 1971–1988 of trends for 1950–1970, for mortality by lung cancer, according to different models, compared with actual trends: Males, 60–84 years (data on England and Wales).

diovascular diseases (Fig. 75–3). These trends have the same form as that for total mortality, and no model was able to isolate it. The quick pace of this decline was unexpected and coincided with major therapeutic advances in the early 1970s. Thus, the "period" effects had an important role to play on trends in mortality rates.

The decline in lung cancer mortality was much more gradual and was lagged in time as the age of the population increased. First, this involved generations who throughout their adult lives could benefit from antismoking campaigns, either moderating or avoiding smoking altogether (Fig. 75–4). It was only later that other generations, who had been smokers and thus exposed for a lengthy period, began to be involved. Mortality trends by age for this cause clearly illustrate this process (Fig. 75–5). It is interesting to see that, at least with regard to lung cancer, the APC model offers the best insight into mortality trends because this approach, at the extrapolation stage, emphasizes the importance of the cohort component in mortality trends.

IV. HYPOTHESES FOR MEDIUM AND LONG-TERM FORECASTS

We now turn to critical hypotheses, the baseline data series, and the starting date of the forecasts. Figure 75–6 compares the results of forecasts made for English males aged 60 to 84 years between 1989 and 2050 based on two different data series: 1950–1988 and 1970–1988.

As may be expected, mortality levels to 2050 are fairly similar for each of the models used when based on the same reference period. Perceptible differences arise, however, depending on the length of the data series: mortality levels are lower when based on 1970–1988 data, coinciding with the date when a marked decline in mortality for these ages first began.

As further proof, and to focus on the importance of the starting date set for forecasts, we present a summary of forecasts made in the late 1980s in certain European countries (Table 75–1). First, let us consider the forecasts for the 1980s by national statistics bureaus, beginning with a comparison of the forecasts made by France's National Institute for Statistics and Economic Studies (INSEE) and Italy's National Institute of Statistics (ISTAT). Both these bodies adopted a "trend" hypothesis of changes in the mortality rate. ISTAT, which made its forecasts in 1989, took as the point of departure the year 1988, whereas INSEE began with 1985. This difference of only a few years, together with the different lengths of the baseline series (in France, 1950–1985; in Italy, 1980–1988), meant that the French forecast could not fully reflect recent mortality trends, and thus, their results were more pessimistic (Fig. 75–7).

Figure 75–7 helps to provide a focus on the problem of what baseline period to choose when making a forecast, by providing a global overview of mortality trends from 1900–2020; the effect of different assumptions on the results can be observed. For the middle age groups, French data show a steep decline in rates for adults and the elderly, although it is not even

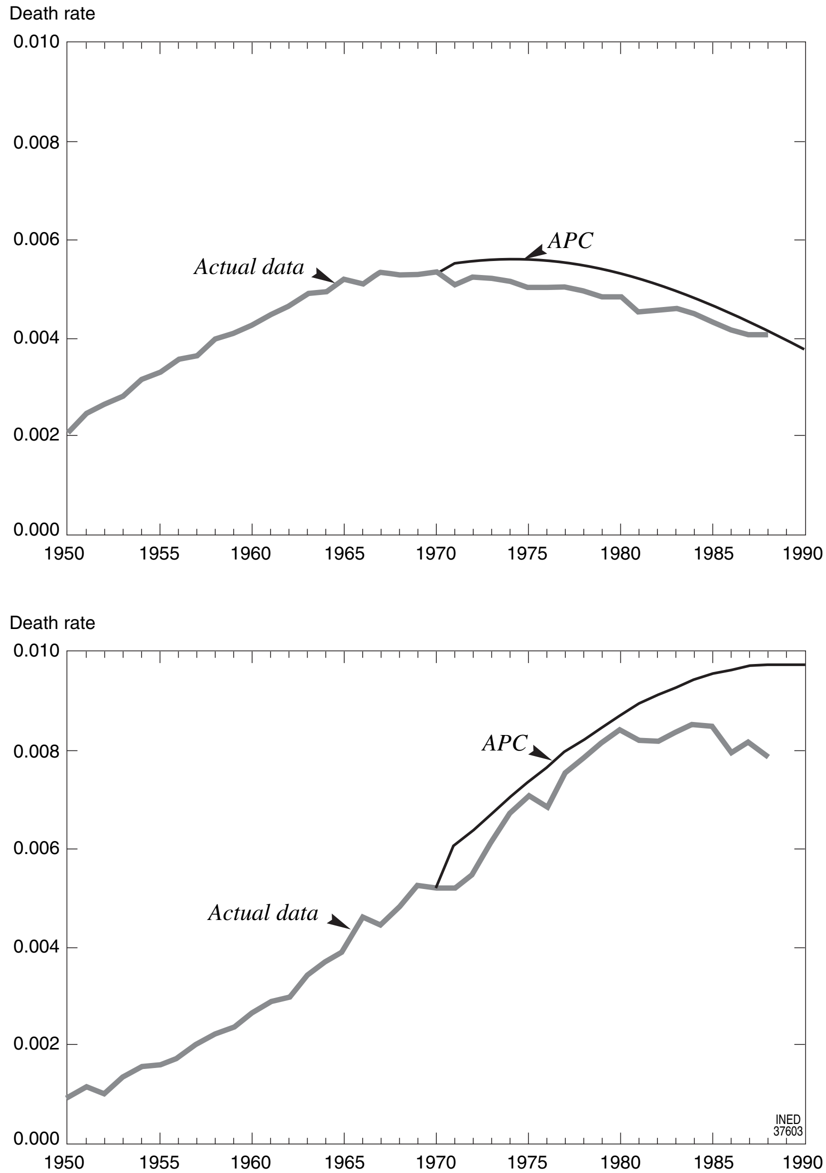


FIGURE 75–5 Extrapolations for 1971–1988 of trends for 1950–1970, for mortality by lung cancer, according to different models, compared with actual trends: Males, 65–69 and 80–84 years (data on England and Wales).

slightly apparent in the Italian forecast (Caselli, 1993–1994). An ex post assessment of these results is that the elderly mortality drop after the 1980s was not foreseen for either of the two countries. Moreover, this drop has recently involved the upper age limits, for both men and women (Vallin and Meslé, 2001).

Thus, it is not surprising that each country has entirely different life expectancies for the year 2020: 74.0 years for French men compared with 76.3 years for Italian men (the latter had an advantage of 1 year in 1985), and 82.4 years for French women compared with 83.7 for Italian women (who had roughly a 6-month advantage with respect to French women in 1985). If we consider the fact that in 2001 French men already enjoyed an average life expectancy of 75.5 years, it is extremely pessimistic to believe that over the next 19 years 1.5 years of this would be lost. The same consideration applies to women, whose estimated life expectancy in the 1980s for 2020 is the same as that already achieved in 1999. Survival gains already achieved at the dawn of the 21st century in Italy and France are far removed from those estimated in the 1980s for the year 2020. These figures show that, although theoretically tractable, the extrapolation of

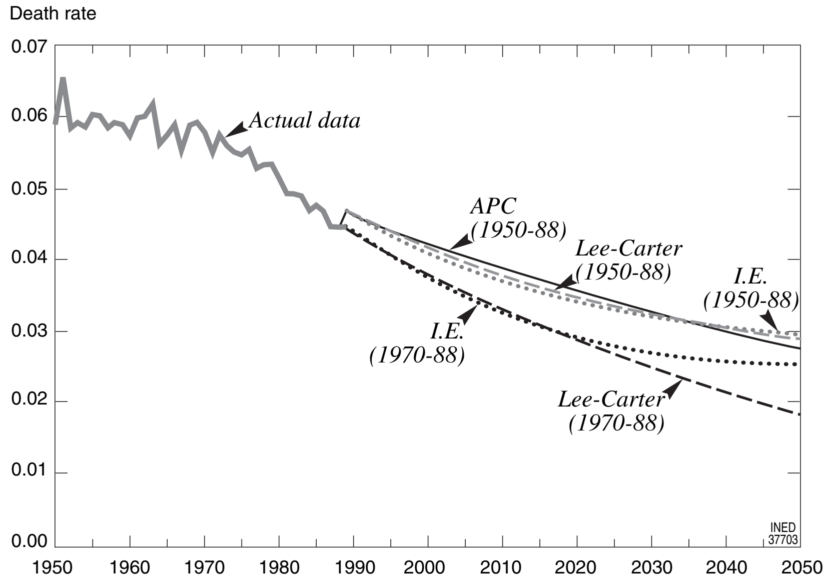


FIGURE 75–6 A comparison of forecasts of standard mortality rates for males aged 60–84 years, according to three models and different periods of reference (1950–1988 and 1971–1988) (data on England and Wales).

TABLE 75–1 Life expectancy at birth observed and forecasted for selected European countries

Countries	Men			Women		
	Observation (1) 1999	National statistics bureaus 2020 forecast		Observation (1) 1999	National statistics bureaus 2020 forecast	
		Observed in 1980–1988 (2)	Observed in 1992–2001 (3)		Observed in 1980–1988 (2)	Observed in 1992–2001 (3)
Austria	75.1	74.1	78.3	81.0	80.9	84.0
Belgium	74.4	74.3	77.1	80.8	81.8	83.6
Denmark	74.2	71.5	76.5	79.0	79.0	81.0
Finland	73.8	71.5	77.3	81.0	79.0	83.1
France	74.9	74.0	79.2	82.4	82.4	86.7
Germany	74.7	71.6[a]	[b]	80.7	78.6[a]	[b]
Italy	75.6	76.3	79.6	82.3	83.7	86.2
Luxemburg	74.7	74.1	[b]	81.2	80.9	[b]
Netherlands	75.3	75.0	78.0	80.5	81.5	81.7
Portugal	72.0	72.9[c]	75.4	79.1	80.0[c]	82.0
Spain	75.3	73.6	76.0	82.5	83.7	83.7
Sweden	77.1	72.4	79.6	81.9	78.9	83.8
United Kingdom	75.0	75.2	78.5	79.8	80.3	82.7

(1) Data from Eurostat, 2002; (2) data from national statistics bureaus, different countries; (3) data from Eurostat, 2001.
[a] RFA; [b] data missing; [c] in 2015.

past trends risks being disproved each time a sudden change occurs as was the case in the 1980s.

Apart from the prudent attitude generally shown by most European countries (Table 75–1) when the forecasts were made (some maintained that mortality levels were to remain constant, whereas others fixed a threshold which could not be surpassed), recent mortality trends have rendered obsolete all estimations made in the 1980s, as none of these were able to anticipate the present decline. In 1981 and 1988, respectively, the Finnish and Swedish statistical bureaus made forecasts to the year 2020. Finnish women would have had no further advance after 1987, the year in which they achieved survival levels foreseen for 2020

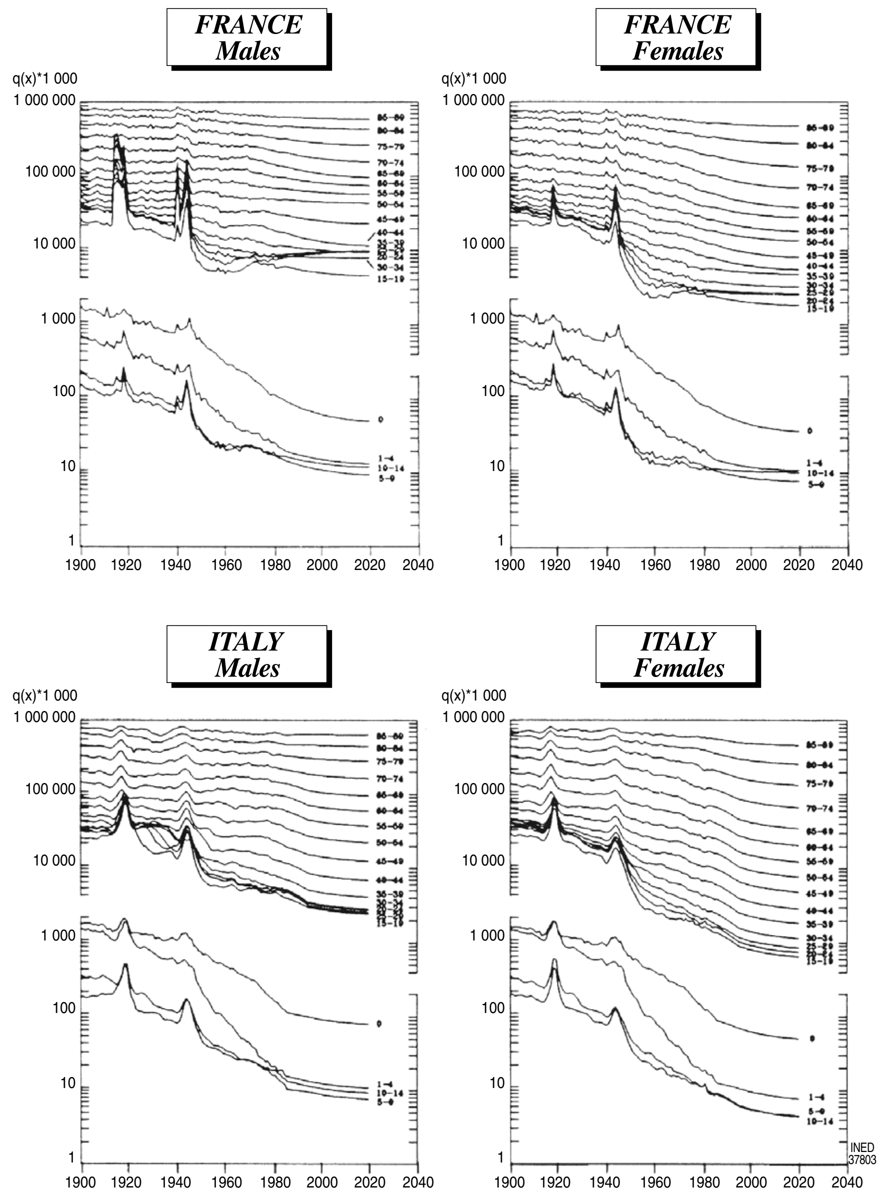


FIGURE 75–7 Changes in mortality rates for different age-groups, observed and forecast, from 1900–2020 in France and Italy.

(79 years), which was easily surpassed by 1999 with an 81-year life expectancy. Likewise, Swedish women were forecast to achieve a life expectancy of 78.9 by 2020; by 1999 they had achieved 81.9. Generally speaking, Table 75–1 shows that for most countries, the figures forecast for the year 2020 were already either reached or surpassed for both sexes by 1999.

In recent years forecasts of future European mortality levels across all ages have become more optimistic. This is the trend referred to by national statistics bureaus in their most recent estimations, such that they now proclaim a much higher life expectancy than was previously the case (Table 75–1), beginning in the 1980s and prolonging mortality trends observed in 1980s and 1990s. At a quick glance, there is about a 3-year difference in forecasted life expectancy to the year 2020 when using more current baseline data than when data from before the 1980s only is considered. The widest gap between both estimates is in Sweden, with a difference of 6.5 years for men and 5 years for women.

CONCLUSIONS: FACT OR FICTION?

At this stage it is impossible to say whether Eurostat forecasts for 2020 will in fact become a reality or whether, as in the past, they will be unable to capture latent trends that will emerge over the next few years. An approach that makes use of cause-specific mortality may be able to yield more satisfactory results because it uses more dynamic indicators. On the other hand, such an approach can generate quite volatile outcomes, as is seen in Fig. 75–2 for English males. In the early 1980s, Bernard Benjamin and Elizabeth Overton (1985) made forecasts for England and Wales, taking into account trends in mortality for eight major causes of death, which were much more optimistic than those contained in Table 75–1. They estimated a life expectancy of 81.3 and 87.1 years for males and females, respectively, for the year 2017, 3 and 4 years higher than foreseen in the 1990s for 2020.

The life expectancy values forecast by extrapolating the changes in cause-specific mortality might appear too high were it not for the comparable results obtained by Ward Kingkade (1994), who has calculated the lowest mortality figures ever observed in the developed countries in his life tables corresponding to the mortality rates based on age and major groups of causes of death. This table, although purely theoretical, has the characteristic of being based on the minimum values observed in practice of 81.6 years for men and 87.2 years for women, figures that come astonishingly close to those obtained for England and Wales by projecting recent trends in the mortality rate disaggregated by cause of death.

By altering the controlling hypothesis, forecasted results may be completely changed. Only the future can tell which will be the most accurate. Thus, no assumption should be taken lightly: even clear and well-established trends can change, as they did in the developed world in the 1980s. Experience has shown the arbitrariness of preimposing limits to mortality trends, so perhaps we should begin to seriously consider the hypothesis that in the long term the only acceptable limit is that mortality will all but disappear until the end stage of the life course, resulting in a full rectangularization of the survival curve. Nor should this goal be seen as too distant on the horizon as far as developing countries are concerned.

Many experts suggest that the hypothesized rectangularization of the mortality curve for the wealthier nations may be accompanied by even greater longevity at the tail-end of the human life cycle (Vaupel *et al*., 1979, 1998; Duchene and Wunsch, 1994; Valkonen, 1994; Caselli and Vallin, 2001a,b). These aspects of human survival are covered in Chapter 21 (Volume I); Chapter 78 covers long-term future trends. For the time being, the question belongs to the realm of fantasy and not of forecasting. At this stage it would be advisable for demographers to begin a discussion with colleagues in other fields, particularly biologists and epidemiologists, so as to reach a better understanding of extremely complex processes that could be increasingly influenced by new discoveries in the field of biogenetics (Walford, 1984; Fries, 1989; Dulbecco, 1995; Vaupel *et al*., 1998).

References

Alho Juha M. and Spencer Bruce D. 1985. Uncertain population forecasting, *Journal of the American Statistical Association*, vol. 80(390), p. 306–314.

Beaumont Paul M. and Isserman Andrew M. 1987. Comment, *Journal of the American Statistical Association*, vol. 82(400), p. 1004–1009.

Benjamin Bernard and Overton Elizabeth, 1985. Prospects for mortality decline in England and Wales. *Population Trends*, vol. 42(400), p. 22–29.

Box George E.P. and Jenkins Gwilym M., 1970. *Time series analysis: forecasting and control*. San Francisco, Holden-Day, 575 p.

Brass William, 1974. Perspectives in population prediction: illustrated by the statistics of England and Wales, *Journal of the Royal Statistical Society A*, vol. 137(4), p. 532–583.

Burgio Alessandra and Frova Luisa, 1995. Projections de mortalité-par cause de décès: extrapolation tendancielle ou modèle âge-période-cohorte. *Population*, vol. 50(4–5), p. 1031–1052.

Caselli Graziella, 1993. L'évolution à long terme de la mortalité en Europe, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European Population, II: Demographic Dynamics*, p. 111–164. Paris, INED and John Libbey, 526 p.

Caselli Graziella, 1994. *Long-Term Trends in European mortality*. Londres, OPCS, 32 p.

Caselli Graziella, 1996. Future longevity among the elderly, *in:* Graziella Caselli and Alan Lopez (ed.), *Health and Mortality among Elderly Populations*, p. 235–265. Oxford, Clarendon press, 330 p.

Caselli Graziella and Capocaccia Riccardo, 1989. Age, period, cohort, and early mortality: an analysis of adult mortality in Italy, *Population Studies*, vol. 43(1), p. 133–153.

Caselli Graziella and Vallin Jacques, 1999. Est-il utile de prendre en compte les causes de décès pour extrapoler les tendances de la mortalité?, *in:* Christine Wattelar and Josiane Duchêne (eds.), *Le Défi de l'Incertitude: Nouvelles Approches en Perspectives et Prospectives Démographiques*, p. 85–110. Louvain-la-Neuve, Institut de démographie de l'UCL, Academia-Bruylant, L'Harmattan, 436 p.

Caselli Graziella and Vallin Jacques, 2001a. Demographic trends: beyond the limits, *Population, An English Selection*, vol. 13(1), p. 41–72.

Caselli Graziella and Vallin Jacques, 2001b. Une démographie sans limite?, *Population*, vol. 51(special), p. 51–84.

Coale Ansley I. and Demeny Paul, 1966. *Regional Model Life Tables and Stable Population*. Princeton, University Press, 871 p.

Coale Ansley I., Demeny Paul, and Vaughan Barbara, 1983. *Regional Model Life Tables and Stable Population*, 2nd ed. New York, Academic Press, 496 p.

Duchêne Josianne and Wunsch Guillaume, 1994. Population aging and the limits to human life, *in:* Wolfgang Lutz (ed.), *Future Demographic Trends in Europe and North America: What Can We Assume Today?*, p. 27–40. London, Academic Press, 573 p.

Dulbecco Renato, 1995. *I Geni e il Nostro Futuro*. Milan, Sperling & Kupfer Editori, 219 p.

Eurostat, 2001. *Statistiques Sociales Européennes. Démographie*. Luxembourg, Eurostat, 175 p.

Eurostat, 2002. *Eurostat Yearbook: A Statistical Eye on Europe*. Luxembourg, Eurostat, 532 p.

Fries James F., 1989. The compression of morbidity: near or far? *The Milbank Quarterly,* vol. 67(2), 208 p.

Gomez de Leon Jose, 1990. *Empirical DEA Models To Fit and Project Time Series of Age-Specific Mortality Rates*. Unpublished manuscript of the Central Bureau of Statistics, Norway.

Gompertz Benjamin, 1825. On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Society,* no. 115, p. 513–585.

Heligman Larry and Pollard John H., 1980. The age pattern of mortality, *Journal of the Institute of Actuaries,* vol. 107(434), p. 49–80.

Hobcraft John, Menken Jane, and Preston Samuel H., 1982. Age, period, and cohort effects in demography: a review, *Population Index,* vol. 48(1), p. 4–43.

Keyfitz Nathan, 1981. The limits of population forecasting, *Population Development Review,* no. 7, p. 579–593.

Kingkade Ward W., 1994. Soviet mortality by cause of death: an analysis of years of potential life lost, *in:* Wolfang Lutz, Sergei Schermbov, and Andrei Volcov (eds.), *Trends and Patterns in the Soviet Union before 1991,* p. 339–358. London, Routledge, 496 p.

Lee Ronald, 1998. Probabilistic approaches to population forecasting, *in:* Wolfang Lutz, James W. Vaupel, Dennis A. Ahlburg (eds.), *Frontiers of Population Forecasting, no 4: Population and Development Review,* p. 156–190. 199 p.

Lee Ronald and Carter Lawrence, 1992. Modelling and forecasting U.S. mortality, *Journal of the American Statistical Association,* vol. 87(419), p. 659–675.

Lee Ronald, Carter Lawrence, and Tuljapurkar Shiripad, 1995. Desegregation in population forecasting: do we need it? And how to do it simply, *Mathematical Population Studies,* vol. 5(3), p. 217–257

Lutz Wolfang, Vaupel James W., and Ahlburg Dennis A., 1998. *Frontiers of Population Forecasting, no 4: Population and Development Review*. 199 p.

Makeham William M., 1867. On the law of mortality, *Journal of the Institute of Actuaries,* vol. 13, p. 325–358.

McNown Robert and Rogers Andrei, 1988. *Time Series Analysis and Forecasting of Parameterized Model Schedules: Mortality*. Boulder (CO), University of Colorado, Population Program Institute of Behavioural Science, 43 p.

McNown Robert, Rogers Andrei, and Little Jani, 1995. Simplicity and complexity in extrapolative population forecasting models, *Mathematical Population Studies,* vol. 5(3), p. 235–257.

Olshansky S. Jay, 1987. Simultaneous multiple cause-delay (SIMCAD): an epidemiological approach to projecting mortality, *Journal of Gerontology,* vol. 42(4), p. 358–365.

Olshansky S. Jay, 1988. On forecasting mortality, *The Milbank Quarterly,* vol. 66(3), p. 482–530.

Osmond Clive, 1985. Using age, period, and cohort models to estimate future mortality rates, *International Journal of Epidemiology,* vol. 14(1), p. 124–129.

Osmond Clive and Gardner Martin, 1982. Age, period, and cohort models applied to cancer mortality rates, *Statistics in Medicine,* vol. 1(245), p. 259.

Perks Wilfred, 1932, On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries,* no. 63, p. 12–57.

Pollard John H., 1988. Projection of age-specific mortality rates, *Population Bulletin of the United Nations,* no. 21–22, p. 55–69.

Rogers Andrei, 1995. Population forecasting: do simple models outperform complex models?, *Mathematical Population Studies,* vol. 5(3), p. 187–202.

United Nations [UN], 1995. *World Population Prospects. The 1994 Revision*. New York, Nations unies, Division de la population, 886 p.

Valkonen Tapani, 1994. Assumptions about mortality trend in industrialized countries, *in:* Wolfgang Lutz (ed.), *Future Demographic Trends in Europe and North America. What Can We Assume Today?,* p. 3–26. London, Academic Press, 573 p.

Vallin Jacques and Meslé France, 2001. *Tables de Mortalité Françaises pour les XIXe et XXe Siècles et Projections pour le XXIe siècle*. Paris, INED, 102 p. + CD-rom.

Vaupel James W., Carey James R., Christensen Kaar, Johnson Thomas E., Yashin Anatoli I., Holm Niels V., Iachine Ivan A., Kannistö Vaïno, Khanzaeli Aziz A., Liedo Pablo, Longo Valter D., Zeng Yi, Manton Kenneth J. and Curtsinger James W., 1998. Biodemographic trajectories of longevity, *Science,* no. 280, p. 855–860.

Vaupel James W., Manton Kenneth J., and Stallard Eric, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography,* vol. 16(3), p. 439–454.

Van Poppel Frans and de Beer Joop, 1996. Evolution of standard mortality projections for the elderly, *in:* Graziella Caselli and Alan Lopez (ed.), *Health and Mortality among Elderly Populations,* p. 288–312. Oxford, Clarendon press, 330 p.

Walford Roy, 1984. *Maximum Life Span,* New York, Avon. Translated into French: *La Vie la plus Longue*. Paris, Laffont, 250 p.

Wilmoth John, 1993. *Computational Methods for Fitting and Estrapolating the Lee-Carter Model of Mortality change,* Technical report, Department of Demography. Berkeley, University of California, 19 p.

Wilmoth John, 1996. Mortality projection for Japan: a comparison of four methods, *in:* Graziella Caselli and Alan Lopez (eds.), *Health and Mortality among Elderly Populations,* p. 266–287. Oxford, Clarendon press, 330 p.

Wilmoth John, Vallin Jacques, and Caselli Graziella, 1989. Quand certaines géné rations ont une mortalité différente de celle que l'on pourrait attendre, *Population,* vol. 44(2), p. 335–376.

Wilmoth John, Vallin Jacques, and Caselli Graziella, 1990. When does a cohort's mortality differ from what we might expect?, *Population, An English Selection,* vol. 45(2) p. 93–126.

CHAPTER

76

Hypotheses for Migration Projections

ODO BARSOTTI AND ALBERTO BONAGUIDI

Dipartimento di Statistica e Matematica applicata all' Economia, Università degli Studi di Pisa, Pisa, Italy

INTRODUCTION: CAUSAL MECHANISMS, FORECASTING METHODS, MIGRATION

Demographic forecasts have been developed primarily through the search for regularities in previous trends and improvement in the quality of available data.

"Demographic forecasting is seen as the search for functions of population that are constant through time, or about which fluctuations are random or small," wrote Nathan Keyfitz (1972: 347). It is in fact an approach that is essentially "guided by the data" (Willekens, 1991). Even the impetus to improve the corpus of methods and models adapted to the work of forecasting was provided by the need to make the best possible use of past trends in order to forecast the future.

It has nevertheless been proved that the growing quantity of available data and the increasing effectiveness of analytical tools have hardly improved on the quality of population projections, and their precision depends more on a real understanding of demographic behavior than on the sophistication of the methods used (Keilman, 1990). Forecast quality can only be improved if we increase our ability to define the causal factors and the processes that determine the level, the sequence, and the timing of demographic events, and we improve our understanding of the mechanisms of the determining factors of observed population dynamics (Willekens, 1990, 1991).

After considering the manner in which knowledge can improve demographic forecasting, Nathan Keyfitz (1982) concluded that research on population tends above all to show conditioned causal associations (if . . . , so . . . , all other things being equal), which are useful in understanding the past but inadequate for making forecasts. Such associations are necessarily based on partial theories, namely, theories that explain one or several of the associations that make up the whole system that determines population dynamics. If, to improve forecast quality, we were to move from forecasts based on data trends to those based on causal processes, demographic research should progress from quantitative associations between variables to the causal mechanisms that produce these associations. Research of this type implies that demographic events that occur at different periods are linked as much with the characteristics of the social context, with the successive phases of the development process of which these events are a part, as with the individual characteristics of those experiencing them.

The new generation of population projection models should show explicitly how contextual changes (socioeconomic, cultural, technological, and legal) are mediated by the characteristics of the process at the microlevel (individual and family). All this calls greater attention to life cycle transitions and the influence of exogenous factors on these transitions. In other words, an association between the most advanced conceptual framework of demographic analysis (i.e., the age-period-cohort approach), and the dynamic behavioral theories should be established. The shift in the focus from observed data and trends toward subjacent processes constitutes a considerable challenge. "The data we have traditionally been working with may prove to be inadequate and the models need to be revised or need to be developed anew. The real challenge is to develop simulation and forecasting models that are based on theories which reveal the casual mechanisms determining human behavior and which explain demographic change in terms of these mechanisms" (Willekens, 1991).

The search for subjacent processes is far more complicated for migration than for natural events, as migration is more sensitive than are fertility and mortality to the variations, even in the short term, of economic, social, and political variables. Furthermore, migratory phenomena interact more directly with the same phenomena that influence them. In reality, migration is by nature closer to a *process* than to a simple event, such as a birth or a death. This obviously results in it being less foreseeable. The difficulty in forecasting migration results fundamentally from the limits and the uncertainty of our theoretical knowledge of the complexity of the factors affecting the phenomenon and its manifestations. Beyond this uncertainty of the concepts and nature of the data is the added complication of the *instrumental* character that surrounds the migratory event, different to other events in the life cycle.

These difficulties concern internal migration as much as international migration, but there are particularities in each case that lead us to discuss the two phenomena separately.

I. INTERNAL MIGRATION FORECASTS

1. Problems

The difficulties in formulating hypotheses on the evolution of internal migratory movements stem from the fact that the propensity to migrate from one geographic area to another depends on a complex set of variables acting and interacting at the macrolevel—such as the general economic and social conditions of the country, characteristics specific to the departure zone and to the destination zone (labor market, housing market, quality of the environment)—and on the microlevel, such as the individual characteristics of the potential migrants and the current phase of their personal and family life cycle (see Chapters 59, 60 of Volume II). Formulating a hypothesis by taking into consideration the different aspects of the variables involved (or at least a considerable number of them) and their interrelations requires the reorientation and investigation of the concepts on which forecasting migratory processes is based.

However, at the present, we are too far from an ideal theory of migration—that is, a series of logically arranged concepts that can explain the entire constellation of the migration phenomena at different scales—to be able to account for past and present trends and forecast the future. We have to be satisfied with many partial theories that give preference, according to the case, to a particular approach to the phenomenon: a historical or economic approach, a sociological or psychological approach, a macro approach that gives priority to explanations by variables acting at an aggregate level, or the micro approach that associates individuals' migratory behavior with their personal characteristics, their past experiences, and their expectations for the future.

The current theoretical corpus on migration owes much to the pioneering work of Edward Ravenstein (1885), who greatly influenced subsequent theoretical approaches with his *laws* (see Chapter 59 of Volume II). The most important law is the one inspired by a well-known law of physics, according to which the number of persons moving between two geographic areas is directly proportional to the size of the population in the departure and destination zones and inversely proportional to the distance between them. Another fundamental contribution to the study of internal migration comes from the more recent work of Everett S. Lee (1966), who provided an interpretation method in which four categories of factors intervene: factors associated with the departure zone, factors associated with the destination zone, obstacles to movement, and personal characteristics (e.g., age, sex, level of education) of potential migrants.

Another very important contribution was made by Wilbur Zelinsky's *mobility transition* theory (1971), according to which the migratory phenomenon is situated in a perspective of historical evolution. According to this theory, during the most advanced phase of economic and social development, traditional rural to urban mobility gives way to other forms of mobility, such as interurban and intraurban migration, which become the most important components in the process of population distribution.

The micro point of view transfers the focus on to the individual and the reasons that incited him/her to migrate (e.g., improvement in economic and social conditions, construction of family ties, the search for a certain lifestyle). According to Larry A. Sjaastad's *human capital* theory (1962), the decision to migrate is taken when the current value of the gain expected in the destination zone, less the costs (financial and psychological) of migrating, exceeds the net gain of staying in the zone of origin. The instrumental nature of migration implies that the decision whether or not to migrate is conditioned by other decisions through which the individual attempts to satisfy all of his/her aspirations. Migratory behavior is the outcome of a decision-making process comprising various phases, during which all of the consequences of the decision are evaluated. Gordon De Jong and James Fawcett's "value-expectancy model" (1981) proposes a conceptual model for studying the reasons for migrating and the decision to migrate. This approach is even more useful when it is considered from a dynamic perspective, that is, when it is linked to the context in which the individual lives (Willekens, 1985).

To illustrate the complex interweave of variables on which migratory behavior depends, we can refer to the phenomena of context (the coexistence of significant surpluses in the supply and demand of work in different regions of the same country), which, considered in isolation, could be deemed key determinants of large internal migratory movements but, which under the influence of other factors (increase in family income available through public transfers, role of the family in estimating guarantees, cost of moving in to new accommodation) and the social and cultural characteristics of the new generations (increase in level of education, refusal of migration as a solution to employment problems), may not determine any of these internal movements that, in a partial vision, would have been considered inevitable.

The main limitation to formulating hypotheses on internal migration is owing to the impossibility of defining all the variables which affect this phenomenon. The variables most often considered in selecting hypotheses for the future evolution of internal migration are, as we know, the differential growth between sections of population of working age, territorial imbalances in the supply and demand of work, industrial implantation policies, development of infrastructure and services, and the consequences of past movements. These are undoubtedly extremely important factors, yet they only represent a partial vision of a far more complex system of interaction.

Moreover, the instrumentation of migration in relation to other key life events, such as marriage, entry into the labor market, and the birth of a child, has obvious consequences on their future evolution. Therefore, forecasts made without reference to other life events may go against the existing relations between migration and the life cycle. There are phases in life when the propensity to migrate is greater than in others. These phases are precisely those during which other important choices are made (e.g., to work, get married, have a child). The changes occurring in the development of the life cycle that are translated by postponement or prescheduling of the most important milestones in life (e.g., completion of education, finding work, marriage, birth of a child) undoubtedly affect migration projects.

Among the specific problems that forecasting internal migration poses with regard to the two other components of population dynamics (mortality and fertility) is that of definition and measurement. Migration is generally considered as the transfer of the place of residence from one administrative unit to another. This reference to crossing administrative borders introduces the spatial dimension of the migratory phenomenon, which in itself constitutes an extra element of complication. It is then duplicated by a supplementary problem: the occurrence of the event depends strictly on the administrative division of the territory.

The second large problem is that of measurement. The migratory event can, as we know (see Chapter 13 of Volume I), be directly observed through the data in population registers (registration and removal of individuals on population registers), or indirectly by comparing the place of residence observed at two different dates during surveys or censuses. In this latter case, the period of time between the two dates is generally fixed (often 1, 5, or 10 years). The first source, called *anagraphic*, collects data on the migratory *event*; the second, called *censitary*, records the migrants *status* and not the migratory event. For the second source, the difference between migration and migrants is negligible only if the time lapse is short. The two sources obviously need to be processed differently. Contrary to appearances, the anagraphic source is not necessarily more accurate than is the censitary source: a migratory event only becomes such once the new place of residence has been declared, but this declaration can naturally be made with a more or less significant delay; conversely, some declarations can be purely fictitious.

2. Methods

The methods for forecasting internal migration vary according to how the three fundamental characteristics are considered: level (number of migrations during a given period of time), calendar (distribution of the migrations in the interval), and direction (departure and destination zones) (Willekens, 1990). If the interval is defined in terms of calendar years, by *period*, the *cross-sectional* approach is adopted, as is the case in most forecasting models. If, on the contrary, the interval is defined in terms of years of life, by *cohort*, the *longitudinal* approach is adopted. The forecasting methods that take the relation between the migratory event and other life events into consideration are naturally situated in a longitudinal perspective. The age variations in the propensity to migrate can be modeled on profiles, which have proved to be relatively stable in time and space. This applies to the *migration schedule model* (a mathematical model that linearly combines double-exponential functions), the parameters of which are useful for description and for interpretation and forecast (Rogers and Castro, 1981). In Great Britain, for example, this model was widely used for the projection of internal migration between localities (Bates and Bracken, 1987). In the Netherlands, it is an integral part of the mltiregional demographic analysis (MUDEA) model that was adopted for subnational population projections. However, this model complicates the formulation of hypotheses of the evolution in the relation between migration and age, as specific demographic significance cannot be attributed to its parameters.

Forecast methods also change according to the way in which the *direction* of the flows are treated. Whereas traditional methods were mostly based on *net balance*, the most recent studies more frequently consider the *flows*, thus distinguishing the direction of migration from the departure zone to the destination. Indeed, it is only in the case of flows that the issue of causation makes any sense (the balances have no direct cause). The flows are determined by behavior, and it is on these that the formulation of evolution hypotheses should be based. The systems of flows are the basis for multiregional models for forecasting by cohort, which have greatly reformed national and regional population projection methods. To make population projections, one can also usefully resort to methods that express, in terms of the parameters of log-linear models, the system of internal flows by geographic zone (region) of origin and destination (e.g., in the recent projections of the National Institute of Statistics [ISTAT]). These methods enable a reduction in the complexity of the data and an indirect estimation of the missing data.

The extrapolation of historical series as a projection method of the origin-destination flow system is hardly used owing to the limited availability of historical data and the complexity of the data. However, the original and interesting method of treating and projecting historical series proposed by Frans Willekens and Nazli Baydar (1986) should be mentioned. This method has several phases. First of all, the internal flow system is separated into three distinct components: a *level* component (total number of internal migrations), a *sending* component (the proportion of a given zone within the total emigration flow), and a *distribution* component (the proportion of emigration from the zone for each of the destinations). Each of these components is weighed and then subjected to an exploratory analysis with a view to locating constants and irregularities in the data structure. The structure is described by using a generalized linear model, which enables the parameters dependent on time, that is, the variable components of the migratory system, to be identified and then extrapolated.

For projecting the spatial dimension of internal migrations, it is also possible to resort to explanatory models, notably spatial interaction models or gravitational models. The most modern version of these models was formulated by William Alonso (1978) in his general mobility theory. Alonso's model has proved to be very useful in studying internal migration and has also shown its validity for forecasting the origin-destination flow system (Ledent, 1986).

II. INTERNATIONAL MIGRATION FORECASTS

1. Problems

In (national) population projections, international migration is generally treated as a subsidiary of fertility and mortality. The reasons given for this are as follows:

1. Low intensity of the phenomenon in many cases—"For most countries, the volumes of immigration and emigration are either too small, or the balance of the two (net migration) is too insignificant to justify separate treatment." (George and Perreault, 1992: 87).
2. Inadequacy and paucity of data that are difficult to compare owing to the nature, the plurality, and the heterogeneity of the sources used: "The administrative nature of the data sources used is the factor largely responsible for the lack of international comparability." (United Nations, 1998: 6).
3. Uncertainty in the definition of an international migrant: "The slow progress made in enhancing comparability coupled with a worldwide trend toward the stricter regulation of international migration and the expansion of migrant categories considered policy-relevant suggests that a new approach must be taken. International comparability may be enhanced by recognizing that, from the policy perspective, there is a set of migrant categories that can be distinguished statistically on the basis of administrative sources because they are subject to distinct treatment in terms of admission and residence rights." (United Nations, 1998: 6).
4. Low foreseeability of flows that are more affected by the economic, social, and political situation: "Recent immigration trends in Western Europe clearly demonstrate the volatility of migration trends. During the

early 1970s West Germany had an annual net migration gain of more than 300,000; 5 years later this number had declined to only 6,000 and 3,000 during the early 1980s. During 1985–1989, however, the annual net gain increased sharply to 378,000, 100 times that of the previous period. Few other countries have these extreme fluctuations, but the traditional immigration countries (USA, Canada, and Australia) show remarkable ups and downs." (Lutz *et al*., 1996: 369).

5. Difficulty in formulating hypotheses on the future dynamic of international migration as a result of the profound changes in its nature: "International migration therefore concerns nearly all the countries of the world, and recent changes have overturned the traditional classifications between countries of emigration and countries of immigration . . . In a parallel direction with the globalisation of the international movement of people . . . we are also observing a multiplication in the different forms of migration."[1] (Guengant, 1996: 110).

Whatever the reasons leading to minimizing the issue of international migration in population projections, the most recent developments in demographic analysis have, on the contrary, clearly highlighted its importance:

1. The number of international migrants and refugees is currently estimated at approximately 120 or 130 million, and the phenomenon has increased considerably over the course of the past 30 years, as the estimated number at the beginning of the 1960s was 70 million (Guengant 1996: 107–108). In the 1985–1990 period, the net rate of international immigration of the most developed regions increased to 1.6 per thousand, which was equivalent to more than a quarter of population growth (6 per thousand). In the 1990–1995 period, it increased to 1.8 per thousand, or 45% of population growth which itself had fallen to 4 per thousand (United Nations, 1998: 14).

2. The significance of the phenomenon in the industrial countries is reinforced by the idea that it could thwart the effect of fertility decline (often well below the replacement level) on the size and composition of the native population and labor force.

3. The influence of migration on the social, ethnic, and cultural structures of the receiving countries has been increasing (George and Perreault, 1992).

4. Through the services, material or intangible, that migrants render to their country of origin, international migration plays an important role in the development process of a growing number of countries that export manpower: "At the macrolevel, remittances are usually a major source of foreign exchange earnings for countries of origin and, in some cases, contribute a sizeable share of GDP. For the world as a whole, the level of remittances estimated on the basis of data compiled by the International Monetary Fund (IMF) rose from US$ 43 billion in 1980 to US$ 71 billion in 1990 (Russel, 1992). In 1989, the estimated remittances amounted to US$ 61 billion worldwide, a figure that compares favourably with the US$ 47 billion provided as official development assistance by Organisation for Economic Cooperation and Development (OECD) member States to developing countries that year (Russel, 1992; IOM, 1994)." (United Nations, 1998: 154–155).

The total integration of the international migration component in projection work is naturally confronted by problems of uncertainty in definition, heterogeneous data sources, erratic movements, and, consequentially, the above-mentioned phenomena that are difficult to foresee.

These problems are even more acute for international migration, but even for internal migration, the fundamental issue that needs to be resolved in order to make reasonable hypotheses on their future evolution remains the comprehension of its causal determinants.

Various theories have been proposed to explain why international migration occurs and how migratory flows grow with time and spread into space. Each of them tries to explain the same thing, although through concepts, hypotheses, and analytical frameworks that are often radically different. The various interpretations are not necessarily contradictory, but their results are only partial and considerably influenced by the angle of approach adopted (economic, sociological, political, systemic) and the method used (micro, macro) (Massey *et al*., 1993: 432; United Nations, 1998: 141–147).

Their complexity and their multiform nature require, on the contrary, "A sophisticated theory that incorporates a variety of perspectives, levels, and assumptions. . . . We are skeptical about both atomistic theories that deny the importance of structural constraints on individual decisions, and of structural theories that deny agency to individuals and families." (Massey *et al*., 1993: 432, 455).

If we look closely, the theory on which research for an explanation of international migration is based is

[1] "Les migrations internationales concernent donc désormais la quasi totalité des pays du monde, et les évolutions récentes bouleversent les classifications traditionnelles entre pays d'émigration et pays d'immigration. . . . Parallèlement à la globalisation des mouvements internationaux des personnes . . . on assiste également à la multiplication des formes de migrations."

not substantially different from that which explains internal migration. The explanatory variables are essentially the same; even if some of them, notably demographic, political, and geopolitical variables, have a more important role in international migration. Contrary to internal migration, international migration is not always a voluntary choice; it can even be forced or can constitute a way of fleeing from political persecution, ethnic conflict, serious environmental crises, or intolerable poverty.

The macrostructural factors responsible for international migratory pressure are numerous and interconnected: economic imbalances, demographic and social imbalances, political crises, ethnic conflicts, ecological deterioration, revolution of the communications system, standardization trends in the content and structure of education systems in countries with different levels of development, increasing levels of education, and inappropriate development strategies and policies, often resulting in privileges for tiny social categories. Macrostructural attraction factors also have an important role to play in the migratory choice of the potential migrant: a request from the most industrialized countries for workers to immigrate, varying degrees of accessibility, the degree of permeability of borders, the stringency of controls and rigidity in expelling irregular or illegal immigrants, and the development of formal or informal reception and support networks (Stahl, 1995).

The problem then becomes more complicated because it is not sufficient to try to interpret current events only, but future trends must also be explored. We should, in fact, try to imagine the specific weight of each structural push and pull factor; if and how the interaction between these factors can be modified; the effect they will have on migratory pressure; and how, through the mediation of individual and family factors, they could be interpreted as a migratory option and, consequently, as a flow.

2. Methods

The paucity of data, the uncertainty of definitions, and the complexity of the phenomenon that makes the simplification in one global explanatory model difficult have generally impeded the formulation of hypotheses on the future of international migration based on sophisticated models and extrapolation techniques. The aforementioned work of M. V. George and Jeanine Perreault (1992: 99) illustrates the results of a survey carried out in 30 industrial countries that show that nearly all the developed countries apply rather simple methods to measure and extrapolate external migration: the vast majority of countries assume that future migrations will occur at the same level as previous ones, unless they believe that the level will be zero. Nevertheless, the two authors draw attention to the need for a general framework of analysis and interpretation of international migration, as a result of the scale of the problems raised by the phenomenon at different levels of public action (national, regional, local).

The estimation of future international migratory movements should take the following events into consideration: (1) the past evolution of the phenomenon and current trends, (2) the migratory policies in the host country and the country of origin, and (3) the push and pull factors that determine migratory pressure and the propensity to migrate (Oberg, 1996; Zlotnik, 1999).

There are three fundamental types of error in population forecasts (Lutz *et al*., 1996: 35). The same types of error, considerably accentuated, also apply to international migration:

1. *Measurement errors*—As we have already stated, basic data on international migration are by their very nature uncertain (much more so than data on births and deaths and more than internal migration data), and the category of international migrants, for which in practice these forecasts are made, often only consists of the most visible component made up of the migrants authorized to stay for a prolonged period or permanently in the host country (on the contrary, temporary, clandestine, and illegal migrants are excluded, even though usually, a wider definition of migrants is theoretically used) (Oberg, 1996: 339).
2. *Trend misevaluation*—International migration is characterized by great fluidity and considerable fluctuation from one year to another, much more so than other demographic events; the extrapolation of past trends can therefore lead to erroneous predictions of future changes.
3. *Unexpected events*—Each event, such as a particularly severe and prolonged food shortage, the explosion of war, the blaze of ethnic conflicts, or a drastic change in migratory policies in the host or departure countries, can provoke severe changes in trends and therefore cause the extrapolated image of the future to be inconsistent.

In Europe, various international organizations have endeavored to fill the gaps in the information on the scale and volume of migratory flows. The primary objective is to harmonize the data sources; to reexamine the practical concepts, definitions, and measurements used in order to improve the comparability of migration statistics; and to assemble into one international database the data collected in each country (EUROSTAT, 1992, quoted by Willekens, 1994: 20).

All this, with the utilization of techniques that are increasingly capable of combining data sources, and the use of increasingly sophisticated forecasting models, should make it possible to reduce the first two types of error (measurement and trend evaluation) that international migration is subject to.

But although these improvements are important, they only slightly affect the third type of error, which concerns unexpected events. To reduce this type of error, it may be useful to exploit a source of information that has remained in the shadows for a long time: that which is constituted by the various experts and practitioners who for various reasons work directly or indirectly on the issue of international migration and who for years have been observing and interpreting the flows and study national migratory policies. Their judgment, opinions, and intuitions can capture signals of changes and can foresee transformations in behavior and models that are difficult to record objectively, and therefore provide a more concrete view of reality than that which statistical measurements can supply (Willekens, 1994: 25).

The estimation of future international migration, based on this type of expertise, has benefited from the development of new techniques and projection methods, in a probabilistic approach.

One of these methods was used by International Institute for Applied Systems Analysis (IIASA) (Lutz, 1996) to calculate world projections (see Chapter 77). The approach (by scenario), based on hypotheses proposed by experts, enables hypotheses on the effect of unexpected events to be made and seems to be the best method of evaluating the influence of structural changes on the dynamics of demographic parameters, taking into account the uncertainty caused by random factors. The opinion of experts can indeed be based on recent experiences that incite them to investigate unexpected but possible alternative trends and allows them to be converted to parametrically specific hypotheses (Lutz *et al*., 1996: 37).

The predictive distribution of probabilities is entirely based on the subjective opinions of the practitioners, opinions that are rooted in their understanding of past trends; their knowledge of the determinants of fertility, mortality, and migration; and the ability of these experts to imagine potential future changes at the level of the determinants themselves. By combining the subjective distribution of probabilities of a sufficient number of experts, an aggregate predictive distribution of probabilities can be obtained, which makes it possible to produce reliable confidence intervals for estimating demographic parameters (Lutz *et al*., 1996: 37).

In the area of migration, a matrix of interregional migratory flows (the world having been divided into 13 regions defined on the basis of geographic proximity and socioeconomic criteria) was constructed based on practitioners' opinions. Three levels of annual net migratory balance were identified (high, low, and intermediate), and it is presumed that the distribution is normal and that 90% of the cases are situated between the high and low levels. To determine the age distribution of migrants, age-specific rate profiles of Rogers and Castro's migratory model were used (Lutz *et al*., 1996). With this type of method, an evaluation of the degree of uncertainty of the estimated demographic parameters can thus be obtained, taking into account the possibility of unexpected events.

CONCLUSIONS

The main conclusion of this short review of possible hypotheses for migration projections is doubtless that the conceptual device developed to study internal migrations can be used for international migration. Moreover, this is an argument that has for a long time been supported by migration specialists and that appears increasingly realistic today. On the one hand, local factors that are at the root of internal migration are increasingly influenced by the international and even global context; on the other hand, local factors themselves have more and more influence beyond national borders.

Finally, the distinction between internal and international migration has less and less meaning; both forms of migration are now expressions of phenomena governed by the same logic. The remaining differences, which mainly relate to the greater influence of political and geopolitical factors on international migration and which naturally should be taken into consideration in formulating evolution hypotheses, are not of a nature to justify treating them separately.

In contrast, international migration has characteristics that reflect, above all, the action of regional factors. First, more so than between countries, migratory flows are formed between a given region of a departure country and a given region of a host country. Second, foreign immigrants are concentrated in certain specific geographic zones of the host countries. Third, international migration supplies internal migration, either because foreign immigrants substitute locals in the labor market or because their presence incites certain elements of the local population to migrate as they refuse to reside in close contact with the foreign immigrants.

All these characteristics of international migratory movements increasingly reinforce the hypothesis according to which this form of mobility can be taken to the same conceptual scheme as internal migration. Ultimately, in a context such as Europe on the path to integration, the disappearance of national borders even causes the idea of international migration being distinct from internal migration to disappear.

Acknowledgment

This chapter was translated from the French version by Accenta, Ltd.

References

Alonso William 1978. A theory of movements, *in:* Niles M. Hansen (ed.), *Human Settlements: International Perspective on Structure, Change and Public Policy*, p. 197–211. Cambridge (MA), Ballinger, 297 p.

Bates John and Bracken Ian 1987. Migration age profiles for local authority areas in England, 1971–1981. *Environment and Planning A*, vol. 19, p. 521–536.

De Jong Gordon F. and Fawcett James T. 1981. Motivation for migration: an assessment and a value-expectancy research model, *in:* Gordon F. de Jong and Robert W. Gardner (eds.), *Migration Decision Making: Multidisciplinary Approaches to Microlevel Studies in Developed and Developing Countries*, p. 13–58. New York, Pergamon Press, 394 p.

George M. V. and Perreault Jeanine, 1992. Methods of external migration projections and forecasts, *in:* Nico Keilman and Harri Cruijsen (eds.), *National Population Forecasting in Industrialized Countries: A Survey Carried Out by the Netherlands Central Bureau of Statistics*, Amsterdam, Swets & Zeitlinger, 364 p.

Guengant Jean-Pierre, 1996. Migrations internationales et développement: les nouveaux paradigmes, *Revue Européenne des Migrations Internationales*, no. 2, p. 107–121.

OM (International Organization For Migration [IOM]), 1994. Migration and development, *in: Population Distribution and Migration. Proceedings of the United Nations Expert Group Meeting on Population Distribution and Migration, Santa Cruz, Bolivia, January 18–22, 1993, Convened in Preparation for the International Conference on Population and Development, Cairo, September 5–14, 1994*, p. 471–476. New York, United Nations, Population Division, 502 p.

Keilman Nico, 1990. *Uncertainty in National Population Forecasting: Issues, Backgrounds, Anzalyses, Recommendations*. Amsterdam, Swets & Zeitlinger, 211 p.

Keyfitz Nathan, 1972. On future population, *Journal of the American Statistical Association*, vol. 67(338), p. 347–363.

Keyfitz Nathan, 1982. Can knowledge improve forecasts? *Population and Development Review*, vol. 8(8), p. 729–751.

Ledent Jacques, 1986. Forecasting interregional migration. An economic-demographic approach, *in:* Andrew M. Isserman (ed.), *Population Change and the Economy: Social Science Theories and Models*, p. 53–77. Boston, Kluwer-Nijhoff Publishing, 269 p.

Lee Evrett S., 1966. A theory of migration, *Demography*, vol. 3(1), p. 47–57.

Lutz Wolfgang (ed.), 1996. *The Future Population of the World: What Can We Assume Today?* London, Earthscan Publications Ltd, IIASA, 500 p.

Lutz Wolfgang, Goldestein Joshua R., and Prinz Christopher, 1996. Alternative approaches to population projection, *in:* Wolfgang Lutz (ed.), *The Future Population of the World: What Can We Assume Today?*, p. 14–44. London, Earthscan Publications Ltd, IIASA, 500 p.

Lutz Wolfgang, Sanderson Warren, and Scherbov Serguei, 1996. Probabilistic population projections based on expert opinion, *in:* Wolfgang Lutz (ed.), *The Future Population of the World: What Can We Assume Today?*, p. 397–428. London, Earthscan Publications Ltd, IIASA, 500 p.

Lutz Wolfgang, Sanderson Warren, Scherbov Serguei, and Goujon Anne, 1996. World population scenarios for the 21st century, *in:* Wolfgang Lutz (ed.), *The Future Population of the World: What Can We Assume Today?*, p. 361–396. London, Earthscan Publications Ltd, IIASA, 500 p.

Massey Douglas S., Arango Joaqim, Hugo Graenne, Kouaouci Ali, Pellegrino Adella, and Taylor J. Edward, 1993. Theories of international migration: A review and appraisal. *Population and Development Review*, vol. 19(3), p. 431–466.

Oberg Sture, 1996. Spatial and economic factors in future South-North Migration, *in:* Wolfgang Lutz (ed.), *The Future Population of the World: What Can We Assume Today?*, p. 336–357. London, Earthscan Publications Ltd, IIASA, 500 p.

Ravenstein Edward G., 1885. The laws of migration, *Journal of the Royal Statistical Society Part 2*, vol. 48, p. 167–227.

Rogers Andrei and Castro Luis, 1981. *Model migration schedules.* Laxenburg (Austria), IIASA, 150 p.

Russel Sharon S., 1992. Migrant remittances and development, *International Migration*, vol. 30(3–4), p. 267–288.

Sjastaad Larry A., 1962. The costs and returns of human migration, *Journal of Political Economy*, vol. 70(5), p. 80–93.

Stahl Charles W., 1995. Theories of international labor migration: an overview, *Asian and Pacific Migration Journal*, vol. 4(2–3), p. 211–232.

United Nations, 1998. *World Population Monitoring 1997. International Migration and Development.* New York, United Nations, Population Division, 204 p.

Willekens Frans J., 1985. *Migration and Development. A Micro Perspective*. Voorburg, NIDI, 18 p. (Paper presented at the Seminar on internal migration and regional development, Montreal, April 1–3, 1985, IUSSP).

Willekens Frans J., 1990. Demographic forecasting; state-of-the art and research needs, *in:* Cornelius A. Hazeu and Gerard A. B. Frinking (eds.), *Emerging Issues in Demographic Research*, p. 9–66. Amsterdam, Elsevier Science Publishers, 314 p.

Willekens Frans J., 1991. *Theoretical Foundations for Simulation and Forecasting in Demography.* Voorburg, NIDI, 15 p. (Paper presented at the ZUMA-Conference "Formale Modelle in der Demographie," Mannheim, December 4–6, 1991).

Willekens Frans J., 1994. Monitoring international migration flows in Europe. *European Journal of Population*, vol. 10(1), p. 1–42.

Willekens Frans J. and Baydar Nazli, 1986. Forecasting place-to place migration with generalized linear models, *in:* Robert Woods and Philip Rees eds.), Population Structures and Models, p. 203–244. London,. Allen & Unwin, 439 p.

Zelinsky Wilbur T., 1971. The hypothesis of the mobility transition. *Geographical Review*, vol. 61(2), p. 219–249.

Zlotnik Hania, 1999. *World Population Prospects: The 1998 Revision.* Rome, ISTAT, 25 p.

C H A P T E R

77

The United Nations' World Population Projections

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études demographiques (INED), Paris, France

Dipartimento di Science Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION

According to Thomas Frejka (1981), the first attempt at projection of the world's population dates more than 300 years ago, when Gregory King (1695, 1973), drawing on William Petty's (1682) political arithmetic at the end of the 17th century, forecasted that the total world population, which he estimated at 630 million in 1695, would reach 780 million in 2050. However, the question was of little interest to the scientific community at the time. In fact, Gregory King's manuscript, which was published for the first time in 1973, was largely ignored by his contemporaries as it did not provoke any debate or critique. Yet the questions asked by both William Petty and Gregory King are recurrent ones: How many men can the world feed, and when will humanity reach its sustainable limit?

William Petty wrote: "The present 320 millions . . . will within the next 2000 years . . . increase . . . to give one head for every 2 acres of land in the habitable part of the earth" and will then "exceed . . . 20,000 millions" (Petty, 1682:19).[1] Gregory King, for his part, calculated that "If fully peopled, [the World] would sustain 6,257,000,000 which is 10 times the number of people now in being, and it is not possible to maintain more than double that number or 20 times more than the number of people now in being . . . and if it were possible to sustain 20 times the number of people now in being, it could not increase to that number till about the year of Christ 10,000, which is above 8,000 years to come." (King, 1973:2).

Although the debate concerning the nature of population growth has always been of interest to the scientific community, especially from the end of the 18th century, we know of no other attempt at a world population projection until the 1920s, when the issue returned to the forefront. In 1924, Edward East, a geneticist of vegetables who converted to the social sciences, declared that having made a thorough study, he was able to conclude that "A reasonable maximum for the world's future population is one person for each 2.5 acres on 40% of the land area of the globe. This gives a figure of 5.200 million, a population which at the present rate of increase would be reached in just a little over a century." (East, (1924) 1977:69). In fact, as Thomas Frejka mentions, Edward East used George Knibbs' estimate of a doubling of the population every 80 years, a doubling that would lead rapidly to intolerable population numbers: 3.9 billion in 2008, 7.8 billion in 2089, 15.6 billion in 2169, 31.2 billion in 2250, and 62.4 billion in 2330. According to George Knibbs (1928), at this rate, the sustainable limit would be reached before the end of the 21st century. However, these population projections were based, as was Gregory King's, on the rough application of a

[1] Although Gregory King used William Petty's work as a basis, he significantly revised the world population estimates upward at the end of the 17th century, as he calculated that this figure was twice the size William Petty had thought.

doubling rate; as a result, they were rather anachronistic and far removed from the numbers one would expect to find with a modern demographic approach that had been enlightened by considerations of carrying capacity.

Raymond Pearl attempted to improve on East's projection by applying a logistic model that was used widely in zoology. This method took into consideration the initial size and carrying capacity according to production methods and consumption at the time, population increases at a variable rate, and that fact that the population tends to stabilize as the maximum is approached. In 1924 Pearl calculated a curve that, according to him, made a world population projection possible. This study was later revised in collaboration with Sophia Gould (1928). Of course, such an approach depends entirely on the supposed compliance between the logistic curve and the population limit imposed by the carrying capacity. Although this compliance is verifiable at some point, it changes with time, and as the Pearl himself suggests, the projection should be revised with each significant change in the relation between the carrying capacity and the available technological resources.

The modern era of world population projections actually began with the League of Nations' belated initiative, which was picked up and developed by the United Nations (UN) after World War II.

In 1941, the League of Nations asked Princeton University's *Population Research Office* to undertake a broad analysis of demographic trends and world population problems, and it was through this program that a first volume on Europe and the Soviet Union (Notestein *et al.* 1944) was published in 1944. Even though this first volume does not give a world population projection in the strictest sense, it does reveal all the ingredients that Frank Notestein (1945) then used to form the basis that would later result in the UN projections: as reliable a database as possible, a summation of independently obtained data for each country or region, and national and regional projections calculated according to what is generally known as the "component method" (see Chapter 72). As Thomas Frejka (1981) wrote, "The actual size of the world population in the year 2000 will—absent catastrophe—certainly be much larger than the then projected 3.3 billions, but the ideas, approach and methods Notestein applied provided a valuable base for a generation of world projections to follow." (Frejka, 1981:512).

Indeed, since the beginning of the 1950s, the UN's Population Division has regularly published world population projections based on the principles proposed by Frank Notestein. The endpoint of the recent projections[2] (UN, 2001) is no longer the year 2000, which is now the starting year, but 2050. World population growth should continue and result in an increase from the 6.06 billion reached in 2000 to 9.32 billion in the middle of the 21st century if the medium-scenario hypotheses are confirmed.

The results of this projection are presented in this chapter. We will then see—by analyzing their development and the underlying hypotheses and raw data—if the history of the UN projections teaches us anything about their plausibility and reliability. We will also compare the results obtained by the UN with those of other projections developed at the same time, either by independent researchers or by rival institutions. Finally, we will briefly discuss the possible repercussions of these world projections in terms of specific categories, such as large age groups or urban and rural populations.

We conclude that the endpoint (2050) of the UN regular projections is too soon to supply a correct projection if all the variables (e.g., demographic transition) are to be taken into account. Because of this problem, and notably following the work of Thomas Frejka, at the end of the 1970s the UN published longer-term projections that forecasted as far ahead as 2100 and even 2150.

I. BETWEEN 8 AND 11 BILLION IN 2050?

The aforementioned figure of 9.32 billion for 2050 is the result of the medium hypothesis in the last UN projection. This is accompanied by other scenarios and is bound by low and a high hypotheses of 7.9 and 10.9 billion, respectively, at the same date.

1. Source Data

One of the most difficult tasks of the UN experts in charge of population projections is assembling, evaluating, and sometimes even completing the necessary source data. This task is also one of the most decisive elements for the projection's credibility, at least in the short and medium term. Ideally, it consists of an

[2] Since this chapter was written (autumn 2002), the United Nations' Population Division has produced a new version (United Nations, 2003b,c,d). However, it seemed futile to delay the publication of this volume in order to update the chapters on the basis of this new version. The only major innovation of this new version is that it includes a hypothesis of an extension in fertility decline under the replacement level in certain developing countries. Nevertheless, the resulting global difference does not in any way affect the reasoning of this chapter.

accurate estimate of population by age on the January 1 of the starting year (in this case, January 1, 2000), as well as a comprehensive estimate of the trends in mortality, fertility, and migration by age observed during the previous decades. The former obviously constitutes the starting point for the projections, and the latter enables the formulation of hypotheses on the future evolution of mortality, fertility, and migration. In reality, the first is often lacking, except in countries that regularly produce annual population estimates by age and sex based on reliable observations. More often than not, notably in the least developed countries, the most recent population estimates date from the last census, which may sometimes precede the starting date of the projection by several years. In such cases, recent mortality, fertility, and migration estimates are crucial to estimate the initial population.

These are all the data that the UN Population Division strove to systematically collect for the 2000 projections, as it did for all the preceding projections, for each country or at least for the 187 countries and territories (out of 228) with a population over a certain level (140,000 inhabitants for the last projection). Below this level the benefits of a projection using the component method are low because of the risks affecting age-specific rates.

The source data collected in this manner are of variable quality and accuracy in different countries. For example, when the initial population is estimated, nearly all the developed countries have not only a recent census but often a good estimate of population by age at January 1, 2000. On the contrary, very few developing countries have such estimates and a recent census or the necessary fertility, mortality, and migration estimates to produce one. Only 135 of the 173 developing countries held a census during the 1990s.

Whether it is a question of the retrospective data used to make hypotheses of future developments (traditionally, the UN endeavors to retrace the trends since 1950) or the more recent data necessary to evaluate the population at January 1, 2000, the fertility, mortality, and migration estimates are subject to even greater disparity. Since 1950, the developed countries often have nearly complete series of age-specific mortality and fertility rates since 1950. However, there are significant gaps in migration data. As for developing countries, a significant number of them still have only fragmented recent data, which is often crucial to consolidate the starting point, As an example, only 63% of Africa's population lives in countries that have an evaluation of fertility by age for 1995–1999. Yet this score should be considered a valuable asset compared with the one observed for infant and child mortality (57%) and particularly for adult mortality (11%). In this latter case, the total lack of information is the lot of the majority of the African population (73%). The situation is better in Asia and Latin America, where the majority of the population resides in countries that have recent information on fertility. Knowledge of infant mortality is much more widespread in Latin America than in Africa (but not in Asia), and there is more information in Asia than in Latin America for adult mortality (UN, 2002).

Hardly any country supplies data by age and sex on migration entries and departures. The best that one can expect, with some very rare exceptions, is an estimate of net migration. In many cases, one must be satisfied with often not very recent net migration data obtained by subtraction of the natural balance from growth between censuses.

Added to these gaps is that the estimates available are not always exempt from major errors, as internal and external consistency analyses show. Preliminary analysis and reconstruction work is therefore essential as much for constituting the basis of the projections as for helping with reflection on hypotheses of the trend factors in population change.

After half a century of experience and the remarkable progress of national and international statistical institutions, the UN Population Division now produces a demographic database by country for the 1950–2000 period, which, although far from perfect, has the unique advantage of being exhaustive and comparable. This prerequisite in the production of world population projections is in itself a very valuable tool in the comparative analysis of past global trends, provided that we do not lose sight of the heterogeneity of the quality of the estimates behind it.

2. Method

The projection method used by the UN has hardly changed since the 1950s. It is based on the dual principle already recognized as essential by Frank Notestein: an independent study of each country (or, if this is not possible, of each region) and the use of the component method by cohort (using an age structure at the start and projections of fertility, mortality, and migration by age). In fact, as we have just stated, countries with less than 140,000 inhabitants are not subject to the second principle, as the conjunction of errors related to small numbers with the paucity of essential information may make this refinement counter-productive. In these cases, rough projections based on no less rudimentary hypotheses of the trend in the global growth rate are the best that can be done and certainly do not affect the world projection.

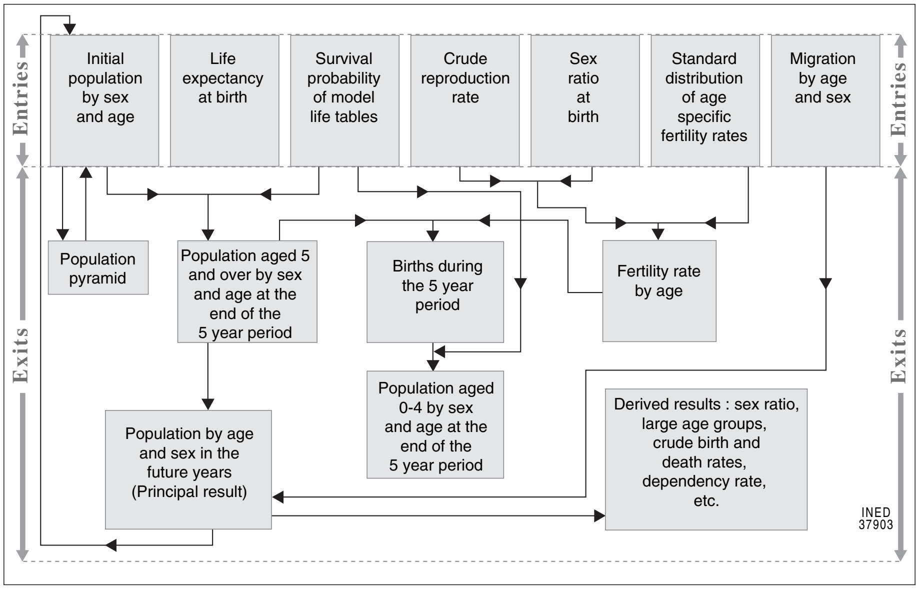


FIGURE 77–1 Major stages of the United Nations' projection for the countries where the mortality and fertility hypotheses are formulated in terms of age-adjusted indicators (life expectancy at birth and total fertility rate) (Data adapted from United Nations, 1971).

A comprehensive description of the algorithm used to transcribe the component method was given at the end of the 1950s (UN, 1957). Even though, computers make further refinements possible and the UN takes studies made by the national statistical institutes of the different countries into consideration, the process remains more or less the same. Figure 77–1 summarizes the progression of one frequent case in which model life tables and fertility models are used to translate into age-specific rates the hypotheses of trends in life expectancy and total fertility. This process is obviously simplified if mortality and fertility have been initially projected directly by age (and by sex for mortality).

3. Hypotheses and Scenarios

The choice of the hypotheses of the trend of the three factors of population dynamics (fertility, mortality, migration) is crucial. During the first years of the projection, the estimate of the initial data, particularly the population's age structure, is an essential factor, but in time this factor is erased and becomes nearly negligible and the evolution of fertility, mortality, and migration rates by age make their mark. Because of the generalization of Lotka's theory (see Chapter 20 of Volume I), we know that the initial age structure disappears completely and that two populations with entirely different age structures converge into one, sometimes new, structure when they are subjected to the same fertility, mortality, and migration rates for a sufficiently long period. The main concern of investigators of relatively long term population projections is therefore that of making the right choice, which is unfortunately, as we have seen in the previous chapters, far from easy.

It would be ideal to refer to a theory to anchor the hypotheses for the three components of demographic change. However, an investigator must simply make the most reasonable choice possible. One classic measure of caution consists of establishing several scenarios by combining different hypotheses for each component of change. The UN gives the "medium scenario" (or medium hypothesis), a scenario that, if not the most probable, is the least improbable.

In the context of the UN projections, these hypotheses, as we will see below, have varied. In the 2000 projection, the hypotheses reveal a subtle combination of reasoning, pragmatism, intuition, and the arbitrary, never really leaning toward one theory in particular. The UN experts have adopted five fertility hypotheses, of which three are currently used; two mortality hypotheses; and two migration hypotheses. Of the mortality and migration hypotheses, only the one called "normal" is currently used.

a. Fertility Hypotheses

The first choice made by the UN experts is to consider that the main possible source of error relates to fertility. There are three main fertility hypotheses that determine the difference between the three scenarios currently in use; they are called high, medium, and low as they are a direct result of the high-, medium-, and low-fertility hypotheses. Each of these scenarios is based on the same "normal" mortality and migration hypotheses. In addition to these three major fertility hypotheses, the 2000 projection includes "constant" and "instant replacement" fertility hypotheses. These last two hypotheses are only used as demonstrations to better define the evolutionary target in fertility behavior, do not depend on a preliminary data analysis, and do not require any particular choice. They are self-defining.

Constant fertility hypothesis—For each country the age-specific fertility rates are kept constant during the whole projection phase at the level estimated for the 1995–2000 period.

Instant replacement fertility hypothesis—For each country and each 5-year period of the projection, the total fertility rate is fixed at a level that, taking into consideration the mortality rate, ensures a net reproduction rate of one, or an intrinsic growth rate of zero (see Chapter 19). This does not mean that the resulting growth rate for the 5-year period will be zero but that the corresponding stable population will have a growth rate of zero; in other words, it will be stationary (see Chapter 20).

The three main fertility hypotheses require the most attention, especially the medium hypothesis as it determines the way the other two frame it. The choice is made in a complex way, beginning with distinguishing three categories of countries, by referring to the demographic transition theory (the only theory that is explicitly evoked) or, more exactly, to fertility transition, one of its components. Referring to the theory does not, however, mean that it is followed blindly. Different countries of the world are at different stages of their fertility transition, and the rate of fertility decline that is likely to be observed during the coming decades depends on this fact. The UN therefore distinguishes three categories of countries: the high-fertility countries (the fertility level estimated for 1995–2000 remains close to the highest level reached between 1950 and 2000), the medium-fertility countries (the fertility decline has already begun but still remains above the replacement level), and the low-fertility countries (fertility is lower than replacement level).

Medium-fertility hypothesis—In the countries with high fertility (Fig. 77–2A), the medium-fertility hypothesis supposes that the total fertility rate will decline at a rhythm of one child per woman per decade. The starting point of this decline, however, has a variability of 5 years. It is supposed that fertility will remain constant during the first stage (2000–2005) if no sign of decline has been detected at the start, but it is applied at the beginning for countries in which a slight decline has already been identified. In most cases (13 countries out of 17), these countries will still be above the replacement level in 2045–2050, at the end of the projection. However, whether they start from a lower level or whether they have started a minor decline, the other countries (4 of 17) will attain this replacement level.

In the countries with medium fertility (Fig. 77–2B), the medium-fertility hypothesis supposes that the replacement level will be reached before the end of the projection and that from then on fertility will remain constant until the end of the projection.[3] Everything therefore depends on the date at which the objective of 2.1 children per woman is reached and the rhythm of decline that enables this objective to be reached. Each case is treated in a reasoned manner according to the country's context, but while fertility remains quite high (with rare exceptions), it is presumed that it will decline rather rapidly (at the rate of one child per woman in 15 years) and this rate declines significantly with the approach of the replacement level. For these two groups of countries, the fertility rates by age[4] are interpolated between the rates observed at the beginning of the period and the rates combining 2.1 children per woman with three fertility schedules by age (early, late, or medium) chosen according to the national context.

For the low-fertility countries (Fig. 77–2C), the medium hypothesis forecasts that fertility will rise but will remain, barring exceptions, under the replacement level. This result is obtained in every country for which data are available by assuming that the targeted fertility level attains the levels reached by the cohorts born at the beginning of the 1960s. In countries where these data are not available, fertility is supposed to increase up to 1.9 if the starting level is more than 1.5 and up to 1.7 if the starting point is below 1.5. The idea of a systematic return to replacement level has been rejected as unrealistic for the duration of the projection, but the idea of a systematic rise has

[3] However, in the 2002 revision of its projection, the United Nations for the first time retained the idea that the decline could continue below 2.1 children per woman in some developing countries. (United Nations, 2003a,b,c).

[4] To be precise, the distributions of fertility by age are interpolated, not the rates; these distributions by age are then multiplied by the total fertility rates (themselves interpolated) to determine the fertility rates by age, which comes to more or less the same.

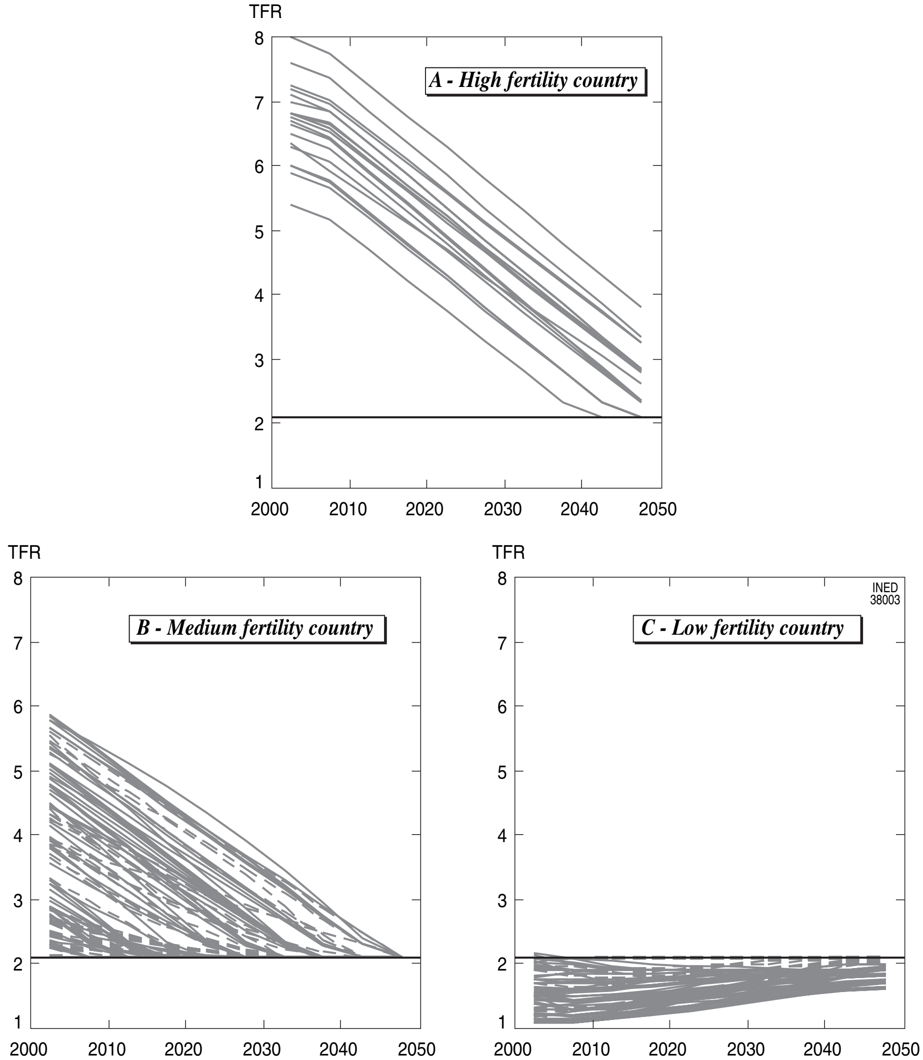


FIGURE 77–2 Hypotheses of the trend in total fertility by country in the case of the medium variant hypothesis according to the category of initial fertility.

nevertheless been accepted over the hypotheses of a deepening in the decline or of variable case by case trends.

High- and low-fertility hypotheses—While the criteria for choosing the medium hypothesis are varied and complex, the high- and low-fertility hypotheses are simply deducted from the first. In countries with high fertility, they evolve systematically at 0.5 children above or below the medium hypothesis. In countries with low fertility, the same occurs, but with a gap of 0.4 children.

b. Mortality Hypotheses

The 2000 hypothesis includes a so-called normal mortality hypothesis and a constant mortality hypothesis.

In the constant hypothesis, mortality is presumed to remain for the duration of the projection at the levels observed in 1995–2000 for each country.

The normal hypothesis is based on an estimate in the evolution of life expectancy at birth. This is estimated in different ways, according to the public health context of each country. For most countries it was considered that life expectancy would increase regularly during the whole period. However special attention has been given to the 45 countries particularly affected by the AIDS epidemic.

For the countries where mortality is expected to decline regularly, the rate of increase in life expectancy has been determined according to three reference models (rapid, medium, and slow increase). In each of these models, it is supposed that the rise in life expectancy will be accompanied by a decline in its rate of increase

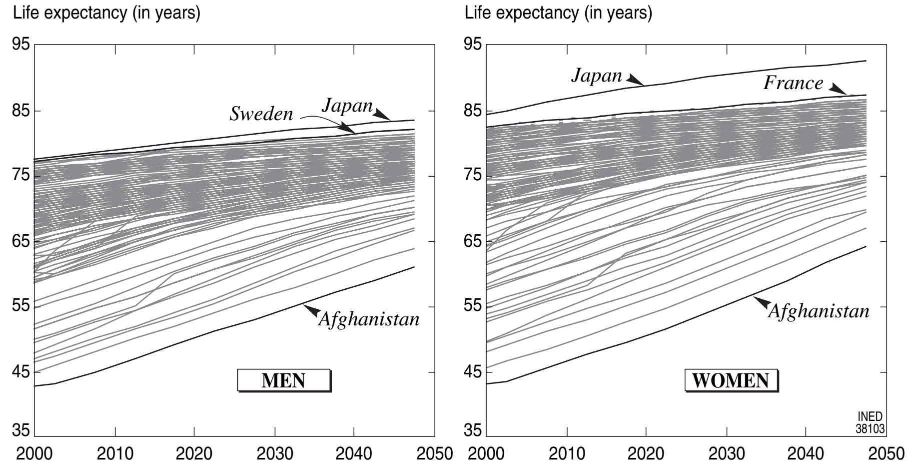


FIGURE 77–3 Trend by country of expected life expectancy at birth in the context of the "normal" hypothesis, in countries where AIDS has no major incidence.

and that this rise will become almost negligible at a certain threshold of life expectancy. In fact, this threshold, 82.5 years for men and 87.5 years for women, will not be reached by any country (except Japan) during the projection period. In the rapid or medium mortality decline models, if life expectancy does not yet exceed 60 years, it will increase by 2.5 years every 5 years for both sexes; this increase is only 2 years in the slow decline model. Afterward, the rate of decline decreases more or less rapidly according to the model and in different ways according to sex. Thus, when life expectancy is between 67.5 and 70.0 years, it is supposed to increase by 1.5 years every 5 years for men and by 2.3 years for women in the rapid mortality decline model, but it only increases from 1.2 to 2.0 years in the medium model and 1.0 and 1.5 years in the slow model. Finally, with the approach of the aforesaid maximum, the increase is only between 0.5 and 0.3 years per 5-year period, according to the model.

The authors of the projection nevertheless do specify that they are not opting for the theory of an absolute limit on life expectancy (see Chapters 48, 57), stating in their report that: "These life expectancies were not meant to represent a ceiling for human longevity, but they were high enough so that no country was expected to achieve them before 2050." (UN, 2002:213). This precaution in language is not only a little hypocritical, as everything leads to the belief that the evolution models appear to have been created to set a limit, but also incorrect, as according to published figures, Japanese women should already have a life expectancy that slightly exceeds the 87.5 years in 2015 and should even reach 92.4 in 2045–2050. Thus, the hypothesis really selected for Japan seems to be more likely than that which this report would suggest, as Japanese women have already surpassed the 85 years, and there are no indications that this progress will cease in the near future. What is most surprising here is that only Japan benefits from this unspoken preferential treatment. For example, the gap between Japanese and French women (the second country in terms of life expectancy for women), which was only 1.8 years, will reach 5.1 years in 2050 (Fig. 77–3). There is no justification for this choice.

Although less spectacular, the same anomaly is observed for men; Japanese men are the only ones to exceed the "limit" by reaching a life expectancy of 83.5 years, thus increasing the gap with Swedish men, their closest challengers (0.2 years of gap in 1995–2000 and 1.4 years in 2045–2050).

The mortality rates by age were obtained by interpolation between the estimated life table for the 1995–2000 period and the theoretical table accepted for maximum life expectancies of 82.5 and 87.5 years.[5] These rates are defined in age groups of 5 years from 5 to 100 (however, childhood mortality rates under 5 years were separated into infant mortality before age 1, and child mortality at age 1–4).

For the countries most affected by AIDS (35 African countries; 4 Asian countries, including India; and 6 Latin American countries),[6] the UN proceeded with country by country evaluations of the anticipated consequences of the epidemic according to available esti-

[5] The report does not specify how the rates were obtained for Japan beyond these limits.

[6] Countries where HIV prevalence at 15 to 49 years was more than 1.9% in 1999 plus India and Brazil, where the prevalence is significantly lower (0.7 and 0.6%) but where the absolute number of infected persons is very high (United Nations, 2002:215).

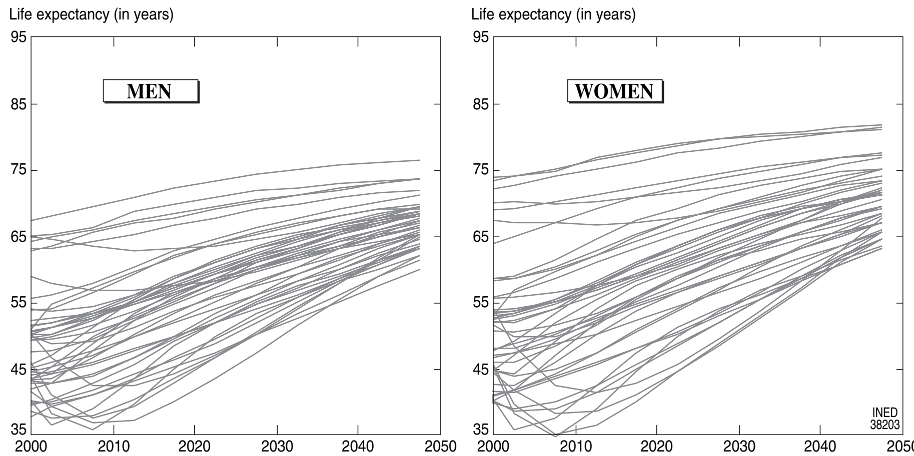


FIGURE 77–4 Trend in expected life expectancy at birth by country according to the "normal" hypothesis, for the 45 countries most affected by AIDS.

mates of the prevalence and incidences of HIV and a probable increase in the development of the epidemic. These evaluations are the subject of a special chapter in the report (UN, 2002:chapter 3). It is expected that in the most affected countries not only will life expectancy decline significantly during the coming years but also the epidemic will soon be suppressed and that all countries will experience an accelerated increase in life expectancy, which will enable many of them to catch up (Fig. 77–4).

c. Migration Hypotheses

The 2000 projection includes a normal migration hypothesis and a zero migration hypothesis.

The zero hypothesis does not mean a total absence of migration, but a zero net migration for each age over the whole of the projection period.

The normal hypothesis is based on a case by case estimate of past net migration and on currently known or predictable future policies regarding international migration. No explicit reference is made to any particular theory. The hypotheses are quantified in terms of absolute numbers of net migrants by 5-year projection period. The report does not specify age distributions of these totals.

In concrete terms, these hypotheses are basically translated by an extremely large inflow toward the United States, a share that is actually meant to decrease during the first decades before becoming constant from 2015 onward (Fig. 77–5). For most of the other countries, the inflow and outflow are supposed to remain constant (in absolute numbers) during the whole of the projection period. The only spectacular changes owing to national political contexts are those for countries that are currently experiencing exceptional circumstances, such as Afghanistan (return of refugees) and Pakistan (departure of the Afghans). The largest negative net migration is forecast for China, but it is three times lower than the positive net migration forecast for the United States.

d. Combination of Hypotheses Used for the Seven Published Scenarios

From these various hypotheses, $5 \times 2 \times 2 = 20$ scenarios are conceivable. However, the UN has explored the results of only seven of them, of which three are meant to realistically forecast the possible future; the others only serve to illustrate the importance of ongoing changes in behavior.

The three main scenarios combine the three principal fertility hypotheses (*high, medium, and low*) with the normal mortality and migration hypotheses. The first two illustrative scenarios combine the *constant* fertility hypothesis and the *instant replacement* hypothesis with these same mortality and migration hypotheses. Inversely, the latter two successively combine the *constant* mortality hypothesis and the *zero* migration hypothesis with the *medium* fertility hypothesis, and with the *normal* migration and *normal* mortality hypotheses (Table 77–1).

Priority Given to Fertility

The priority given to fertility is such that only one illustrative scenario appears in the publications, that of *constant* fertility, destined to show the gap between the three basic scenarios and what would occur if fertility remained fixed at its current level. It is however possible to obtain the results of the last

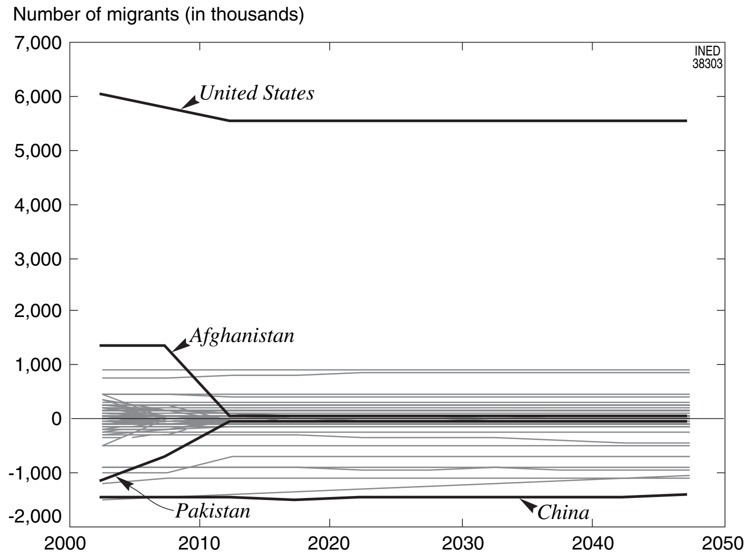


FIGURE 77–5 Trends in total net migration (absolute numbers) by 5-year period, projected according to the "normal" migration hypothesis.

TABLE 77–1 The Seven Scenarios Used for the Publication of the 2000 Projection Results

Scenarios and variants	Hypotheses of:		
	Fertility	**Mortality**	**Migration**
Principle scenario			
Low variant	Low	Normal	Normal
Medium variant	Medium	Normal	Normal
High variant	High	Normal	Normal
Illustrative scenarios			
Constant fertility scenario	Constant	Normal	Normal
Instant replacement fertility scenario	Instant replacement	Normal	Normal
Constant mortality scenario	Medium	Constant	Normal
Zero migration scenario	Medium	Normal	Zero

Data from United Nations, 2002:9.

three scenarios on a computer file from the UN's Population Division.[7]

Figures 77–6 and 77–7 illustrate the results obtained for the total population, in the prolongation of the trend observed since 1950, according to the five fertility hypotheses: high, medium, low, constant, and instant replacement, which are combined each time with the *normal* mortality and *normal* migration hypotheses. Fig. 77–6 represents the population of developing and developed countries on the same scale as world population, and Fig. 77–7 shows only the results of the developed countries, which are crushed in Figure 6 by the disproportion with the rest of the world.

Figure 77–6 shows that when viewed from a global perspective, the demographic situation and change in the developed countries are of little importance. This fact has had a considerable effect, as we will see below, on the extraordinary success of the UN projections.

In the central scenario, world population will increase from 6.06 billion in 2000 to 9.32 billion in 2050. Despite the deceleration in growth rate already established for several decades, a deceleration that is supposed to continue and even increase in the central scenario, an even greater growth in numbers should be expected between now and the middle of the century. It is a real possibility that the leaders of the planet have an tendency to forget that the great demographic scare of the 1970s no longer exists. The stake is even more considerable as the total burden of this future growth will be carried only by the developing countries, as is shown in the second part of Fig. 77–6, which according to the medium hypothesis would see their populations increasing from 4.86 to 8.14 billion (an absolute growth greater than that of the total world population) as, at the same time, the population in developed countries will decrease from 1.19 to 1.18 billion. In other words, the relative growth of the developing countries (67%) is significantly greater to that of the world as a whole (54%).

Furthermore, this burden that weighs so heavily on the future of the developing countries is far from being fairly distributed as the population growth will be

[7] Our thanks to Hania Zlotnik for kindly supplying these data.

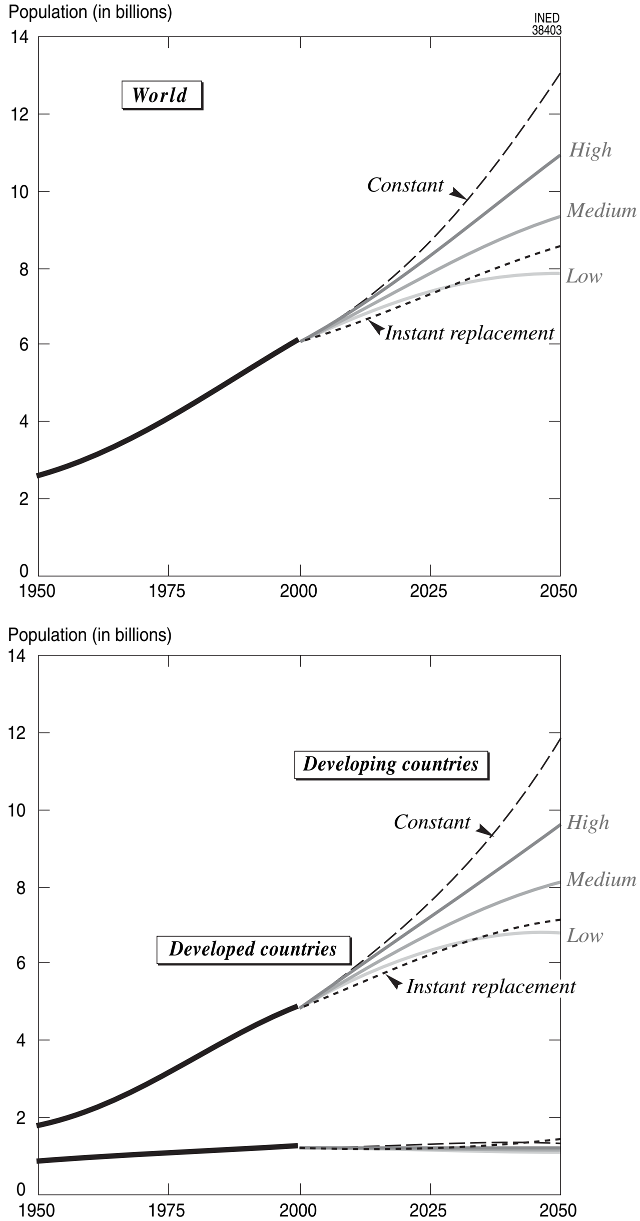


FIGURE 77–6 World population projections according to the five fertility hypotheses combined with the *normal* mortality and *normal* migration hypotheses.

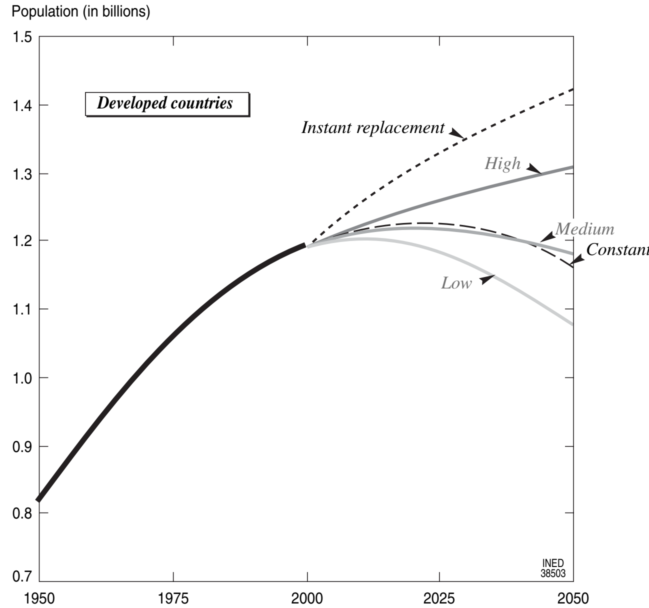


FIGURE 77–7 Zoom on the population projection for developed countries according to the five fertility hypotheses combined with the *normal* mortality and *normal* migration.

much faster among those who are economically the poorest and politically the most vulnerable. In their categories, the UN has for some time distinguished between the least developed countries and the other developing countries. According to the central scenario, the population of the least developed countries will increase from 660 million inhabitants to 1.8 billion, an increase of 178%, or nearly tripling. Demographic growth thus remains a major concern for the coming half century. But who is concerned with such destitute populations that have today, it must be admitted, become a minority at the global level and even at the level of that which used to be called the Third World?

What is the likelihood (or the risk) that this medium scenario occurs? We will be better placed to judge at the end of the following section. Let us say at this stage that in the UN projection system, the high- and low-fertility hypotheses are supposed to cover the plausible reality. In other words this system supplies us with a rough margin of error. In 2050 the world population will probably be between 7.9 and 10.9 billion. In absolute numbers, the margin is quite wide, as the gap between the extremes is more than 3 billion, but the relative error (±16%) is a very modest number for a 50-year forecast. And there is no comparison with other economic and social forecasts. Which economist would dare to make such a long-term forecast? Only the inertia of demographic phenomena gives this type of exercise a respectable credibility.

The two supplementary illustrative hypotheses make it possible to judge both the considerable impact and the limits of behavioral changes taken into account in the UN's population projections. The constant fertility scenario shows that if fertility had suddenly stopped declining in 2000, then the world population would exceed 13 billion in 2050: 2 billion more than in the high hypothesis and nearly 4 billion more than in the medium hypothesis. Instead of a growth of 50% expected in this central hypothesis, the projected population will double in the absence of fertility decline.

This shows the strength of the growth deceleration represented by current changes. Here again, let us emphasize that most of this change occurs in the south: without fertility decline, the population of developing countries will multiply by 2.5 and that of the least developed countries by nearly 5.

But the instant replacement fertility scenario also makes it possible to measure the limits of this type of change. Even with a drop in fertility resulting eventually in a strict replacement of generations all over the world from the year 2000, the world population would not immediately stop growing. On the contrary, it would continue to increase for several decades, reaching 8.6 billion in 2050, an increase of 40% from the year 2000. This is why, owing to past trends, the structure of world population is still young, and even with severely reduced mortality, the mothers of tomorrow, who are more numerous than are their elders, will have more children than their elders did (see Chapter 4). Of course, what is true for the world population is even more so for the population of developing countries, which despite an immediate transition to replacement fertility will grow by nearly 50%; the population of the least developed countries will grow even more, increasing by 60%.

If, seen from the world level of Fig. 77–6, the population projection of the developed countries appears particularly lacking in flavor, this does not mean that they do not need to be concerned with future demographic trends at their level. On the contrary, as Fig. 77–7 shows, the future is most uncertain for them. If only the fertility variants are looked at, the future of their population is open to evolution in the opposite sense. Although the central hypothesis clearly shows the threat of a general population decline in the second quarter of the century, the low hypothesis would introduce this decline during the first decade. In contrast, the high hypothesis would ensure that a significant population growth is maintained, which would enable the population of developed countries to reach 1.3 billion in 2050 instead of the current 1.2.

In the case of developed countries, strictly maintaining fertility at its year 2000 level would hardly change the result shown in the central scenario. This is because the expected changes in the latter would compensate each other, with fertility declining in the countries where it is still close to replacement level (more than 1.9 children per woman) and increasing where it is already below. On the contrary, the instant replacement fertility scenario would provoke a significantly greater population growth than that of the high hypothesis, whereas in the rest of the world, it would give a similar result to that of the low hypothesis as average fertility in developed countries is so low that a sudden increase would be necessary in order to reach replacement level.

The Importance of Mortality Decline

By comparing the constant mortality scenario to the medium variant of the principal scenario, mortality decline in the growth forecast for world population can be seen clearly, as the only difference between the two scenarios is the expected decline in mortality.

With constant mortality, world population would reach only 8.1 billion in 2050 instead of 9.3 (Fig. 77–8). In other words, in the expected growth of 3.3 billion between now and 2050, a little more than a third is owing to mortality decline.

This phenomenon obviously massively affects the developing countries, which in absolute numbers would benefit from the most of the bonus associated with mortality decline, but relatively speaking, the developed countries would be decisively affected. If their mortality stopped declining, the population of the latter would fall to 1.08 billion rather than 1.18. Instead of a slight decrease of −10 million in 50 years, they would lose 100 million inhabitants. In other words, according to the principal scenario hypothesis, it is only through mortality decline that the developed countries would succeed in maintaining their population at more or less the current numbers. But migration also affects the principal scenario.

The Role of International Migration

The future of world population as a whole is obviously indifferent to the UN migration hypotheses, although they could change the growth dynamic of the regions of the world (Fig. 77–9). In reality, they have little effect on the developing countries, as the expected flows are marginal in comparison with the size of the population, but they weigh heavily on the future of developed countries. In the absence of migration, the population of developed countries would only be 1.05 billion in 2050 (instead of 1.18), that is, 130 million less. In other words the anticipated net immigration weighs more than does mortality decline. The UN has not yet explored a scenario without mortality decline and without migration, but barring any interference effects, it can be said that without these two factors the population of developed countries in 2050 would be lower than the 230 million expected in the medium hypothesis. This signifies a loss of nearly one quarter of the current population.

4. Toward a New Demographic Pattern

The geographic distribution of world population has already changed much during the past half century

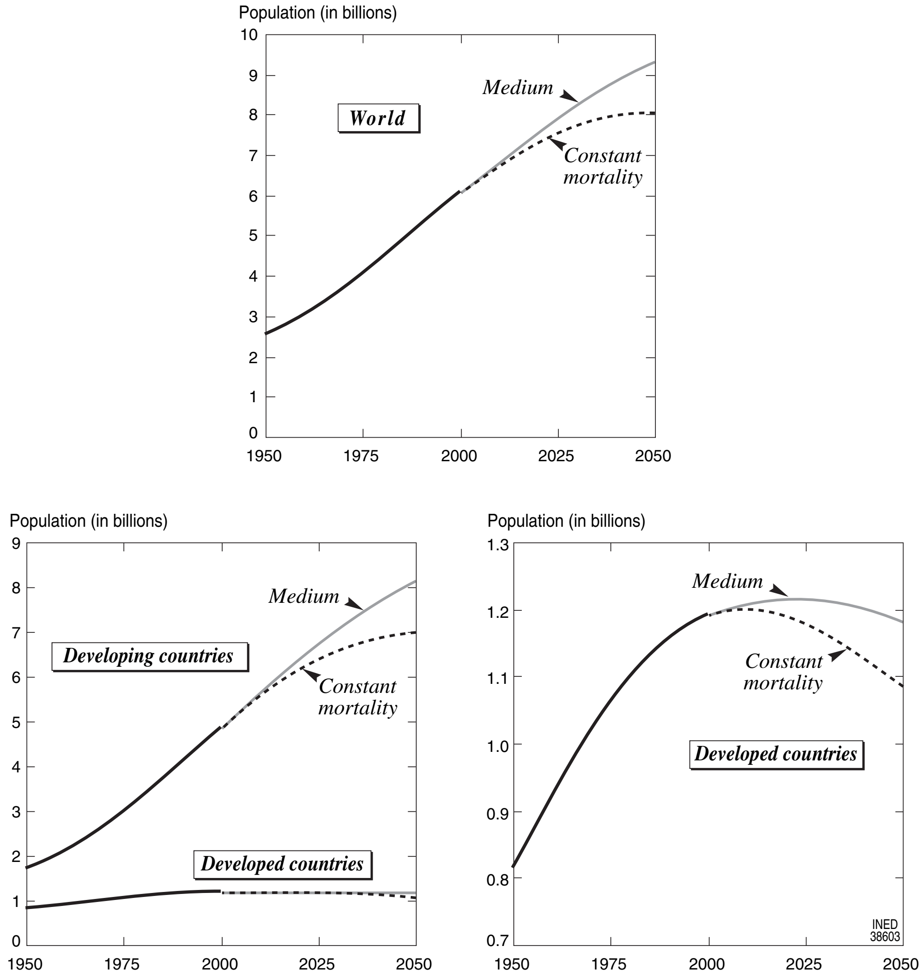


FIGURE 77–8 Comparison of the medium variant of the central scenario with the constant mortality scenario hypotheses.

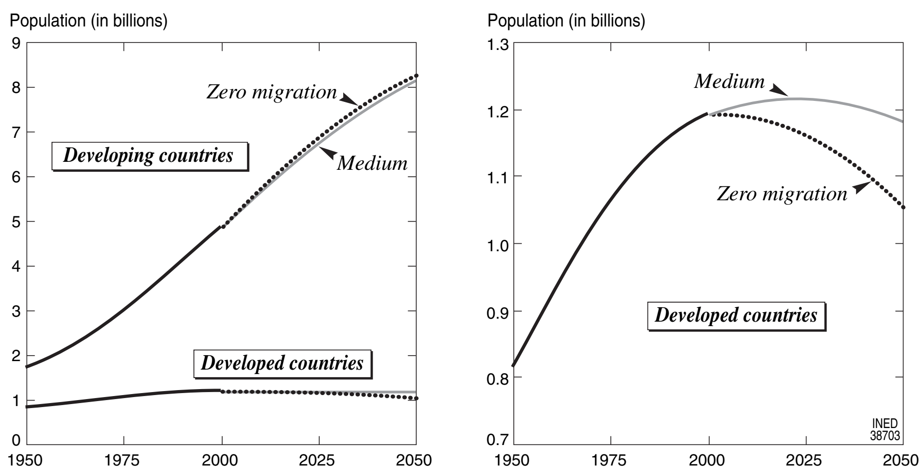


FIGURE 77–9 Comparison of the medium variant of the central scenario to the scenario without migration.

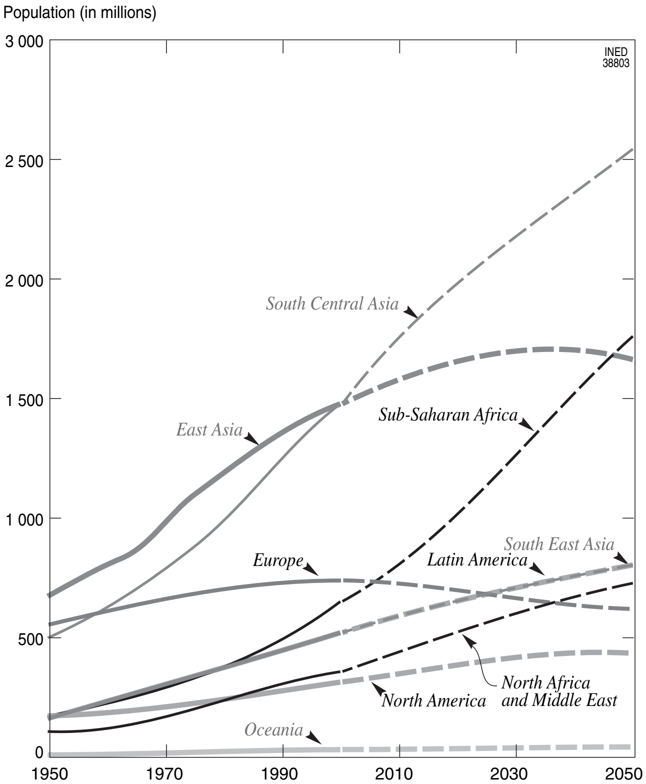


FIGURE 77–10 Trend since 1950 and projection up to 2050 of population by large world regions according to the United Nations' medium scenario (2000) (From United Nations 2001.)

(see Chapters 69 and 71). The redistribution will again be accentuated. To be convinced of this, one should simply look at the UN medium scenario results. Figure 77–10 traces the 1950–2050 trajectories of eight large regions, closely following the UN demarcation.[8]

a. The Large Regions

Despite AIDS, it is sub-Saharan Africa that will witness the largest increase in relative population as, although it will experience a severe loss in life expectancy at the beginning of the period, it will experience a more rapid growth rate than any other region between now and 2050. Consequently, a little before 2050 the population of sub-Saharan Africa will have overtaken that of western Asia. According to the UN central scenario, before the middle of the century, there will be more sub-Saharan Africans than Chinese, Japanese, and Koreans together. In absolute numbers, however, the growth of south central Asia (India, Pakistan, Bangladesh, etc.) will be nearly as spectacular. Although the rate of growth will certainly be slower than in sub-Saharan Africa, the initial population is so high that the increase in population will be almost as spectacular. Sub-Saharan Africa, which in 1950 had only 177 million inhabitants, will increase from more than 600 million in 2000 to 1.8 billion in 2050, whereas south central Asia, which had nearly 500 million inhabitants in 1950, will increase from 1.5 billion in 2000 to 2.5 billion in 2050. It will therefore gain only a billion inhabitants compared with the 1.2 billion gained by sub-Saharan Africa.

On the contrary, if east Asia is about to be overtaken by sub-Saharan Africa, it is also because it should experience a reflux in the second quarter of the century according to the UN central scenario. If the rapid fertility decline, below replacement level, that the entire region has experienced has not yet reversed population growth, it is because of the potential for growth that remains within its age structure. This phenomenon will soon be over, and the region's population could drop to 1.6 billion in the middle of the century after reaching 1.7 during the 2030s.

Latin America and southeast Asia's trajectories are almost perfectly identical during the hundred years from 1950–2050. Both these regions are due to experience growth rates comparable to that of south central Asia and should soon overtake Europe, Russia included. Later, a similar situation should be seen in the north Africa–Middle East block. The future dynamism of these regions contrasts violently with the reflux forecast for Europe. Because of a surprising reshuffle, the current population of Europe will be nearly that of Latin America (or southeast Asia) in 2050, whereas the population of Europe in 2050 will be close to that of Latin America (or southeast Asia) today.

North America should maintain a moderate growth rate, although with a slight deceleration, during the whole of the projection period, increasing from 315 to 440 million in 50 years.

Two supplementary observations clearly expose the demographic change in the large regions. In 1950 five of the eight large regions—sub-Saharan Africa, southeast Asia, Latin America, north Africa—Middle East, and North America—had a population of very similar size, around 170 million inhabitants. Since then, major differences have occurred, but this divergence is only the beginning of a spectacular explosion expected in 2050, from the 440 million in North America to the 1.7 million in sub-Saharan Africa. Yet more remarkable is the expected divergence between Europe and south central Asia. While these two units were the same size in the middle of the 20th century, central Asia is twice more populated than Europe today (1,400 million against 700) and will be four times in 2050 (2,500 against 600).

[8] For obvious reasons (cultural community, comparable demographic situations), we have regrouped the United Nations' North African subregions and western Asia, under North Africa and the Middle East, in order to isolate sub-Saharan Africa. For the rest of the world, UN subregions were used for Asia only.

b. *At a National Level*

The expected recomposition in the large regions is often more obvious at a national level. However, here we will only hint at it by mentioning the trajectories of 23 countries that currently have more than 50 million inhabitants.

Of the three most populated countries of the planet, two belong to the developing world (India and China) and the third to the developed world (United States). It is therefore not surprising that the aforementioned divergences should have such an impact on the top three (Fig. 77–11). China and India are currently rivaling for first position, but India will finally overtake China around 2040. The United States, despite its unstinted growth, lags far behind and will continue to do so. Although in 1950 the 160 million inhabitants of the United States could be compared to the 360 in India and the 550 in China, in 2050 it will not, with 400 million, be comparable to the 1.5 billion Chinese, themselves overtaken by 1.6 billion Indians. As to the other 20 "large" countries, the scale must be changed in order to spot their trajectories (Fig. 77–11b) as they are crushed by the first two giants (Table 77–2).

But the recomposition is even more obvious with the ascending trajectories crossing the descending trajectories. Thus, a large country such as Russia, currently numbering a little more than 150 million inhabitants will have no more than 100 million in 2050, whereas Brazil, which was at exactly the same level as was Russia 10 years ago, will have 250 million by the middle of the century. Another spectacular example is Italy, which was more populated than was Pakistan in 1950 (47 million against 40 million) and will have only 43 million inhabitants in 2050 (less than in 1950), whereas Pakistan will have more than 340 million (eight times more, and nearly as many as the United States). The Congo, which was the least populated of the list in both 1950 and 2000, will overtake 15 countries of this short list in 2050, including reunified Germany, Japan, Russia, and Mexico. In 2050 the two "large" European countries (France and the United Kingdom) with permanent places on the UN Security Council are likely to be among the last on this list (just ahead of Italy).

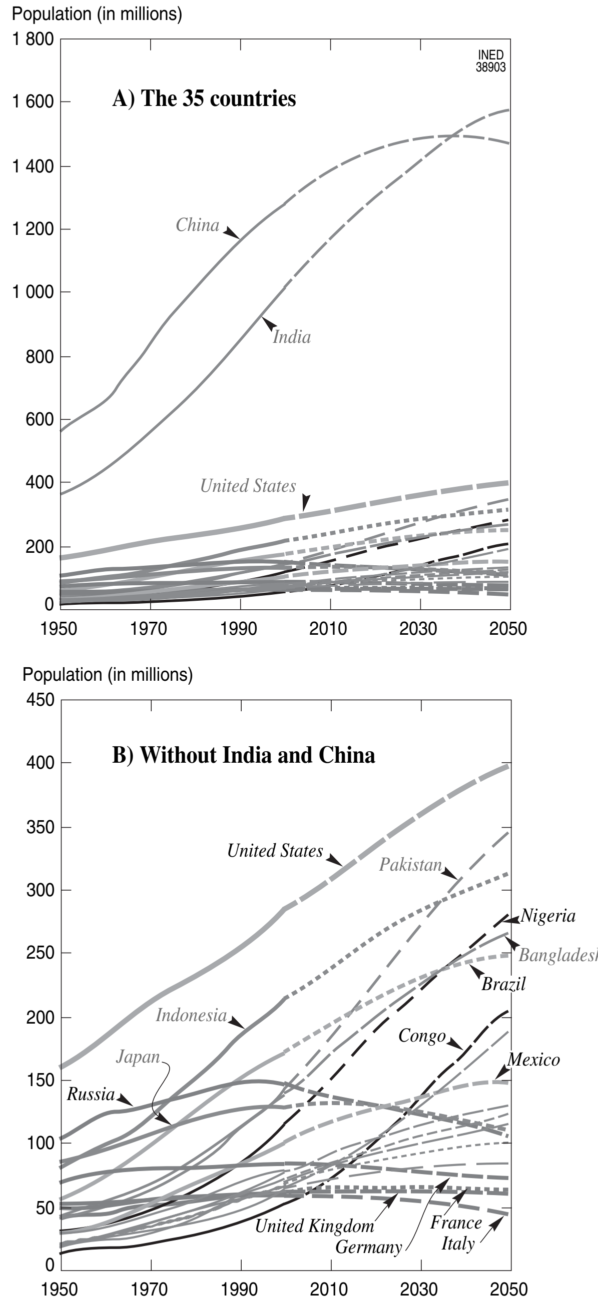


FIGURE 77–11 Population growth since 1950 and projection up to 2050 (medium hypothesis) of the population of 23 countries with more than 50 million inhabitants in 2000 (From United Nations, 2001).

II. DATA IMPROVEMENT, TECHNIQUE REFINEMENT, AND EVOLUTION OF THE HYPOTHESES

To fully appreciate the robustness of the UN projections, we can look at previous projections and compare them to the scenarios that actually occurred. By doing this, we are immediately struck by the success of the forecasts for the year 2000: at the end of the 1950s, the central scenario of the Population Division had forecast the 6 billion that was actually reached. Yet the data quality and techniques were far from the standard of today, and the hypotheses have significantly changed with time. But paradoxically, despite this early success and despite the improvement

TABLE 77–2 Trend Since 1950 and the United Nations' Projection up to 2050 of the Population of 23 Countries with More Than 50 Million Inhabitants in 2000

Country	1950	1960	1970	1980	1990	2000	2010	2020	2030	2040	2050
China	555	657	831	999	1155	1275	1366	1446	1485	1490	1462
India	358	442	555	689	845	1009	1164	1291	1409	1503	1572
United States	158	186	210	230	255	283	309	334	358	379	397
Indonesia	80	96	120	150	182	212	238	262	283	299	311
Brazil	54	73	96	122	148	170	191	211	226	239	247
Russia	103	120	130	139	148	145	137	130	121	113	104
Pakistan	40	49	62	81	110	141	181	228	273	309	344
Bangladesh	42	52	66	85	110	137	168	198	223	246	265
Japan	84	94	104	117	124	127	128	126	121	116	109
Nigeria	30	37	48	64	86	114	147	184	220	249	279
Mexico	28	37	51	68	83	99	113	125	135	142	147
Germany	68	73	78	78	79	82	81	80	78	74	71
Vietnam	27	34	43	53	66	78	89	100	110	118	124
Philippines	20	27	37	48	61	76	90	101	113	122	128
Iran	17	22	29	39	58	70	81	94	104	114	121
Egypt	22	28	35	44	56	68	79	90	99	107	114
Turkey	21	28	35	45	56	67	75	83	90	95	99
Ethiopia	18	23	29	35	48	63	80	101	127	157	186
Thailand	20	27	36	46	55	63	70	75	80	82	82
United Kingdom	51	52	56	56	58	59	60	61	61	60	59
France	42	46	51	54	57	59	61	62	63	63	62
Italy	47	50	54	56	57	58	56	54	51	47	43
Congo	12	15	20	27	37	51	71	99	133	170	204

Data from United Nations, 2001.

in methods, it must be admitted that the future is more uncertain today than it was in the past.

1. The Extraordinary Success of the 2000 Forecast

Table 77–3 lists the results of the 16 successive world population projections published by the UN since 1958. The first ones were based on data estimated in 1955, 1963, 1968, 1973, and 1978, after which this exercise has been repeated regularly every 2 years. In 1958 (on the 1955 database), the Population Division forecast a population of 6.27 billion for 2000 in its central scenario; since then, in each successive projection, this result has fluctuated a little but never greatly from this initial vision, which is close to today's reality of 6.06 billion for the year 2000. It must be emphasized that the high point of 6.5 billion corresponding to the 1968 calculation could be considered an accident, the experts of that period having given in to slight but unfortunate pressure from the apocalyptical talk of the time on the population explosion of the Third World (Vallin, 1986). Except for this high point, the trend revealed in Fig. 77–12 is eloquent. At the end of the period, a slight reduction in the estimates leading to the reality observed in 2000 is just visible.

Of course, the closer the limit the more likely that the range straddling the estimated final result is reduced; the most remarkable fact is that the range appears today as being nearly excessively cautious. Only in 1968 did the result of the medium hypothesis really vary from the center of any of the preceding estimates. In retrospect and taking into account the difficulties currently faced in formulating good hypotheses of a population projection, this result is hardly believable and cannot be put down to sheer luck as, although repeated 18 times with increasingly effective methods, the exercise has never contradicted the first attempt.

2. The Evolution of the Projections

It is true that the 1958 attempt was not exactly the first. It was preceded by two attempts, the results of which appear decidedly less brilliant. The first was published in 1951 based on 1946–1948 data; the second in 1954, on 1950 data. At the time, they did not dare to examine the 2000 horizon, but this caution was certainly genuine in the context of the fragility of the data (and also the hypotheses). At the end date of these first projections, 1980, the world population was estimated to be at 3.3 billion in the 1951 projection and barely

TABLE 77–3 Population Estimates on January 1, 2000, According to Successive Projections Published by the United Nations Since 1958

Hypothesis	Publication year														
	1958[a]	**1966**	**1973**	**1979**	**1979**	**1981**	**1985**	**1986**	**1989**	**1991**	**1993**	**1994**	**1997**	**1999**	**2001**
	Starting year of the projection														
	1955	**1963**	**1968**	**1973**	**1978**	**1980**	**1982**	**1984**	**1988**	**1990**	**1992**	**1994**	**1996**	**1998**	**2000**
World total															
High	6900	6994		6638	6508	6336	6366	6341	6411	6420	6301	6235	6123	6082	
Medium	6280	6129	6494	6254	6199	6119	6127	6122	6251	6261	6229	6158	6091	6055	6057
Low	4880	5448		5840	5855	5837	5899	5827	6089	6093	6151	6081	6062	6028	
Developed countries															
High	1495	1574		1434	1319	1304	1316	1309	1283	1286	1295	1198	1192	1193	
Medium	1438	1441	1454	1360	1272	1272	1276	1277	1262	1264	1278	1186	1186	1188	1191
Low	1241	1293		1307	1229	1233	1238	1247	1244	1246	1260	1176	1183	1184	
Developing countries															
High	5405	5420	5650	5204	5189	5033	5050	5031	5128	5134	5005	5037	4931	4889	
Medium	4842	4688	5040	4894	4926	4847	4851	4845	4989	4997	4950	4973	4904	4867	4865
Low	3639	4155	4523	4531	4626	4604	4660	4680	4844	4847	4891	4906	4879	4843	

[a] The 1958 projection does not present results for all of the developed and developing countries. To estimate the population in the first category, we have regrouped the results given for North America, Europe, the Soviet Union, Japan, and Australia—New Zealand and obtained that of the developing countries by the difference with the world total.

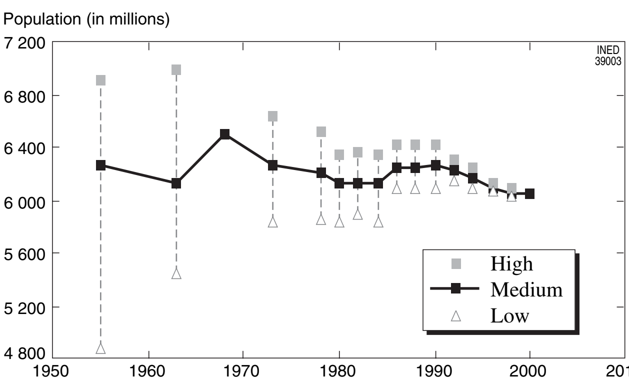


FIGURE 77–12 Estimates of world population on January 1, 2000, according to successive projections published by the United Nations since 1958. High, low, and medium variants of the principle scenario.

higher (3.6 billion) in the projection from 1954. The population actually observed in 1980 was 4.4 billion. Although the trajectories suggested by the central scenario of the following projections are close, nearly exactly superimposing any estimates made today using real data, these two first attempts appear as failures (Fig. 77–13). Not only were the available data terribly fragmentary and often deficient, but in particular (and this partly explains the failure), the importance of the ongoing mortality decline and real fertility levels in the Third World had not yet been taken into account.

Significant progress has been made in several directions since then. The projection method and the technical tools for producing it have improved, but in particular, the quantity and quality of the data have greatly progressed and the hypotheses have become more reliable.

a. Technological Refinement

There is no need to discuss the technical aspects in detail. We have already mentioned that with the adoption of the component method, the basics were in place since 1958 in terms that hardly differ from the algorithm currently used, even though the efficient use of computers has greatly facilitated its application. Even the failed attempts of 1951 and 1954 were basically using the same method. The essentials are shown in the two following points.

b. Improvement in Basic Data

Considerable efforts have been deployed since World War II in the direction of the developing countries (as much by international organizations as in the context of bilateral cooperation) for collecting data aimed at providing information on the demographic situation in most of the Third World countries (partic-

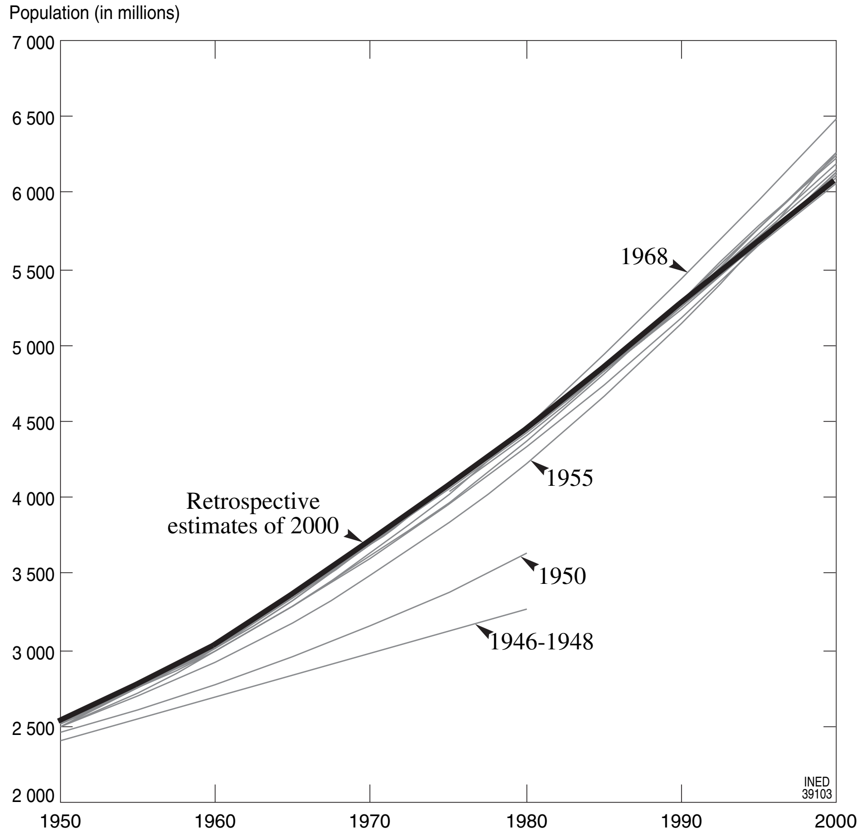


FIGURE 77–13 Trajectories of the first attempts of 1951 and 1954 compared with those of projections made from 1958 and with estimates made today for the 1950–2000 period (The projections are marked by their starting year).

ularly those of sub-Saharan Africa and several countries in Asia and Latin America) owing to the paucity of censuses and civil registration systems enabling births and deaths to be continuously followed. From the 1950s the United Kingdom and France tried to fill these gaps with a series of retrospective surveys in their African colonies, in which more sophisticated approaches were used such as multiround surveys and in-depth studies into women's fertility histories (see Chapter 121). At the same time, indirect analysis and estimation methods were perfected in order to get the best of deficient and fragmentary data (see Chapter 127). From the 1960s, significant efforts were also made by the UN so that the majority of countries could conduct a census every 10 years. This is how numerous Third World countries were able to hold their first population censuses at the beginning of the 1970s and the 1980s. From the 1970s, the World Fertility Survey program, succeeded since the 1980s by the Demographic and Health Survey program, has supplied an invaluable wealth of data on fertility and child mortality, as well as on their determinants. Let us also remember the huge effort of training statisticians and demographers that has enabled southern countries to contribute increasingly to the production and analysis of this demographic data.

Very rapidly, these efforts bore their first fruit, which were decisive in understanding the demographic situation at the start of a population projection, and here no doubt is one of the main sources of the impressive qualitative leap observed between the 1954 and the 1958 projection. Of course, in 1958 there were still enormous gaps that have only since been progressively filled, but in less than 5 years much had been learned about the mortality decline that had started in the poor countries and their levels of fertility. However, the correct trend hypotheses had yet to be made.

c. The Transition Theory

The experts were greatly assisted in this by the emerging theory of demographic transition (see Chapters 68, 69). As soon as the first data announcing this formidable population dynamic that would develop in the Third World countries began to take shape, the authors did not fail to hastily commit themselves in the description of apocalyptic perspectives based on the simple extrapolation of growth rates, themselves in full growth, as Edward East and George Knibbs had already done in the 1920s. The component method in itself gave the experts more confidence, but in order not to give in to the pressure of catastrophism, they needed to be able to formulate reasonable hypotheses not only on the progress of mortality decline but also on the approaching fertility decline in the countries in which fertility had been high and was likely to increase under the effects of progress in health and behavioral changes regarding breast feeding and *postpartum* sexual intercourse (see Chapter 69).

This is what enabled the demographic transition theory to be applied (which seems to us to be nearly a truism today): with the modernization of societies and behaviors, nearly all populations experience a phase of rapid growth that is not permanent, as mortality decline ends up being followed and even overtaken by fertility decline, which restores the balance between the two components of natural growth.

Of course, the authors of the UN projection in 1958 were careful not to refer to a future generalization of the demographic transition theory model or even to deduct the future stabilization of the world population that could result from it (this would appear with the first long-term attempts at population projection), but they used it to imagine the starting dates and the rates of fertility decline, which made it possible to announce such a reasonable population number for 2000 that it indeed became reality.

It would obviously be naïve to believe that all the starting dates and all the rates of decline chosen (varied according to country categories) were correct, but two major phenomena gave a consistency to the whole. On the one hand, the component method enabled a balance between errors of opposite signs (whether it be on the estimate of initial population or on the evolution of fertility and mortality), thus limiting the global error. On the other hand, limiting the exploration to a 40-year period meant a reduced global error was less likely to have a major effect.

But the third reason for the success of the 1958 projection was that the proportion of the world population likely to suffer from extremely erroneous hypotheses was also the population that was about to experience a considerable reduction in weight as a result of the ongoing evolution.

3. A Future More Uncertain Than Ever

Figure 77–14 shows how poor the quality of the 1958 projection would have been if it had forecast the future of the developed countries only. The result finally

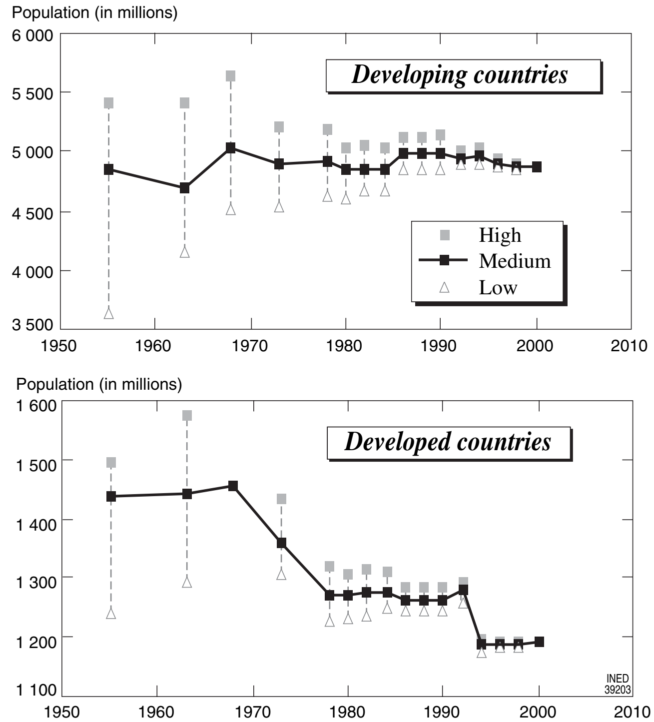


FIGURE 77–14 Population estimates of developing and developed countries on January 1, 2000, according to successive projections published by the United Nations since 1958. High, low, and medium variants of the principal scenario.

observed in 2000 for these countries at 1.2 billion is lower by 300 million than the expected figure in the medium hypothesis—exactly a quarter less! Although it was easy to predict the fertility decline of developing countries through the demographic transition theory, this theory was incapable of predicting the fertility trends observed in developed countries after 1970. On the contrary, at that time everything seemed to indicate that far from being a simple short-term phenomenon associated with the war, the *baby boom* era of the 1940s reflected a permanent behavioral change, in other words, the end of secular fertility decline and its near stabilization around the replacement level.

From the end of the 1970s, the UN experts attempted to adjust their aim, but it was insufficient. It was only with the 1994 projection that they finally approached the reality of 2000 as closely as they had done for world population from 1958. Since 1958 the projection for the developing countries has been consistently very close to the final result.

This observation calls for great caution concerning projections made today for the coming half century. If for a certain fraction of the world population that is still experiencing relatively high levels of fertility and therefore quite far from achieving its demographic transition the theory of the same name continues to provide quite

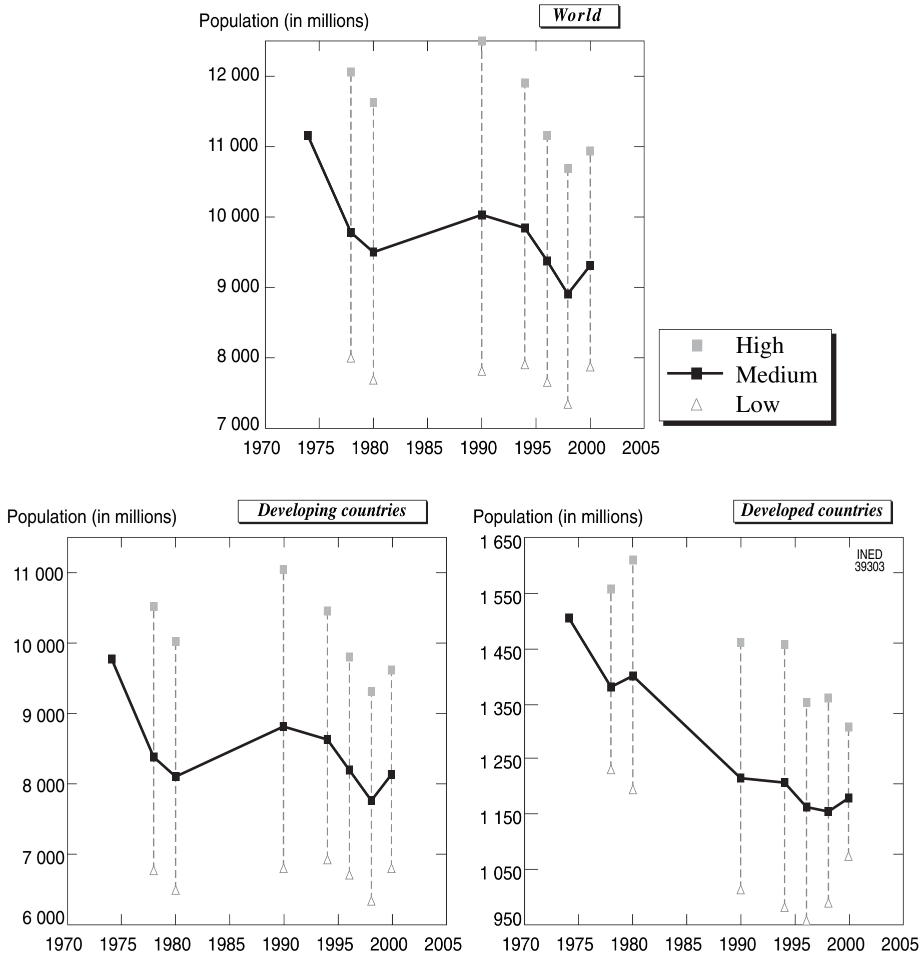


FIGURE 77–15 Population estimates for January 1, 2050, according to successive projections published by the United Nations since 1975. High, low, and medium variants of the principle scenario.

a reliable framework for proposing realistic fertility hypotheses that largely override all other considerations, the section of the world population that it represents is increasingly dwindling, so much so that the future of the greatest numbers tends to depend more and more on the same uncertainties that previously affected the so-called developed countries. Not only is there no longer a general theory capable of indicating the direction in which fertility should significantly evolve, but the future of mortality, which as we have seen weighs heavily on the projections, is the subject of a lively debate, without mentioning migration.

It is only since 1994 that the UN has been systematically exploring the 2050 horizon in their biannual projections. We can, however, connect the results published for the same period in their various long-term projection attempts, which we will discuss below. Table 77–4 assembles the available data; the first United Nation publication announcing a world population for 2050 dates from 1975.

We obviously do not know what the real result will be, but the evolution in numbers successively announced in the context of the medium scenario has already been significantly revised downward, from the 11.1 billion announced in 1975 to the 9.3 billion resulting from the current 2000 projection (Fig. 77–15). We can certainly consider that the 1975 projection was too long term and that slight errors in the increase of the fertility and mortality trajectories were enough to cause significant gaps in the forecast numbers. What is most worrying is that during the 1990s only the forecast had to be reduced by nearly a billion (despite the upward revision of 2001, which thus appears a little artificial). One wonders if the UN is not seriously underestimating the potential fertility decline beyond the replacement level, not only in the European countries for which a recovery in fertility is systematically expected but also in the numerous developing countries close to this level for which a drop below the replacement level is never forecast (in the medium

TABLE 77–4 Population Estimates on January 1, 2000, According to Successive Projections Published by the United Nations Since 1975

Hypotheses	Publication year[a]							
	1974	1981	1982	1992	1994	1997	1999	2001
	Reference year							
	1974[b]	1978	1980	1990[c]	1994	1996	1998	2000
World total								
High		12,076	11,629	12,506	11,912	11,156	10,673	10,934
Medium	11,163	9,775	9,513	10,019	9,833	9,367	8,909	9,322
Low		8,004	7,687	7,813	7,918	7,662	7,343	7,866
Most developed regions								
High		1,559	1,610	1,460	1,456	1,352	1,361	1,309
Medium	1,454	1,381	1,402	1,217	1,207	1,162	1,155	1,181
Low		1,232	1,194	1,013	981	959	990	1,075
Least developed regions								
High		10,517	10,018	11,047	10,457	9,805	9,313	9,625
Medium	9,709	8,394	8,111	8,802	8,626	8,205	7,754	8,141
Low		6,773	6,493	6,800	6,937	6,703	6,353	6,791

[a] The 1994 to 2001 projections are the current ones on the 2050 horizon. The preceding projections are long-term projections, which will be discussed in the following point.

[b] The 1974 projection does not distinguish a northern group (3084 million); therefore, east Asia (1760), not including Japan (130), must be subtracted in order to find the developed countries and a southern group (8079), which must be inversely increased to obtain the population of the developing countries.

[c] The 1990 projection does not distinguish the developed countries from the developing countries but a group I and a group II assemble the entire continents. For comparison, we have excluded Japan from group II and added eight Asian republics of the former Soviet Union and inversely for group I. The 1990 projection does not include results per country; we have therefore taken the 1994 projection estimates for these countries.

scenario).[9] But nothing prevents us from making a radically opposite criticism of this scenario tomorrow as we are currently incapable of devising a reliable theoretical framework to support a medium-term prediction.

III. DO ALTERNATIVE APPROACHES EXIST?

Another way to judge the pertinence of the UN work on projections is to look at the competition. During the 1950s and the first half of the 1960s, the UN's Population Division was the only producer of world population projections. However, at the end of the 1960s, two authors, Aron Boyarsky and Donald Bogue, took up the issue and were each emboldened to produce a potential alternative to the UN's work, even though they never really convinced anyone. Shortly afterward, Thomas Frejka picked up the torch, not only proving that it was possible to usefully explore the future beyond 2000 but also conceiving extremely long-term projections, based entirely on the idea of a generalization of the demographic transition conducted to the very end, and attracting in his wake the first attempts of the same type by the UN experts. During the 1970s and 1980s, institutions other than the UN also attempted to compete with the Population Division's projections.

1. Boyarsky and Bogue

Aron Boyarsky's (1965) and Donald Bogue's (1967) approaches differed significantly from that of the UN. The quality requirement of the source data used and the calculation procedures are much less elaborate.

Aron Boyarsky worked at the level of large regions, dividing the world into 12 more or less homogenous areas with respect to their political and social systems and their degree of economic development. He estimated growth coefficients for each region for the 1960–2000 period based on the idea that mortality and fertility depended largely on the economic and social

[9] In its last revision, the United Nations' projection (2003bc,d) acknowledges the possibility of a decline in fertility below 2.1 in some developing countries.

context. He thus estimated, for example, that the expected growth in the advanced capitalist countries would be less than that in the socialist countries, benefiting from better conditions for a reduction in mortality and the maintenance of higher fertility. Of course, he estimated that the growth capacity of developing countries was greater where the potential of mortality reduction was much higher and the perspective of controlling fertility more distant. By combining the results thus obtained by region, he estimated that in 2000 the world population would be between 4.2 and 5.0 billion, significantly less than the 6 billion forecast in the UN's medium perspectives, which turned out to be correct. Besides this obvious error of appreciation, the major fault in Aron Boyarsky's attempt was to follow a path that rejected the component method and lent itself much more to the rough growth rate projections of the past. The only improvement with regards to past projections was that he adopted a regional rather than a global approach, but it still remains far behind the UN's methodical approach by country.

Donald Bogue (1966, 1967) based his work on the general principle that the observed mechanisms that determine fertility in developing countries would soon be subject to a rapid change. He believed that the large majority of these countries' citizens wished to reduce their births, that the national elite and the political leaders were increasingly in favor of developing programs to limit births, and that modern contraception methods—soon to benefit from new technological progress—would easily spread among the population and be adopted by the majority. He also thought that further reductions in mortality would be increasingly difficult to obtain. He concluded that the population growth rate would decrease so rapidly that zero growth could be attained by the year 2000 and that the problem of rapid population growth would only affect a few small backward regions. According to his projection, world population would only reach 4.5 billion in 2000. Even less so than Boyarsky's work, Bogue's attempt could not compete with the UN projections: not only was his prediction of the future no better, but the arguments used were quite naïve and based on an equally coarse method. Donald Bogue, with Ami Ong Tsui, took up the question again in 1979 with more vigor and the same desire to expose the UN overestimate and did not hesitate to write, "We predict that by the year 2025, the world will have nearly achieved zero population growth. It is estimated that this equilibrium will be achieved with a world population of about 7.4 billions." (Bogue and Tsui, 1979:99). This new publication attracted a lot of attention and incited a certain controversy. Giving an estimate of between 5.75 and 5.97 for the year 2000, far below the final result and only 20 years before that date was not particularly perspicacious.

2. Thomas Frejka

Thomas Frejka's contribution (1973) was much more decisive. He convinced the UN to look beyond the year 2000 for the first time, and he exposed all the potential advantages of a very long term logic to better base the hypotheses useful to short- and mid-term projections.

Even though they more or less integrated the teachings of the demographic transition theory, the UN's projections have always given priority to the analysis of past trends confronted with the current political and social context in order to deduct the most reasonable possible future trends of mortality and fertility without too many theoretical apriorisms. On the contrary, Thomas Frejka began with the idea that the demographic transition theory could usefully be applied up to its supposed limit, a stabilization based on a generalized convergence toward zero growth. Instead of considering the future toward which such and such an evolution considered plausible of components of population dynamics could lead us, he considered which trends of these components could lead to this stabilization and in what time limit. Thomas Frejka also accentuated the benefit of producing one or more scenarios considered very plausible side by side with extremely improbable scenarios, this made possible the illustration of the force of current changes and the unlikely results of hypotheses sometimes retained by politicians or public opinion. By describing the improbable as plausible, his intention was for each person to be able to form a considered opinion on the demographic future of the planet. In concrete terms he showed that in the year 2000, which is of interest here, the world population could be 4.7, 5.1, 5.9, 6.4, or 6.6 billion if the net reproduction reached 1 in 1970 (his projection beginning at 1965), in 1980, in 2000, in 2020, or in 2040. It was up to each person to determine at what point reaching this objective was most plausible in order to deduct the estimated populations at the chosen year (2000 here, but it could as easily be 2050 or 2100 as the exercise was resolutely based on the long term, as we will discuss further below). This was doubtless less comfortable than the UN scenarios and no doubt the author can be reproached for this excess of freedom to the reader, but it is perhaps more stimulating because of it. Thomas Frejka's work is a complement to the UN's work and an encouragement to widen the field rather than a real alternative.

TABLE 77–5 Population Projections for 2000 Published before 1980, According to Author

Authors	Year of		Hypotheses						
	Publication	Reference	Classic			Frejka: if R = 1 from			
			High	Medium	Low	2040	2020	2000	1980
Pearl	1924	1914		1963					
Knibbs	1928	1928		3900					
Pearl and Gould	1936	1931		2459					
Notestein	1945	1945		3300					
UN	1957	1955	6900	6280	4880				
UN	1963	1960	6994	6130	5449				
Boyarsky	1965	1960	5036	4626	4216				
Bogue	1966	1960		4527					
UN	1968	1965	7104	6494	5977				
Frejka	1973	1970				6670	6422	5923	5116
UN	1973	1970	6638	6254	5840				
UN	1974	1970		6406					
Littman-Keyfitz	1977	1975		5882					
UN	1978	1975	6509	6199	5856				
World Bank	1978	1975		6064					
Bogue and Tsui	1979	1975	5972	5883	5756				
Census Bureau	1979	1975	6797	6350	5921				
UN	1980	1975	6340	6121	5838				
World Bank	1980	1980		6015					
Frejka and Mauldin	1980	1980				6357	6234	5930	5333

R, net reproduction rate; UN, United Nations.

3. The World Bank, the Census Bureau

The attempts made from the end of the 1970s by the World Bank and the United States Bureau of the Census to compete with the UN projections are entirely different. Indeed, they are more variations on the same theme than truly original contributions.[10] The repetition of the same type of protocols on the same data produced very similar results, which from a scientific point of view was not without interest. In a slightly more impertinent interpretation, one could say that although the World Bank (Zacharia and Vu, 1978, 1980; Vu, 1985) did good work in predicting 6.02 billion in 2000, the Census Bureau (1979) exaggerated a little with their 6.35 billion in a medium hypothesis for reasons that were perhaps not totally innocent.

The contemporary work of G. Littman and Nathan Keyfitz (1977) was no more innovative and concluded with a result for 2000 that was a little further from reality than that of the UN (5.88 billion).

Table 77–5 and Fig. 77–16 synthesize the results of population projections for the year 2000 published before 1980, according to author and hypothesis. The medium hypothesis of the UN projections have always been closest to reality, with one exception, the World Bank's 1975 projection, which did barely more than pick up on the basic estimates of the UN's hypotheses.

Donella Meadows *et al.* (1972)[11] and then Mihajlo Masarovic and Eduard Pestel (1974) attempted a radical innovation by developing complex models that combined demographic projections with economic, social, environmental, and political trends, with the objective of revealing that the population trends predicted by the UN would necessarily lead to an explosion of the world system during the 21st century, causing an increase in mortality and a rapid population decline. The authors concluded that if this explosion did not finally occur, it would be because fertility would have declined far more rapidly than predicted by the UN. Although the holistic approach deserves to be emphasized, the ideological presumptions significantly reduce its scientific interest.

[10] The U.S. Bureau of the Census (McDevitt 1996; McDevitt *et al.*, 1999) was the first to retain a hypothesis of an extension in fertility decline below replacement level for Arab and Muslim countries (Morocco, Tunisia, Turkey, Iran, etc.), even before the United Nations decided to take this type of hypotheses into account in its last revision (United Nations, 2003b).

[11] See also Delaunay (1972; in French).

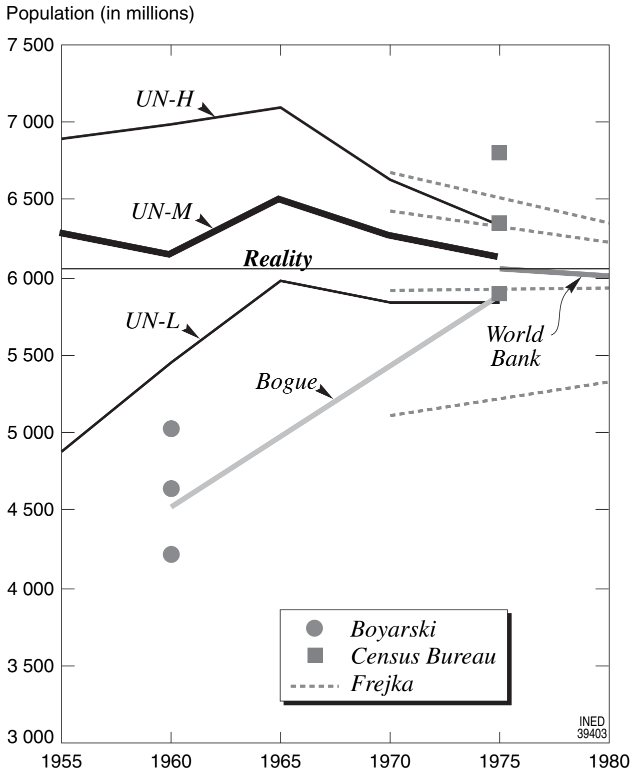


FIGURE 77–16 Comparison of the results for 2000 from the various projections published between 1950 and 1980.

4. International Institute for Applied Systems Analysis

In 1994 International Institute for Applied Systems Analysis (IIASA) published world population projections based on a multiregional model of components by cohort (Lutz, 1994). One of the advantages of these projections is no doubt that a larger number of scenarios was explored by combining different migration, mortality, and fertility hypotheses. Table 77–6 summarizes the results for 2050 published in 1994. By taking the central scenario and combining it with the central hypotheses of migration, mortality, and fertility, a total world population of 11.3 billion was found, against 14.9 billion in the scenario of high fertility and mortality and low mortality and against 8.3 billion in the inverse scenario. The central estimate is much higher than what the UN predicted in its medium scenario in its 1994 projection (11.3 billion instead of 9.8).

Two years later, IIASA (Lutz, 1996) published a new version of these projections, giving clearly different results in which the central scenario (9.9 billion in 2050) was considerably closer to the medium scenario of the UN 1996 projection (9.4 billion). The declared margin of uncertainty was greater as the high fertility scenario (with constant mortality and migration) gave a much higher result than did the preceding scenario (19.0 billion); the low-fertility scenario (still with constant mortality and migration) gave an even lower result (5.1 billion).

More recently, IIASA revived its projection work on a new foundation aimed at completely reforming projection methodology based on an approach of probability (Lee and Tuljapurkar, 1994; Alho, 1997; Lee, 1998; Lutz and Scherbov, 1998; Lutz *et al.,* 1998). The first results were featured in an article published in *Nature* (Lutz *et al.*, 1997) and are available on the Web (*www.iiasa.ac.at*). The previous results were also revised downward significantly: this time the central scenario gives an estimate for 2050 (8.8 billion) that is significantly lower than that of the UN's 2000 projection (9.3 billion). This promising work is, however, too recent to be compared retrospectively to reality, as has been done here with the UN's projections.

IV. FROM MEDIUM TO LONG TERM

As stated above, the credit for showing the benefit and feasibility of long-term world population projections goes to Thomas Frejka (1973). Immediately after the publication of his work, the UN integrated this new dimension into their projection work and published their first attempt in 1974 (UN, 1974a,b 1975). The Population Division then repeated the operation several times, but less regularly than for mid-term

TABLE 77–6 Summarized Results of the International Institute for Applied Systems Analysis's World Population Projection (1994) for 2050

	Hypotheses								
Migration	Central	High	High	High	High	Low	Low	Low	Low
Mortality	Central	High	High	Low	Low	High	High	Low	Low
Fertility	Central	High	Low	High	Low	High	Low	High	Low
Developing countries	9,958	11,037	7,142	13,144	8,495	11,326	7,359	13,473	8,748
Industrialized countries	1,349	1,655	1,181	1,812	1,328	1,387	947	1,542	1,092
World total	11,307	12,691	8,323	14,957	9,823	12,713	8,307	15,015	9,840

All figures are in millions.

projections (UN, 1981, 1982a,b 1992, 1998). In 1974 and in 1982, the targeted year was 2100, the end of the 21st century; in 1992 and 1998, it was extended to 2150.[12]

The priority is no longer to propose as realistic a view as possible on the short- and mid-term future of world population but to illustrate the consequences of a more long-term extension of current trends or those expected in the coming decades. As Thomas Frejka demonstrated, this exercise enables us to imagine what could be the end of the current demographic transition currently generalized at a global level (see Chapter 69).

This is to some extent the role played here by the medium scenario, which shows all the countries of the world toward their maximum life expectancy, with fertility stabilizing at a level close to replacement in 2050–2055 at the latest and supposing that all net international migration would be zero after 2025.

The projection contains seven scenarios that only differ in their fertility hypotheses. In all countries mortality should fall and life expectancy is supposed to increase to the maximum level of 87.5 years for men and 92.5 years for women, at the rate already established in the medium scenario of the projections for 2050. Let us note in passing that for long-term projections, the reference limits of the projection for 2050 (82.5 and 87.5 years, according to sex) were increased, which is proof that the limit is not considered as an absolute but as a pragmatic reference point in order to converge life expectancies. Similarly, there is only one migration hypothesis, which is the application of the trends chosen in the projection for 2050 until 2025 and afterward a constant balance of zero.

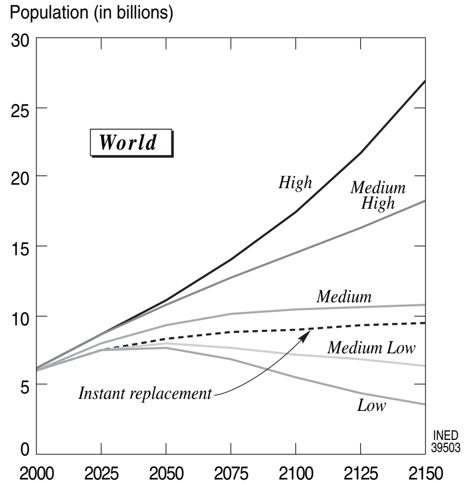
In contrast, fertility is the subject of seven different hypotheses. The medium, high, and low hypotheses are a prolongation of those established for the projection to 2050 in the context of the *1996 revision.* In this case the medium hypothesis differs significantly in this aspect from the *2000 revision*, which was presented at the beginning of this chapter, as it supposes that all countries will converge toward the replacement level in 2050–2055 at the latest and that fertility will then remain constant until the end of the projection period (UN, 1998:8). This provision results in the medium scenario forecasting a direction of *stationarity* (see Chapter 20 of Volume I), in other words toward complete stabilization in the number and the structure of the population. In addition to these three basic hypotheses, the 1998 long-term projection included four other hypotheses. Two of them (called medium-low and medium-high) were justified by the fact that in time the divergence would have increased so much that it was interesting to look at the intermediate trajectories. There are thus five exploratory hypotheses of potential future realities. The objective of the two last hypotheses, as in the 2050 projection, is pedagogic, enabling a comparison of the results of the first five either with fertility maintained constantly at the initial level or with fertility reaching the replacement level immediately (at the beginning of the projection period).

In the context of the medium scenario, the world population would be 10.8 billion in 2150, a period at which it would be practically stable, 10 billion having been exceeded in 2075. The range between the low and high hypotheses obviously widens exponentially (Fig. 77–16). In 2150 it would rank from 3.6 billion to 27 billion, the intermediary hypotheses for the same date being 6.4 and 18.3 billion, respectively. Thus, it is clearly seen that, with time, relatively modest differences in fertility[13] can produce huge gaps in terms of population. But this range, large as it may be, is negligible compared with the result if current fertility was held constant: 296 billion in 2150, including 295 billion for Asia, Africa, and Latin America and only 1 billion for the rest of the world. This number is entirely unlikely but is nevertheless of great pedagogic value: it shows the huge importance of the changes currently occurring. Yet, the calculation is based on fertility levels from 1990–1995 when average world fertility was already only 2.96 children per woman. The same exercise carried out by the UN in 1992, with a slightly higher world TFR (3.45 children per woman) gave a result of 700 billion human beings in 2150 (UN, 1992). The other lesson of such a perspective is to raise awareness of the vast disparities observed in demographic behavior. If the medium growth rate of the world population observed at the start of the projection (1.7% per year in 1985–1990, for 1992) had been forecast, world population would have certainly increased exponentially but would have reached only 70 billion in 2150 and not 700. This results from the fact that, with time, the most fertile proportion of the population increases considerably and the average world growth rate increases, progressively approaching the rates observed in the most fertile countries (Vallin, 1994).

Inversely, the instant replacement fertility scenario reminds us, as we have already stated regarding projections to 2050, of the existence of the growth potential concealed by a relatively young world population,

[12] The United Nations' Population Division (2003a) is currently planning to explore the demographic future of humanity over 3 centuries. A first meeting of experts was held in New York in June 2003.

[13] It results from the fertility hypotheses forecasting that in 2050–2055 the maximum observed gap between total fertility rates would only be between 1.6 and 2.6 children per woman.



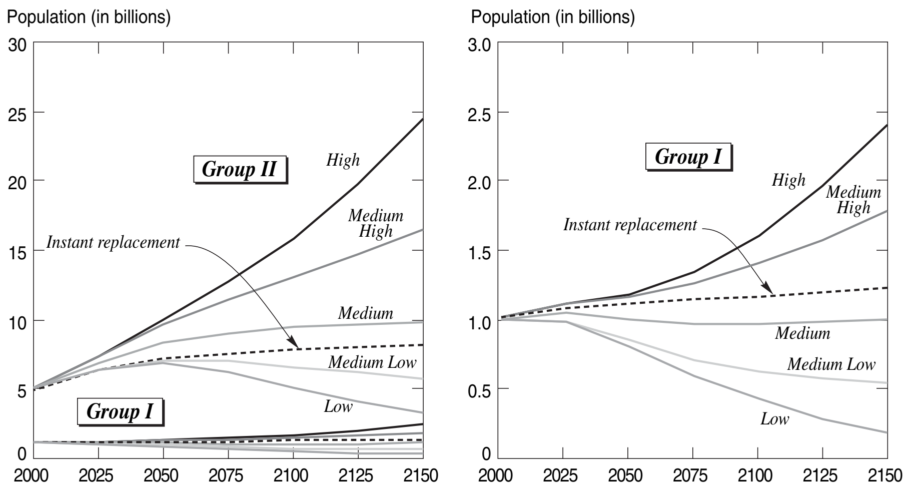


FIGURE 77–17 Population projection up to 2150 for the world and for country groups I and II, according to the United Nations' 1998 projection (group I: Europe, North America, and Oceania; group II: Asia, Africa, and Latin America).

but from this viewpoint, the long-term projection does not yield any more than does the projection to 2050, which is enough to deplete most of this potential.

Contrary to its usual practice, the UN did not divide the world into developed countries and developing countries in this projection but distinguished two groups: group I, which included Europe, North America, and Oceania; and group II, which included Asia, Africa, and Latin America. This does not facilitate comparison,[14] but the general idea is the same: to obtain a classic division between present-day developing countries and developed countries all that is needed is to transfer Japan from group II to group I and do the opposite for Melanesia, Micronesia, and Polynesia.

Figure 77–17 shows the evolution in total population from 2000 to 2150 for both groups according to the six first fertility hypotheses. In the context of the medium-hypothesis, the population of group II would stabilize at around 10 billion inhabitants in 2150, thus doubling its current number, but most of this would already have been acquired by 2075 with more than 9 billion. Group I would just about preserve its current population level of 1 billion, after a slight decline followed by an increase.

More striking than the huge demographic imbalance that will exist between these groups is its susceptibility to fertility hypotheses. Indeed, in the

[14] Not only with mid-term projections but also with the preceding long-term projection, which includes the whole of the former Soviet Union in group I, whereas in the 1998 projection only the European republics of the former Soviet Union were included in Europe.

context of the high hypothesis, group II would reach a population of 24.5 billion in 2150 compared with 2.4 billion for group I. The relative gap is the same in the preceding case, but in absolute numbers, the difference increases from 8 to 22 billion. In the low hypothesis, the population of group II would be reduced to 3.3 billion in 2150, whereas that of group I would literally crumble to 270 million. This time the relative gap widens slightly, but the absolute gap is much less (3 billion). But because all these scenarios are considered as being possible, cases in which both groups follow different scenarios must also be considered. At the extreme case therefore, a group II population of 24.5 billion (high hypothesis) in 2150 next to a group I population of 270 million (low hypothesis) could be imagined, as well as a world in 2150 consisting of a group II population of 3.3 billion and a group I of 2.4 billion. In other words, the so-called north–south demographic ratio of the planet could equally be 1 to 1.5 as 1 to 100 in the middle of the 22nd century.

Beyond these two groups of countries, the UN has also published results by large geographic areas, including India and China as the two largest countries in the world. Figure 77–18 illustrates population trends from the medium hypotheses for each of these large areas.

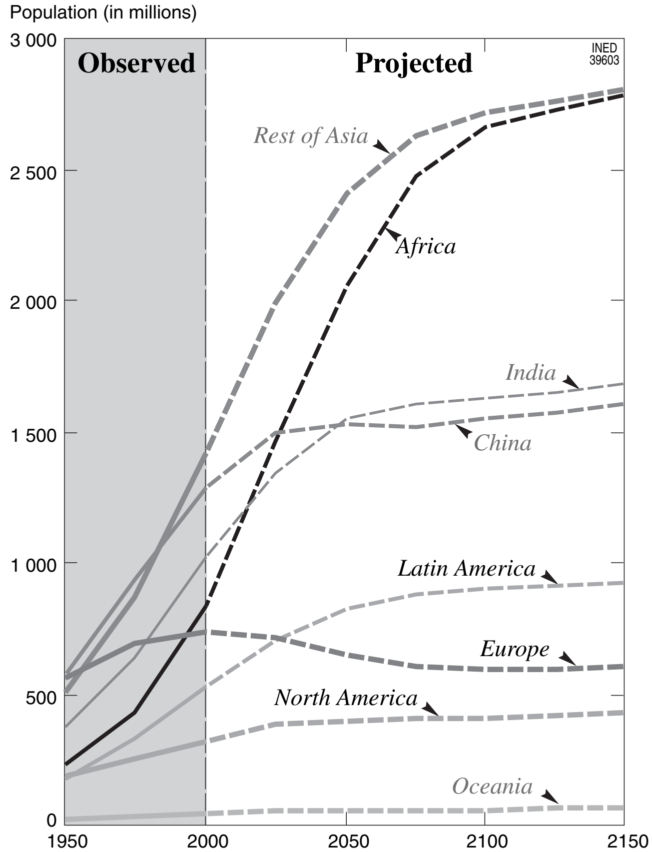


FIGURE 77–18 Trend since 1950 and projection up to 2150 of the population of eight large geographic areas of the world in the medium hypothesis of the United Nations' 1998 long-term projections.

If the general stabilization presumed by this hypothesis occurs, it is clear that all, or nearly all, of it would have taken place in the majority of these areas by 2050. Only Africa and, to a lesser extent, Asia excluding China and India will still experience a substantial growth after that. The population of Africa would increase by another 800 million inhabitants between 2050 and 2150, and Asia excluding China and India would increase by 400 million, with the two latter countries both reaching 2.8 billion inhabitants each.

As previously done for the projections to 2000, Table 77–7 and Fig. 77–19 compare the results for 2001 of the five long-term projections produced by the UN with the various projections published by other authors and organizations.[15] It is obviously not as yet possible to compare them to the real results awaiting us nearly a century from now.

Thomas Frejka's results can not be compared to the others as they do not provide one central scenario that is more probable than the others. The World Bank's results, which were limited to only one scenario (except in 1994) appear once more to be nearly a copy of the UN's medium hypothesis, a little shaky, however, if we judge by the fluctuations surrounding the trend indicated by the Population Division. In its central scenario IIASA's result is, as we mentioned earlier, significantly revised downward. But what is most striking here (Fig. 77–18) is that the range of imprecision varies enormously, notably under the effect of the strong fluctuation of the high hypothesis. All things considered, this comparison reinforces the idea that barring proof to the contrary, the UN's projections give the most complete and reliable context to speculations regarding the future of world population.[16]

[15] Two results from Thomas Frejka (Frejka, 1973; Frejka and Mauldin, 1980), six from the World Bank (Vu, 1984, 1985; Zacharia and Vu, 1988; Bulatao *et al.*, 1990; Bos*et al.*, 1992, 1994), and three from the International Institute for Applied Systems Analysis (Lutz, 1994, 1996; Lutz *et al.*, 1997).

[16] At least at a global level. Indeed, these projections are subject to significant errors at the level of countries considered individually, the global result depending largely on the compensation of opposing local errors.

TABLE 77–7 World Population in 2100 According to Various Long-Term Projections Compared with the United Nations Estimates (in Billions of Inhabitants)

Fertility hypothesis	Frejka	UN	WB	Frejka	UN	UN	WB	WB	WB	UN	WB	IIASA	WB	IIASA	UN	IIASA
	Publication year															
	1973	1974	1980	1980	1981	1982	1985	1988	1990	1992	1992	1994	1994	1996	1998	2000
	Reference year															
	1970	1973	1975	1980	1975	1980	1980	1987	1989	1990	1992	1994	1994	1996	1996	2000
Classic hypotheses																
High[a]					14.18	14.20				19.16		16.09	12.22	18.95	17.50	14
Medium-high										17.19					14.59	
Medium[b]		12.26	9.87	9.208	10.53	10.18	10.42	10.41	11.33	11.19	11.65	12.56	10.96	10.35	10.41	8.41
Medium-low										6.41					7.23	
Low[a]					8.03	7.72				6.01		9.13	9.38	5.13	5.58	4.2
Specific hypotheses																
Constant										10.94					57.18	
Replacement										8.09					9.04	
Growth						14.93										
Decline						7.25										
R = 1 in 2040	15.10			12.35												
R = 1 in 2020	11.17			13.43												
R = 1 in 2000	8.39			10.64												
R = 1 in 1980	6.42			8.54												
R = 1 in 1970	5.69			6.99												

More often than not, the projections include only one mortality and one migration hypothesis. When this is not the case, we have chosen the hypothesis called normal, central, and medium accordingly. WB, World Bank; UN, United Nations; IIASA, International Institute for Applied Systems Analysis.

[a] Slow decline and Rapid decline of fertility for the World Bank 1994; "Fractiles" 0.985 and 0.025, in IIASA's probabilistic projections.

[b] Unique hypothesis for the World Bank from 1980–1982, "standard" for the World Bank 1994, hypothesis called "plausible low" by Frejka, "central" by IIASA in 1994 and 1996, and "median" in IIASA's probabilistic projections.

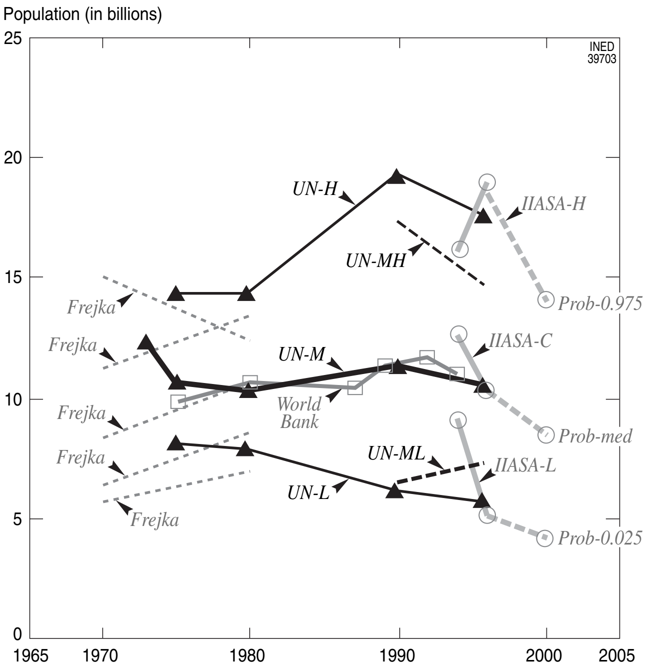


FIGURE 77–19 World population estimates for 2100 according to different sources compared with the United Nations' successive estimates (in billions).

V. PROJECTIONS BY AGE AND SEX AND DERIVED PROJECTIONS

Population projections obviously do not only aim at imagining the total population at a particular point. Reflection on the demographic future or, further still, using population projections for economic and social planning requires, for example, consideration of the sex and age structure, place of residence, and social and occupational category of this future population. The production of "derived perspectives" that enable these results to be reached requires work in the prolongation of the production of principal projections. When it is used, as is the case here of the component method, the projection of the age and sex structure is in fact a direct result of the main projection as it passes by the application of fertility, mortality, and migration data to population by age and sex calculated during successive stages of the process. To derive the principal projection toward other differentiation criteria, a special program must be developed according to new hypotheses, as was done in Chapter 73 by Marlène Lamy with regard to projections of population schooling. Here we will merely give an outline of the results by age and sex of the UN's projection and mention the specific projections that the UN develops quite regularly concerning urbanization.

1. Age and Sex Structure

After some examples of the possible transformations of the population pyramid between the present day and 2050, we will illustrate some hypothetical trajectories by large age groups up to 2150.

a. Evolution of the Population Pyramid by Age and Sex Up to 2050

Figure 77–20 compares the population pyramid of the world population in 2000 with what it could become in 2050 according to which of the three fertility variants of the principal scenario of the UN's 2000 projection occurs. While the world population in 2000, dominated by the developing countries that are still experiencing high growth, presents a young pyramid, wide at the base and tapering at the top, its profile would change completely in the medium hypothesis and even more so in the low hypothesis. With medium fertility the structure of the pyramid does not disappear entirely, but at its base the frequencies of the young age groups are only slightly higher than those of the adult age group; the top of the pyramid is considerably reinforced. In the context of the low hypothesis, fertility no longer ensures a replacement of the generations, the young age groups are on the contrary smaller than the adult age groups and only after age 70–74 years do we find frequencies lower than that of the age group 0–4. In the context of the high hypothesis, the pyramid structure becomes even clearer than the one for 2000, which is already very marked at the base by the decline in fertility. It is, however, wider at the top, becoming nearly triangular. Even though the hypothetical fertility for 2050 is lower than it currently is, it would actually be higher than the replacement level and nearly constant for several decades, which produces this structure of a nearly regular triangle.

All the following pyramids are based on the medium hypothesis to show that even in this scenario in which preference is given to a general stabilization, a wide variety of different situations will be observed in 2050. All these pyramids illustrate populations of unequal size[17] and are drawn on a relative scale (population proportions and not total frequencies by age group) in order to highlight the structural differences.

Figure 77–21 begins by comparing the world to the two large traditional groups, namely, developing countries and developed countries. Although the developing countries would have a pyramid hardly

[17] Including the pyramids in Figure 77–20 in order to be able to compare them with the ones following.

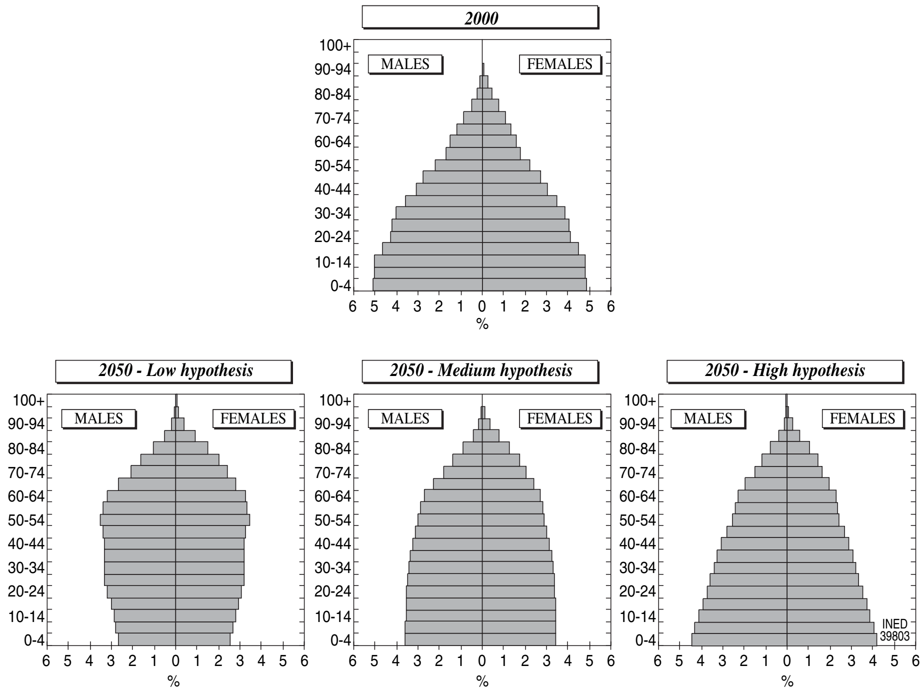


FIGURE 77–20 Transformation of the world population's age pyramid from 2000 to 2050 according to the three fertility variants of the United Nations' principal scenario.

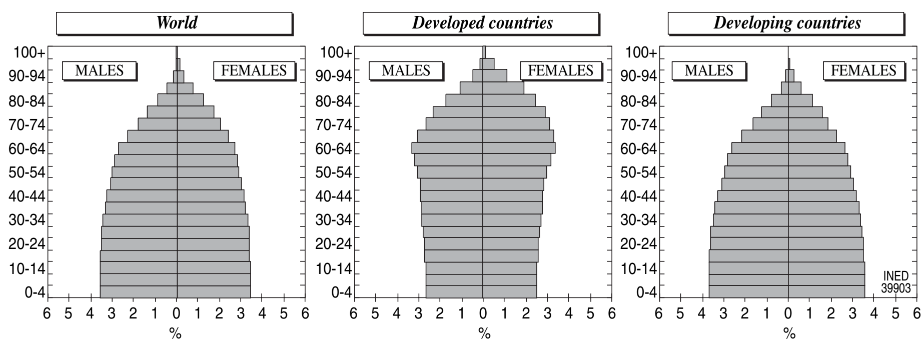


FIGURE 77–21 The world population pyramid in 2050 and those of the developed countries and developing countries, according to the medium variant of the United Nations' principal scenario.

different from the world pyramid (barely wider at the base), which is not surprising in view of their comparative weight, the developed countries would have an age structure dominated by the 50- to 75-year age groups. This population aging may cause serious economic and social problems.

However, this distinction between developing countries and developed countries will not make much sense in 2050 as the diversity of the first group will be so wide. This can be seen by comparing the continents or the large continental regions (Fig. 77–22). Although Africa would still have a very young pyramid, that will not be the case of any of the other subgroups of developing countries. Furthermore, organizing them according to profile as in Fig. 77–22 shows that the developed countries alternate with the younger structures of some developing countries; east Asia (which, although it includes Japan, is massively dominated by China) can be inserted between North America and Europe.

Figure 77–23 illustrates the population pyramids of seven countries that in 2050 will in the hypothesis (as Fig. 77–11 shows) become the most populated countries in the world (India, China, United States, Pakistan, Nigeria, Bangladesh and Indonesia), to which the European Union and Russia were added.

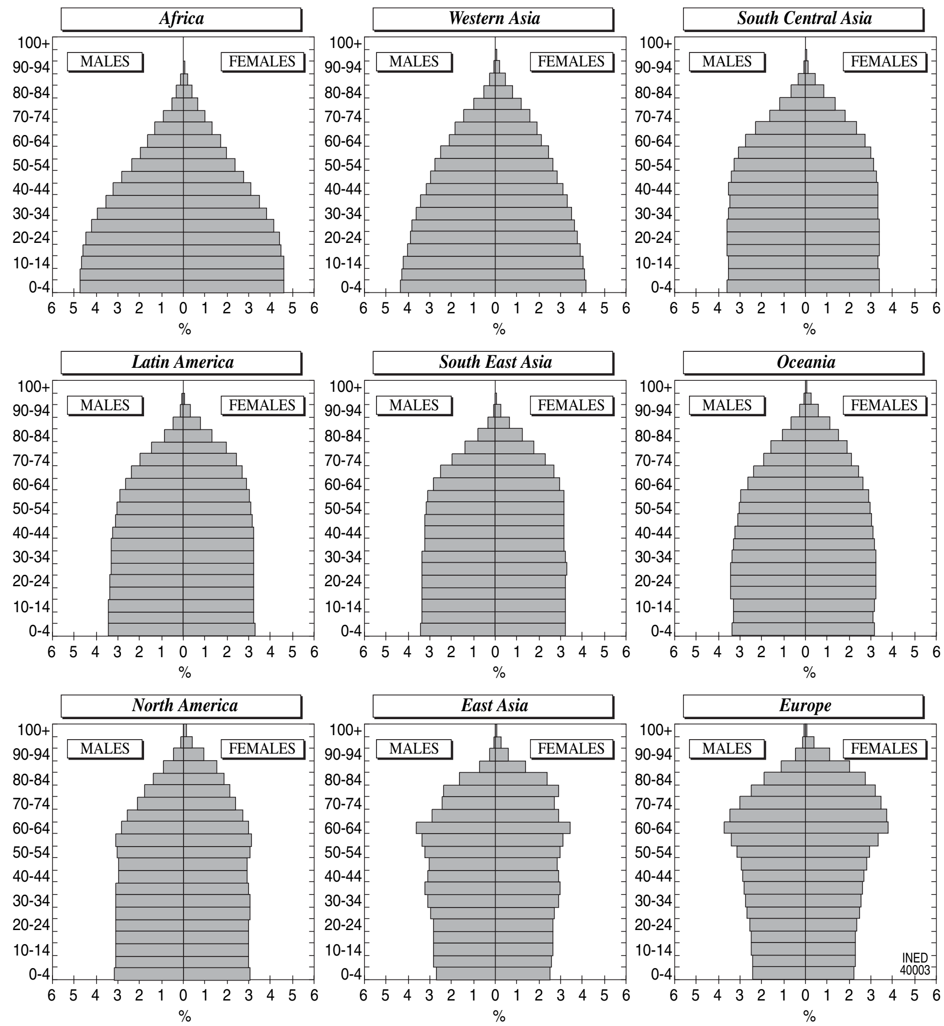


FIGURE 77–22 Population pyramid in 2050 of the population of the large regions of the world, according to the medium-fertility variant of the United Nations' principal scenario.

From Nigeria, which yet presents quite a different profile to the typical African one, to Russia, the diversity is even greater. Here again there is no systematic border between developed countries and developing countries: China is more similar to Europe than to India, whereas the United States resembles Indonesia more than it does the European Union.

b. Evolution by Large Age Groups Up to 2150

The UN does not publish the results of its very long-term projection by 5-year age groups. However, these results by large age groups are at our disposal and thus enable us to visualize the various possible trajectories for the young (0 to 14 years), adults (15 to 64 years), and the old (65 years and over). Table 77–7, illustrated by Fig. 77–24, presents the results of the three variants of the principal scenario.

In all three cases, the proportion of the young declines significantly, whereas that of the old increases rapidly. This general phenomenon is so strong that only in the high hypothesis does the proportion of the 65 years and over group remain constantly below the under-15 group to the end of the projection period. In the medium hypothesis, the two proportions become equal shortly before 2080, each consisting of 20% of the total population. In the low hypothesis, this would occur from 2045; the proportion of the young would fall very rapidly to 15%, whereas the proportion of the old would not yet have had the time to exceed this level (Fig. 77–24).

At the end of the projection period, the world population would be made up of 25% old people to 18% young people in the medium hypothesis, and these proportions would reach 35% against 12% only in the low hypothesis.

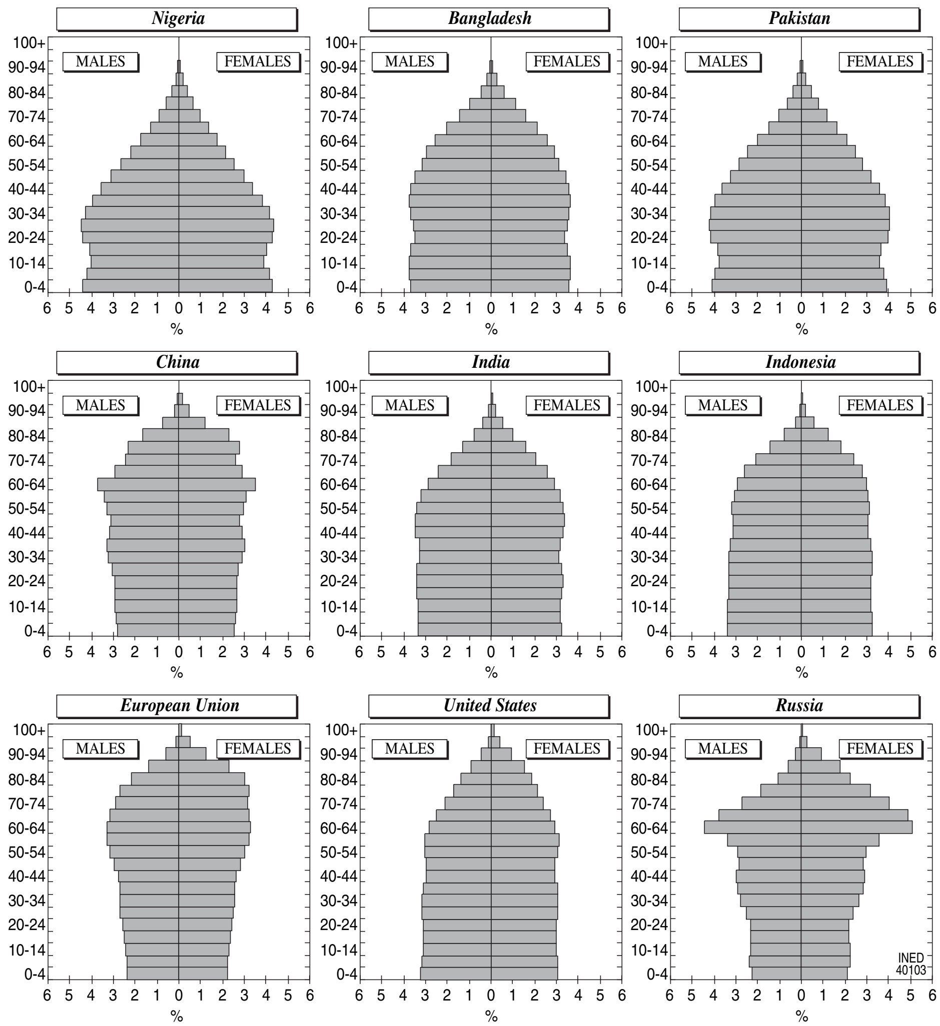


FIGURE 77–23 Population pyramid of the seven countries that will be the most populated plus the European Union and Russia in 2050, according to the medium-fertility variant of the United Nations' principal scenario.

Paradoxically, regardless of the hypothesis, the near symmetry in the evolution of these proportions of the young and old are accompanied by a strong fluctuation within the adult population (Fig. 77–25). With population aging it begins by increasing more or less strongly depending on the selected scenario, opening the way to a kind of golden age of population in which a high proportion of adults creates an excellent opportunity for economic development. In the medium hypothesis, the proportion of adults would reach nearly 66% in 2025 before falling relatively rapidly below the starting level and arriving at 58% at the end of the projection period. This phenomenon would obviously be less significant in the high hypothesis but much more marked in the low hypothesis, with a maximum of 69% around 2025 followed by a drop to 53%. This would apply to the developing countries in particular, of course. In the developed countries, whatever the scenario, the proportion of the adult population decreases from the beginning to the end of the projection period.

This phenomenon is similar to an oxygen supply for the developing countries, giving them an opportunity to invest today in social protection systems that will enable them in future to cope with the problem of population aging with less difficulty. But this is only an interlude, a window of hope that must not be missed if the coming risk of failure is to be avoided. It is even more important to build on it if the fertility decline is rapid. China is a typical example in which this demographic window of economic opportunity is large, as the proportion of adults will soon beat records, but the golden age will not last long since fertility declined very rapidly.

TABLE 77–8 2000–2150 Projection of the Proportions (%) of Children (0 to 14 Years), Adults (15 to 64 Years), and Old People (65 Years and Over) in the World Population and by Country Groups, According to the Three Fertility Hypotheses of the United Nations' Principal Scenario

Year	0–14 Years			15–64 Years			65+ Years		
	High	Medium	Low	High	Medium	Low	High	Medium	Low
World total									
2000	30.3	30.0	29.6	62.9	63.2	63.5	6.8	6.8	6.9
2025	27.1	24.3	20.8	63.5	65.7	68.5	9.4	10.0	10.7
2050	24.9	20.5	15.1	62.4	64.4	66.4	12.7	15.1	18.5
2075	24.2	19.0	13.1	61.2	61.9	60.7	14.6	19.1	26.2
2100	23.7	18.3	12.4	60.0	59.7	56.7	16.3	22.0	30.9
2125	23.4	17.9	12.0	59.3	58.5	55.0	17.3	23.6	33.0
2150	23.2	17.5	11.7	58.6	57.6	53.5	18.2	24.9	34.8
Group I (Europe, North America, Oceania)									
2000	19.1	18.8	18.6	67.1	67.4	67.5	13.8	13.8	13.9
2025	19.2	16.9	13.7	62.4	63.7	65.7	18.4	19.4	20.6
2050	21.1	17.2	11.4	58.3	58.8	59.6	20.6	24.0	29.0
2075	22.3	17.7	10.5	58.8	58.3	54.9	18.9	24.0	34.6
2100	22.5	17.6	10.1	58.8	57.9	52.4	18.7	24.5	37.5
2125	22.3	17.3	9.8	58.3	57.2	50.8	19.4	25.5	39.4
2150	22.0	17.0	9.5	57.8	56.3	49.4	20.2	26.7	41.1
Group II (Africa, Latin America, Asia)									
2000	32.7	32.3	32.0	62.0	62.4	62.6	5.3	5.3	5.4
2025	28.4	25.5	21.9	63.7	66.0	69.0	7.9	8.5	9.1
2050	25.4	20.9	15.6	62.9	65.1	67.3	11.7	14.0	17.1
2075	24.4	19.1	13.4	61.5	62.4	61.3	14.1	18.5	25.3
2100	23.9	18.3	12.6	60.0	59.9	57.1	16.1	21.8	30.3
2125	23.6	17.9	12.2	59.3	58.7	55.3	17.1	23.4	32.5
2150	23.3	17.6	11.8	58.7	57.7	54.0	18.0	24.7	34.2

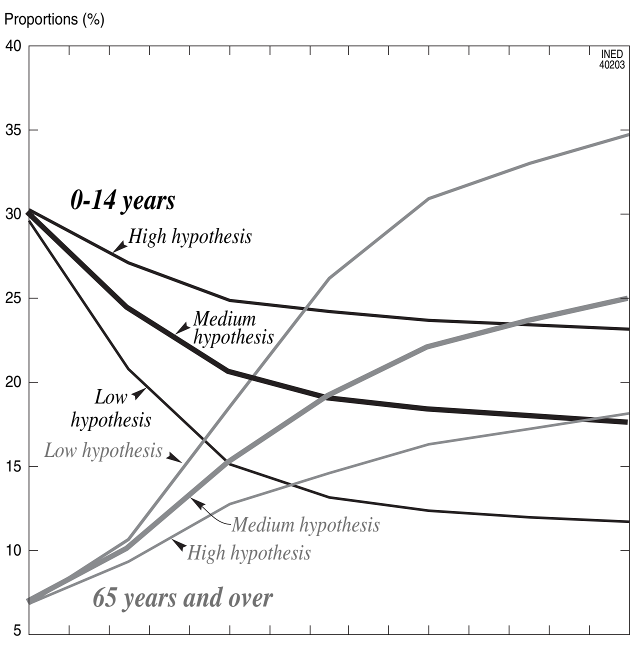


FIGURE 77–24 Trend up to 2150 of the proportion (%) of young (0 to14 years) and old (65 years and over) persons in the total world population according to the three fertility variants of the United Nations' principal scenario.

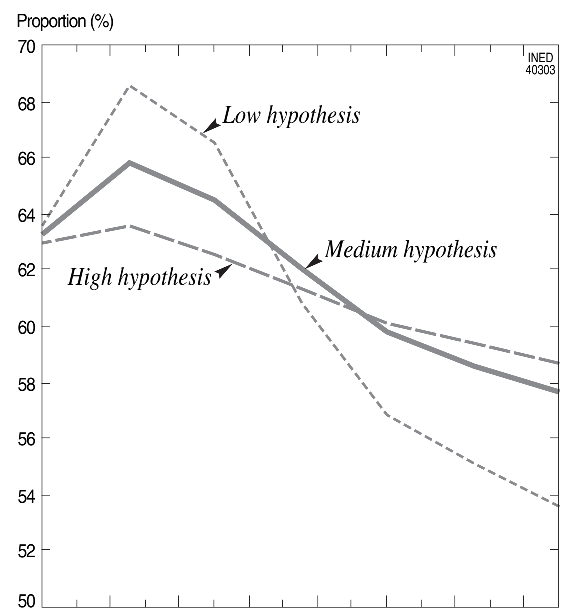


FIGURE 77–25 Trend up to 2150 of the proportion (%) of adults (15 to 64 years) in the total world population according to the three fertility variants of the principal scenario.

TABLE 77–9 1970–2000 Trend and 2000–2030 Projection of the Urban Population Proportion (%) of the World and of the Large Regions

World and large regions	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030
World total	36.7	37.9	39.6	41.4	43.5	45.2	47.0	49.0	51.1	53.4	55.7	58.0	60.3
Developed countries	67.6	70.0	71.5	72.7	73.8	74.9	76.0	77.2	78.4	79.7	81.1	82.3	83.5
Developing countries	25.1	26.8	29.3	32.1	35.1	37.4	39.9	42.5	45.2	48.0	50.8	53.5	56.2
Africa	23.1	25.2	27.3	29.5	32.1	35	37.9	40.8	43.7	46.5	49.2	51.8	54.5
Latin America	57.4	61.2	64.9	68.1	71.0	73.3	75.3	77	78.6	79.9	81.1	82.2	83.2
North America	73.8	73.8	73.9	74.7	75.4	76.2	77.2	78.3	79.6	80.9	82.1	83.3	84.4
Asia	23.4	24.7	26.9	29.4	32.4	34.4	36.7	39.2	41.9	44.7	47.6	50.6	54.6
Europe	64.5	67.3	69.4	70.9	72.3	73.6	74.8	76	77.3	78.6	80.0	81.3	82.6
Oceania	70.8	71.8	71.2	70.8	70.6	70.3	70.2	70.2	70.5	71.2	72.2	73.3	74.4

Data from United Nations, 2001c:10, 96, 164–171.

2. Urbanization

To deduct urban and rural population projections from the main projection, a definition of the border between these two categories must be available (see Chapter 61), as well as estimates of their current size and hypotheses of the specific dynamics of their trends. To do this, in its most recent projection the UN (2001) made some very rudimentary choices, admitting that any attempt in this field is risky and that it is better not to project too far into the future or to overly refine the data or the hypotheses.

The definition of an urban population retained for each country is simply the one used by statistical bodies. In other words, data that are often of a different nature from one country to another are added up on a world level. Moreover, the starting point is also one estimated by each country, at least in terms of the proportion of the urban and rural population; these are then applied to the Population Division's estimated total initial population. Finally, urban and rural population projections are obtained by applying future proportions to the principal projection (in this case the 1998 projection), estimated by using a very simple hypothesis. This hypothesis states that the evolution of the urban population will continue in each country at the same rate as that observed between the two most recent estimates available from census data. Yet again, the conditions for the projection are heterogeneous for all countries. It is not too much to say that the care taken with this type of derived projection is far from the efforts made for the principal projection. It can only therefore give very approximate results, just the general size. It is true that the projection is relatively short term, only up to 2030. We will only show three very basic tables here.

a. Evolution Up to 2030 of the Proportion of Urban Population

The first table (Table 77–9) shows the evolution since 1970 and the projection up to 2030 of the urban population proportion in the world and by continent. In the developed countries, it is already very high (76% in 2000), but according to this projection, it will continue to increase until it reaches nearly 84%. In the developing countries, it was only 40% in 2000 but should increase even more rapidly to reach 56% in 2030. However, the situation in Latin America is closer to the situation in the developed countries than in the other developing countries, with 76% in 2000 and 83% expected in 2030. In contrast, Africa and Asia are far less urbanized (around 37% in 2000) and could each remain separated in this context from the rest of the world (barely 55% in 2030).

b. The Large Cities in 2015

The last two tables describe the very large cities. However, the results for these are given only up to 2015, no doubt to avoid the failures of previous projections that were bolder but completely contradicted by the facts.[18]

Table 77–10 gives the number of agglomerations that have over a million inhabitants according to three size categories, their cumulated population, and their

[18] In their 1990 urban projections, the United Nations (UN, 1991b) had predicted an incredible growth for Mexico City that would increase the population from 14.5 million in 1980 to 25.6 million in 2000. However, the 1992 urban projections (UN, 1993) quickly corrected this, predicting a more modest increase from 13.9 million in 1980 to 16.2 million in 2000 and was guilty of underestimation. In the 2001 urban projections, the UN (2002) estimated the population of Mexico City at 18.1 million in 2000.

TABLE 77–10 Number and Population of the Agglomerations with 5 Million or More Inhabitants in 1975, 2000, and 2015

Agglomeration size	Number of agglomerations			Population (in millions)			Proportion (%) of total population		
	1975	2000	2015	1975	2000	2015	1975	2000	2015
World total									
10 million and over	5	19	23	68	262	375	1.7	4.3	5.2
5 to 10 million	17	22	36	127	155	248	3.1	2.6	3.5
1 to 5 million	173	370	505	327	704	1006	8.0	11.6	14.1
Developed countries									
10 million and over	2	4	4	36	67	69	3.0	6.0	10.0
5 to 10 million	8	6	7	62	45	51	6.0	4.0	4.0
1 to 5 million	75	109	127	145	219	250	14.0	18.0	21.0
Developing countries									
10 million and over	3	15	19	33	195	306	1.0	4.0	5.0
5 to 10 million	9	16	29	64	110	197	2.0	2.0	3.0
1 to 5 million	98	261	378	182	485	756	10.0	10.0	13.0
Africa									
10 million and over	0	2	2	0	24	37	0.0	3.0	3.0
5 to 10 million	1	1	4	6	5	25	1.0	1.0	2.0
1 to 5 million	7	40	62	13	75	135	3.0	10.0	13.0
Latin America									
10 million and over	2	4	4	21	59	66	7.0	11.0	10.0
5 to 10 million	2	3	5	17	19	34	5.0	4.0	5.0
1 to 5 million	17	44	62	32	89	133	10.0	17.0	21.0
North America									
10 million and over	1	2	2	16	30	32	7.0	10.0	9.0
5 to 10 million	2	1	2	16	7	13	7.0	2.0	4.0
1 to 5 million	28	38	44	52	84	96	21.0	27.0	28.0
Asia									
10 million and over	2	11	15	31	150	241	5.3	11.1	12.4
5 to 10 million	7	12	20	51	86	137	8.6	6.3	7.1
1 to 5 million	77	183	267	142	333	510	24.0	24.6	26.2
Europe									
10 million and over	0	0	0	0	0	0	0	0	0
5 to 10 million	5	5	5	37	38	38	5.0	5.0	5.0
1 to 5 million	42	59	64	82	111	118	10.0	15.0	16.0
Oceania									
5 to 10 million	0	0	0	0	0	0	0	0	0
1 to 5 million	2	6	6	6	12	13	29.0	40.0	36.0

Data from United Nations, 2001c.

proportion in the total population for 1975, 2000, and 2015. In 1975, only five cities had over 10 million inhabitants. There are currently 19, and that number could increase to 23 in 2015. These cities would then have 375 million inhabitants, 10% of the total urban population. If agglomerations of more than 5 million are considered, then 59 cities, 600 million inhabitants, and a quarter of the urban population would be counted in 2015.

Table 77–11 gives the list of the agglomerations that could number more than 10 million inhabitants in 2015 and their size in 1950, 1975, 2000, and 2015. It gives an idea of the diversity in the growth rates of the large agglomerations. Cities that in 1950 had hardly more than a million inhabitants (Karachi, Istanbul, Hyderabad, etc.), others that had already more than 5 million (Tokyo, Shanghai, Buenos Aires) and even over 10 million (New York) are found in this list. There is not one European agglomeration included in this table, not even those that approached 10 million inhabitants during the 1950s such as London or Paris.

TABLE 77–11 Agglomerations that Could Have More than 10 Million Inhabitants in 2015 and the Number of Inhabitants in 1950, 1975, 2000, and 2015

Agglomeration	Population in millions			
	1950	1975	2000	2015
Tokyo	6.9	19.8	26.4	26.4
Bombay	2.9	6.9	18.1	26.1
Lagos	0.3	3.3	13.4	23.1
Dhaka	0.4	2.2	12.3	21.1
Sao Paulo	2.4	10.0	17.8	20.4
Karachi	1.0	4.0	11.8	19.2
Mexico City	2.9	11.2	18.1	19.2
New York	12.3	15.9	16.6	17.4
Jakarta	1.5	4.8	11.0	17.3
Calcutta	4.4	7.9	12.9	17.3
New Delhi	1.4	4.4	11.7	16.8
Manila	1.5	5.0	10.8	14.8
Shanghai	5.3	11.4	12.9	14.6
Los Angeles	4.0	8.9	13.1	14.1
Buenos Aires	5.0	9.1	12.6	14.1
Cairo	2.4	6.1	10.6	13.8
Istanbul	1.1	3.6	9.5	12.5
Beijing	3.9	8.5	10.8	12.3
Rio de Janeiro	2.9	7.9	10.6	11.9
Osaka	4.1	9.8	11.0	11.0
Tien Tsin	2.3	6.2	9.1	10.7
Hyderabad	1.1	2.1	6.8	10.5
Bangkok	1.4	3.8	7.3	10.1

Data from United Nations, 2001c.

CONCLUSION

Like any projection, the World Population Projections published by the UN give hypothetical results. Since the end of the 1950s, however, they have always given surprisingly reliable results for the 2000 horizon. Initially, this was made possible only by the improvement in the source data and the adoption of a method capable of taking into account the diversity of situations (projection by country) and the principles of population dynamics (component method). But it is also owing to the existence of a general theory (the demographic transition) robust enough to express most of the current trends, at least in the mid term. In fact, no other rival attempt has produced better results.

This is, however, not the case for very long term projections, which predict very different futures because slight variations in the evolution of fertility and mortality can have radically different consequences when they are projected over a century or over a century and a half.

But without even going as far as exploring the long-term future, the mid-term future is today as difficult to predict as it was in the past. In an era in which the great historical change, called the demographic transition, is coming to an end, the paradigm of the same name is of no assistance in predicting what will follow. As to what will follow, we can imagine multiple scenarios leading to consequences that, as extravagant as they may seem, are no less plausible than the improbable prediction of a permanent stabilization.

Acknowledgment

This chapter was written with the collaboration of Francesca D'Onorio, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Italy, and was translated from the French version by Accenta Ltd.

References

Alho Juha, 1997. Scenarios, uncertainty, and conditional forecast of the world population, *Journal of the Royal Statistical Association Series A*, no. 160, p. 71–85.

Bogue Donald J., 1966. *The prospects for the world population control.* Chicago, University of Chicago.

Bogue Donald J., 1967. The end of the population explosion, *The Public Interest*, no. 7, p. 11–20.

Bogue Donald J. and Tsui Amy Ong, 1979. Zero world population growth? the public interest, no. 55, p. 99–113.

Bos Edward, Vu My T., Levin Ann, and Bulatao Rodolfo, 1992. *World population projections, 1992–2993 edition: estimates and projections with related demographic statistics.* Baltimore, The Johns Hopkins University Press, 516 p.

Bos Edward, Vu My T., Massiah Ernest, and Bulatao Rodolfo, 1994. *World population projections, 1994–1995 edition: estimates and projections with related demographic statistics.* Baltimore, The Johns Hopkins University Press, 522 p.

Boyarsky Aron Y., 1967. A contribution to the problem of the world population in the year 2000, *in: Proceeding of the World Population Conference. Belgrade, 30 August–10 September 1965 Volume II: fertility, family planning, mortality,* p. 5–12. New York, United Nations, 509 p.

Bulatao Rodolfo, Bos Edward, Stephens Patience W., and Vu My T., 1990. *World population projections, 1989–1990 edition*. Baltimore, The Johns Hopkins University Press, 422 p.

Delaunay Jeanine, Meadows Donella H., Meadows Dennis L., Randers Jorgens, and Behrens William W. III, 1972. *Halte à la croissance? Enquête sur le Club de Rome & Rapport sur les limites de la croissance.* Paris, Fayard, 314 p.

East Edward M., 1977. *Mankind at the crossroads.* New York, Arnold Press, 368 p. (new edition).

Frejka Thomas, 1973. *The future of population growth: alternative paths to equilibrium.* New York, John Wiley and Sons, 268 p.

Frejka Thomas, 1981. World population projections: a concise history, *in:* IUSSP (ed.), *International Population Conference, Manila, 1981, vol.3,* p. 505–527. Liège, Derouaux, 696 p.

Frejka Thomas, 1996. Long-range global population projections: lessons learned, *in:* Wolfgang Lutz (ed.), *The future population of the world: what can we assume today?* p. 1–13. London, Earthscan Publications Ltd and IIASA, 500 p.

Frejka Thomas and Mauldin W. Parker, 1980. *Long-term global and regional population growth prospects*. New York, The Population Council.

King Gregory, 1973. The L. C. C. Burns journal: a manuscript notebook containing workings for several projected works (composed c. 1695–1700), *in:* Peter Laslett (ed.), *The earliest classics John Graunt [. . . and] Gregory King,* p. 1–292. Weastmead (England), Gregg International Publishers Limited, 457 p.

Knibbs George H., 1928. *The shadow of the world's future or the earth's population possibilities and the consequences of the present rate of increase of the earth's inhabitants.* London, Ernest Benn, 132 p.

Lee Ronald D., 1998. Probabilistic approaches to population forecasting, *Population and Development Review,* vol. 24(suppl), p. 156–190.

Lee Ronald and Tuljapurkar Shripad, 1994. Stochastic population forecasts for the U.S.: beyond high, medium and low,*Journal of the American Statistical Association,* vol. 87(419), p. 659–671.

Littman G. and Keyfitz Nathan, 1977. *The next hundred years.* Cambridge, Harvard University, Center for Population Studies, 101 p.

Lutz Wolfgang (ed.), 1994. *The future population of the world: what can we assume today?* London, Earthscan Publications Ltd and IIASA, 484 p.

Lutz Wolfgang, Sanderson Warren, and Sherbov Serguei, 1997. Doubling of world population unlikely, *Nature,* no. 387, p. 808–805.

Lutz Wolfgang, Sanderson Warren, and Scherbov Serguei, 1998. Expert-based probabilistic population projections, *Population and Development Review,* p. 139–155.

Lutz Wolfgang, Vaupel James W., and Ahlburg Dennis A. (eds.), 1998. *Frontiers of population forecasting*. New York, The Population Council, 200 p

McDevitt Tomas, 1996. *World population profile 1996. With a special chapter focusing on adolescent fertility in the developing world*. Washington, U.S. Agency for International Development and U.S. Department of Commerce, Bureau of the Census, 183 p.

McDevitt Tomas, Stanecki Karen, and Way Peter O., 1999. *World population profile 1996: with a special chapter focusing on HIV/AIDS in the developing world.* Washington, U.S. Agency for International Development and U.S. Department of Commerce, Bureau of the Census, 172 p.

Meadows Donella H. and Club of Rome, 1972. *The limits to growth: a report for the Club of Rome's project on the predicament of mankind.* New York, Universe Books, 205 p.

Mesarovic Mihajlo and Pestel Eduard, 1974a. *Mankind at the turning point: the second report of the Club of Rome*. London, Hutchinson, 210 p.

Mesarovic Mihajlo and Pestel Eduard, 1974b. *Stratégies pour demain: deuxième rapport du Club de Rome.* Paris, Le Seuil, 204 p.

Notestein Frank, 1945. Population, the long view, *in:* T. Paul Schultz (ed.), *Food for the world,* p. 36–57. Chicago, University of Chicago Press, 367 p.

Notestein Frank W., 1950. The population of the world in the year 2000, *American Statistical Association Journal,* vol. 45(251), p. 235–249.

Notestein Frank W., Taeuber Irene B., Coale Ansley J., Kirk Dudley, and Kiser Louise K., 1944. *The future population of Europe and the Soviet Union: population projections 1940–1970*. Geneva, League of Nations, 332 p.

Pearl Raymond, 1924. *Studies in human biology.* Baltimore, Williams and Wilkins, 653 p.

Pearl Raymond and Gould Sophia A., 1936. World population growth, *Human Biology,* vol. 8(3), p. 399–419.

Petty William, 1682. *An essay concerning the multiplication of mankind.* London, Robert Clavel at the Peacock and Henry Mortlock at the Phoenix, 50 p.

United Nations [UN], 1958. *The future growth of world population.* New York, UN, Department of Economic and Social Affairs, 75 p.

United Nations [UN], 1966. *Les perspectives d'avenir de la population mondiale évaluées en 1963.* New York, UN, Département des affaires économiques et sociales, 168 p.

United Nations [UN], 1971. *Demographic projection with the aid of computers.* New York, UN, 22p.

United Nations [UN], 1973. *World population prospects as assessed in 1968*. New York, UN, Department of Economic and Social Affairs, 167p.

United Nations [UN], 1974a. *Concise report on the world population situation in 1970–1975 and its long-range implications.* New York, UN, 77p.

United Nations [UN], 1974b. *La situation démographique dans le monde en 1970–1975 et ses incidences à long terme.* New York, UN, Département des affaires économiques et sociales, 77 p

United Nations [UN], 1979a. *Les perspectives d'avenir de la population mondiale évaluées en 1973.* New York, UN, Département des affaires économiques et sociales, 179 p.

United Nations [UN], 1979b. *World population trends and prospects by country, 1950–2000. Summary report of the 1978 assessment.* New York, UN, Department of Economic and Social Affairs, 98 p.

United Nations [UN], 1981a. *Long-range global population projections, based on data as assessed in 1978.* New York, UN, Population Division, 70 p.

United Nations [UN], 1981b. *World population prospects as assessed in 1980*. New York, UN, Department of Economic and Social Affairs, 101 p.

United Nations [UN], 1982a. *Demographic indicators of countries: estimates and projections as assessed in 1980.* New York, UN, Department of International Economic and Social Affairs, 437 p.

United Nations [UN], 1982b. Long-range global population projections, as assessed in 1980, *Population Bulletin of the* United Nations, no. 14, p. 17–30.

United Nations [UN], 1985. *World population prospects. Estimates and projections as assessed in 1982*. New York, UN, Department of International Economic and Social Affairs, 523 p.

United Nations [UN], 1986. *World population prospect. Estimates and projections as assessed in 1984*. New York, UN, Department of International Economic and Social Affairs, 330 p.

United Nations [UN], 1988. *World demographic estimates and projections, 1950–2025.* New York, UN, Department of International Economic and Social Affairs, 386 p.

United Nations [UN], 1991a. *World population prospect as assessed in 1990*. New York, UN, Department of International Economic and Social Affairs, 607 p.

United Nations [UN], 1991b. *World urbanization prospects 1990. Estimation and projections of urban and rural populations, and of urban agglomerations*. New York, UN, Department of International Economic and Social Affairs, 224 p.

United Nations [UN], 1992. Long-range world population projections: two centuries of population growth 1950–2150. New York, UN, 35 p.

United Nations [UN], 1993. *World urbanization prospects: the 1992 revision: estimates and projections of urban and rural populations, and of urban agglomerations*. New York, UN, Department of International Economic and Social Affairs, 164 p.

United Nations [UN], 1994. *World population prospect: the 1994 revision*. New York, UN, Department for Economic and Social Information and Policy Analysis, Population Division, 886 p.

United Nations [UN], 1997. *The sex and age distribution of the world populations: the 1996 revision.* New York, UN, Department for Economic and Social Information and Policy Analysis, Population Division, 884 p.

United Nations [UN], 1998a. *World population prospects: the 1996 revision.* New York, UN, Department for Economic and Social Information and Policy Analysis, Population Division, 839 p.

United Nations [UN], 1998b. *World population projections to 2150.* New York, UN, Department for Economic and Social Affairs, Population Division, 41 p.

United Nations [UN], 1999a. *World population prospects: the 1998 revision. Volume I.* New York, UN, Department for Economic and Social Affairs, Population Division, 627 p.

United Nations [UN], 1999b. *World population prospects: the 1998 revision. Volume II.* New York, UN, Department for Economic and Social Affairs, Population Division, 883 p.

United Nations [UN], 2001a. *World population prospect: the 2000 revision. Volume I.* New York, UN, Department for Economic and Social Affairs, Population Division, 745 p.

United Nations [UN], 2001b. *World population prospect: the 2000 revision. Volume II.* New York, UN, Department for Economic and Social Affairs, Population Division, 919 p.

United Nations [UN], 2001c. *World urbanization prospect prospects: the 2001 revision: data, tables and highlights.* New York, UN, Department for Economic and Social Affairs, Population Division, 260 p.

United Nations [UN], 2002a. *World population prospects: the 2000 revision. Vol.3: analytical report.* New York, UN, Population Division, 263p.

United Nations [UN], 2002b. *World urbanization prospects: the 2001 revision: data, tables and highlights.* New York, UN, Department for Economic and Social Affairs, Population Division, 182 p.

United Nations [UN], 2003a. *United Nations population projections to 2300.* New York, UN, Population Division, 36 p.

United Nations [UN], 2003b. *World population prospects: the 2002 revision. Vol. 1: comprehensive tables* New York, UN, Population Division, 781 p.

United Nations [UN], 2003c. *World population prospects: the 2002 revision. Vol. II: sex and age* New York, UN, Population Division, 939 p.

United Nations [UN], 2003d. *World population prospects: the 2002 revision. Vol. III: analytical report* New York, UN, Population Division, 261 p.

US Bureau of the Census, 1979. *Illustrative projections of the world population to the 21st century.* Washington, Bureau of the Census, 116 p.

Vallin Jacques, 1976. Les perspectives de population mondiale des Nations unies, *Population,* vol. 31(3), p. 561–592.

Vallin Jacques, 1986. *La population mondiale.* Paris, La Découverte, 128 p.

Vallin Jacques, 1994. *Réflexions sur l'avenir de la population mondiale.* Paris, CEPED, 28 p.

Vu Mi Thi, 1984. *World population projections 1984: short- and long-term estimates by age and sex with related demographic statistics.* Washington, The World Bank, 452 p.

Vu Mi Thi, 1985. *World population projections 1985: short- and long-term estimates by age and sex with related demographic statistics.* Washington, The World Bank, 452 p.

Vu Mi Thi and Zachariah Kunniparampil C., 1983. *Short-term population projections, 1980–2020, and long-term projection, 2000 to stationary stage by age and sex for all countries of the world.* Washington, The World Bank, Population Health and Nutrition Department, 386 p.

Zachariah Kunniparampil C. and Vu My T., 1978. *Population projections: 1975–2000.* Washington, World Bank.

Zachariah Kunniparampil C. and Vu My T., 1988. *World population projections, 1987–1988 edition: short- and long-term estimates.* Baltimore, The Johns Hopkins University Press, 440 p.

Zachariah Kunniparampil C., Vu My T., and Eluan, Ann, 1981. *Population projections: 1980–2000 and long-term (stationary population).* Washington DC, The World Bank, Population Health and Nutrition Department.

C H A P T E R

78

The Future of Mankind

Looking Ahead after the Transition

JACQUES VALLIN AND GRAZIELLA CASELLI

Institut national d'études démographiques (INED), Paris, France

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION

As we have seen in the previous chapter, the hypothesis of widespread and conclusive stabilization of the world's population, which in the mid-1990s still remained the focus of projections of the United Nations (UN, 1995), is of course no more than a reference model—that of the *demographic transition.* It is fairly plausible that, initially, populations will tend to follow this model. However, this is not certain, and it is unlikely that after this transition, things will continue as before. Regarding the first point, the UN projections comprise, on both sides of the average hypothesis, low and high hypotheses (the difference with the former lies in the level to which fertility would tend), with mortality remaining unchanged (UN, 1995). The UN was studying the consequences of a convergence of fertility levels toward 1.7 children per woman and toward 2.5 as an alternative to the 2.1 that ensures strict generation replacement. Of course, in both cases the assumption that the size of the world's population will stabilize is no longer valid. Over the very long term, a level of 2.5 children per woman corresponds to infinite growth, whereas with 1.7 children, growth would soon give place to decrease and the population would tend to become extinct. Toward the end of the 1990s, the UN, within its central hypothesis, dismissed the assumption of a systematic return to the fertility level necessary to ensure replacement in all northern countries, where fertility has already reached even lower levels.

However, whether we are dealing with the central hypothesis of a population size stabilization or the hypotheses of continuous growth or decline, the world's population would become a stable population in the strict sense, that is, a population with constant parameters and an invariable age structure. In other words, even if the size of the population keeps changing, it would be, at the end of the transition, only under the effect of completely frozen behaviors. Of course, this second postulate can also be questioned; in fact, it is very likely that it will be questioned.

Not only is it unlikely that some day human behaviors might definitely freeze, but it does seem to us that we can already observe the early beginnings of new changes that might be the sign of new demographic revolutions to come. Indeed, some authors have already formulated the idea of a second demographic transition (see Chapter 69) in order to account for the changes recently observed in northern countries in terms of marriage and fertility (van de Kaa, 1987). On the one hand, the recent changes in the mortality of the very old reinforces the idea held by a number of researchers that human life expectancy could increase far beyond the limit of 85 years adopted in the 1994 UN projections.[1] On the other hand, the current variations observed in fertility, particularly the recent changes in its tempo, could lead to the adoption of fertility models by age that are completely different from those that have been referred to up to now. In other words, not only is the current transition probably not

[1] In more recent UN projections, this threshold of 85 years (82.5 for males and 87.5 for females) is no longer a strict limit, since the UN foresee that some countries could overpass it. Indeed female Japanese life expectancy already reached 85 years as soon as 2002.

the only foreseeable demographic trend, but new trends might appear even before the generalization of the current transition referred to in UN projections. Thus, it is not wholly ludicrous to suggest that the ostensible limits to the main population parameters of life expectancy at birth, average parity, or sex ratio at birth may not be fixed as they seem. The underlying idea is that there may be some value in exploring the demographic impacts of biological or social behavior that could extend some of these limits sufficiently to give rise to demographic regimes that are radically different from those referred to in the UN's current long-term projections.

However, we do not mean to assert that the social and biological behaviors of our species can endlessly change. Although this exercise requires some imagination, it remains within strict limits, compared with scenarios of infinite life expectancy, zero fertility, or hermaphroditic reproduction.

Tempting as it may be to imagine situations wildly beyond things as they are or even beyond what is generally allowed as possible, we have no wish to take the exercise further than stretching a few more or less established facts or trends. For the same reason, we will limit our investigations to changes in fertility and mortality, because although migratory scenarios can modify the trends observed for a given population, they do not have an impact on the world population as such, not unless we begin to consider interplanetary migrations, which are far from likely to happen in the imaginable future. Let us thus consider a few unconventional but not completely unthinkable possibilities for each of our three indicators of natural increase (fertility, mortality, sex ratio at birth). What would then happen in terms of population size and age structure?

But before moving on to wild conjecture, we must set the baseline.

I. THE BASELINE: WORLD POPULATION IN 2050

The baseline itself has to be resolutely probable. Unlike what follows, however, it will be set in the most conventional possible way: at that mythical cusp when the entire world population will have completed its demographic transition. So mythical is that moment, in fact, that it requires some ingoing massaging of the facts.

The planet's various populations are today at very different transitional stages. Most countries in the northern world completed this long historical process over half a century ago (see Chapter 68), although no precise point in time can be pinpointed. The restored fertility–mortality balance of the interwar years was thrown out of kilter by the baby boom, and it was not until the 1970s or 1980s that it evened up again, or, more accurately, temporarily redressed, as a protracted fertility decline has brought it well below replacement level (see Chapter 69). At the same time, the southern world displays a wide mix of situations: some countries have already joined the northern world in an early post-transitional stage, whereas others remain in high fertility, rapidly rising life expectancy scenarios. Granted, transition theory is borne out by across-the-board, and in most cases rapid, fertility decline; but it will still be some decades (perhaps close to a century) before what is still a large percentage of the world population chiefly in sub-Saharan Africa will achieve the stabilization predicted by transition theory (see Chapters 69, 77).

Strictly, then, we should wait on this culmination of demographic globalization when all the world's populations have completed their transition. But the rub is that no evidence points to a general, post-transition stabilization. Although all the world's populations seem likely to adopt increasingly convergent demographic behavior (which vindicates our world approach), there are no grounds for suggesting that such behavior will meet the stationary population criteria. On the contrary, recent 10-year trends in northern and some southern countries show sharp deviations from this model, which could soon exceed anything we plan to hypothesize.

Our baseline, then, will be the year 2050 a point when the UN *medium variant* population projection estimates that the overall world population will more or less stabilize with fertility of 2.1 children per woman and a life expectancy of 78 years. This is cheating on two scores, because not only do we know that this transition point will not in any sense be a baseline for stabilization but, for practical reasons, we are taking the 1994 projections,[2] which, at the time, were still projecting an early return for the northern world to fertility of 2.1 children per woman (UN, 1995). In the 1998 projections, this hypothesis has itself been dropped in favor of a less unrealistic variant of below replacement fertility for several decades to come (UN, 1999, 2001, 2003).

This baseline is read subject to trend projections for our three key indicators. Average parity will remain indefinitely fixed at 2.1 under the UN age-specific fertility model. Life expectancy will continue to rise at its

[2] We could have also used the 1996 projections, which are based on the same hypothesis, but they were not yet available at the time we were working on the paper for the Siena seminar (Vallin *et al.*, 1996). Because there was very little difference at this level between the 1994 and the 1996 projections, we did not consider it necessary to make an entire new set of computations.

current rate to 85 years for both sexes combined (i.e., 82.5 years for men and 87.5 years for women), which is regarded by the UN as the uppermost limit toward which all the world's populations are moving (see footnote 1, page 235). Lastly, the sex ratio at birth will remain unchanged at its initial value of 1.05 males per female. Fortunately, taking the whole world population as our baseline lets us off factoring in migration projections, as this exercise in science fiction will not be pushed as far as interplanetary flows.

The UN 1994 medium variant projects a world population of 9.8 billion inhabitants by 2050. By then, however, although stationarity may be on the way to being achieved, two population growth factors will continue to have an influence for some time. On one hand, past developments mean that the age structure still contains a significant growth potential. On the other hand, as long as life expectancy increases, the population also continues to grow. So, the total world population would rise to 12.2 billion by 2150 and even to 13 billion in 2250 and 13.4 billion by 2300 (Appendix 1). But its age structure would flatten quite quickly once the rapid aging process begun in past decades had completed (Fig. 78–1). Whereas the UN projection gives 28% aged under 20 years, 52% aged 20 to 59 years, and 20% aged 60 years and over in 2050, the almost final age structure of 24% "young people" to 46% "adults" and 30% "older people" is reached by 2100 (Appendix 5).

II. FROM GENERATION REPLACEMENT TO THE ONE CHILD WORLD

The recent long-term trends observed in the northern world prompt two types of questions about where the fertility trend is heading. First, far from following the (until recently) UN consistent scenarios of a return to the crucial 2.1 children per woman, fertility rates have stayed resolutely below it or continued to decline. Today, Italy and Spain are below 1.2 children per woman (Council of Europe, 1999). Some northern Italian regions have even fallen to 0.8 (Liguria) or 0.7 (Emilia-Romagna) (Santini, 1995; Zanatta and De Rose, 1995). Instead of a two-child family, are we not heading toward the one-child world? Is the one-child model that Chinese leadership has been trying to impose on its people for the past 20 years now being willingly embraced as the European ideal and perhaps spreading worldwide? This is not beyond the bounds of

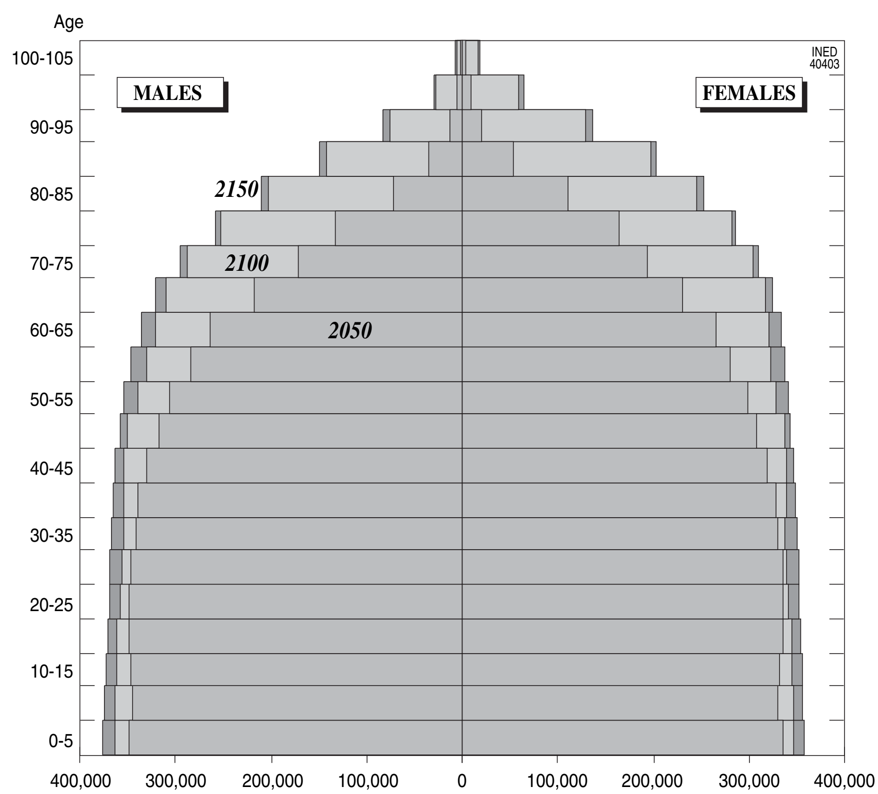


FIGURE 78–1 Age structure of the world population in 2050, 2100, and 2150 in the baseline variant (population in thousands; 2.1 children per woman, life expectancy of 85 years, and sex ratio at birth of 1.05 males per female).

possibility, and it deserves to be explored. But it would be pointless to explore anything below that level, which would just exacerbate a situation already amply catered for by the one-child model.

But recent decades have also seen a slower, more unobtrusive general trend emerging toward postponement of childbirth. In Italy, for example, mean age at first birth rose from under 25 years in 1976 to nearly 28.2 years in 1996 (ISTAT, 2000). So very low fertility may simply result from this later family formation. Delayed motherhood may drive period fertility very low by combining the now very low fertility of women who achieved their desired family size when younger with the still very low fertility of younger women who have put off family formation until later. Overall, desired family sizes may have remained unchanged, but period fertility could in these circumstances actually fall to very low levels. To take this reasoning a step further, if the age of menopause could one day itself be deferred, a new fertility transition might occur that ultimately returns to replacement level. However, the shift from an early model with a normal age around 25 years (UN model) to a late model with a normal age nudging 50 years takes period fertility into a very low phase, possibly even sinking to zero for the space of a few years.

Mean age at motherhood is not only rising, but age at childbearing is also growing more diverse. Some women are still having their children very young, others increasingly later, whereas others are testing out a two-stage fertility, perhaps foreshadowing a model postulated by Jean Bourgeois-Pichat back in the mid-1960s, although admittedly in an altogether different context. Later-onset menopause, he thought, ought to allow women to spread out family formation over a longer time. The day might even come when couples could have two children first between the ages of 25 and 30, then a later pair around the age of 50 after raising the earlier family. That, he said, would create the serious problem of over-rapid population growth, putting society in the position of having to apportion the right to this latter stage of reproductive life (Bourgeois-Pichat, 1966). Twenty years on, Jean Bourgeois-Pichat revisited this idea and updated it to reflect the declining European fertility. The issue was not now fear of overpopulation, but using this two-stage fertility as the means of minimum generation replacement (Bourgeois-Pichat, 1987). It is this last possibility that will be explored below.

1. The Transition to the One-Child World

Many one-child fertility models are clearly possible. We have deliberately chosen here to use a relatively early model, patterned on the tempo referred to by the

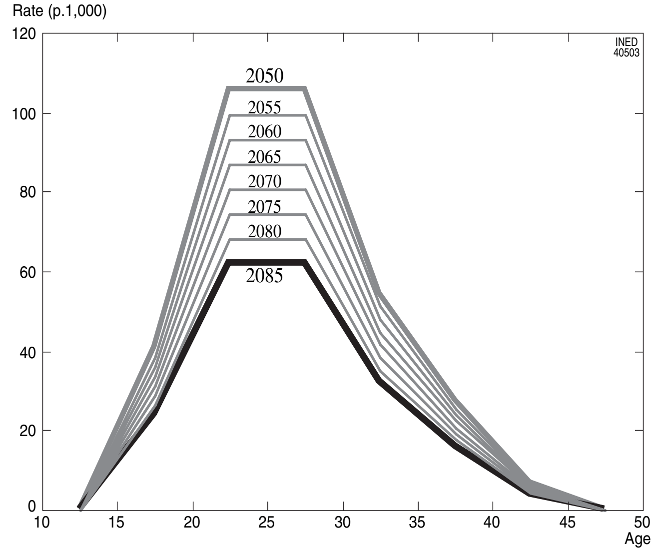


FIGURE 78–2 Age-specific fertility rates in the transition from the United Nations' low fertility variant (1.7 children per woman) to the one-child model over a period of 35 years.

1994 UN's fertility hypothesis at 1.7 children per woman. Thus, the transition to the one-child model imagined here spreads over 35 years, on the basis of the latter UN model. The intermediary stages are obtained through linear interpolation (Fig. 78–2).

For consistency's sake, our baseline in this case is the UN's (according to the 1994 projection) *low variant* world population by 2050 of 7.9 billion inhabitants (instead of the 9.8 billion of the medium variant).[3] A transition to the one-child world with constant fertility at that level over the long-term clearly widens the gap with our baseline variant to a staggering extent. The world population would decline after a century to 2.4 billion by 2150. A century later (2250), it would be no more than 200 million, barely higher than the end of the Neolithic era (see Chapter 66). In 200 years, humankind would have regressed down the same population road along which it had advanced over the past 6000 years. At this rate, of course, total extinction would not be far away. Like a supernova, after flaring up in a breakneck 20th century growth spurt, world population would die out with the same speed.

Meanwhile, it would obviously have aged to a considerable extent (Fig. 78–3). By 2150 the age structure would have stabilized at 8% under 20 years of age, 36% aged 20 to 59 years, and 56% aged 60 years and over (Appendix 5).

[3] Rather than projecting long-term population stability by tending toward stationarity (United Nations' medium scenario), the low scenario provides a more suitable baseline for the transition to the one-child world in 2050 by assuming a drop in fertility below the generation replacement level of 2.1 children per woman.

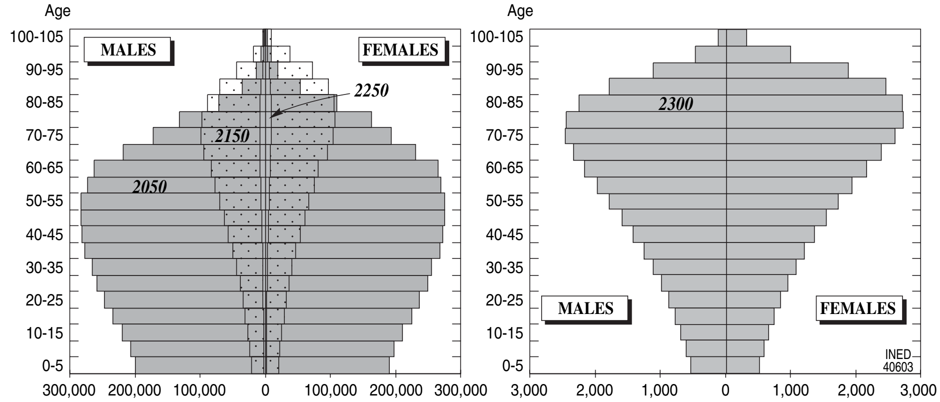


FIGURE 78–3 Long-term age structure trend (in thousands) in the one-child world scenario (1 child per woman, life expectancy of 85 years, sex ratio at birth of 1.05 males per female).

2. Toward New Constant Replacement Level Fertility Scenarios

But the transition to the one-child world is far from the only possible extrapolation of recent trends observed in the northern world. Consider the other two scenarios mentioned above: the bimodal timing of family formation and postponed childbirth, both of which would in the long term restore the balance of births and deaths.

a. Two-Stage Fertility?

Here again, there are an infinite variety of possible bimodal scenarios and many ways of moving from the current scenario of 2.1 children per woman (UN medium scenario) to whatever new model is chosen. We hypothesize that women's lifetime births occur by combining an initial period of reproductive life, with a mean number of 1.4 children born per woman at a modal age of 25 years, and a second period, 30 years later, with a mean number of 0.7 children ever born; the slight deferral of the menopause allows a modal age of 55. Around each of these two modes, the timing of family formation is patterned on the age distribution taken from the UN medium scenario.

The transition from the baseline scenario (in this case the UN medium variant) to the new bimodal variant is based on an assumed gradual take-up of this new schema by younger generations.[4] So earlier reproductive life fertility starts to decrease before the second period of later reproductive life begins (Fig. 78–4). Thus, the total fertility rate first decreases sharply, and then after a period of 50 years (Table 78–1), it reaches its level of equilibrium (set here at 2.06 to settle at the

[4] A more detailed exposition of the fertility scenarios used here can be found in our earlier article (Vallin and Caselli, 1997).

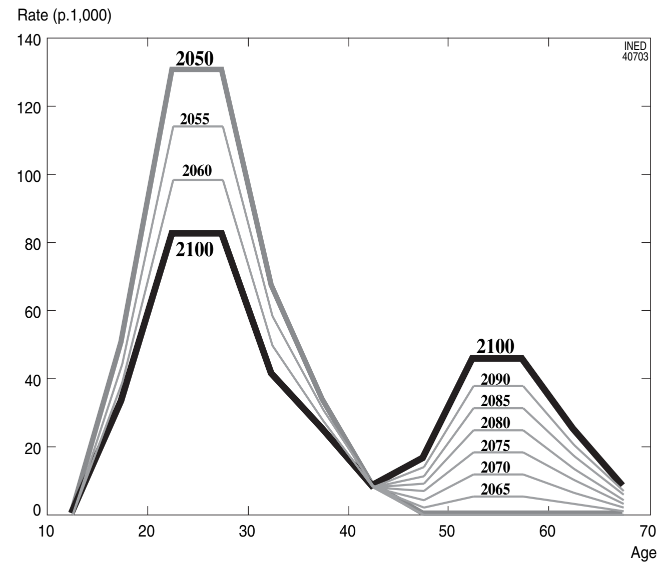


FIGURE 78–4 Age-specific fertility rates in the transition from the United Nations' medium variant (2.1 children per woman) to a bimodal model.

virtual disappearance of mortality below the reproductive ages). From its putative 2.1 in 2050, the total fertility rate drops to 1.4 by 2065, before rising again until 2100, where it stabilizes at the generation replacement level then set at 2.06 to settle at the virtual disappearance of mortality below the reproductive ages entailed by a life expectancy of 85 (Vallin and Caselli, 1997).

Far-fetched as may seem, the transition to this bimodal fertility would not produce major demographic upsets. On the contrary, it would act as a sort of regulator on the total world population beyond 2050 by absorbing the surplus growth expected in the baseline variant. Starting from 9.8 billion human beings in 2050, this would give us 9.4 billion instead of 12.2

TABLE 78–1 Trends in Fertility Rates by 5-Year Age Groups in the Transition from the United Nations' Medium Variant to a Bimodal Model

Age (years)	Calendar year										
	2050	2055	2060	2065	2070	2075	2080	2085	2090	2095	2100
15–19	0.252	0.223	0.194	0.165	0.165	0.165	0.165	0.165	0.165	0.165	0.165
20–24	0.651	0.571	0.492	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412
25–29	0.651	0.571	0.492	0.412	0.412	0.412	0.412	0.412	0.412	0.412	0.412
30–34	0.336	0.293	0.249	0.206	0.206	0.206	0.206	0.206	0.206	0.206	0.206
35–39	0.168	0.153	0.138	0.124	0.124	0.124	0.124	0.124	0.124	0.124	0.124
40–44	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.041	0.041	0.041	0.041
45–49	0.000	0.000	0.000	0.010	0.022	0.034	0.046	0.057	0.069	0.081	0.082
50–54	0.000	0.000	0.000	0.028	0.061	0.093	0.125	0.158	0.190	0.223	0.227
55–59	0.000	0.000	0.000	0.028	0.061	0.093	0.125	0.158	0.190	0.223	0.227
60–64	0.000	0.000	0.000	0.015	0.033	0.051	0.068	0.086	0.104	0.121	0.124
65–69	0.000	0.000	0.000	0.005	0.011	0.017	0.023	0.029	0.035	0.040	0.041
Total fertility rate	2.1	1.85	1.61	1.45	1.55	1.65	1.75	1.85	1.95	2.05	2.06

billion by 2150, and 9.2 billion rather than 13.4 billion by 2250. This should not come as a surprise, as what is being projected is the long-term stability of fertility at replacement level, whereas the shift between scenarios involves a transitory decrease in the total fertility rate. This low fertility period simply allows the twofold potential inherent in the age structure by 2050 and the transition to a life expectancy of 85 to be absorbed without increase.

It is clear that, ultimately, the age structure will be identical to the baseline variant, the only difference being the less consistent path taken to achieve stability. Population aging will first exceed long-term expectations and then rejuvenate to settle at the proportions indicated above. The proportion aged 60 years and over, in particular, rises sharply from 20% to 34% in 50 years before ultimately slipping back gradually to 30% (Vallin and Caselli, 1997).

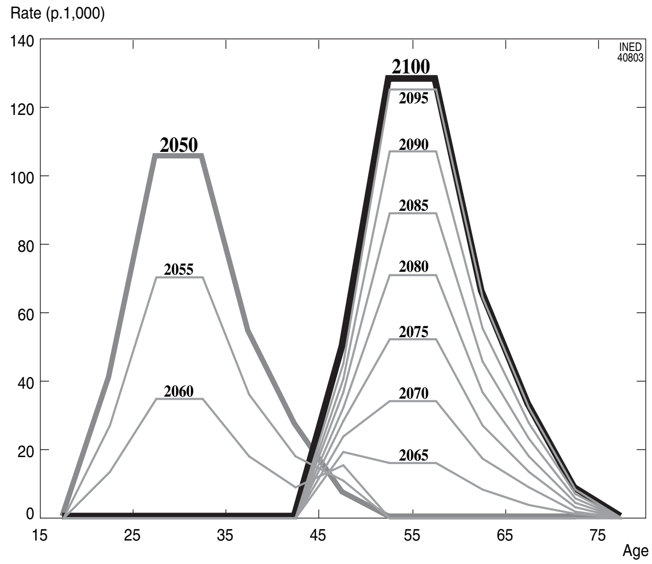


FIGURE 78–5 Age-specific fertility rates according to the United Nations' low variant (1.7 children per woman) in the transition to a postponed childbirth model ensuring equilibrium (2.06 children per woman).

b. Or Postponed Childbirth?

Suppose that recent fertility decline in the northern world—or, more generally, the convergence of the different populations worldwide toward the UN low scenario (1.7 children per woman for the world as a whole in 2050)—is the result only of delayed motherhood rather than a reduction in the number of desired births, and take the reasoning to the extreme by positing later-onset menopause, a much later reproductive life with a mode at age 45 to 54 years instead of age 20 to 29 years. Taking as our baseline the UN low scenario of 1.7 children per woman in 2050, we would again, as in the preceding model, end up 50 years later with the rate of 2.06 children, which will ensure equilibrium but after a much more testing transition period for the total fertility rate. Here, all fertility rates under age 40 years fall rapidly to zero, and rates among those more than 40 years start from zero but still remain low (Table 78–2 and Fig. 78–5). As a result, the total rate falls to 0.3 in 2065 before returning to 2.06 in 2100 (Vallin and Caselli, 1997). In fact, the paradox could easily have been taken further to make total period fertility zero for a period of several years.

So, the most salient outcome of such a change is to cause wide swings in the trend of different demographic parameters, starting with the growth rate (Fig. 78–6) and considerably extending the time frame to

TABLE 78–2 Fertility Rates by 5-Year Age Groups in the Transition from the United Nations' Low Variant to the Postponed Childbirth Model

Age (years)	Year										
	2050	2055	2060	2065	2070	2075	2080	2085	2090	2095	2100
15–19	0.204	0.136	0.068	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20–24	0.527	0.351	0.176	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25–29	0.527	0.351	0.176	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30–34	0.272	0.181	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
35–39	0.136	0.091	0.045	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40–44	0.034	0.055	0.077	0.098	0.119	0.141	0.162	0.183	0.205	0.226	0.247
45–49	0.000	0.000	0.000	0.080	0.171	0.262	0.354	0.445	0.536	0.627	0.639
50–54	0.000	0.000	0.000	0.080	0.171	0.262	0.354	0.445	0.536	0.627	0.639
55–59	0.000	0.000	0.000	0.041	0.088	0.135	0.182	0.230	0.277	0.324	0.330
60–64	0.000	0.000	0.000	0.021	0.044	0.068	0.091	0.115	0.138	0.162	0.165
65–69	0.000	0.000	0.000	0.005	0.011	0.017	0.023	0.029	0.035	0.040	0.041
Total fertility rate	1.70	1.17	0.63	0.32	0.60	0.89	1.17	1.45	1.73	2.01	2.06

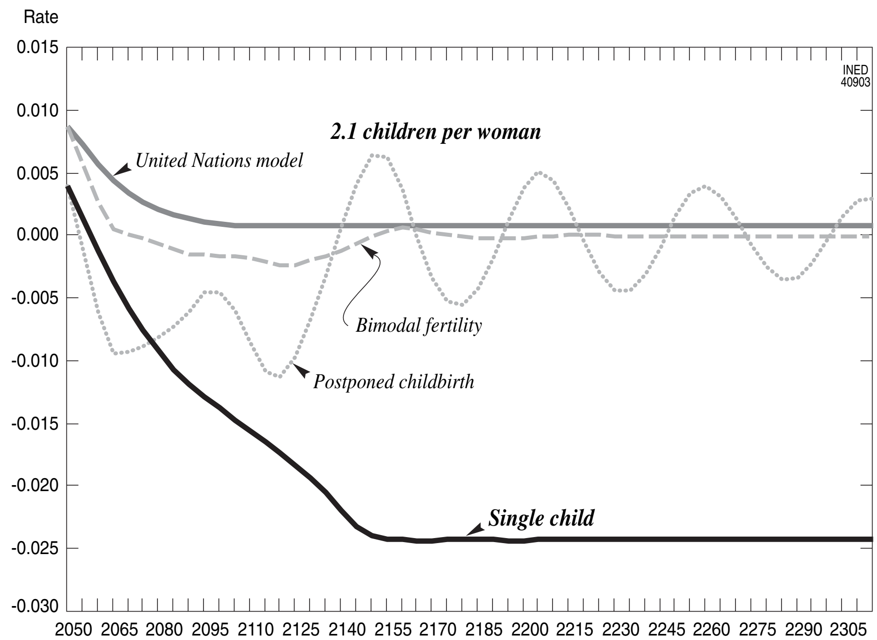


FIGURE 78–6 Population growth rate trend in different fertility scenarios (life expectancy of 85 years, sex ratio at birth of 1.05 males per female).

achieve a stationary population once stationarity conditions are fulfilled. Although the transition to the one-child world or even bimodal fertility scenario required only a hundred years for the growth rate to become constant, this scenario will take several centuries: not until over 5 centuries later, after 2570, would the amplitude of oscillations diminish to less than 1%.

Throughout this entire period of fluctuations, the age structure concertinas wildly. So, for example, the proportion aged 60 years and over would rise from 25% in 2050 to 44% by 2090, dropping to 25% in 2145, rising 36% in 2175, dropping again to 26% in 2200, and so on for over 5 centuries (Vallin and Caselli, 1997). By way of illustration, Fig. 78–7 compares the population pyramids for 2150 for fertility following the UN medium scenario and the postponed childbirth variant considered.

As these wild swings are caused by a short-lived collapse in fertility, they would also start with a sharp fall in total population size, which would never be made up. Of the 7.9 billion in 2050, only 4.3 billion would remain a hundred years later in 2140, rising

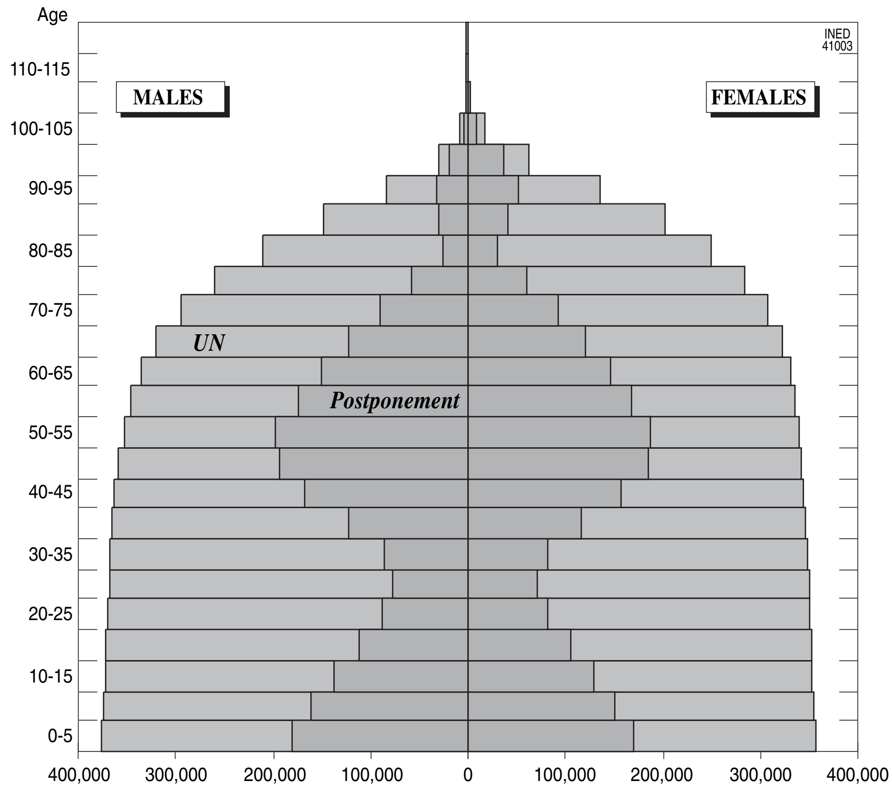


FIGURE 78–7 Age structures in 2150 (population in thousands) for fertility following the United Nations' medium variant and the postponed childbirth model (2.1 children per woman, life expectancy of 85 years, sex ratio at birth of 1.05 males per female).

back up only to 4.8 billion around 2165, at the beginning of a long series of oscillations around its future stabilization level of 4.45 billion inhabitants.

Finally, the least improbable of these three fertility scenarios, the one-child variant (which requires no biological change), would also be the most detrimental. But what about mortality?

III. LIFE EXPECTANCY: 85 OR 150 YEARS?

As mentioned in Chapter 48, the prospects for increasing longevity are a hotly debated issue. There is general agreement, in light of recent steadily rising trends in the most advanced countries, that the mean length of life is likely to increase further in the coming decades. But to what point is where opinions diverge.

In the very early 1950s, Jean Bourgeois-Pichat (1952) calculated a *biological limit life table*, giving the life expectancy that might be attained by removing all the *exogenous* causes of death that, being owing to clearly identified pathogens or external causes, could be avoided if the benefits of medical progress were extended to the world population. He calculated maximum life expectancies of 76 years for men and 78 years for women. The UN world population projections (UN, 1982), indeed, long assumed an age close to this maximum (75 years) as the level toward which all populations would tend.

It is clear that reality has now overtaken fiction, at least as far as women are concerned, because female life expectancy is already above 80 years in a score of industrialized countries, topping 83 years in France and 85 years in Japan. The question is whether that is owing to the discovery of hitherto unknown or ill-identified pathogens and ways of defeating them, or whether recent advances in life expectancy are actually owing to a retreat in *endogenous* mortality itself, which, for Jean Bourgeois-Pichat, was owing to the biological makeup of the human organism and would be much more difficult to influence.

Revisiting the distinction between endogenous and exogenous mortality 30 years later, but in the light of the most recent knowledge about the nature of disease factors, Bernard Benjamin (1982) calculated significantly higher life expectancy limits of around 85 years of age for both sexes together. This is very close to the result obtained if *degenerative diseases* are taken as the only cause of death (Vallin and Meslé, 1988). More recently, Jay Olshanski (1990), drawing on the work of the biologist James Fries (1980, 1989), finds it unlikely that life expectancy can ever exceed 85 years: it will continue to rise for a time through conventional rectangularization of the survival curve but will necessarily reach a maximum that, because of the inviolable longevity of humankind, cannot be above the age of 85 years. This maximum of 85 years (82.5 for men and 87.5 for women) is what the UN projections have been using since the early 1980s.

But there is another take to be had on recent developments. Have we not now developed the means of effectively slowing down people's biological aging and increasing their longevity? In other words, where we have so far only allowed a growing number of individuals to near the biological limit, are we now not raising the bar? First, the main reason why life expectancy has been lengthening in the northern world since the 1970s is the reduction in cardiovascular diseases, which (along with cancers, in Jean Bourgeois-Pichat's time) were the main age-dependent diseases. And what has mainly prevented a sharp drop in cancer deaths is smoking, although this has been declining (among men) for some time now, especially in the United Kingdom, the United States, and now France and Italy. Also, recent decades have seen a rapid increase in the rate of mortality decline among the oldest old (Kannisto, 1996). In France, for example, the female probability of dying between ages 80 and 89 years was still 72% in the early 1970s, but had fallen to 54% by the mid-1990s (Meslé, 1995; Vallin and Meslé, 2001). Declining mortality is also occurring among 90- to 99-year-olds, but life expectancy among centenarians has risen among women from 1.5 years in the 1950s to 2.2 years in the 1990s (Allard *et al*., 1996; Meslé *et al*., 2001; Meslé and Vallin, 2002). That does not necessarily mean that the upper limit of life is rising, but it seems increasingly obvious that, even were it to be otherwise, life expectancy at birth would be very likely to considerably exceed the 85 years that some have taken as an all-but absolute limit. For example, by simply extrapolating recent age-specific mortality trends by reference to a constant life expectancy of 101.5 years, it has been possible to show that French men could have a life expectancy of 84.2 years by the turn of the century; French women, 91.3 years (Vallin and Meslé, 1989).[5] But what if longevity itself is already increasing? For a start, the oldest age at death is rising steadily year over year in the most advanced countries (Wilmoth and Lundström, 1996). Jeanne Calment is probably the first woman ever to have lived to over 120 (Allard *et al*., 1994; Meslé *et al*., 2000). Also, the clearer research into mortality at extreme ages becomes, the more it appears that exponential growth in age-specific mortality rates, generally interpreted as the direct result of the process of biological aging which leads to death, slows down beyond a certain age. This slowdown can be construed as the product of extreme selection operated by the high death rate at these ages on an initially very heterogeneous population growing more homogeneous around the hardiest individuals (Vaupel *et al*., 1979, 1998; Barbi *et al*., 2003). How long will it be before we can identify why some individuals are so long-lived, so as to extend the benefits to the greatest number? In other words, are we on the threshold or even at the start of an effective revolution in the sphere of human longevity? It was that belief that prompted Roy Walford (1984a,b) to announce a life expectancy of close to 150 years during the next century as quite within the bounds of possibility. The fact that there is little likelihood of that happening during this century should not stop us exploring some of the ramifications.

But simply setting the new limit of life expectancy at 150 years does not tell us the new bounds of human life. We must still know what survival curve will take us to that figure, which, as yet, is beyond comprehension. There are two extreme possibilities. Either a life expectancy of 150 years is achieved through a technological quantum leap that could immediately benefit the greatest number—in other words, one that is consistent with the rectangularization of the survival curve within which we have evolved for 2 centuries —or the same result is achieved by an accelerated mortality decline at the oldest ages, entailing a rapid expansion in ages at death and a derectangularization of the survival curve. Let us consider both.

To illustrate the first hypothesis, we simply step-shifted the survival curve corresponding to the UN upper limit table toward increasingly higher ages. So, survival at 15 in the UN table becomes, 10 years later, survival at 20; 10 years after that, survival at 25; and so on. However, the share of survival under 15 years had to be considered differently: so we gradually increased this to completely eliminate child mortality. Above the age of 100 years, which is where the UN table stops, we have expanded it by extrapolating the quinquennial death probabilities so as to achieve one at 135 years before making changes in the corresponding proportions of survivors. This takes us from a UN life expectancy of 85 years to a life expectancy of 150

[5] According to recently updated calculations, based on more recent data that take into account the decline of mortality owing to road accidents and the increase of life expectancy at very old ages, life expectancy could reach 91.1 years for males and 94.9 years for females in 2100 (Vallin and Meslé, 2001).

years for both sexes in 140 years' time. In other words, taking the UN projection for 2050 as the baseline, the new upper limit sanctioned by technological advance will not be achieved until 2190. It is a revolution, certainly, but a very gradual one. We favor this conservative projection to a more rapid evolution. In fact, for population projections, we must work on a sex-specific basis. This leads us to raise the male life expectancy of 82.5 (UN upper limit) to 147.4 years and the female one from 87.5 to 152.5 years (Fig. 78–8; Table 78–3).

The alternative scenario, a derectangularization of the survival curve, is portrayed here no less simply: all the quinquennial death probabilities are reduced by 10% every 5 years so that the age at which the probability is equal to one is lowered by 5 years each time. Clearly, this proportional reduction more greatly benefits ages in which mortality is highest, which is more

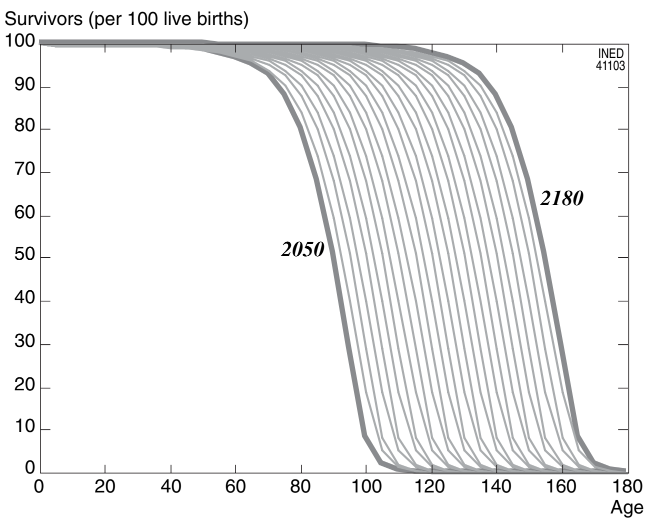


FIGURE 78–8 Female survival curve in the transition from the United Nations' maximum life expectancy (87.5 years) to a life expectancy of 153 years, with rectangularization.

than enough to amply derectangularize the survival curve (Fig. 78–9; Table 78–3). This brings us to a life expectancy of 150 years within the same time as before (150.5 years for men and 153.0 years for women after 140 years), but with a very different extreme age: the final deaths occurring between ages 270 and 275 years instead of 190 and 195 years in the preceding case. Figure 78–10 illustrates this approach for women.

The first result of the shift to a life expectancy of 150 is to upset the conditions of population growth for a century, regardless of the fertility scenario (Fig. 78–11). In the UN generation replacement scenario, a life expectancy of 150 years irrespective of the shape of the survival curve would almost double the world population in 2320 to nearly 25 billion rather than 13.5 billion. The only difference between the maintained survival curve rectangularization and the expansion of age at death scenarios is that this doubling of the final result comes earlier in the former case than in the latter case.

In the case of postponed childbirth, the lengthening of life expectancy to 150 years compensates, either immediately (rectangularization) or in the longer term, the population fall attributable to the low fertility period.

Last, in the scenario of a transition to the one-child world, lengthening life expectancy, although it clearly cannot stave off the threat of ultimate extinction, can at least defer the drop by several decades (in the event of rectangularization, there is even a short-lived increase) such that by around 2225, the world population is still above 2 billion, whereas with a limit of 85 years, it has already fallen below 400 million.

Figure 78–12 shows in more detail how, throughout the century in which life expectancy increases, the growth rate, although quickly turning negative, does not collapse as it would with a maximum limit of 85 years. It is also clear how the maintenance of survival

TABLE 78–3 Life Expectancy Trends in the Hypothesis of a Transition to 150 Years, Depending on the Type of Transformation Undergone by the Survival Curve

Sex	Calendar year														
	2050	2060	2070	2080	2090	2100	2110	2120	2130	2140	2150	2160	2170	2180	2190
Maintaining the rectangularization of the survival curve															
Males	82.5	87.5	92.5	92.5	97.5	102.5	107.4	112.4	117.4	122.4	127.4	132.4	137.4	142.4	147.4
Females	87.6	92.6	97.6	97.6	102.6	107.5	112.5	117.5	122.5	127.5	132.5	137.5	142.5	147.5	152.5
Total	85.1	90.0	95.0	95.0	100.0	105.0	110.0	115.0	120.0	125.0	130.0	135.0	140.0	145.0	150.0
With expansion of ages at death															
Males	82.5	85.1	87.8	90.7	94.0	97.7	101.9	106.6	111.9	117.8	124.2	131.2	138.6	146.4	150.5
Females	87.6	90.0	92.5	95.3	98.5	102.1	106.1	110.7	115.8	121.4	127.6	134.3	141.5	149.1	153.0
Total	85.1	87.5	90.1	93.0	96.3	99.9	104.0	108.6	113.8	119.6	125.9	132.7	140.0	147.7	151.7

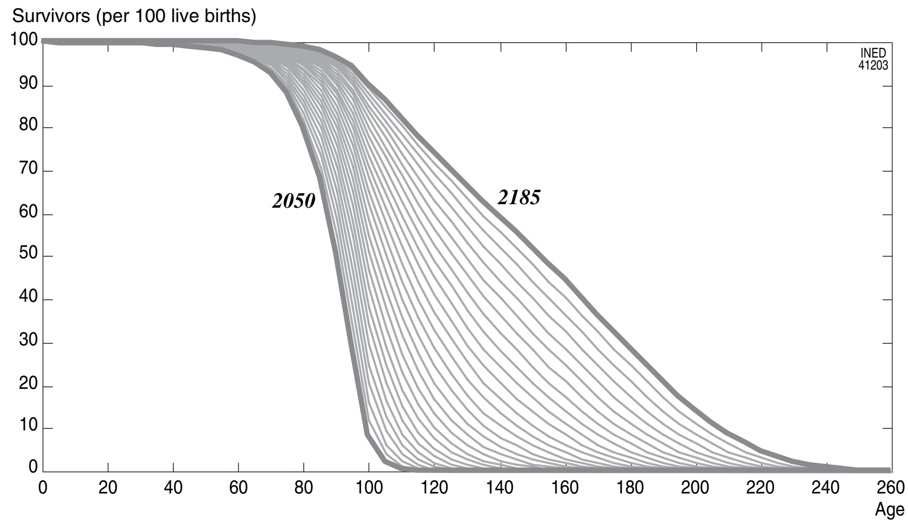


FIGURE 78–9 Female survival curve in the transition from the United Nations' maximum life expectancy (87.5 years) to a life expectancy of 153 years, with expansion of ages at death.

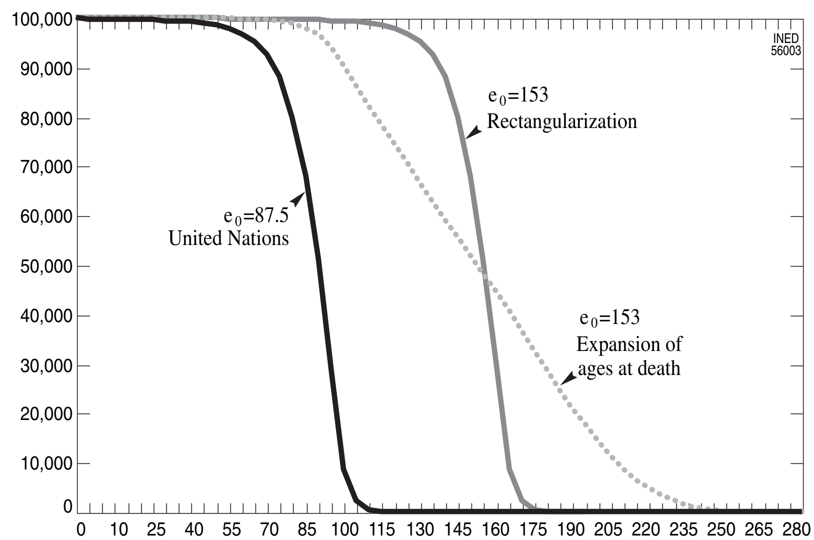


FIGURE 78–10 Comparison of two female survival curves for a life expectancy of 153 years with that for the United Nations' limit of 87.5 years.

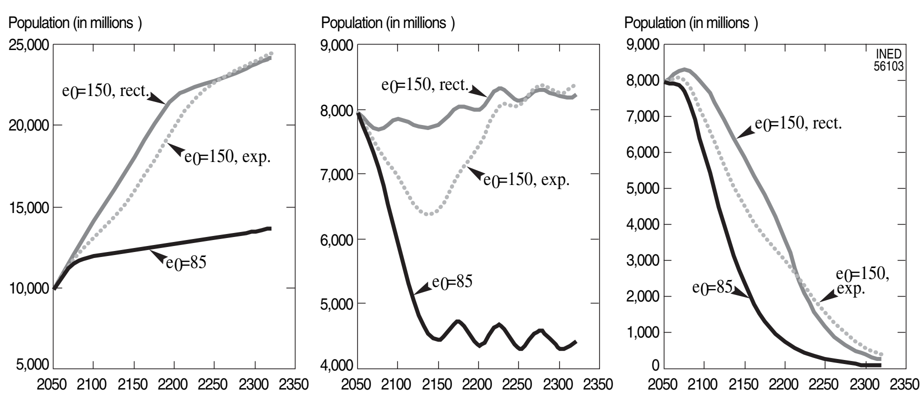


FIGURE 78–11 Comparative world population trends (2050–2320) in three mortality scenarios with fertility following the United Nations' medium variant **(left)**, returning to 2.1 in a postponed childbirth scenario **(center)** or shifting to the one-child world **(right)**.

curve rectangularization initially staves off that collapse. But going to the other extreme, this scenario is one in which the growth rate collapses abruptly once the period of lengthening life expectancy has ended.

The second major consequence of a rise in life expectancy to 150 years is to radically transform the age distribution of the population; the shape of the survival curve is clearly crucial. Consider just two cases of fertility: the UN model with 2.1 children per woman and the transition to the one-child world. In the former (Fig. 78–13), there is a clear trend toward a stationary population, and the pyramid shape corresponds exactly to the survival curve. If life expectancy rises to 150 years and the survival curve remains rectangular, the age structure increases by 70 increments but the shape remains essentially unchanged. By contrast, with an expansion of age at death, the pyramid top reverts to a tapered profile that has long since disappeared from the landscape of the northern world.

The question then arises which of the two pyramids so obtained is the "oldest." Ostensibly, it could be the second (the third in Fig. 78–13), which rises much higher in age. But what if greater longevity can only be achieved if there is an age-for-age improvement in health and if old age-related problems are clustered in the final third of the maximum human life span (currently, between 80 years and Jeanne Calment's 120 years, i.e., 9% of the population). With a life expectancy of 150 years in the first case (rectangularization), our calculations give survivors up to 190 years, so the same age bracket would rise from 130 to 190 years, i.e., 13% of the population. In the second (expansion of ages at death), the final survivors reach 270 years of age, widening this age bracket from 180 to 270 years but compressing it to just 4% of the total population. In this definition of the truly "old" population, the transition to a life expectancy of 150 clearly ages the population if the survival curve remains rectangular but would have the reverse effect of "rejuvenating" the population in the event of an expansion of age at death, to some extent turning centenarians into the "youngsters" of the age force.

The situation would be worse in the one-child world variant; with a life expectancy of 150 years, aging would assume such a magnitude that the population would contain little more than centenarians and supercentenarians (Fig. 78–14).

The effect of the one-child world variant on a population with a maximum life span of 85 was seen earlier: ultimately, 56% of the population would be more than 60 years of age and 24% over 80 years. But only 1% would be over 100 years. With a life expectancy of 150, more than 90% of the population would be over 60 years (91% or even 95% with an expansion of age at death), and centenarians alone would make up 74% of

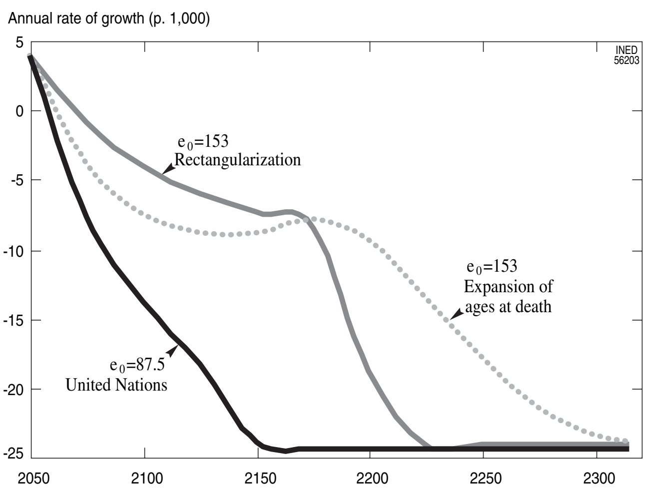


FIGURE 78–12 Comparative trends in the annual rate of growth (2050–2315) according to the three life expectancy hypotheses in the transition to the one-child world scenario.

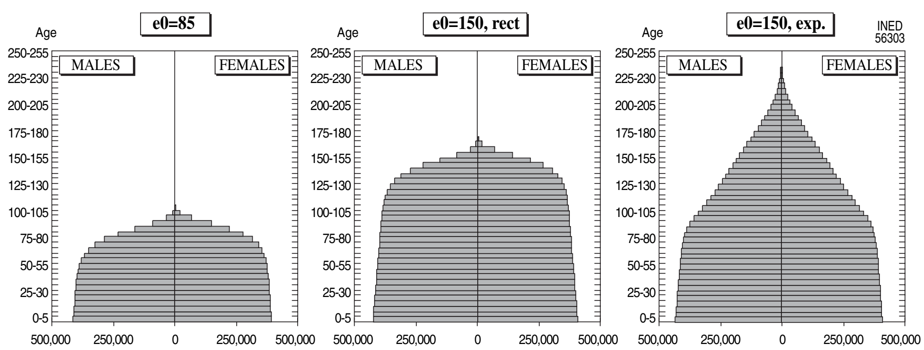


FIGURE 78–13 Age structures in 2300: effect of the transition to a life expectancy of 150 by type of survival curve, in the case of the United Nations' fertility variant of 2.1 children per woman (population in millions).

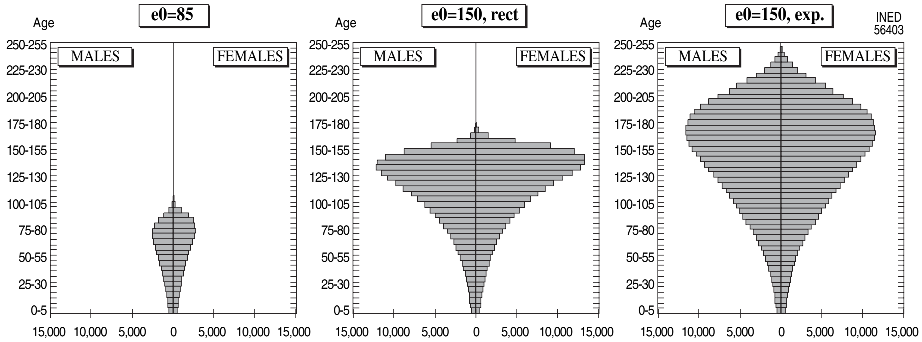


FIGURE 78–14 Age structures in 2300: effect of a rise in life expectancy to 150 years by type of survival curve, combined with a transition to the one-child world (population in millions).

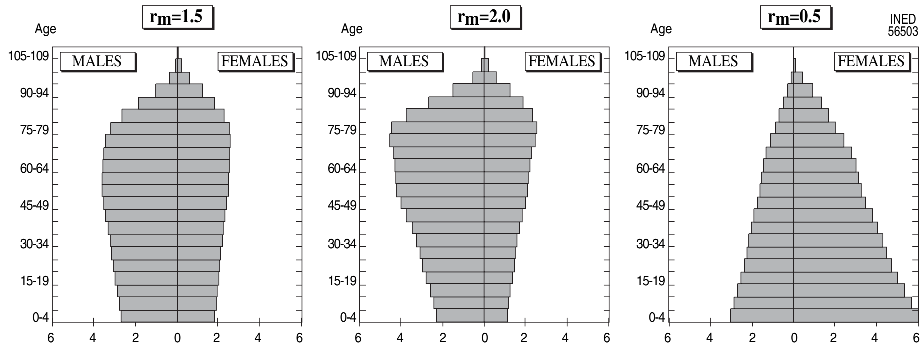


FIGURE 78–15 Population pyramids in 2150: effect of the change in the sex ratio at birth (r_m) with United Nations' scenario fertility of 2.1 children per woman and a maximum life expectancy of 85 years (proportions %).

the total population in the event of rectangularization and even 84% in the event of an expansion of ages at death! In either case, young people under the age of 20 would account for less than 2%.

IV. THE EFFECTS OF A RADICAL CHANGE IN SEX RATIO AT BIRTH

Returning to the classic scenario of 2.1 children per woman and a life expectancy of 85, what would be the result of a long-term change in the sex ratio at birth? To gain some insights into this, we explored three assumptions that deviate to differing extents from the 1.05 males per female norm:

At 1.5, the level is barely above that extractable from the current situation in a handful of Chinese and Indian provinces (Zhang *et al.*, 1983; Calot and Caselli, 1988; Miller, 1996).[6]

At 2.0 the level merely extends slightly the consequences of a strong son preference.

At 0.5, the level finally portrays the consequences of a female birth resurgence.

In all three scenarios, even when unceremoniously applying the new norm as of the projection baseline year of 2050, stabilization is almost entirely achieved within a bare century, by 2150. However, each of these three assumptions radically changes not just the sex structure but also the age structure of the population (Fig. 15). Were the sex ratio (r_m) to rise to 1.5 at birth, then in the stabilized population the share of women would fall from just over 50% (with $r_m = 1.05$) to under 42% and the population would have aged considerably: the proportion of young people under 20 would fall from 24% to 19%, whereas the proportion aged 60 and over would rise from 30% to 36%. Obviously, with a sex ratio of two males per female, this twofold development is accentuated: the proportion of women falls to 35% and that of young people to 15%, and the share of 60 years or more rises to 42%.

In both variants, the baseline replacement-level population declines at varying rates according to the ratio chosen. From a baseline of 9.8 billion in 2050, the

[6] The increase in the sex ratio at birth is a phenomenon that has very recently affected three countries of the Caucasus, with the sudden and perfectly simultaneous rise observed since 1995 in: Armenia, Azerbaidzhan, and Georgia (Badurashvili *et al.*, 2003).

world population would be no more than 3.6 billion by 2300 ($r_m = 1.5$) or even 1.2 billion ($r_m = 2$) instead of the 13.4 billion projected by the benchmark scenario. After 100 years, the growth rate levels off at a markedly negative level: –0.6% with $r_m = 1.5$ and even –1.2% with $r_m = 2$. That is because as the proportion of women declines at equal fertility rates, the birth rate stabilizes at a level below that of the benchmark scenario: instead of 12 per 1000, it rises to no more than 9 in the first case and even to less than 7 in the second (Appendix 2). At the same time, as the population ages, the death rate stabilizes conversely at the higher level of 15 or even 18 per 1000 instead of 11, according to the sex ratio used (Appendix 3).

Conversely, if the sex ratio at birth favors females, then not only does the proportion of women in the total population increase but the population also rejuvenates. With a ratio of 0.5, after stabilizing (from 2150), women's share of the total population rises to 68%, and the proportion of young people increases from 24% to 33% and that of the 60 and older falls from 30% to 21%. So, although in the benchmark scenario women's life expectancy gain makes them more contributory than men to the aging of the population, the excess of females to males at birth would produce an unhoped-for rejuvenation.

Obviously, that would come at the price of very rapid population growth. With women far outnumbering men, at constant fertility, the birth rate stabilizes at a much higher level than in the benchmark scenario (just over 18 per 1000 against 12). However, with a rejuvenated population, the crude death rate stabilizes at 7 per 1000 instead of 11. With a clearly positive rate of increase of 1.1%, the 9.8 billion population of 2050 would be 113 billion by 2300.

Would this high growth and rejuvenation potential be enough to offset the reverse effects of the transition to the one-child world? As Fig. 78–16 clearly shows, from this point of view, the result obtained with a sex ratio at birth of two females per males, is much better than those stemming from a rise in the sex ratio at birth, which actually worsens the baseline situation. Even so, a sex ratio at birth of two females per males remains manifestly too low to restore balance.

After stabilization in 2150, there would be 13.5% young people under 20 years of age (instead of 8.5% with $r_m = 1.05$) and 46% aged 60 and over (instead of 56%). Population loss would be substantially delayed, with the world population still at 400 million instead of 60 million in 2300 (Appendices 1 and 5).

What sex ratio at birth, then, would fully offset the effects of the transition to the one-child world? There could not be one, because it would entail women bearing only girls and reproducing without men. Only parthenogenesis or cloning could achieve that, bar establishing a huge sperm bank before allowing the last man to be extinguished.

The same sex ratios were combined with a rise in life expectancy to 150 years (in the sole case of a rectangularized survival curve). The results (see chapter appendix) are illustrated in Figs. 78–17 and 78–18, which speak for themselves: unsurprisingly, the effects specific to greater longevity (longer stabilization time, rise in the age structure) intersect with those of the sex ratio variation, whether 2.1 children per woman or the one-child scenario.

To add one final wild imagining, suppose that the sex life expectancy gap were to widen such that only women lived to 150 years (more precisely here 152.5 years) and the men's maximum longevity was 85 years (more precisely, 82.5 years). To cap it all, let us suppose that happens in a one-child world.

Figure 78–19 shows two final pyramid variants: one with a sex ratio at birth of two males per female, the other with two females per male. Both cases clearly end with an overwhelmingly female population and, perhaps less immediately evident, identical proportions in both cases: 87.1% when the population

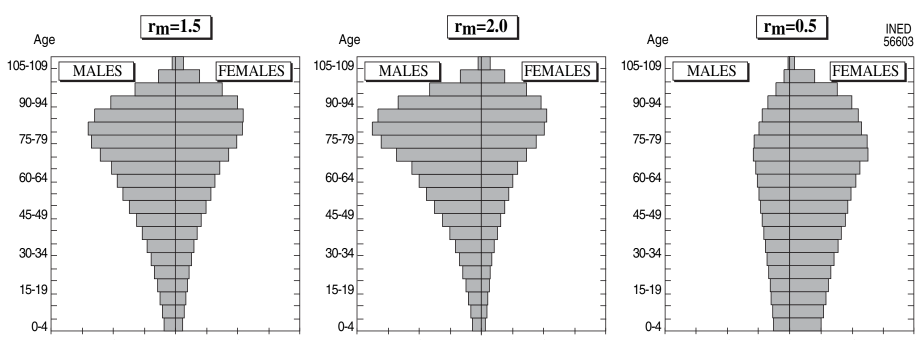


FIGURE 78–16 Population pyramids in 2150: effect of change in the sex ratio at birth (r_m) under a one-child fertility regime with a maximum life expectancy of 85 years (proportions %).

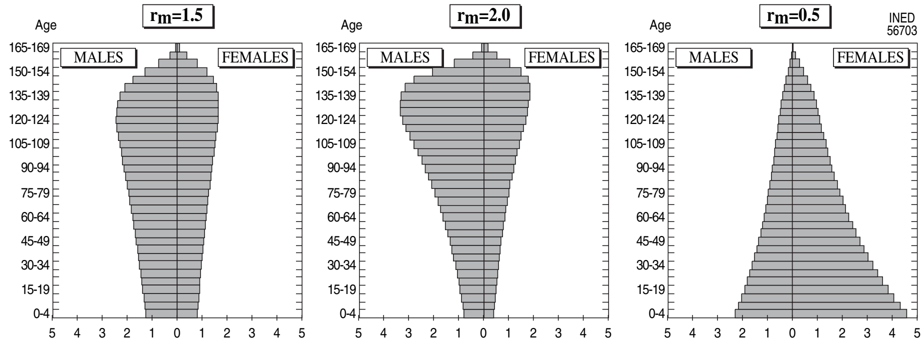


FIGURE 78–17 Population pyramids in 2250: effect of the change in sex ratio at birth (r_m) with United Nations' variant fertility of 2.1 children per woman and life expectancy of 150 years (rectangularized survival) (proportions %).

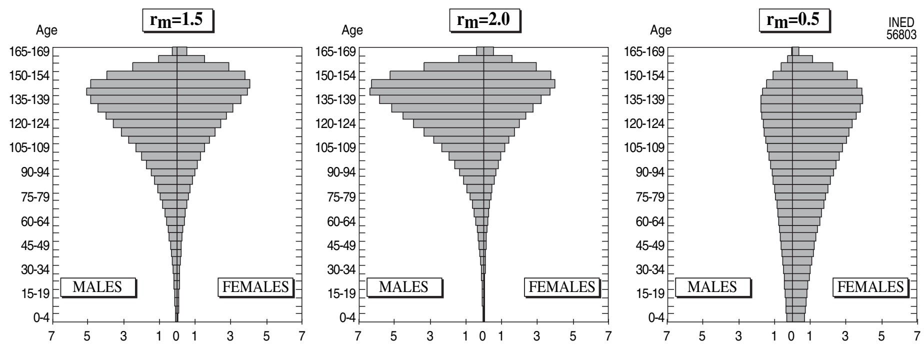


FIGURE 78–18 Population pyramids in 2250: effect of the change in sex ratio at birth (r_m) under the one-child fertility variant and life expectancy of 150 years (rectangularized survival) (proportions %).

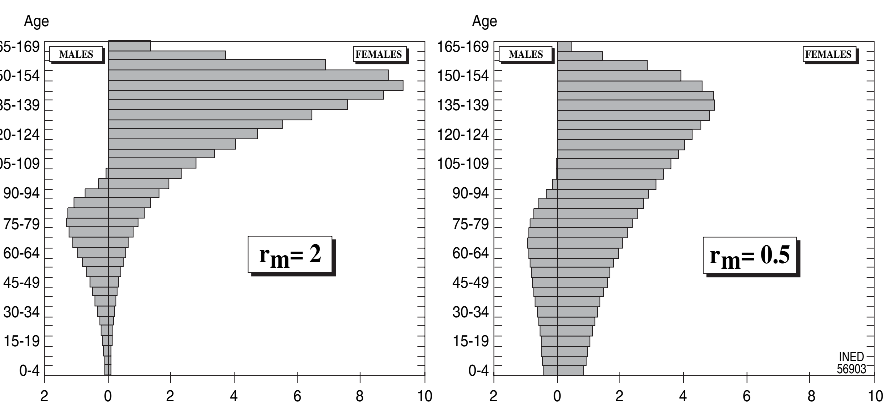


FIGURE 78–19 Population pyramids in 2250: effect of the change in sex ratio at birth (r_m) under the one-child fertility variant and female life expectancy of 152.5 years (rectangularized survival) and maximum male longevity of 82.5 years (proportions %).

stabilizes. However, a male dominance at birth again produces a much more marked aging of the population than does a female dominance. A sex ratio of two males per female gives 93% aged 60 years and over, and even 76% aged over 100 years, whereas two females per male gives proportions of just 77% and 52%.

In the first case, it is clear that a very large number of women aged between 130 and 150 years would have to live with a very small number of comparatively youthful 75-year-old men. However, working on the earlier male health assumptions, these men would not be younger than them. Such a situation would not endure, however, because, just to complicate matters, we have kept the current tempo of fertility for this scenario, so that the left-hand pyramid in Fig. 78–19 clearly shows that the population is likely to be extinguished very quickly for want of women of reproductive age: on stabilization, the annual rate of population increase is –3.7%, such that by 2300, the world population would be just 33 million (instead of 60 million with $r_m = 1.05$ and a life expectancy of 85 years for both sexes, or even 331 million at 150 years for both sexes). The right-hand pyramid leads a little less quickly but no less inexorably to disaster (the final rate of growth would be –1.4%, and the world population would be still 1 billion by 2300).

CONCLUSION

Life expectancy is clearly set to continue rising in many countries over the coming decades. The UN projects are likely to converge around a maximum world limit of 85 years. Should world fertility itself converge toward the absolute minimum replacement level of just under 2.1 children per woman with a life expectancy of 85, all the world's populations would stabilize not just in size but also age structure, at 24% under 20 years of age, 46% aged 20 to 59 years, and 30% aged 60 years and older. But although this life expectancy of 85 years is an absolute limit for some authors, for others it is a downbeat prognosis on two counts. First, even if longevity increases no further, there is a sufficiently wide gap between this putative limit and Jeanne Calment's age of 122 years to give the average person hope of passing it. But not only is there no proof that human longevity is an intractable constant, but there is good evidence that its boundaries are not fixed. As a result, there are no limits to what is possible. We have taken a conservative tack, hypothesizing a rise in life expectancy to around 150 years. All other things being equal, that would double the final prestabilization population size, but obviously at the cost of an unprecedented population aging: at 2.1 children per woman, there would be only 14% of young people aged under 20 years, and 27% of adults aged 20 to 59 years , but 59% aged 60 years and older (and even 30% aged 100 years or more). But there is no evidence that fertility is set to level off permanently at 2.1 children per woman. Were the sporadically observable one-child trend to spread, then the world population would be doomed to die out within a matter of centuries but not before its age structure had altered even more spectacularly: with a life expectancy of 150, the under-20s would account for less than 2%, the 20- to 59-year-olds under 7%, and the 60-year-olds or older more than 91%. Even more acutely, 74% would be centenarians.

However, this brief incursion into an imaginary future clearly lifts only a very tiny corner of the veil on the infinity of scenarios that could be hypothesized. Taking the world population as a whole as the baseline for speculation, we painlessly sidestepped the need for projections on migration, even though that is a phenomenon that is likely to loom large in coming decades. We nevertheless thought the exercise would be more productive if confined to a few basic parameters with a clearly identified role in population dynamics. Migration is a much more complex case that brings into play the population's relations with its exterior. That not only raises the problem of defining that exterior but the nature of the relations between the population and its exterior: in the case of immigration, for example, does the host population attract migrants, or are they imposed on it by exogenous factors. In other words, what population is the benchmark for devising rates from which to extrapolate trends for population forecasts?

For example, by keeping to the world level, we could have gone as far as projecting interplanetary migrations. But not knowing the supply of migrants, how can we apply to it migration rates to the Earth? Granted, departures could have been allowed for, including toward as-yet uninhabited planets.

But, many other equally far-fetched scenarios could have been devised: for instance, a 22nd century raiding party of extraterrestrials abducting large numbers of young women to regenerate their reproductive potential in a latter-day rape of the Sabine women. What impact would this have on the population of our planet? What if it were repeated at regular intervals like the Norman incursions in 9th and 10th century Europe what would the consequences of a permanent regime that included this type of periodic culling be?

In the same vein, one could clearly also muse on all kinds of events likely to set back life expectancy, such as the AIDS epidemic in Africa or the health crisis in Eastern Europe. And why not consider more drastic occurrences such as thermonuclear war or a meteorite crash?

In a more unassuming vein, however, the fertility and mortality trends sketched out and explored here are just a few of the infinite number of possibilities. The field remains wide open for those who wish to imagine other demographic never-never lands.

Of course, in themselves, none of the hypotheses explored here have much of a chance of ever really occurring, and the same is true for the various combinations that can be derived from them. Nonetheless, the possibility cannot be totally excluded, and our aim in studying their demographic consequences is to show that these scenarios are no more improbable than is perfect stabilization in strict terms of generation replacement.

Acknowledgments

This chapter was written with the collaboration of Rosa-Maria Lipsi and Alessandra Reale, Dipartmento di Scienze Demografiche, Università degli Studi di Roma, "La Sapienza," Italy. It is based on an initial article that developed more extensively the hypotheses selected for fertility and mortality trends, and provided more results in terms of birth rates, mortality, and population increase. The article was first published among the Proceedings of the international seminar in Siena (Vallin and Caselli *et al*., 1996), and a second time in English by the IPSEN Foundation (Vallin and Caselli, 1997). More recently, for the special issue of the journal *Population* on human longevity, we completed this essay, adding new developments, in particular the hypotheses concerning the sex ratio at birth, and we presented additional findings concerning changes in the age structure (Caselli and Vallin, 2001a,b). This chapter represents a synthesis of these various publications.

This chapter was translated from the French version by Accenta Ltd.

References

Allard Michel, Lèbre Victor, and Robine Jean-Marie, 1992. *Les 120 ans de Jeanne Calment, doyenne de l'humanité*. Paris, Le Cherche Midi, 160 p.

Allard Michel, Vallin Jacques, Andrieux Jean-Michel, and Robine Jean-Marie, 1996. In search of the secret of centenarians: a French demographic and medical survey about centenarians, *in:* Graziella Caselli and Alan Lopez (eds.), *Health and mortality among elderly populations*, p. 61–86. Oxford, Clarendon Press, 360 p.

Badurashvili Irina, Mesle France, and Vallin Jacques, 2003. Brusque élévation du rapport de masculinité à la naissance dans les pays du Caucase, *in:* Jacques Vallin (ed.), *Homage à Thérèse Locoh*, p. 275–288. Paris, INED, 296 p.

Barbi Elisabetta, Caselli Graziella, and Vallin Jacques, 2003. *Trajectories of extreme survival in heterogeneous population*. *Population E*, vol. 58(1), p. 43–66.

Benjamin Bernard, 1982. The span of life. *Journal of the Institute of the Actuaries*, vol. 109, p. 319–340.

Bourgeois-Pichat Jean, 1952. Essai sur la mortalité biologique de l'homme, *Population*, vol. 7(3), p. 381–394.

Bourgeois-Pichat Jean, 1966. Intervention, *in: Comptes rendus analytiques des séances: évolution de la fécondité, ses causes et ses conséquences, Documents officiels de la conférence*, vol. IV, CR2, p. 6–7. Strasbourg, Council of Europe.

Bourgeois-Pichat Jean, 1977. Le dilemme de la révolution démographique: croître ou vieillir, *in: Actes du congrès de Vichy sur le vieillissement. Vol. sur "Future Age Projection Society*, p. 260–278. Paris, Institut de la vie à Vichy.

Bourgeois-Pichat Jean, 1978. Future outlook for mortality decline in the world. *Population Bulletin of the United Nations*, no. 11, p. 12–41.

Bourgeois-Pichat Jean, 1987. La demografia e le altre scienze: interazioni di metodo e cotenuto, *in:* Eugenio Sonnino, Antonella Pinnelli, Dionysia Maffioli, and Anunciata Nobile (eds.) *Demografia: scienza insegnamento professione*, p. 53–67. Rome, Franco Angeli, 574 p.

Bourgeois-Pichat Jean, 1988. Du XX[e] au XXI[e] siècle: l'Europe et sa population après l'an 2000, *Population*, vol. 43(1), p. 9–44.

Calot Gérard and Caselli Graziella, 1988. *La mortalité en Chine d'après le recensement de 1982: I. Analyse selon le sexe et l'âge au niveau national et provincial*. Paris, INED, IUSSP.

Caselli Graziella and Vallin Jacques, 2001a. Une démographie sans limite?, *Population*, vol. 56(1/2), 51–84.

Caselli Graziella and Vallin Jacques, 2001b. Demographic trends: beyond the limits, *Population, An English Selection*, vol. 13(1), p. 41–72.

Caselli Graziella, Vaupel James W., and Yashin Anatoli I., 2000. Longevity, heterogeneity and selection. p. 49–69. *Proceedings of the XL Riunione Scientifica della Società Italiana di Statistica*, Florence, SIS., 437 p.

Fries James F., 1980. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, vol. 303, p. 130–135.

Fries James, 1989. The compression of morbidity: near or far? *The Milbank Quarterly*, vol. 67, p. 208.

Guibert-Lantoine Chaterine and Monnier Alain, 1995. La conjoncture démographique: l'Europe et les pays développés d'Outre-mer, *Population*, vol. 50(4/5), p. 1185–1210.

Kannisto Väinö, 1994. *Development of oldest-old mortality, 1950–1990: evidence from 28 developed countries*. Odense, Odense University Press, 108 p.

Kannisto Väinö, 1996. *The advancing frontier of survival*. Odense, Odense University Press, 144 p.

Meslé France, 1995. La mortalité en France: le recul se poursuit, *Population*, vol. 50(3), p. 743–776.

Meslé France and Vallin Jacques, 2002. Évolution de la mortalité aux âges élevés en France depuis 1950, *in:* Joël Ankri, Patrice Burdelais, Francis Derriennec, Myriam Khlat Pierre Mormiche and Catherine Sermet (eds.), *La santé aux grands âges*, p. 5–48. Paris, INED, 140 p.

Meslé France, Vallin Jacques, and Robine Jean-Marie, 2000. Vivre 100 ans en France, *Gérontologie et Société*, no. 94, p. 101–120.

Miller Barbara, 1996. Changing equality: the politics of sex selective abortion in Asia, *in:* IUSSP (ed.), *Socio-cultural and political aspects of abortion in a changing world*, p. 1–15. Liège, IUSSP, nonsequential pagination.

National Institute of Statistics [ISTAT], 2000. *Nascite: caratteristiche demografiche e sociali. Anno 1996. Annuario no. 5*. Rome, ISTAT, 132 p.

Olshansky S. Jay, Carnes Bruce A., and Cassel Christine, 1990. In search of Methuselah: estimating the upper limits to human longevity. *Science*, vol. 250, p. 634–640.

Ssantini Antonio, 1995. *Continuità e discontinuità nel comportamento riproduttivo delle donne italiane nel dopoguerra: tendenze generali della fecondità delle coorti nelle ripartizioni tra il 1952 e 1991*. Florence, Dipartimento di Statistica G. Parenti, 62 p.

United Nations [UN], 1982. Projections démographiques mondiales à long terme effectuées en 1980, *Bulletin démographique des Nations unies*, no. 14, p.17–31.

United Nations [UN], 1992. *Long-range World population projections: two centuries of population growth 1950–2150*. New York, UN, Department for Economic and Social Affairs, 32 p.

United Nations [UN], 1994. *The sex and age distribution of the world populations, the 1994 revision*. New York, UN, Population Division, 898 p.

United Nations [UN], 1995. *World population prospects, the 1994 revision*. New York, UN, Population Division, 886 p.

United Nations [UN], 1999. *World population prospects, the 1998 revision, Volume I*. New York, UN, Department for Economic and Social Affairs, Population Division, 627 p.

United Nations [UN], 2001. *World population prospects: the 2000 revision. Volume 1: Comprehensive tables*. New York, UN, Population Division, 746 p.

United Nations [UN], 2003. *World population prospects: the 2002 revision. Volume 1: Comprehensive Tables* New York, UN, Population Division, 781 p.
Vallin Jacques, 1989. De la mortalité endogène à l'allongement de la vie: hommage à Jean Bourgeois-Pichat, *in:* Jacques Vallin (ed.), *L'avenir de l'espérance de vie*, p. 23–35. Paris, INED and PUF, 106 p.
Vallin Jacques, 1994. *Réflexions sur l'avenir de la population mondiale.* Paris, CEPED, 28 p.
Vallin Jacques and Caselli Graziella, 1997. Toward a new horizon in demographic trends: the combined effects of 150 years life expectancy and new fertility models, *in:* Jean-Marie Robine *et al.* (eds.), *Longevity: to the limits and beyond*, p. 29–68. Berlin, Springler-Verlag, Fondation IPSEN, 180 p.
Vallin Jacques, Caselli Graziella, and Reale Alessandra, 1996. L'avenir de l'humanité à plus long terme: après la transition démographique, *in:* Graziella Caselli (ed.), *Demografia: analisi e sintesi. Cause e conseguenze dei processi demografici, Actes du séminaire international. Demography: analysis and synthesis. Siena 22–24 avril 1996, volume1*, p. 103–141. Rome, Università di Roma, Dipartimento di Scienze Demografiche, 252 p.
Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978.* Paris, INED and PUF, 608 p.
Vallin Jacques and Meslé France, 1989. À long terme, l'écart d'espérance de vie entre hommes et femmes devrait diminuer, *Population*, vol. 44(6), p. 1244–1251.
Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle*. Paris, INED, 102 p + 1 CD-rom.
Van de Kaa Dirk J., 1987. Europe's second demographic transition, *Population Bulletin*, vol. 42(1), p. 1–59.
Vaupel James W., Carrey J. R., Christensen K., Johnson T. E., Yashin Anatoli I., Holm N. V., Iachine Ivan A., Kannistö Vaïno, Khanzaeli A. A., Liedo P., Longo V. D., Zeng Yi, Manton Kenneth J., and Curtsinger J. W., 1998. Biodemographic trajectories of longevity, *Science*, no. 280, p. 850–860.
Vaupel James W., Manton Kenneth J., and Stallard Eric, 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, vol. 16(3), p. 439–454.
Walford Roy, 1982. Immunoregulation and aging, *in:* Samuel H. Preston (ed.), *Biological and social aspects of mortality and the length of life: proceedings of a seminar at Fiuggi, Italy, May*.
Walford Roy, 1984a. *Maximum life span*, New York, Norton, 256 p.
Walford Roy, 1984b. *La vie la plus longue*. Paris, Laffont, 250 p.
Wilmoth John R. and Lundström Hans, 1996. Extreme longevity in five countries: presentation of trends with special attention to issues of data quality, *European Journal of Population/Revue Européenne de Démographie*, vol. 12(1), p. 63–93.
Zanata Anna Laura and De Rose Alessandra, 1995. *Il figlio unico in Italia: frequenza e determinanti di scelta.* Rome, Dipartimento di Scienze Demografiche, Università degli Studi di Roma La Sapienza, 24 p.
Zhang W. *et al.* 1983. Yinger xingbili shitiao yao qieshi jiuzheng [Efficient measures must be taken against the abnormal level of sex ratio at birth]. *Shesshui [Society]*, 1983, no. 2.

APPENDIX Different World Population Parameter Trends 2050 to 2315 According to Two Fertility Scenarios, Three Mortality Projections, and Four Sex Ratios at Birth (r_m)

Projections	1. Population (millions)											
	2050	2075	2100	2125	2150	2175	2200	2225	2250	2275	2300	2315
UN maximum life expectancy (85 years)												
Fertility: UN = 2.1												
r_m = 1.05	9.833	11.342	11.808	12.007	12.201	12.393	12.587	12.785	12.985	13.189	13.396	13.522
r_m = 1.50	9.833	11.327	11.262	10.452	9.216	7.903	6.775	5.808	4.979	4.268	3.659	3.336
r_m = 2.00	9.833	11.315	10.852	9.381	7.344	5.398	3.966	2.913	2.140	1.571	1.154	959
r_m = 0.50	9.833	11.375	12.933	15.689	20.392	27.119	36.073	47.987	63.843	84.939	113.005	134.120
Fertility: Single child												
r_m = 1.05	7.918	7.672	5.887	3.972	2.360	1.282	696	377	205	111	60	42
r_m = 1.50	7.918	7.666	5.752	3.700	1.979	913	421	194	89	41	19	12
r_m = 2.00	7.918	7.662	5.651	3.507	1.722	684	272	108	43	17	7	4
r_m = 0.50	7.918	7.684	6.164	4.588	3.305	2.331	1.642	1.157	815	575	405	328
Life expectancy 150 years with survival curve rectangularity												
Fertility: UN = 2.1												
r_m = 1.05	9.833	11.966	14.014	16.004	17.989	20.014	21.631	22.234	22.695	23.164	23.646	23.940
r_m = 1.50	9.833	11.950	13.469	14.455	15.008	15.212	14.654	12.855	11.073	9.536	8.216	7.513
r_m = 2.00	9.833	11.938	13.058	13.388	13.139	12.473	11.027	8.432	6.221	4.588	3.386	2.823
r_m = 0.50	9.833	11.998	15.141	19.676	26.178	35.399	47.845	63.975	85.540	114.375	152.934	182.052
Fertility: Single child												
r_m = 1.05	7.918	8.269	7.783	6.877	5.829	4.782	3.537	2.045	1.115	607	331	230
r_m = 1.50	7.918	8.264	7.649	6.611	5.460	4.341	3.043	1.543	713	329	152	96
r_m = 2.00	7.918	8.259	7.547	6.421	5.212	4.057	2.734	1.240	492	196	78	45
r_m = 0.50	7.918	8.281	8.060	7.482	6.745	5.968	4.955	3.592	2.544	1.802	1.277	1.038
Life expectancy 82.5 years (males) and 152.5 years (females) (survival curve rectangularity)												
Fertility: Single child												
r_m = 2.00	7.918	7.956	6.602	4.965	3.350	2.120	1.228	530	210	84	33	19
r_m = 2.00	7.918	7.979	7.118	6.075	5.228	4.588	3.855	2.820	1.998	1.415	1.002	815

(continues)

APPENDIX (*continued*)

Projections	2. Birth rate (per 1000)											
	2050	**2075**	**2100**	**2125**	**2150**	**2175**	**2200**	**2225**	**2250**	**2275**	**2300**	**2315**
UN maximum life expectancy (85 years)												
Fertility: UN = 2.1												
$r_m = 1.05$	13.92	12.34	12.08	12.08	12.09	12.09	12.09	12.09	12.09	12.09	12.09	12.09
$r_m = 1.50$	13.92	11.41	9.96	9.19	8.93	8.92	8.93	8.93	8.93	8.93	8.93	8.93
$r_m = 2.00$	13.92	10.71	8.33	7.03	6.60	6.60	6.60	6.61	6.61	6.61	6.61	6.61
$r_m = 0.50$	13.92	14.20	16.31	17.89	18.37	18.43	18.43	18.43	18.42	18.42	18.42	18.42
Fertility: Single child												
$r_m = 1.05$	10.45	6.26	4.43	3.57	3.30	3.30	3.30	3.30	3.30	3.30	3.30	3.31
$r_m = 1.50$	10.45	5.79	3.57	2.54	2.22	2.23	2.23	2.23	2.23	2.23	2.24	2.23
$r_m = 2.00$	10.45	5.42	2.94	1.85	1.52	1.53	1.53	1.53	1.53	1.53	1.54	1.53
$r_m = 0.50$	10.45	7.22	6.22	5.93	5.86	5.87	5.87	5.87	5.87	5.87	5.87	5.87
Life expectancy 150 years with survival curve rectangularity												
Fertility: UN = 2.1												
$r_m = 1.05$	13.92	11.57	10.07	9.00	8.17	7.50	7.13	7.09	7.09	7.10	7.09	7.09
$r_m = 1.50$	13.92	10.70	8.22	6.56	5.43	4.62	4.18	4.12	4.12	4.12	4.12	4.12
$r_m = 2.00$	13.92	10.04	6.83	4.85	3.62	2.82	2.39	2.33	2.33	2.33	2.33	2.33
$r_m = 0.50$	13.92	13.32	13.82	14.25	14.38	14.24	14.12	14.11	14.10	14.10	14.10	14.10
Fertility: Single child												
$r_m = 1.05$	10.45	5.71	3.27	2.01	1.29	0.86	0.65	0.62	0.62	0.62	0.62	0.62
$r_m = 1.50$	10.45	5.28	2.62	1.38	0.77	0.45	0.31	0.29	0.29	0.29	0.29	0.29
$r_m = 2.00$	10.45	4.95	2.15	0.98	0.48	0.25	0.15	0.14	0.14	0.14	0.14	0.14
$r_m = 0.50$	10.45	6.59	4.66	3.57	2.82	2.27	1.97	1.93	1.93	1.93	1.93	1.93
Life expectancy 82.5 years (males) and 152.5 years (females) (survival curve rectangularity)												
Fertility: Single child												
$r_m = 2.00$	10.45	5.18	2.48	1.28	0.76	0.48	0.34	0.32	0.32	0.32	0.32	0.32
$r_m = 0.50$	10.45	6.90	5.33	4.43	3.65	2.95	2.52	2.46	2.46	2.46	2.46	2.46

Projections	3. Death rate (per 1000)											
	2050	**2075**	**2100**	**2125**	**2150**	**2175**	**2200**	**2225**	**2250**	**2275**	**2300**	**2315**
UN maximum life expectancy (85 years)												
Fertility: UN = 2.1												
$r_m = 1.05$	5.42	9.84	11.32	11.44	11.48	11.47	11.46	11.46	11.46	11.46	11.46	11.46
$r_m = 1.50$	5.42	9.88	11.94	13.37	14.93	15.05	15.09	15.10	15.10	15.10	15.09	15.09
$r_m = 2.00$	5.42	9.91	12.46	15.10	18.48	18.83	18.96	18.98	18.96	18.95	18.95	18.95
$r_m = 0.50$	5.42	9.77	10.20	8.47	7.11	7.04	7.01	7.00	7.00	7.00	7.00	7.00
Fertility: Single child												
$r_m = 1.05$	6.68	13.96	18.35	22.00	27.37	27.76	27.78	27.75	27.73	27.72	27.73	27.73
$r_m = 1.50$	6.68	13.99	18.85	23.94	32.25	33.11	33.31	33.27	33.20	33.17	33.13	33.10
$r_m = 2.00$	6.68	14.01	19.24	25.52	36.73	38.25	38.65	38.57	38.43	38.34	38.27	38.23
$r_m = 0.50$	6.68	13.91	17.40	18.53	20.00	20.02	19.89	19.84	19.85	19.86	19.86	19.86
Life expectancy 150 years with survival curve rectangularity												
Fertility: UN = 2.1												
$r_m = 1.05$	5.42	4.75	4.45	4.12	3.77	3.42	5.67	6.25	6.28	6.27	6.27	6.27
$r_m = 1.50$	5.42	4.77	4.64	4.58	4.55	4.55	8.52	10.09	10.10	10.07	10.08	10.09
$r_m = 2.00$	5.42	4.78	4.80	4.97	5.22	5.60	11.49	14.56	14.50	14.44	14.46	14.49
$r_m = 0.50$	5.42	4.72	4.08	3.31	2.55	1.89	2.48	2.46	2.48	2.49	2.49	2.48
Fertility: Single child												
$r_m = 1.05$	6.68	6.69	7.39	8.03	8.74	9.47	19.56	24.76	24.81	24.83	24.85	24.85
$r_m = 1.50$	6.68	6.70	7.53	8.37	9.36	10.52	23.14	31.11	30.96	30.98	31.04	31.04
$r_m = 2.00$	6.68	6.71	7.64	8.63	9.83	11.32	26.10	37.08	36.65	36.69	36.83	36.81
$r_m = 0.50$	6.68	6.66	7.11	7.35	7.49	7.45	13.47	15.51	15.70	15.74	15.73	15.71
Life expectancy 82.5 years (males) and 152.5 years (females) (survival curve rectangularity)												
Fertility: Single child												
$r_m = 2.00$	6.68	10.21	12.46	14.85	19.21	18.13	30.97	37.15	36.70	36.88	37.05	36.96
$r_m = 0.50$	6.68	10.13	11.38	10.95	8.90	8.24	13.03	16.06	16.23	16.27	16.26	16.24

(continues)

APPENDIX (*continued*)

Projections	4. Rate of natural increase (per 1000)											
	2050	**2075**	**2100**	**2125**	**2150**	**2175**	**2200**	**2225**	**2250**	**2275**	**2300**	**2315**
UN maximum life expectancy (85 years)												
Fertility: UN = 2.1												
r_m = 1.05	8.50	2.49	0.76	0.64	0.61	0.62	0.62	0.62	0.62	0.62	0.62	0.62
r_m = 1.50	8.50	1.53	–1.98	–4.19	–5.99	–6.12	–6.16	–6.17	–6.17	–6.16	–6.16	–6.16
r_m = 2.00	8.50	0.80	–4.13	–8.07	–11.88	–12.23	–12.35	–12.37	–12.36	–12.34	–12.34	–12.35
r_m = 0.50	8.51	4.43	6.11	9.41	11.25	11.38	11.42	11.43	11.42	11.42	11.42	11.42
Fertility: Single child												
r_m = 1.05	3.77	–7.70	–13.93	–18.43	–24.07	–24.46	–24.48	–24.45	–24.43	–24.42	–24.43	–24.42
r_m = 1.50	3.77	–8.21	–15.28	–21.40	–30.03	–30.89	–31.08	–31.04	–30.97	–30.94	–30.89	–30.86
r_m = 2.00	3.77	–8.59	–16.30	–23.67	–35.21	–36.72	–37.12	–37.04	–36.90	–36.81	–36.73	–36.69
r_m = 0.50	3.77	–6.69	–11.18	–12.60	–14.14	–14.14	–14.02	–13.97	–13.98	–13.99	–13.99	–13.99
Life expectancy 150 years with survival curve rectangularity												
Fertility: UN = 2.1												
r_m = 1.05	8.50	6.82	5.62	4.88	4.40	4.08	1.47	0.85	0.81	0.82	0.82	0.82
r_m = 1.50	8.50	5.94	3.58	1.98	0.88	0.06	–4.34	–5.97	–5.98	–5.95	–5.95	–5.97
r_m = 2.00	8.50	5.26	2.02	–0.12	–1.60	–2.78	–9.09	–12.23	–12.16	–12.11	–12.12	–12.15
r_m = 0.50	8.51	8.61	9.74	10.94	11.83	12.35	11.63	11.65	11.62	11.61	11.61	11.62
Fertility: Single child												
r_m = 1.05	3.77	–0.97	–4.12	–6.03	–7.45	–8.61	–18.91	–24.14	–24.19	–24.20	–24.23	–24.23
r_m = 1.50	3.77	–1.42	–4.91	–6.99	–8.59	–10.07	–22.83	–30.83	–30.67	–30.69	–30.76	–30.75
r_m = 2.00	3.77	–1.76	–5.50	–7.66	–9.36	–11.08	–25.95	–36.94	–36.51	–36.55	–36.69	–36.68
r_m = 0.50	3.77	–0.07	–2.45	–3.78	–4.67	–5.18	–11.51	–13.58	–13.77	–13.81	–13.80	–13.78
Life expectancy 82.5 years (males) and 152.5 years (females) (survival curve rectangularity)												
Fertility: Single child												
r_m = 2.00	3.77	–5.03	–9.98	–13.57	–18.46	–17.66	–30.63	–36.83	–36.39	–36.56	–36.73	–36.64
r_m = 0.50	3.77	–3.24	–6.05	–6.53	–5.25	–5.28	–10.51	–13.61	–13.78	–13.82	–13.81	–13.78

(continues)

APPENDIX (*continued*)

Projections		**5. Population shares (percentage)**											
r_m	**Age**	**2050**	**2075**	**2100**	**2125**	**2150**	**2175**	**2200**	**2225**	**2250**	**2275**	**2300**	**2315**
UN maximum life expectancy (85 years)													
Fertility: UN = 2.1													
1.05	0–19	27.7	24.5	24.0	23.9	23.9	23.9	23.9	23.9	23.9	23.9	23.9	23.9
	20–59	52.3	47.6	46.4	46.4	46.4	46.4	46.4	46.4	46.4	46.4	46.4	46.4
	60+	20.0	27.9	29.6	29.7	29.7	29.7	29.7	29.7	29.7	29.7	29.7	29.7
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1.50	0–19	27.7	24.4	20.9	19.4	18.9	18.9	18.9	18.9	18.9	18.9	18.9	18.9
	20–59	52.3	47.7	48.0	46.6	45.1	45.0	45.0	45.0	45.1	45.1	45.1	45.1
	60+	20.0	27.9	31.1	34.0	35.9	36.0	36.0	36.0	36.0	36.0	36.0	36.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2.00	0–19	27.7	24.4	18.4	15.8	15.0	15.0	14.9	14.9	14.9	14.9	14.9	14.9
	20–59	52.3	47.7	49.3	46.4	43.1	42.9	42.9	42.9	43.0	43.0	43.0	43.0
	60+	20.0	27.9	32.2	37.8	41.9	42.1	42.1	42.1	42.1	42.1	42.1	42.1
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	0–19	27.7	24.8	29.5	32.2	32.8	32.8	32.8	32.8	32.8	32.8	32.8	32.8
	20–59	52.3	47.5	43.5	44.9	46.3	46.4	46.5	46.5	46.5	46.4	46.4	46.4
	60+	20.0	27.8	27.0	22.8	20.9	20.8	20.7	20.7	20.8	20.8	20.8	20.8
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Fertility: Single child													
1.05	0–19	21.2	16.3	11.5	9.3	8.5	8.5	8.5	8.5	8.5	8.5	8.5	8.5
	20–59	53.9	46.5	45.2	39.5	36.0	35.9	35.9	35.9	35.9	35.9	35.9	35.9
	60+	24.8	37.1	43.2	51.2	55.5	55.6	55.6	55.6	55.6	55.6	55.6	55.6
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1.50	0–19	21.2	16.3	9.9	7.2	6.2	6.2	6.2	6.2	6.2	6.2	6.2	6.2
	20–59	53.9	46.6	45.9	38.0	32.4	32.2	32.2	32.3	32.3	32.3	32.3	32.3
	60+	24.8	37.2	44.3	54.8	61.4	61.6	61.6	61.6	61.6	61.6	61.5	61.5
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2.00	0–19	21.2	16.2	8.6	5.6	4.6	4.5	4.5	4.5	4.5	4.5	4.5	4.6
	20–59	53.9	46.6	46.4	36.7	29.1	28.8	28.8	28.8	28.9	28.9	29.0	29.0
	60+	24.8	37.2	45.1	57.7	66.4	66.7	66.7	66.6	66.6	66.6	66.5	66.4
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	0–19	21.2	16.5	14.7	13.9	13.5	13.5	13.5	13.5	13.5	13.5	13.5	13.5
	20–59	53.9	46.4	44.0	41.5	40.9	41.0	41.0	41.0	41.0	41.0	41.0	41.0
	60+	24.8	37.1	41.3	44.6	45.6	45.6	45.5	45.5	45.5	45.5	45.5	45.5
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

(continues)

APPENDIX (*continued*)

Projections		5. Population shares (percentage)											
r_m	Age	2050	2075	2100	2125	2150	2175	2200	2225	2250	2275	2300	2315
Life expectancy 150 years with survival curve rectangularity													
Fertility: UN = 2.1													
1.05	0–19	27.7	23.3	20.2	18.0	16.4	15.0	14.2	14.1	14.1	14.1	14.1	14.1
	20–59	52.3	45.3	39.4	35.2	31.9	29.2	27.6	27.4	27.4	27.4	27.4	27.4
	60–99	19.9	30.7	34.8	33.3	30.8	28.2	26.6	26.4	26.4	26.4	26.4	26.4
	100+	0.1	0.7	5.6	13.5	21.0	27.6	31.6	32.1	32.1	32.1	32.1	32.1
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1.50	0–19	27.7	23.2	17.5	14.1	11.7	10.0	8.9	8.8	8.8	8.8	8.8	8.8
	20–59	52.3	45.4	40.5	34.1	28.1	23.8	21.3	20.9	20.9	20.9	20.9	20.9
	60–99	19.9	30.7	36.2	36.8	35.0	30.1	27.0	26.5	26.5	26.5	26.5	26.5
	100+	0.1	0.7	5.8	15.0	25.2	36.1	42.8	43.9	43.9	43.8	43.8	43.8
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2.00	0–19	27.7	23.1	15.4	11.2	8.4	6.6	5.5	5.3	5.3	5.3	5.3	5.3
	20–59	52.3	45.4	41.3	32.9	24.5	18.9	15.8	15.3	15.3	15.3	15.3	15.3
	60–99	19.9	30.7	37.3	39.8	38.3	30.6	25.7	24.8	24.7	24.8	24.8	24.8
	100+	0.1	0.7	6.0	16.2	28.8	43.9	53.0	54.7	54.7	54.7	54.7	54.7
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	0–19	27.7	23.5	25.2	25.8	25.8	25.4	25.2	25.2	25.2	25.2	25.2	25.2
	20–59	52.3	45.2	37.4	36.1	36.4	36.1	35.7	35.7	35.7	35.7	35.7	35.7
	60–99	19.9	30.6	32.2	27.1	23.3	22.7	22.3	22.3	22.3	22.4	22.4	22.4
	100+	0.1	0.7	5.2	11.0	14.4	15.7	16.8	16.8	16.8	16.8	16.8	16.8
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Fertility: Single child													
1.05	0–19	21.2	15.2	8.7	5.4	3.5	2.3	1.7	1.6	1.6	1.6	1.6	1.6
	20–59	53.9	43.4	34.5	23.1	14.8	9.8	7.2	6.8	6.8	6.8	6.8	6.8
	60–99	24.7	40.4	47.2	46.1	38.7	25.9	19.1	17.9	17.9	17.9	17.9	17.9
	100+	0.1	1.0	9.5	25.4	43.1	62.0	72.0	73.6	73.7	73.6	73.6	73.6
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1.50	0–19	21.2	15.1	7.5	4.0	2.3	1.3	0.9	0.8	0.8	0.8	0.8	0.8
	20–59	53.9	43.4	34.8	21.6	11.9	6.9	4.6	4.2	4.2	4.2	4.2	4.2
	60–99	24.7	40.4	48.0	48.0	39.8	23.8	15.7	14.3	14.3	14.3	14.3	14.3
	100+	0.1	1.0	9.7	26.4	46.0	68.0	78.9	80.8	80.8	80.7	80.7	80.7
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2.00	0–19	21.2	15.1	6.4	3.1	1.5	0.8	0.5	0.4	0.4	0.4	0.4	0.4
	20–59	53.9	43.4	35.0	20.3	9.8	5.0	2.9	2.6	2.6	2.6	2.6	2.6
	60–99	24.7	40.4	48.7	49.4	40.6	21.7	12.9	11.3	11.3	11.3	11.3	11.3
	100+	0.1	1.0	9.8	27.2	48.2	72.6	83.7	85.7	85.7	85.7	85.7	85.7
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	0–19	21.2	15.3	11.3	8.5	6.7	5.3	4.5	4.4	4.4	4.4	4.4	4.4
	20–59	53.9	43.3	34.0	25.7	20.3	16.3	13.9	13.6	13.6	13.6	13.6	13.6
	60–99	24.7	40.3	45.6	42.4	35.9	28.4	24.0	23.4	23.4	23.4	23.4	23.4
	100+	0.1	1.0	9.2	23.3	37.2	50.0	57.6	58.6	58.6	58.6	58.6	58.6
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

(continues)

APPENDIX (*continued*)

Projections		5. Population shares (percentage)											
r_m	Age	2050	2075	2100	2125	2150	2175	2200	2225	2250	2275	2300	2315
Life expectancy 82.5 (males) and 152.5 years (females) (survival curve rectangularity)													
Fertility: UN = 2.1													
2.00	0–19	21.2	15.6	7.4	4.0	2.4	1.5	1.0	0.9	0.9	0.9	0.9	0.9
	20–59	53.9	44.9	39.8	26.0	15.0	9.4	6.4	6.0	6.0	6.0	6.0	6.0
	60–99	24.7	38.6	45.8	50.0	41.7	26.7	18.4	16.9	17.0	17.0	17.0	17.0
	100+	0.1	0.8	7.0	20.1	40.9	62.4	74.1	76.2	76.1	76.1	76.0	76.0
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.50	0–19	21.2	15.9	12.7	10.5	8.6	6.9	5.8	5.7	5.7	5.7	5.7	5.7
	20–59	53.9	44.8	38.2	31.5	26.0	21.0	17.7	17.2	17.2	17.2	17.2	17.2
	60–99	24.7	38.5	42.5	41.6	39.2	31.3	26.2	25.3	25.3	25.3	25.3	25.3
	100+	0.1	0.8	6.5	16.4	26.2	40.7	50.3	51.9	51.9	51.9	51.9	51.9
	Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

SECTION

II

POPULATION AND SOCIETY

Introduction

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Volume I, Section I focused on the essence of demography as a discipline: the study of the intrinsic mechanisms of population dynamics. Volumes I, Section II, and Volume II each focused on one of the three main forces in population dynamics—fertility, mortality, and migration—to discuss their determinants. This laid the groundwork for Section I in this volume, which gives an overview of the underpinnings of demographic history and geography and considers prospective medium- and long-term trends. With this section we take a further step, placing population dynamics in its biological, socio-economic, and cultural setting. The final volume, following the original plan of the treatise, will cover the history of ideas and policies on population, and the observation methods and analysis tools used in demography.

Section II is in two rather different parts. The nine chapters of Part I discuss the consequences of demographic change, whereas Part II addresses more broadly the interactions between population dynamics, or demography in general, and the main aspects of the life of populations and societies.

The volumes devoted to fertility, mortality, and migration gave in-depth treatment, in connection with each of those three components, to the biological, economic, social, cultural, and political determinants of population change. In Part I of this section, we consider these determinants from a different angle. It will be a more succinct examination, not because they are less important—far from it—but because the demographer's role is more to develop understanding of demographic processes as such than to describe all the effects demographic change may have in fields that are mainly studied by other disciplines. We will systematically address the main areas affected by demographic change (human biology and health, the economy, society, the environment), but others will be left aside, partly because the causal link is less clear-cut and partly because they lend themselves less readily to quantitative approaches. Admittedly, these reasons are not relevant in all cases; for example, we might very well have explored such cultural consequences as the distribution of languages[1] or religions. It would have been more difficult to examine the relationship among population dynamics and political regimes or legal systems, or the psychological consequences of changes in demographic regime.[2] We had to set ourselves reasonable limits, and we hope we have covered the essentials. We have not focused on fertility, mortality, and migration in turn but have considered the combined consequences of changes in these components from the standpoint of the overall demographic changes they produce in terms of population size and composition. In so doing, we will briefly point out what is mainly owing to fertility, mortality, or migration, but rarely in isolation from the rest.

In the previous volumes and in Part I of this section, we have rarely strayed from the basic principle of demographic analysis, which is to consider a population as a set of individuals who, in the course of their life events and experiences, affect the size and structure of the whole. However, although this approach opens the way to a sound understanding of demographic processes and their determinants, it does not give a sufficient grasp of all the interactions between demography and society. For this reason (although making no claim to exhaustive treatment, because here, too, the fields addressed belong primarily to other disciplines), Part II will put a demographic perspective on the biological, economic, social, and cultural aspects of the life of populations, which transcend the reduction of a population to a sum of individual elements. First, we will view the individual and the demographic events of their life in the context of their life cycle as a whole; and then, we will consider families and households, social structures, sex, economic categories, and genetic heritage.

[1] See "La dynamique des langues" by Francois Hêran, Alexandre Filhon, and Christine Deprez, 2002. *Population et sociêtê*, no. 376.

[2] See Jacques Vallin, 1993. "Dêmographie et anxiêtê," *L'exo, tribune de l' anxiêtê*, no. 22, p. 2–4.

P A R T

I

CONSEQUENCES OF DEMOGRAPHIC CHANGE

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Demographic change, particularly the spectacular growth populations have undergone or are still undergoing today through the major historical shift demographers call *the demographic transition* (see Chapters 68, 69 of Volume III), has numerous consequences, affecting almost every aspect of human life from biology and health to the economy, the environment, sociology, and culture. We make no claim to cover all these exhaustively here. More modestly, we have tried to illustrate the diversity of these effects by addressing particular aspects in fields in which they are most striking: biology and health, economy and society, and environment.

The first two chapters discuss consequences for biology and health. The principle of natural selection, which gave rise to the human species as to all others, is based on the evolution of a population's genetic heritage. It is therefore crucial for the future of our species to consider the influence demographic behaviors and changes may have on the evolution of human genetic heritage. This is addressed in Robert Cliquet's chapter on population change and genetic diversity (Chapter 79).

Regardless of the long-term evolution of genetic heritage, all else being equal, a population's health may be affected simply by the fact that medicine has become sufficiently effective to enable many men and women to live to an age which in earlier times only a tiny minority attained. Generally speaking, health declines with age, and if more people live on into old age, the average health of the population is likely to deteriorate. The only possible remedy is for health at any given age to improve systematically as life expectancy increases. Whether we really have this remedy is a matter of some debate. To answer the question, one must be able to make realistic observations in a field that arouses strong subjective reactions. Jean-Marie Robine and Carol Jagger discuss the question methodically and precisely in Chapter 80, "Lengthening of Life and the Population Health Status".

Partly because of its effects on the health situation, the aging of the population also has important consequences in the economic, social and cultural fields. The next two chapters address this in different but complementary ways. Chapter 81, by Antonio Golini, covers "Demographic Trends and Relationships between Generations," whereas Chapter 82, by Jacques Légaré, covers "*Economic, Social, and Cultural Consequences of the Aging of the Population*."

Two chapters deal more particularly with the consequences of migration. We saw in Volume III, Section I how closely migration depends on economic factors, but it clearly impacts on the economy in return. In Chapter 83, Manon Domingues-Dos-Santos addresses "The Economic Consequences of Migration" as such. But migration has other consequences, and besides the genetic mixing it encourages, which is discussed in Chapter 79, it also contributes to the mixing of cultures and social strata, be this successful or conflicting. Although it is far beyond the scope of demography to address this subject as a whole, in Chapter 84 Victor Piché discusses the integration of immigrants.

After these chapters on relatively specific consequences of demographic change in particular fields such as health, or particular aspects of demographic change such as aging and migration, Part I is rounded off with three chapters that address changes in population dynamics more broadly. Chapter 85, by Arnaud Dellis and Pierre Pestieau, reviews theories on the relationship between "Economic Growth and Population Growth." Then in Chapter 86, Jean Coussy discusses the relationship between "Population and Development." In the final chapter, Chapter 87, Françoise Bartiaux and Jean-Pascal van Ypersele discuss "The Relationships between Population and Environment."

CHAPTER

79

Population Change and Genetic Diversity

ROBERT CLIQUET

Population and Social Policy Consultants, Brussels, Belgium

INTRODUCTION

This chapter addresses the effects demographic mechanisms and processes may have on the genetic composition and structure of populations. It is an issue that is, in a long term, of fundamental importance, because the basic demographic variables—mating, fertility, migration, and mortality—are the proximate instruments of population genetic changes, eventually leading to biological evolution. In this perspective, demography appears to be of much greater importance than is often thought, even in the demographic community itself. Whereas demographic variables are routinely included in bio-anthropological and population genetic work (see Cavalli-Sforza and Bodmer, 1971; Adams *et al*., 1990; Lasker and Kaplan, 1995), demography only occasionally deals with these interrelationships (see Léridon, 1973; Sheps and Menken, 1973; Bongaarts and Potter, 1983; Keyfitz, 1984). However, in recent years renewed interest in biodemography seems to be emerging (see Wood, 1994; Wachter and Finch, 1997; Rodgers *et al*., 2000, 2002; Wachter and Bulatao, 2003).

All of the basic demographic mechanisms and processes, as well as the population genetic mechanisms and processes, are closely associated and interact in multiple ways (Fig. 79–1). Every basic demographic mechanism corresponds with one or more of the population genetic mechanisms, and most of the demographic end products are largely the result of population genetic processes. In turn, the basic demographic mechanisms are the channels through which the genetic population composition, that is, the gene or allele frequencies, is changed or maintained. Moreover, the population mating structure and the population size can influence the genetic population structure (i.e. the genotype frequencies).

This chapter first provides a brief overview of the effects the basic demographic variables may have on the genetic composition and structure of populations. Then I will discuss the effects the modern demographic transition has or might have on the genetics of populations.

I. DEMOGRAPHIC AND POPULATION GENETIC INTERRELATIONS IN GENERAL

1. The Behavior of Genes in Populations

In sexually reproducing organisms such as the human species, individuals possess a double set of homologous chromosomes—a set from the female parent and a set from the male parent—on which most of the genes are located that determine or influence hereditary features. With the exception of the features determined by the genes located on the different sex chromosomes of the male (X and Y), every hereditary characteristic is genetically determined by at least two genes. In the case of monogenetic inheritance, only one pair of genes influences a particular feature; in the case

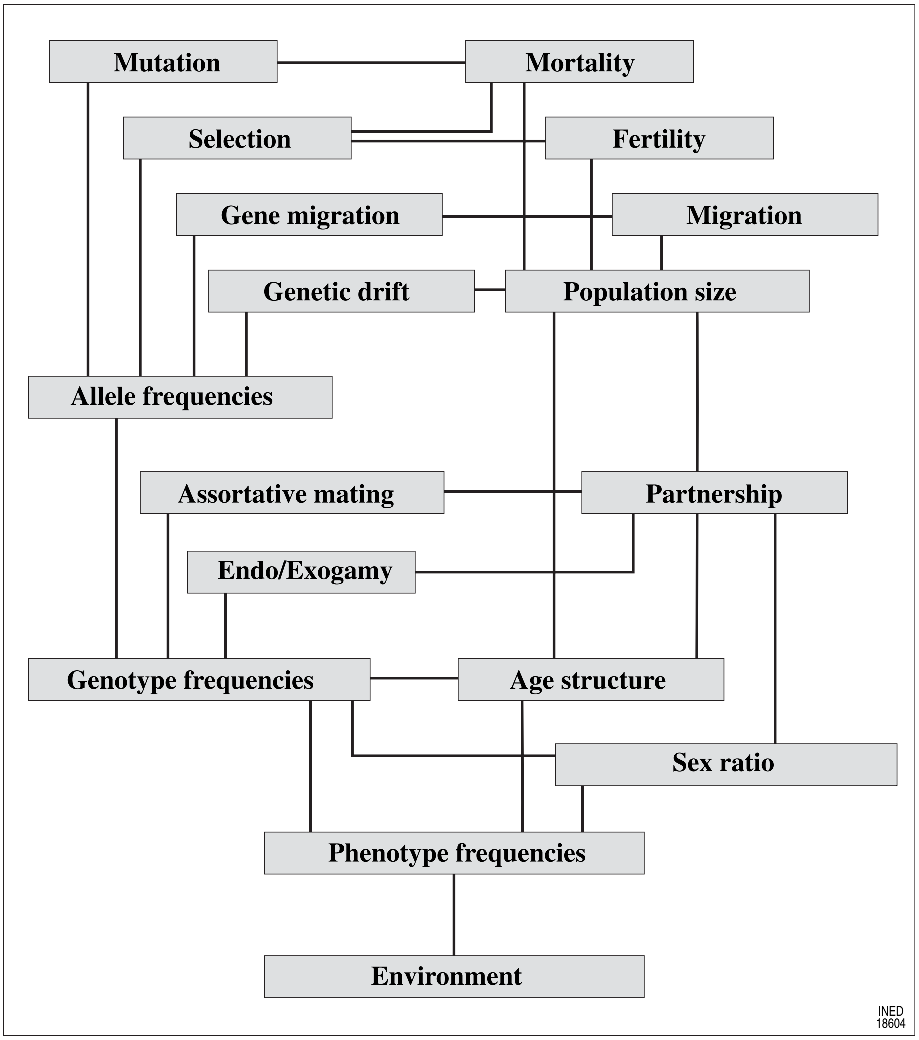


FIGURE 79–1 Demographic and population genetic interrelations. (From Lasker, 1963; Cliquet, 1987.)

of polygenetic inheritance, several pairs of genes influence a particular feature. The corresponding genes on the homologous chromosomes are called alleles. During the process of gamete production, which occurs through a special form of cell division called meiosis, the homologous chromosomes of the germ line cells are distributed over two daughter cells, so that the alleles segregate and the gametes include only one single set of genes. At fertilization, male and female gametes unite, resulting in the reestablishment of a double set of chromosomes with their corresponding alleles for each genetically determined or influenced characteristic. Because a gene for a particular characteristic can take slightly different forms, several allele combinations, named genotypes, can be formed. For a monogenetic feature, determined by two alleles (A and a), three genotypes—AA, Aa, and aa—can be formed. The genotypes with two identical alleles (AA and aa) are called homozygotes, whereas the genotypes with different alleles (e.g., Aa) are called heterozygotes.

The exact genetic composition and structure of a population cannot always be directly deduced from the observation of the phenotypic distribution, that is, from the way the biological features manifest themselves. Some alleles may be dominant, masking the phenotypic effect of recessive alleles. In the case of a dominant allele A and a recessive allele a, the genotypes AA and Aa can not be distinguished phenotypically. Moreover, in many cases (particularly for polygenic characteristics), environmental factors also influence the phenotypic expression and hide or reinforce the genetic effects. With a view of describing correctly the population genetic composition (i.e., the relative proportions of the alleles) and the population genetic structure (i.e., the relative proportions of all of the possible combinations of the alleles in genotypes) for a particular variable in a population in which gene exchange takes place (i.e., in a reproductive community or mendelian population), both the allele frequencies and the genotype frequencies for that variable have to be known. These can be calculated by means of the Hardy–Weinberg law (Hardy, 1908; Weinberg, 1908). In case of a monogenetic variable with two alleles (A and a), respectively, with frequencies p and q, three possible allele combinations or genotypes (AA, Aa, and aa) with the following frequencies can be obtained:

$$(p_A + q_a)^2 = p^2{}_{AA} + 2pq_{Aa} + q^2{}_{aa} = 1$$

The allele frequencies in a population are determined by four evolutionary mechanisms and their interactions: mutation, selection, gene drift, and genetic migration. The genotype frequencies depend on the allele frequencies and the mating system in the population (Wright, 1969; Cavalli-Sforza and Bodmer, 1971; Jacquard, 1974).

Mutation is a change in the chemical structure—the DNA—of a gene or a group of genes. Mutations are at the basis of the genetic variability and, consequently, form the basic condition for possible changes in the genetic composition of the population. Most mutations are deleterious and are eliminated from the gene pool quite quickly by means of selection.

Selection, originally proposed by Darwin (1859) as the major explanatory mechanism for biological evolution, is today usually defined in population genetics as the differential reproduction of carriers of different alleles. Positive selection results in genetic adaptation; negative selection leads to a decrease and eventually the elimination of the targeted genes. At the population level, the selection intensity can be measured by the selection coefficient s, which can vary between zero and one. The degree of genetic change of the population by means of selection can be calculated by introducing the selection values in the Hardy–Weinberg equation. Selection against a recessive allele e.g. changes the Hardy–Weinberg equilibrium to

$$p^2{}_{AA} + 2pq_{Aa} + (1 - s)q^2{}_{aa} = 1 - sq^2{}_{aa}$$

Gene drift, also called Sewall-Wright effect (Wright, 1931), can change the allele frequencies as a result of the accumulation of random fluctuations in the intergenerational transmission of the alleles in small populations. The formation of each new generation is, in fact, a sampling process of the gametes available in the gene pool of the population. The smaller the population, the greater the risk that the allele frequencies in the next generation will deviate from the preceding one. In the absence of the other evolutionary mechanisms, gene drift always leads to the fixation of one allele and the loss of the other(s).

Genetic migration occurs when a genetically differential population section leaves or joins a population. Genetic migration can occur on a small scale, at the level of individual (mate) exchange, or on a large scale, as a massive population invasion. It can occur as a single, non-recurring population move, or as a continuous gene flow between two or more populations. It can be of a unidirectional or a bidirectional nature. It can be merely of a deterministic nature, assuming infinite population size, so that no random elements resulting in drift are included, or it can also be subject to stochastic forces that result in complex interactions between migration and drift (see Piazza, 1990). Genetic migration obviously changes the genetic make-up either of the sending or the receiving population. Genetic immigration leads to an increased heterozygosity and an increased genetic variability within the invaded population, and to a decrease of the between-population variance. It lowers the ratio of the variance between populations to the total (between and within) variance (Wright, 1965). Genetic migration decreases endogamy and increases exogamy (mating within and outside, respectively, the mendelian population). Endogamy is to be distinguished from consanguinity, which refers to the practice of marrying within a geographically or culturally defined population. If the population is small, the risk of consanguinity increases because the number of potential partners is more limited and the chance of meeting a close relative increases. Exogamy is the opposite of endogamy, resulting in an increase of heterozygote genotypes at the expense of homozygotes. The crossing of genetically strongly different populations is an extreme example of exogamy.

Mating can occur at random or not. In the latter case, one can get either a positive or a negative assortative mating. In the first case, the homozygote genotypes (AA and aa) will be favored at the expense of the heterozygotes (Aa) and the population variance increases; in the second case, the opposite occurs. A special case of assortative mating concerns the positive or negative choice of blood relatives. A positive assortative mating for blood relatives leads to inbreeding, whereas a negative results in outbreeding. Inbreeding is a genetic consequence of biologically consanguineous matings, resulting in offspring with a higher than an at random risk to carry a double dose of the genes that were present in a single dose in the common ancestor (Wright, 1921, 1922; Malécot, 1948). Inbreeding changes the Hardy–Weinberg equilibrium to Wright's equilibrium:

$$(p^2 + Fpq)_{AA} + [2pq(1 - F)]_{Aa} + [q^2 + Fpq]_{aa} = 1$$

whereby F represents the inbreeding coefficient which can take values from zero to one. The proportion of the homozygous genotypes (AA and aa) increases and the proportion of the heterozygote genotypes (Aa) decreases according to the size of the inbreeding coefficient F. The average consanguinity in a population is usually measured by the average inbreeding coefficient of its members

$$a = \sum p_i F_i$$

whereby p_i is the relative frequency of individuals with inbreeding coefficient F_i (Wright, 1922). Because many genetic diseases are controlled by recessive genes, inbreeding increases the appearance of such diseases through the formation of homozygote recessive genotypes. The major difference between positive assortative mating and inbreeding is that the first is character specific; the second involves all of the genetically determined characteristics.

The increased genetic risks of consanguineous matings not only are based on population genetic theory but also have been extensively confirmed by empirical studies that systematically show inbreeding to be associated by substantially increased levels of pre- and postnatal mortality and morbidity. These results have been reviewed many times (see Cavalli-Sforza and Bodmer, 1971; Freire-Maia and Elisbão, 1984; Khlat and Khoury, 1991; Bittles, 1994). The inbreeding depression has been shown for major, recessively inherited malformations, for intelligence, and for other biometrical characteristics. For the latter, however, socioeconomic factors may partially be involved because inbreeding often prevails more among the less well to do (Bittles, 1994). Data on fecundity and fertility seem to be inconclusive for the moment. Some studies show reduced levels of primary sterility in consanguineous marriages, but others point to lower fecundability. Fertility levels often are higher among consanguineous couples, but this could be owing to demographic factors, such as younger age at marriage related to cultural traditions or as a reproductive compensatory behavior for increased postnatal mortality (Bittles, 1994). Under high levels of inbreeding, unfavorable recessive alleles become visible and may be subject to negative selection, thus lowering the equilibrium allele frequency for lethal genes (see Khoury *et al*., 1987). The lower incidence of some of those diseases might, however, also be owing to the higher mortality rates in isolates and to the depletion of lethal alleles through inbreeding (Neel, 1992).

Notwithstanding the risks of inbreeding depression, many societies favor modest forms of inbreeding, taking advantage of the socioeconomic profits associated to the maintenance or the increase of the family property (Van den Berghe, 1980). At the same time, the salient negative effects of close inbreeding involving relatives of the first degree (father–daughter, brother–sister relations) are avoided, the latter being the result of Darwinian selected inbreeding avoidance behavior and incest taboo (Westermarck, 1922; Wolf, 1995).

2. The Effects of Demographic Behavior on Genetic Diversity

a. Mating Behavior

Whereas in the past demographers limited themselves often to the study of nuptiality, it has become quite evident that this narrow approach is inappropriate if one wants to comprehensively analyze the demographic effects of relational behavior. A broader perspective is, in any case, indispensable for the study of the genetic effects of partnership. Mating behavior is one of the essential components of the reproductive process. Hence, it can influence the genetic outcome of that process. Broadly, mating behavior can influence the genetic diversity of a population in two respects: it can influence the genetic population composition (i.e., the allele frequencies) and the genetic population structure (i.e., the genotype frequencies).

The first issue to be dealt with is mate selection, the question of whether one chooses a mate or not. From a genetic point of view, mate selection obviously is only relevant when the people who do not choose a mate have genetically distinct features from those who do. In that case (sexual) celibacy constitutes one of the mechanisms by which genes are eliminated from the gene pool. It is a selective mechanism that is often lost out of sight. More refined studies on differential reproduction (see Higgins *et al*., 1962; Bajema, 1963; Waller, 1971; Retherford and Sewell, 1988) have amply shown the methodological and empirical importance of considering mating behavior in a comprehensive way and of not limiting it to nuptiality.

In the second place, mate assortment has to be considered. Assortative mating influences the genotype frequencies: positive assortative mating favors the production of homozygotes and increases the population variance; negative assortative mating results in the increase of heterozygotes and lowers the population variance (see Crow and Felsenstein, 1968). Whereas negative assortative mating is only substantially found for sexual orientation, positive assortative mating has been observed for many particularly polygenetic characteristics: intelligence, body height and weight, constitution type, complexion, physiognomic details of the face, and even sociopathological features such as criminal behavior, alcoholism, and psychiatric disorders (see Suzanne, 1967; Spuhler, 1968; Roberts, 1977; Epstein and Guttman, 1984; Rushton and Nicholson, 1988). More importantly, positive assortative mating is associated with higher relational stability, fecundity, and fertility (see Thiessen and Gregg, 1980; Epstein and Guttman, 1984). The latter findings imply that assortative mating not only influences the genotype frequencies but also, indirectly, can influence the genetic composition of the population by facilitating selection through the differential fertility of extreme genotypes. Assortative mating for polygenetically determined features obviously occurs via the phenotypes. The genetic effect of this phenotypic assortment depends on the heritability of the traits, that is, on the degree to which the population variability of these traits is determined by genetic factors (Mather and Jinks, 1982). Positive assortative mating for blood relationship results in inbreeding, which increases the risk of all of the gene loci to be brought into homozygous condition. Consanguinity, however, increases the risk of inbreeding depression, which is characterized by an increased morbidity and mortality.

b. *Fertility*

Fertility behavior includes two major aspects: timing and intensity of reproduction.

The timing of births is not independent of its genetic outcome. The frequency of a number of genetic impairments is positively correlated to the age of the parents, either of the mother or of the father, owing to the increase of the mutation rate with parental age: genetic conditions, such as Down syndrome and other trisomies, and syndromes owing to nondisjunction of the X chromosome increase with mother's age; conditions such as achondroplasia, Apert syndrome, Marfan syndrome, and myositis ossificans increase with father's age (Vogel and Motulsky, 1982). Some impairments, such as rhesus incompatibility, are related to birth order. Whereas the above-mentioned associations have only a generational or ontogenetic effect, the timing of births can also have an intergenerational genetic effect by increasing or decreasing the generation length of the reproductive period. When genetically different sections of the population are characterized by differences in average generation length through a variation in average birth interval, this can speed up or slow down the intergenerational transmission of genes. Genetic differentials in generation length form one of the mechanisms that result in selective processes (see Cole, 1954; Bajema, 1963).

Differences in the number of children related to genetic diversity obviously are of greater importance. The differential fertility of carriers of different alleles, indeed, forms one of the most powerful mechanisms producing selection. It influences, proximately, the reproductive fitness and, ultimately, evolutionary adaptation (Fisher, 1930).

c. *Mortality*

Mortality is the other major demographic tool of selection. In each generation, a broad genetic variation appears. Genetically differential mortality considerably reduces this initial variation by eliminating genotypes that are not able to meet the challenges of the environment. The largest part of this elimination process occurs prenatally, particularly during the very early stages of embryonic life (see Hertig, 1967; Wilcox *et al*., 1988; Macklon *et al*., 2002). Karyotypological research shows that intrauterine mortality largely consists of a genetic elimination process (see Boué *et al*., 1975; Creasey *et al*., 1976; Clark, 2003; Rubio *et al*., 2003).

In preindustrial living circumstances, mortality continued this elimination process virtually during the whole life cycle, resulting in a survivorship curve that approached closely the constant death-rate logarithmic curve (Comfort, 1979). Although it is probably impossible to document the genetic component of this earlier elimination process properly, it is difficult to imagine that the earlier mortality pattern would not have considerably decreased the genetic diversity, leaving the more fit as survivors up to high age. The genetically differential effect of mortality must have shown up in the population age structure, in which the older age groups may be supposed to have been genetically less variable.

Senescence is the time-dependent accumulation of genetic damage that progresses during the life course, resulting in increasing levels of frailty, impaired function, disease, and ultimately death (see Carnes and Olshansky, 1993). According to some scholars, senescence evolved through the accumulation of deleterious mutations that act only in later life, after the reproductive phase of life or the period of life during which the survival of grandchildren can be influenced, that is, when the force of selection decreases or disappears (Medawar, 1952). Others are of the view that senescence is the result of antagonistic pleiotropy; that is, genes with favorable effects during the early years of life would have unfavorable side effects later in life (Williams, 1957). Whatever mechanism may be at the basis of senescence, genes with late deleterious consequences can only accumulate and show up among individuals who live long enough (Kirkwood and Holliday, 1986).

d. Migration

Demographic migration is only of relevance here if it is associated to genetically differential population movements. Obviously, this is often the case. For example, long-distance migrants usually come from populations that are genetically different from the populations in the immigrant regions (Mascie-Taylor and Lasker, 1988). Although the genetic differences between populations constitute only a small fraction of the total genetic variation in humans (see Lewontin, 1972; Rosenberg *et al*., 2002), they remain important biological markers of group identity and continue to play a crucial role in in-group/out-group relations (Reynolds *et al*., 1987; Thienpont and Cliquet, 1999).

Migrants may differ not only from the populations they join but also from the populations they leave. In the latter case, selective emigration is seen (see Shapiro, 1939; Martin, 1949; Illsley *et al*., 1963; Mascie Taylor, 1984; Steegman, 1985). Migrants may differ from the sedentary population in fecundity, morbidity, longevity, intelligence, energetic drive, etc. (Termotte, 1984). Often, selective migration movements appear to be cases of social assortment, effecting the genetic composition of both the population of destination and the population of origin (Bogin, 1988).

In principle, the effects of genetically differential migration are easy to understand: they change the allele frequencies and genotype frequencies of the populations involved in the directions and the degrees of the migration movements. However, the specific biological characteristics of migrants should be carefully assessed: they may, indeed, be the result of a genetic assortment, but they may also be caused by an assortment of features that are the result of environmental influences (e.g., nutrition, education, infection) or even of a phenotypic adaptation to their new environment (Bogin, 1988).

Migration may change the gene frequencies in two different ways. It may lead to the splitting up of a population in several subpopulations that further evolve in isolation, and it may lead to the fusion of several populations, resulting in hybridization. The splitting up of populations leads to genetic isolation by distance, that is the decrease of genetic exchange between populations with increasing geographical (or social) distance, resulting in decreasing (coefficients of) kinship and genetic similarity (see Wright, 1943). Geographic isolation (lack of gene migration) is an important factor that determines the extent of the genetic distance, that is, the degree of genetic differentiation between populations (see Constandse-Westermann, 1972). The less migration there is, the more relatedness will build up within groups. According to Hamilton (1975), the number of migrants, rather than the size of the colonies, determines the degree of intrarelatedness. Fusion of genetically different populations can occur with different intensity: it may occur on a massive scale in a short period of time, but it may also be achieved by a slow but continuous gene flow, spread over a very long period of time. It may involve neighboring, closely related populations or geographically and genetically more distant populations. Most migration, however, occurs over a short range, among neighboring populations who are often genetically similar. Demographic and population genetic research on marriage distance in traditional societies has shown that the reproductive communities on average only slightly transcended the village or parish (see Cavalli-Sforza *et al*., 1964). The effects of emigration may, moreover, be compensated or counterbalanced by return migration (see Relethford and Lees, 1983).

Almost all known experiences of contacts between human populations, even in agonistic situations, lead to bridal exchanges and/or gene flow. This is the reason why isolation by distance explains most of the genetic variation in large populations settled for a long time and on a large scale (see Morton *et al*., 1972). It is also one of the reasons why interpopulation variation is characterized by genetic gradients or genoclines: these are neighboring populations isolated by distance, forming small overlapping mendelian populations and approaching a stepping-stone model with continuous variation of genetic traits between adjacent groups of people (see Reddy, 1984).

Migration movements may involve much more complex population genetic situations. Genetically differential immigration may lead to the establishment of new, hybrid populations in which the immigrants are genetically integrated. However, in many cases, the immigrants remain in endogamous communities, reproductively isolated from the main, autochthonous populations. In such cases, the geographical population may be composed of a series of endogamous reproductive communities, which each one of them sharing a different gene pool. The African-American population is an interesting case in this respect, because it combines features of both migration processes: partial genetic isolation and partial hybridization. Although the social prevalence of formal marriages between blacks and whites in the United States is extremely rare, the European contribution to the gene pool of the African-American population has been estimated to accumulate to 30% on average, with a strong regional variation going from

10% to 50% (see Reed, 1969; Workman, 1973). The gene flow from the European to the hybrid African-American population, over a period of ±12 generations, has been calculated to be on average 2.5% to 3.5% per generation (Glass and Li, 1953; Roberts, 1955). A continuous gene flow, which is not counteracted by other population genetic processes, inevitably results in population genetic replacement and overall homogenization (see Hiorns *et al.*, 1969; Mascie Taylor and Lasker, 1988). The degree of substitution only depends on the rate of gene transfer. If the European gene transfer to the African-American population continues at its past rate, the genetic identity of the African-Americans will disappear in a few millennia (see Benoist, 1986; Cavalli-Sforza *et al.*, 1994). In their mathematical analysis of the effects of migration on the relatedness between populations, Hiorns *et al.* (1969) obtained the surprising result that, even with modest degrees of exogamy, it takes only a limited number of generations to produce genetic homogeneity in a defined geographical space. The present increasing trends in migratory movements and exogamic behavior definitely stopped the earlier fission in favor of an increasing fusion movement.

The genetic effects of migration are not necessarily limited to population genetic fission or fusion. It may also be related to or induce other demographic processes, resulting in genetic changes. For example, migration may be associated to differential fertility. Immigrants with higher fertility levels will speed up the change of the genetic composition of the population. A typical example is Brussels, where Moroccan and Turkish immigrants have a cohort fertility of more than 4 children, and the autochthone inhabitants and Flemish and Walloon immigrants have, on average, 1.5 children (Schoenmaeckers *et al.*, 1997). Migration may also be associated to genetically differential mortality and hence to selection processes. Migrants may arrive in environments to which they are less well adapted or in which they are freed from their former selective pressures, as has been clearly shown for sickle-cell anemia (see Devoucoux *et al.*, 1991). Migrants may increase the mortality of the autochthones, either by introducing diseases to which the natives are not resistant or by genocide (see Vogel, 1992). Both these migration-induced population genetic effects have been observed in the partial, and in some cases even total, population genetic replacements of the Indian-American populations in North America and the Caribbean by populations of European and African origin during the colonization of the Americas.

Immigration may also induce emigration and strengthen local or regional population genetic substitution processes. The residential segregation of genetically distinguishable ethnic minorities in many European and American metropolitans is a well-known phenomenon (see de Lannoy, 1978; Van der Haeghen, 1995). Last but not least, migration may lead to population fragmentation, resulting in small isolates that then become sensitive to phenomena such as the Sewell-Wright-effect, the founder effect, and inbreeding.

e. Population Size

The demographic size of a reproductive community may have different, although partially interrelated, effects on the genetic diversity of the population. All of them, however, manifest themselves when a reproductive community is small and when there is little or no immigration. The combination of endogamy and a small population size results in the production of isolates, that is, relatively homogeneous populations separated from related groups by geographical or social factors (see Wahlund, 1928; Dahlberg, 1929; Jacquard, 1976).

Usually two major types of isolates are distinguished, the genetic effects of which may be different: (1) primary isolates, tribal populations of ancestral origin closely approaching genetic equilibrium and often characterized by the presence of private polymorphisms (i.e., alleles restricted to a single tribe), and (2) secondary isolates, offshoot populations detached from a large, national population itself usually being an amalgamation of earlier smaller groups. Secondary isolates are usually not (yet) in genetic equilibrium (Neel, 1992).

In the first place, a small population size triggers random genetic drift. Population genetics has knowledge of many and different types of small reproductive communities—isolated island or mountain populations, religious, ethnic, or other social isolates—with (monogenic) allele frequencies that fluctuated or diversified owing to the sampling processes linked to the formation of new generations by a small number of individuals (see Cavalli-Sforza and Bodmer, 1971; Roberts *et al.*, 1992). Hence, the demographic size of a population can influence its genetic composition.

A special case emerges when a population goes through a demographic bottleneck, as a result of emigration, high mortality, low fertility, age, or sex imbalance (see Leslie, 1990). It then becomes susceptible to the founder effect: particular genes, carried sometimes by one single individual, get easily spread over the small population. The presence of bottlenecks leads to a rapid increase of the genetic distance and a reduction in heterozygosity. It produces long branches in phylogenetic trees (Livshits and Nei, 1990).

In combination with the absence of immigration, the small size of a reproductive community inevitably also leads to inbreeding. When the number of potential mates decreases, the chance of meeting a (close) relative increases. As was already explained, consanguinity promotes the formation of homozygote genotypes for all genes and consequently leads to inbreeding depression. Population genetics has documented this phenomenon with amazing examples of small isolates in which many and rare genetic diseases showed up with high frequencies (see Pollitzer *et al*., 1964; Witkop *et al*., 1966; Roberts, 1968). Not only do recessive genes show up in homozygous combination, but also unfavorable dominant alleles can spread easily owing to the close web of kinship (Roberts, 1992).

Population size may also affect the genetic population composition when it becomes too large, particularly with respect to its available resources. Population pressures may then increase the within- and between-population competition and induce selective processes, or it may lead to selective outmigration, for example, in the form of brain drain (UNESCO, 1984). However, the immediate cultural effects of this phenomenon may prove to be of more importance than are the population genetic effects, which may appear only on a longer-term basis thanks to the greater genetic stability of polygenetically influenced characteristics.

II. THE EFFECTS OF THE MODERN DEMOGRAPHIC TRANSITION ON GENETIC DIVERSITY

The human species has experienced several demographic transitions (Muhsam, 1979). Here, only the recent transition is considered, associated to industrialization and other aspects of modernization demographically, the transition is mainly characterized by a substantial acquisition of control over mortality and fertility (Landry, 1934; Kirk, 1996). Although mortality and fertility control are undoubtedly the most important characteristics of the modern demographic transition, it must be acknowledged that virtually all demographic variables are involved in this transition. For the issues at stake in this chapter, the discussion has to deal with mating behavior, fertility, migration, mortality, and population size.

1. Mating Behavior

Modernization is characterized by several societal changes that may influence mating behavior: there is, for example, a shift from a socially controlled toward a more personally determined way of life; the population size and the communication and travel opportunities increased; or scientific knowledge became available about the detrimental genetic effects of inbreeding. All of those changes result in the presence of larger pools of possible mates. Notwithstanding the existing tendencies with respect to in-group behavior, it can be expected that the new opportunities for choosing mates will be taken advantage of. This is what is being observed.

In the first place, the traditional small marriage distance or mate pools that were largely limited to the town and to neighboring villages or towns, increased considerably. On average, people in modern society choose their partner(s) from a much larger geographical or social area (see Walter, 1956; Boyce *et al*., 1968). This broadening of the marriage or mate distance has two distinct population genetic effects. First, isolates are broken up, reducing inbreeding and diminishing the inbreeding depression. Second, exogamic matings increase, leading to interpopulation mixtures and, in some cases, to crossing between genetically strongly distinct populations (see Lionetti and Newell-Morris, 1982), thus resulting in a higher population heterogeneity.

In the second place, assortative mating on the basis of preferred mate characteristics may increase. This increase is difficult to document, because no historical data are available that can be compared with more recent data. The latter, however, clearly show that assortative mating is a common behavioral pattern in modern society (Fig. 79–2). Assortative mating results in increased homozygosity and larger population genetic variance, obviously to the extent that the characteristics involved in mate choice are (partly) genetically determined.

A particular question relates to mate selection. In the past, at least in Western societies, celibacy was not an uncommon phenomenon, either for religious or for economic reasons. In modern culture, religious as well as economic celibacy seems to be regressing. However, the recent decrease in nuptiality may not be considered as a new increase in celibacy. Marriage has simply been replaced by other forms of unions (cohabitation; living apart together [LAT] relations; loose relations) (Villeneuve-Gokalp, 1997). It is difficult to evaluate the genetic effects of changes in mate selection. The earlier forms of celibacy probably had very diverse genetic effects (de Lapouge, 1896). The current strong reduction of those who obey the call in Catholic regions, and who were formerly undoubtedly belonging to the more intellectually gifted individuals of the population, might be considered to have a eugenic effect. But what about the other changes in mate selection? On the whole, one would expect that the decrease of celibacy

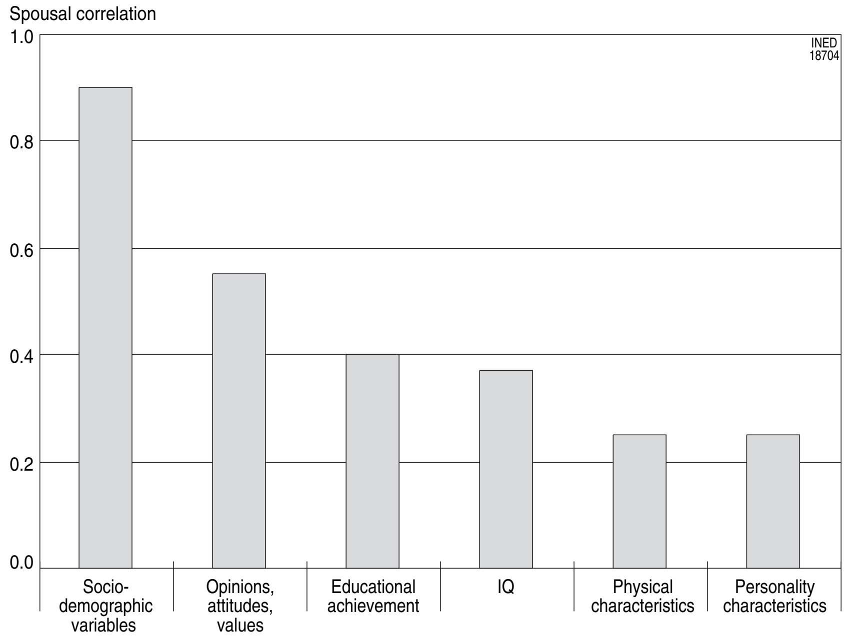


FIGURE 79–2 Spousal resemblance on a variety of characteristics. (From Susanne, 1967; Spuhler, 1968; Roberts, 1977; Thiessen and Gregg, 1980; Mascie-Taylor, 1988; Rushton and Nicholson, 1988.)

might induce a decrease in selection, resulting in an increase in population genetic heterogeneity.

2. Fertility

The modern demographic transition is characterized by profound changes in both the timing and intensity of fertility. As far as the timing is concerned, births are no longer spread over the fecund period of life. With respect to fertility intensity, both the average and the variance decreased. Last but not least, the transition from a high to a low fertility regime has, in many cases, been associated with strong sociobiological differentials in reproductive behavior.

In demographically post-transitional societies, births are avoided at younger ages, in particular during adolescence, and at higher ages, especially after age 35. The latter phenomenon is generally considered to have a eugenic effect, because the prevalence of several genetic impairments increases with parental age and/or birth order and, hence, can be avoided by controlling fertility at higher ages or parity (see Matsunaga, 1966). In recent decades, however, the postponement of births is pushed further up in the life course. In some countries such as The Netherlands, the average age of a woman at her first birth is close to 30 (see Beets *et al.*, 1994), which means that most women will have their children in their 30s instead of their 20s. In addition to increased risks of fecundity problems (primary and secondary sterility start to increase substantially after age 30), relatively older parents also risk higher frequencies of particular genetic disorders. Fortunately, more genetic impairments can be detected prenatally and eliminated by selective abortion, if so desired.

The availability of selective abortion may have two distinct effects on the genetic composition of the population. First, it allows families who have a substantial risk to get genetically impaired children to avoid the birth of seriously handicapped offspring and to replace them with phenotypically normal children. Selective abortion not only avoids genetic impairments but also avoids infertility: it allows couples to build up a normal family with healthy children (see Evers-Kiebooms, 1994). However, the reproductive compensation resulting from the replacement of defective offspring by normal children increases the relative frequency of carriers of the genetic condition in heterozygote individuals, who may transmit the defective allele to future generations and contribute to the increase of the allele frequency in the population. However, it may be expected that more refined future genetic screening techniques will be able to prevent such an increase.

Fertility regulation, resulting in a lower parity combined to an earlier timing, may have other, but perhaps less important, genetic effects. In fact, all genetic phenomena that are differentially related to maternal or paternal age may be supposed to be influenced by fertility regulation. Walter Fuhrmann (1969) has argued that family planning may lead to an initial decrease and a subsequent increase of dizygotic twins. Because the prevalence of dizygotic twins increases with maternal age, the avoidance of births at higher ages must lead to a decrease. However, family planning concerns pregnancies, not births. In the case of twins, a single pregnancy results in two births. In a regime of fertility limitation, the relative share of twin children will become more numerous on the totality of children born, and hence the genes involved will increase and counterbalance the effect of the decreasing maternal age. Fertility limitation at higher ages may also slightly influence the sex ratio because the birth of males decreases with maternal age, but this phenomenon cannot be supposed to have intergenerational effects. Smaller families result in lower numbers and proportions of close relatives and, consequently, may decrease the risk of the formation of consanguineous unions (see Chandrasekar *et al*., 1993).

The modern decreases in fertility average and variance have to be considered together, because the opportunity of selection depends, according to Crow's index of total selection, on the ratio of the variance (V_w) to the square of the mean (w^2) of the number of first generation descendants per individual (Crow, 1958):

$$I = V_w / w^2$$

This index can be separated into two components, I_m and I_f, due to differential mortality and fertility respectively:

$$I = I_m + 1/p_s I_f$$

$$I_m = p_d / p_s$$

$$I_f = V_f / x^2_s$$

where p_d is the proportion of premature deaths, and p_s is the proportion of survivors, with x_s as mean and V_f as variance of the number of births per surviving parent.

Whereas the mortality decline of the modern demographic transition resulted in a decrease of I_m, the fertility decline, contrary to what might intuitively be expected, does not necessarily induce lower opportunities for selection (see Matsunaga, 1966; Spuhler, 1976; Hed, 1987). On the contrary, when the average fertility decreases more strongly than does its variance, the opportunity for selection may even increase. Data from different countries seem to suggest that, at least in the initial stages of the demographic transition, the I_f increased slightly, whereas in later stages it decreased (Fig. 79–3).

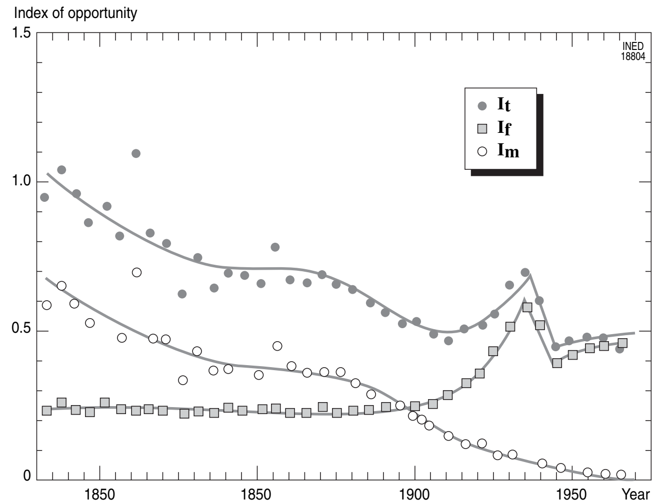


FIGURE 79–3 The changes of the fertility index of opportunity of selection (I_f) in the course of the demographic transition. (From Adams and Smouse, 1985.)

The degree to which selection actually occurs depends on the genetically differential fertility observed. The modern fertility decline in most early transitional societies occurred in a socially differentiated way: birth control first spread among the higher educational and socioeconomic strata, initially resulting in outspoken negative associations between fertility and a variety of indicators of social stratification (see Wrong, 1958; Cochrane, 1979).

Given the substantial positive correlation between educational level or other indicators of socioeconomic status (SES) and intelligence (see Jensen, 1981; White, 1982; Lynn, 1996), the positive correlation between intelligence of parents and children, the abovementioned negative correlation between social status and fertility, as well as direct observations on the negative correlation between intelligence and fertility (see Cattell, 1937, 1950; Vining, 1986) induced the theory of the contraselective effect of modern culture on intelligence, predicting that the intelligence level in industrial societies would start to fall (see Cattell, 1937; Graham, 1970). Contrary to this prediction, however, large-scale surveys on the intelligence level in subsequent generations initially did not show a decrease, but no change, or even an increase (see Cattell, 1950; Flynn, 1984; Lynn and Hampson, 1987). Nevertheless, on the basis of large-scale surveys and data on differential fertility and other demographic mechanisms such as mate selection, differential mortality, and generation length with a view of assessing the net effects of demographic differentials on the intergenerational changes in intelligence levels and interpreting them in the light of findings on the heritability of cognitive

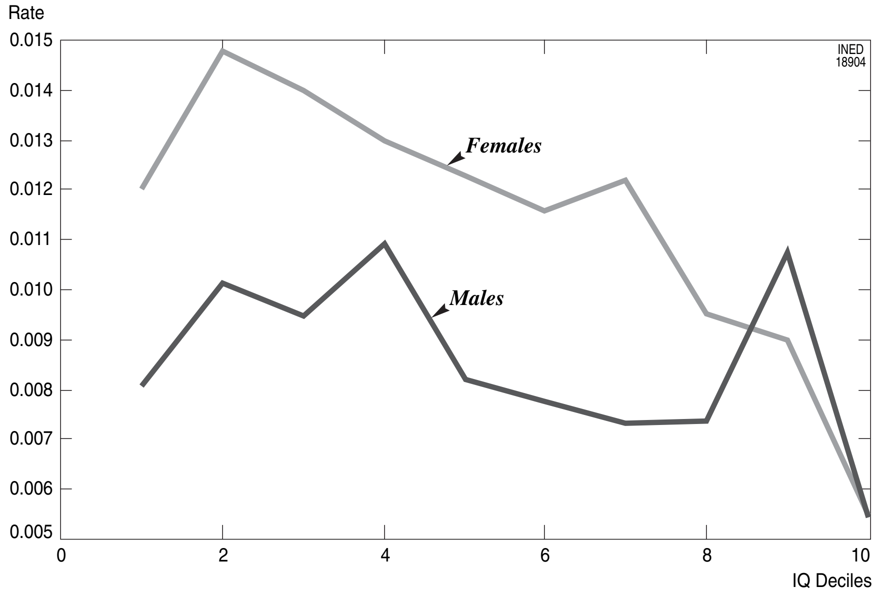


FIGURE 79–4 The intrinsic rate of natural increase *r* by intelligence quotient deciles and sex, derived from the Wisconsin Longitudinal Study. (From Retherford and Sewell, 1988.)

ability,[1] several scholars concluded that the demographic transition had a slight dysgenic effect (see Vining, 1986; Retherford and Sewell, 1988; Loehlin, 1997; Lynn, 1999). For complex features such as cognitive ability that are determined by many monogenetic and polygenetic gene pairs and can, moreover, be influenced by environmental factors, the intergenerational genetic effect of differential reproduction is always smaller than the fertility differentials suggest (see Carter, 1966; Cliquet and Delmotte, 1984). Last, but not least, many industrial societies still experience negative relations between fertility and a variety of SES indicators, but the differentials have decreased substantially because birth control also became a common practice among the lower educational and social-status strata (see Kirk, 1969; Vining, 1986; Perusse, 1993). In some countries the earlier negative association between fertility and SES indicators show the first signs of a possible reversal, confirming Osborn's (1952, 1968) eugenics hypothesis, stating that in a situation of free choice the distribution of births will favor more children being born in the most favorable home environments. This implies that the end of the demographic transition might be characterized by a positive association between reproductive fitness and socially valuable traits.

Although the modern demographic transition is in the field of reproductive behavior mainly characterized by a decrease in fertility, modernization has also increased, in absolute or in relative terms, the fertility of particular population categories. Modern health and welfare care, particularly therapeutic replacement,[2] has not only postponed mortality but also relieved morbidity, in some cases allowing individuals or couples, whose reproductive fitness was strongly reduced in previous times, to get (more) children. In cases in which genetic factors are involved in the etiology of diseases, health care and therapeutic replacement result in selection relaxation. Several mental or physical conditions are known for which medical treatment has a positive effect on reproductive fitness.

[1] Although quantitative genetics (see Mather and Jinks, 1982), and in particular behavior genetics (see Hay, 1985; Plomin *et al.*, 1990), are well-established disciplines, some of their concepts and fields of application such as heritability of cognitive ability continue, partly because of ideological interferences (see Kamin, 1974; Herrnstein and Murray, 1994) and partly because of methodological issues, to raise scientific controversy (see Capron *et al.*, 1999; Hay, 1999; Vogel, 1999; Vetta and Courgeau, 2003). Ronald Fisher (1951) considered the heritability coefficient as one of those unfortunate shortcuts that appear often in biometrics owing to a lack of thorough analysis of the data. Albert Jacquard (1983) has shown that this term covers three different concepts and argued that none of those parameters proposed by geneticists can be used for resolving the questions linked to the transmission of behavioural characteristics such as intelligence quotient. The progress in molecular genetics can be expected to throw more light on the transmission of quantitative trait loci (QTL), just as it is doing for monogenetic traits (see Plomin *et al.*, 1995; Plomin, 1997; Chorney *et al.*, 1998; Mackay, 2001), even if some authors believe that it is really impossible to distinguish the effects of heredity and environment (Gottlieb, 2001).

[2] Medical intervention allowing to supplement a deficient function or to replace a deteriorated organ.

Schizophrenic patients traditionally had and still have a lower reproductive fitness than do other people, owing to lower marriage rates, higher divorce rates, and lower fertility rates (Erlenmeyer-Kimling *et al.*, 1969; Ritsner *et al.*, 1991). In recent decades, however, thanks to medical treatment that allows deinstitutionalization, the fertility rate of schizophrenic patients shows a slight upward trend (Bodmer, 1968; Ödegaard, 1980; Erlenmeyer-Kimling *et al.*, 1969). The same phenomenon has been observed for individuals with diabetes (Aschner and Post, 1956/57). Adams and Smouse (1985) have illustrated the effect of selection relaxation during the demographic transition by calculating the selection coefficients for three hypothetical diseases as examples of traits with different age of onset patterns. The selection intensity decreased, mainly for the trait with an early age of onset, owing to the concomitant effects of mortality decline and a decrease in generation time as a result of a greater proportion of child-bearing being accomplished at earlier ages. Selection relaxation may also act against subfecundity. More and more medical interventions are developed aimed at treating subfecundity conditions (De Jonge and Barratt, 2002; Cetin *et al.*, 2003). Moreover, fertility limitation profoundly changes the parity frequency distribution in favor of the lower parity numbers (Fig. 79–5). The combined effects of replacement therapies and the changed parity distribution might, in the long run, also decrease the fecundity or at least increase the proportion of people facing subfecundity problems (see Graham, 1972; Medawar, 1974).

In recent years an increasing interest can also be observed on the effects of new developments in genetics and evolutionary anthropology on the understanding of human reproductive behavior in modern society with its strong changes in family planning and fertility behavior (Rodgers *et al.*, 2000, 2002; Wachter and Bulatao, 2003).

a. Mortality

One of the major components of the demographic transition is the epidemiologic transition. Mainly because of the control of infectious diseases, the major causes of death in industrial countries have shifted to chronic diseases of late onset (Weiss, 1990). These diseases are much less subject to natural selection, because the force of natural selection decreases with age, more particularly after the reproductive period of life, leading to an accumulation of unfavorable mutations producing senescence and ultimately death (Finch, 1990; Rose, 1990). The control of infectious diseases allows people to reach higher ages, at which the accumulated effects of deleterious mutants can manifest themselves. In addition, the protective nature of modern culture, through improved living circumstances of all sorts and medical replacement therapies, allows more frail individuals to survive and even to reproduce (see Cook, 1956–1957; Vogel, 1992). The modern mortality control is decreasing substantially the mortality component of Crow's (1958) index of opportunity for selection ($I_m = p_d/p_s$) (Fig. 79–3). Modern culture has, in this way, increased the adaptability of persons who would not have lived long and whose reproductive fitness would seriously have been reduced in preindustrial living circumstances. In modern culture, these individuals are better adapted to survive, a nice illustration of the relativity of the normal or abnormal nature of genes and genotypes that changes with medical and social progress. A gene that was, in premodern circumstances, experienced as abnormal, is no longer considered deleterious if the carrier can now have a normal life. However, it also means that such persons depend more on the preservation and further progress of a technologically highly developed culture and society.

The modern mortality transition is rectangularizing the human survival curve and is profoundly changing the quantitative age composition of the population. However, this curve-squaring strategy (Gordon *et al.*, 1979) also influences the quality of life because the gain in years is partially associated to an increase in largely genetically determined chronic diseases (see Verbrugge, 1984; Riley, 1990; Robine *et al.*, 2003) (Fig. 79–6). Future efforts to further rectangularize the survival curve might even worsen this trend. The ideal of a full rectangular survival curve might, however, not be completely reachable owing to the genetic heterogeneity of the human species and its species-specific maximum life span (Weiss, 1990). On the other hand, future progress in molecular genetics might break through the maximum life span and extend the human-specific longevity (Gordon *et al.*, 1979).

Biodemographers have tried to model the age-specific hazard function of frailty (see Manton and Stallard, 1994; Vaupel, 1990; Weiss, 1990). This is, however, extremely difficult, because the life course history of many diseases is not well known. Moreover, both genetic and environmental factors influence many diseases and may be responsible for phenotypic heterogeneity. Last, but not least, the shape of the hazard function is not necessarily uniform for all variables involved, as Weiss (1990) has recently shown for various forms of cancer. A substantial, if not the largest part of the frailty hazard is of a (partial) genetic nature. In modern culture, the genetic component of the frailty hazard is increasing, in relative as well as in absolute

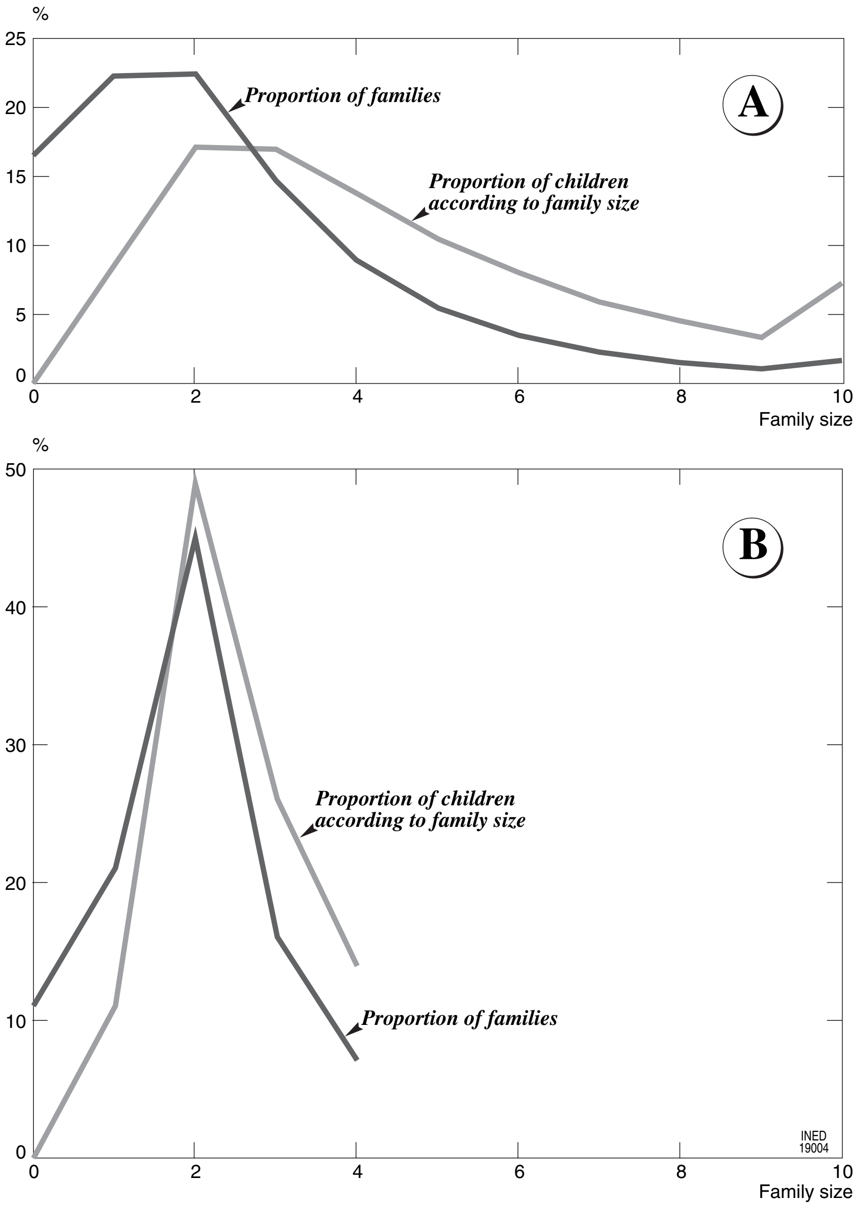


FIGURE 79–5 Family size and offspring number distribution: (A) in a demographically pretransitional regime (France, birth cohort 1881); (B) in a demographically post-transitional regime (Flanders, birth cohorts 1950–1954). (From Vincent, 1946; Callens and Schoenmaeckers, 1993.)

terms. The general protective effect of modern culture and the spreading of replacement therapies results in a selection relaxation (Cavalli-Sforza and Bodmer, 1971; Crow, 1966). The epidemiological shift and the selection relaxation both result in an increase of the genetic heterogeneity of the population. This increasing genetic diversity induces a larger phenotypic vulnerability, which requires a further enhanced medical and welfare care (see Manton *et al*., 1990).

The selection relaxation of mortality control, just as the one of morbidity control, may be reinforced by an increased reproductive fitness. Surviving individuals may find (a) partner(s) and get children. This has already been shown for individuals with several diseases, such as diabetes (Aschner and Post, 1956–1957) and schizophrenia (Erlenmeyer-Kimling and Paradowski, 1966; Ritsner *et al*., 1991), for which replacement therapies or other types of medication have been developed. The reproductive fitness of such patients undoubtedly has been enhanced, so that an increase in the frequency of the alleles responsible for those conditions may be expected to increase.

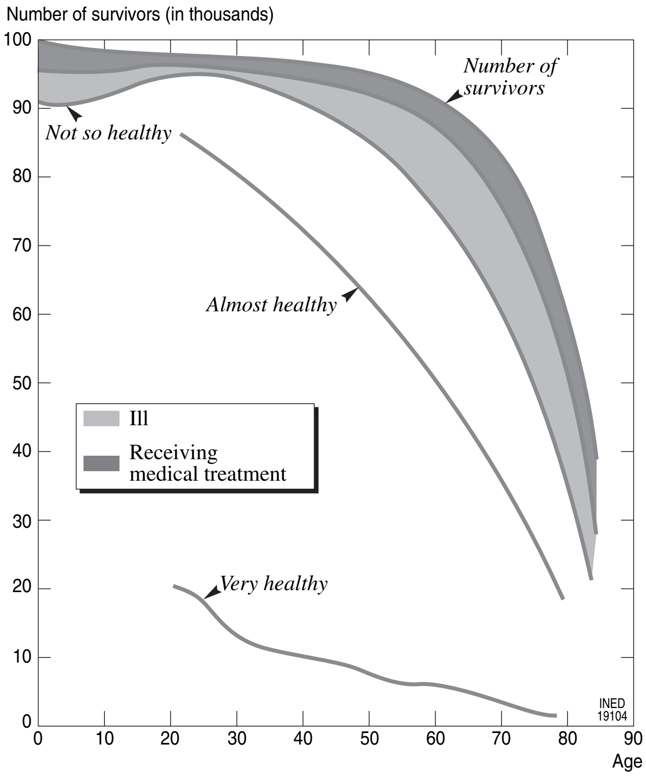


FIGURE 79–6 Survival and frailty curves of females in modern society. (From Koizumi, 1982.)

The quantitative effect of the selection relaxation depends on several factors: the relation between the mutation pressure and the degree of relaxation, the effect of the replacement therapy on the reproductive fitness, and the mode of inheritance (dominant, recessive, polygenetic inheritance) (see Cavalli-Sforza and Bodmer, 1971). Computer simulation shows that in all types of inheritance, the increase of the allele frequency is very slow, especially for recessive alleles or polygenetic conditions, in which most deleterious alleles are hidden in heterozygous combinations.

The possible dysgenic effects of morbidity or mortality control may be counteracted by birth control. In the future, it may be expected that the demographic transition, mainly through the experience of mortality and fertility control, will contribute to shift this control from quantitative to qualitative goals and from phenotypic to genotypic concerns. This way, the human species would control not only its demographic growth and phenotypic development but also its genetic composition, that is, its future evolution (Cliquet, 1996). This proposition may, at the beginning of this century, sound quite daring, if not appalling, just as, in the beginning of the 20th century, quantitative birth control was vigorously opposed. In a longer time perspective, however, it is difficult to image that humans, once they have mastered the necessary technology, will not try to avoid genetically determined diseases and to promote genetic features that are considered socially desirable or advantageous, such as heterosexuality, cognitive ability, sociability, beauty, and mental and physical health in general.

b. Migration

Modernization has increased considerably the opportunities for geographical mobility by enlarging the communication and transportation facilities. This has resulted in a substantial increase in migratory movements, internally as well as internationally.

Modernization not only increased the opportunities for geographical mobility but also influenced the values and norms with respect to in-group behavior. Communication and transportation opportunities and knowledge about and experience with others via a diversity of channels (e.g., scientific discoveries, commercial and industrial activities, mass media, and tourism) have broadened people's minds and made them less averse to strangers, to out-groups. Inevitably, this leads to a further broadening of the mating and reproductive communities and hence to the breaking up of geographical and social isolates, in-groups, etc.

Increasing migration flows make populations genetically more heterogeneous. On the basis of plant and animal experiments, one might expect to witness signs of heterosis or hybrid vigor. Some authors are of the view that the modern secular increase in body height is partly owing to increasing exogamy. However, conclusive proof of the effect of heterosis in the human has not yet been established.

Wherever the increasing migratory movements are of a massive nature and the immigrant populations are not well integrated, new isolates may be formed. Such forms of migration may lead to population genetically stratified societies, in which the immigrant groups often are confined to the lower socioeconomic and occupational strata and are residentially segregated in less attractive quarters (de Lannoy, 1978; Van der Haeghen, 1995). It is a picture that is traditionally well known in the Americas, but which, in recent decades, has also emerged in many Western European metropolitan areas and even in fast-growing economies in the Near East and the Far East.

It is difficult to judge whether the demographic transition has also influenced selective migration processes. There are indications that such forms of migration do occur (Bogin, 1988). Migration policies are, in most cases, selective in one or another way. Guestworkers are usually attracted on a temporary basis in periods of economic boom and are afterward expected to return to their countries of origin. For permanent migrants, on the contrary, selection policies seek to attract highly skilled or economically well-off people. To the extent that such social characteristics are correlated to partially genetically influenced features such as intelligence and physical health, the migration flows may be in part genetically selective. "Brain drain" is an issue that increasingly preoccupies developing countries (UNESCO, 1984). However, because the biological features (intelligence, health) involved in such migratory processes, are largely determined by polygenes, the immediate effects of such migratory processes are largely of a cultural and not of a genetic nature. Such migration flows might have genetic effects only on a long-term basis (Mather and Jinks, 1982).

A fundamentally important effect of the increased opportunities of geographical and social mobility is that it enhances the size of the population, particularly of the breeding unit or effective population size.

c. Population Size

The modern demographic transition is characterized by the broadening of the marriage or mating distances (see Walter, 1956; Lebel, 1983; Walter, 1956; Wijsman and Cavalli-Sforza, 1984), the breaking up of isolates (see Sheets, 1980; Yanase, 1992), the increase of interclass, inter-religious, intertribal, and interethnic matings (see Coleman, 1992; Relethford and Mielke, 1994), and the increase of population genetic exogamy (see Leonetti and Newell, 1982). All of these processes, in fact, involve increases of (effective) population size.

The population genetic effects of increasing population size are obvious: decrease or even disappearance of genetic drift, founder effect, and inbreeding. These changes have been most extensively documented for inbreeding, which is observed not only in developed countries (see Sutter and Tabah, 1955; Imaizumi, 1992) but wherever modernization and industrialization spreads (Fig. 79–7). Studies have documented this phenomenon in developing countries (see Khlat, 1988; Chandrasekar *et al*., 1993). The decreasing inbreeding levels are obviously associated to the decrease, if not the disappearance, of inbreeding depression.

Of course, these changes do not occur everywhere at the same speed; it may, in some cases, even be counteracted by rising fundamentalism, the isolation of immigrant minorities, or the traditional socioeconomic advantages of modest forms of consanguinity (Bittles, 1994). Hence, in some countries the early stages of modernization were characterized by an increase of consanguinity, not because of a lack of available mates but because some population categories wanted to keep their wealth in the family on basis of social status and ethnic identity (Jorde *et al*., 1992). Moreover, the early demographic changes themselves, among others resulting in an increase of surviving children and consequently of available relatives, produced a temporary increase in consanguineous marriages (Sutter, 1969).

CONCLUSIONS

This chapter examined the effects demographic mechanisms and processes may have on the genetic composition and structure of populations and, in particular, the effects the modern demographic transition has or might have on the genetic diversity of populations. These are important issues because the basic demographic variables—mating, fertility, migration, and mortality—are the proximate instruments of population genetic changes, eventually leading to biological evolution. Demography and genetics are closely intertwined: all of the basic demographic mechanisms and processes, on the one hand, and the population genetic mechanisms and processes, on the other hand, interact, often in multiple ways.

The demographic transition influences via all of the major demographic variables—mating, fertility, mortality, migration and population size—the genetic composition and structure of populations.

The broadening of the marriage or mate distance has two distinct population genetic effects: isolates are broken up, reducing inbreeding and hence diminishing the inbreeding depression, and exogamic matings increase, leading to interpopulational mixtures and enhanced genetic heterogeneity. Assortative mating on the basis of preferred mate characteristics increases, leading to a higher population genetic variance. The decreasing levels of celibacy may reduce selection intensity and increase genetic heterogeneity.

Changes in fertility—timing as well as intensity—may have different effects. The avoidance of births at higher ages lowers the frequency of genetic impairments, the appearance of which is correlated to parental age. The practice of selective abortion has the same effect, but may, in the long run, raise the equilibrium level of harmful recessive alleles owing to reproductive compensation. Whereas the mortality decline of the demographic transition resulted in a decrease of I_m, the fertility decline does not necessarily induce lower opportunities for selection. Data from different countries seem to suggest that, at least in the initial stages of the demographic transition, the I_f

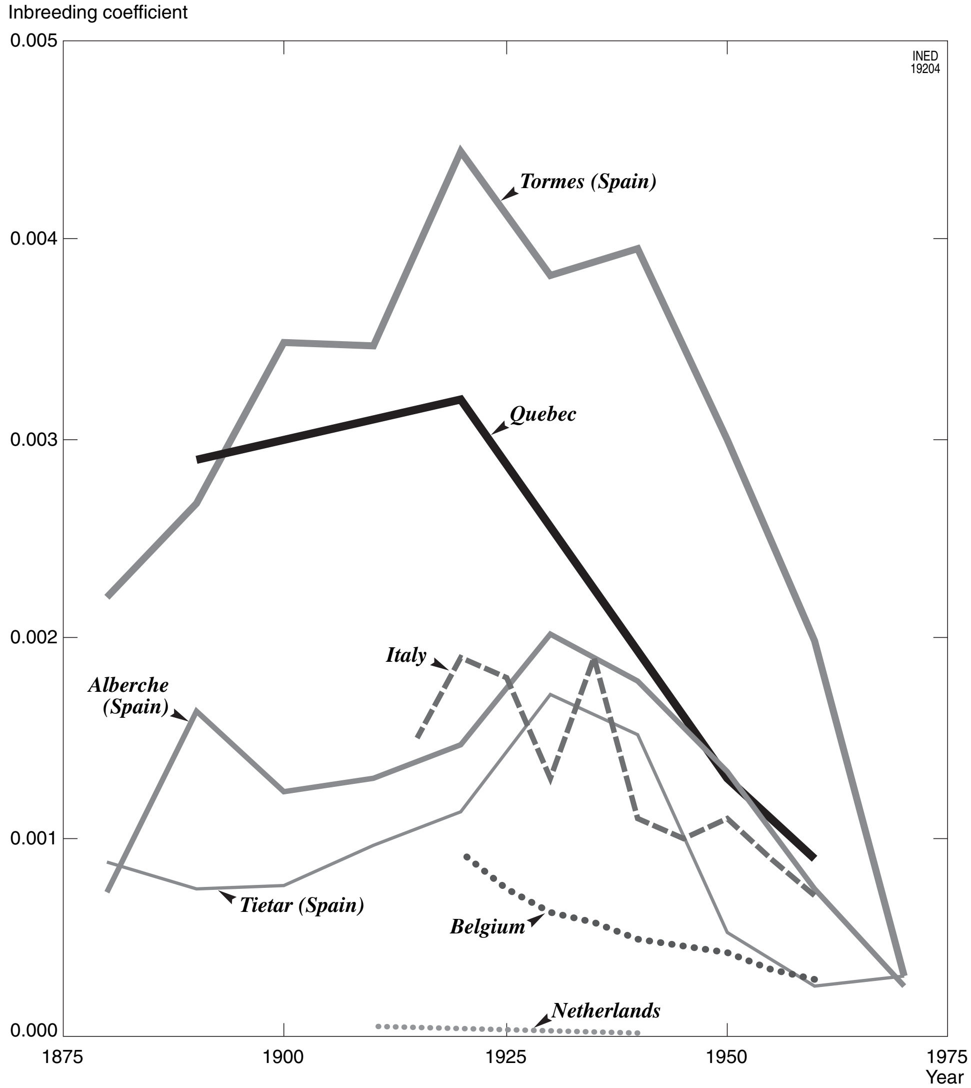


FIGURE 79–7 The change of the coefficient of inbreeding *(F)* in the course of the demographic transition. (From Twisselmann *et al.*, 1962; Moroni, 1969; Cavalli-Sforza and Bodmer, 1971; Fuster *et al.*, 1996.)

increased slightly, whereas in later stages it decreased. Although intelligence-related differential fertility is decreasing, the demographic transition seems so far to have had a slight contraselective effect.

The mortality reduction results in a selection relaxation. Improved living circumstances of all sorts and medical replacement therapies allow more frail individuals to survive and to reproduce. This results in a selection relaxation and in an increased adaptability of persons who would not have lived long. The increased survival rates are associated to an increase of biological frailty and of genetic heterogeneity.

Increasing migration flows also enhance population genetic heterogeneity. In combination with increasing population size, they break up isolates, decrease consanguinity levels, and lower inbreeding depression. The present increasing trends in migratory movements and exogamic behavior stopped the earlier fission in favor of an increasing fusion movement. Modern culture may have strengthened selective migratory processes, for example, by means of brain drain migration.

In the future, it can be expected that the demographic transition, mainly through the experience of mortality and fertility control, will contribute to shifts in life goals from quantitative to qualitative concerns and from phenotypic to genotypic concerns. Some less favorable genetic effects that may have been associated with the demographic transition are probably to be considered as temporary consequences of a major shift in demographic regime. They might be neutralized, if not reversed, by a future progress in genetic knowledge and genetic engineering and by raising expectations about quality of life.

References

Adams Julian, Hermalin Albert I., Lam David A., and Smouse Peter E. (eds.), 1990. *Convergent issues in genetics and demography.* Oxford, Oxford University Press, 361 p.

Adams Julian and Smouse Peter E., 1985. Genetic consequences of demographic changes in human population, *in:* Ranajit Chakraborty and Emoke J. E. Szathmary (eds.), *Diseases of complex etiology in small populations: ethnic differences and research approaches*, p. 283–299. New York, Alan Liss, 423 p.

Aschner Bernhard M. and Post Richard H., 1956–1957. Modern therapy and hereditary diseases, *in:* Tage Kemp, Mogens Hauge, and Bent Harvald (eds.), *First International Congress of Human Genetics*, Copenhagen, 1956, p. 362–369. Basel, S. Karger, 970 p.

Bajema Carl J., 1963. Estimation of the direction and intensity of natural selection in relation to human intelligence by means of the intrinsic rate of natural increase, *Eugenics Quarterly*, vol. 10(4), p. 175–187.

Beets Gijs, te Velde Egbert, Verloove-Vanhorick Pauline, Merkus Hans, and Bruinse Hein, 1994. Medical complications of aging fertility, *in:* Gijs Beets, Hans van den Brekel, Robert L. Cliquet, Gilbert Dooghe, and Jenny De Jong Gierveld (eds.), *Population and family in the low countries 1993. Late fertility and other current issues*, vol. 30, p. 3–23. Amsterdam et Lisse, Swets & Zeitlinger, 246 p.

Benoist Jean, 1986. Le métissage, *in:* Denise Ferembach, Charles Susanne, and Marie-Claude Chamla (eds.), *L'Homme, son évolution, sa diversité*, p. 539–551. Paris, Doin Éditeurs, 572 p.

Bittles Alan H., 1994. The role and significance of consanguinity as a demographic variable, *Population and Development Review*, vol. 20(3), p. 561–584.

Bodmer Walter F., 1968. Demographic approaches to measures of different selection in human progress, *Proceedings of National Academy of Science*, vol. 59(3), p. 690–699.

Bogin Barry, 1988. Rural-to-urban migration, *in:* C. G. Nicholas Mascie-Taylor and Gabriel W. Lasker (eds.), *Biological aspects of human migration*, p. 90–129. Cambridge, Cambridge University Press, 263 p.

Boss Judith A., 1990. How voluntary prenatal diagnosis and selective abortion increase the abnormal human gene pool. *Birth*, vol. 17(2), p. 75–79.

Boué Joëlle, Boué André, and Lazar Philippe, 1975. Retrospective and prospective epidemiological studies of 1500 karyotyped spontaneous human abortions, *Teratology*, vol. 12, p. 11–26.

Boyce Anthony J., Kuchemann C. F., and Harrison G. A., 1968. The reconstruction of historical movement patterns, *in:* Ernest D. Acheson (ed.), *Record linkage in medicine: proceedings of the international symposium, Oxford, July 1967*, p. 303–319. Baltimore, William and Wilkins, 399 p.

Callens Marc and Schoenmaeckers Ronald, 1993. Kinderen: hoeveel en wanneer? *In:* Robert Cliquet and Marc Callens (eds.), *Geszinsvorming in Vlaanderen: hoe en wanneer? Resultaten van de Enquête Gezinsontwikkeling 1991 (NEGO V)*, p. 77–96. Brussels, Centrum voor Bevolkings- en Gezinsstudie, 166 p.

Capron Christiane, Vetta Adrian R., Duyme Michel, and Vetta Atam, 1999. Misconceptions of biometrical iqists, *Current Psychology of Cognition*, vol. 18(2), p. 115–160.

Carnes Bruce A. and Olshansky S. Jay, 1993. Evolutionary perspectives on human senescence, *Population and Development Review*, vol. 19(4), p. 793–806.

Carter G. O., 1966. Differential fertility by intelligence, *in:* James E. Meade and Sir Alan S. Parkes (eds.), *Genetic and environmental factors in human ability: a Symposium held by the Eugenics Society in September–October 1965*, p. 185–200. Edinburgh, Oliver and Boyd, 242 p.

Cattell Raymond B., 1937. *The fight for our national intelligence.* London, King & son, 166 p.

Cattell Raymond B., 1950. The fate of national intelligence: test of a thirteen-year prediction, *Eugenics Review*, vol. 42, p. 136–148.

Cavalli-Sforza Luigi L., Barrai Italo, and Edwards Anthony W. F., 1964. Analysis of human evolution under random genetic drift, *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 29, p. 9–20.

Cavalli-Sforza Luigi L., and Bodmer Walter F., 1971. *The genetics of human populations.* San Francisco, WH Freeman, 966 p.

Cavalli-Sforza Luigi L., Menozzi Paolo, and Piazza Alberto, 1994. *History and geography of human genes.* Princeton, Princeton University Press, 518 p.

Cetin, I., Cozzi V., and Antonazzo P., 2003. Fetal development after assisted reproduction, *Placenta*, vol. 24(Suppl. B), p. 104–113.

Chandrasekar A., Jayraj Jacob S., and Rao PSS Sunder, 1993. Consanguinity and its trend in a Mendelian population of Andhra Pradesh, India, *Social Biology*, vol. 40(3–4), p. 244–247.

Chorney Michael J., Chorney Kathryn, Seese N., Owen Michael, Daniels J., Mcguffin Peter, Thompson L., Detterman Douglas K., Benbow Camillia, Lubinski David J., Eley T., and Plomin Robert, 1998. A quantitative trait locus associated with cognitive ability in children, *Psychological Science*, vol. 9(3), p. 159–166.

Clark David A., 2003. Is there any evidence for immunologically mediated or immunologically modifiable early pregnancy failure? *Journal of Assisted Reproduction and Genetics*, vol. 20(2), p. 63–72.

Cliquet Robert L., 1987. Over het toenemend wederzijds belang van bio-antropologie en demografie in de moderne cultuur, *Bevolking en Gezin*, no. 2, p. 107–129.

Cliquet Robert L., 1996. The demographic future of the human species: from quantity to quality? *in:* Fred Deven and Hans Van den Brekel (eds.), *Population and family in the low countries*, p. 15–42. Brussels, NIDI/CBGS Publications, 244 p.

Cliquet Robert L. and Delmotte Gino, 1984. Possible effects of differences in family size on the genetic and social composition and structure of Western populations, *in:* Heins G. Moors, Robert L. Cliquet, Gilbert Dooghe, and Dirk J. van de Kaa (eds.), *Population and family in the low countries*, p. 73–100. Brussels, NIDI/CBGS publications, 201 p.

Cochrane Susan Hill, 1979. *Fertility and education: what do we really know?* Baltimore, The Johns Hopkins University Press, 175 p.

Cole Lamont C., 1954. The population consequences of life history phenomena, *Quarterly Review of Biology*, vol. 29(2), p. 103–137.

Coleman David, 1992. Ethnic intermarriage, *in:* Alan H. Bittles and Derek F. Roberts (eds.), *Minority populations: genetics, demography and health*, p. 208–240. London, The Macmillan Press, 276 p.

Comfort Alexander, 1979. *The biology of senescence.* Edinburgh, Churchill Livingstone, 414 p. (published in 1956 as *Ageing, the biology of senescence*).

Constandse-Westermann Trinette S., 1972. *Coefficients of biological distance: an introduction to various methods of assessment of biological distances between populations, with special reference to human biological problems.* Oosterhout, Anthropological Publications, 142 p.

Cook R. C., 1956 and 1957. Changing patterns of selection, *in:* Tage Kemp, Mogens Hauge, and Bent Harvald (eds.), *Proceedings of the First International Congress of Human Genetics*, p. 349–353. Basel, S. Karger, 970 p.

Creasy Michael R. C., Crolla John A., and Alberman Eva D., 1976. Cytogenetic study of human spontaneous abortions using banding techniques, *Human Genetics*, vol. 31(2), p. 177–196.

Crow James F., 1958. Some possibilities for measuring selection intensities in man, *Human Biology*, vol. 30(1), p. 1–13.

Crow James F., 1966. The quality of people: human evolutionary changes, *BioScience*, vol. 16, p. 863–867.

Crow James F. and Felsenstein Joseph, 1968. The effect of assortative mating on the genetic composition of a population, *Eugenics Quarterly*, vol. 15(2), p. 85–97.

Dahlberg Gunnar, 1929. Inbreeding in man, *Genetics*, vol. 14, p. 421–454.

Darwin Charles, 1859. *On the origin of species by means of natural selection.* London, J. Murray, 502 p.

Darwin Charles, 1881. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London, J. Murray, 458 p. (New edition).

De Jonge C. J. and Barratt C. L. (eds.), 2002. Assisted Reproductive Technology. Cambridge: Cambridge University Press.

de Lannoy Walter, 1978. *Sociaal-geografische atlas van Brussel-Hoofdstad.* Kapellen, De Sikkel and De Nederlandsche Boekhandel, 99 p.

Devoucoux, R., Hurpin C., Baudon Dominique, Molez Jean-François, Roux Jean F., Guilloud-Bataille M., Carnevale P., and Feingold Josué, 1991. Population genetics of abnormal hemoglobins in Burkina Faso, West Africa, *Annals of Human Biology*, vol. 18(4), p. 295–302.

Epstein Elizabeth and Guttman Ruth, 1984. Mate selection in man: evidence, theory, and outcome, *Social Biology*, vol. 31(3–4), p. 243–278.

Erlenmeyer-Kimling Linda, Nicol Susan, Rainier John D., and Deming W. Edwards, 1969. Changes in fertility rates of schizophrenic patients in New York State, *American Journal of Psychiatry*, vol. 125(7), p. 916–927.

Erlenmeyer-Kimling Linda and Paradowski W., 1966. Selection and schizophrenia, *American Naturalist*, vol. 100, p. 651–665.

Evers-Kiebooms Gerry, 1994. Genetic risk, prenatal testing, and reproductive decision-making, *in:* Gijs Beets, Hans van den Brekel, Robert L. Cliquet, Gilbert Dooghe, and Jenny de Jong Gierveld (eds.), *Population and family in the low countries 1993, late fertility and other current issues*, p. 51–71. Lisse, Swets & Zeitlinger, 246 p.

Finch Caleb E., 1990. *Longevity, senescence, and the genome. Chicago*, The University of Chicago Press, 992 p.

Fisher Ronald A., 1951. Limits to intensive production in animals, *British Agricultural Bulletin*, no. 4, p. 217–218.

Fisher Ronald A. and Bennett J. Henry, 1930. *The genetical theory of natural selection.* Oxford, Clarendon Press, 318 p.

Flynn James R. 1984. The mean IQ of Americans: massive gains 1932–1978, *Psychological Bulletin*, vol. 91, p. 29–51.

Freire-Maia Newton and Elisbão T., 1984. Inbreeding effect on morbidity, III: a review of the world literature, *American Journal of Medical Genetics*, vol. 18(3) p. 391–400.

Fuhrmann Walter, 1969. Genetische überlegungen bei der familienplanung, *in:* Richard Kepp and Helmut Koester (eds.), *Empfängnisregelung und Gesellschaft: dritter Ärztlicher Fortbildungskurs der Medizinschen Fakultät der Justus Liebig Universität Giessen über Fragen der Kontrazeption am 21/22 Juni 1969 in Verbindung mit der pro Familia Deutsche Gesellschaft für Familienplanning.* Stuttgart, Thieme, 130 p.

Fuster Vincente, Morales B., Mesa M. S., and Martin J., 1996. Inbreeding patterns in the Gredos Mountain Range (Spain), *Human Biology*, vol. 68(1), p. 75–94.

Glass Bentley and Li Cno C., 1953. The dynamics of racial intermixture: an analysis based on the American Negro, *American Journal of Human Genetics*, vol. 5(1), p. 1–20.

Gordon Theodore J., Gerjuoy Herbert, and Anderson Mark, 1979. *Life-extending technologies: a technological assessment.* New York, Pergamon Press, 254 p.

Gottlieb Gilbert, 2001. Genetics and development, *in:* Neil J. Smelser and Paul B. Batles (eds.), *International encyclopedia of the social and behavioral sciences*, vol. 9, p. 6121–6127. Amsterdam, Elsevier, 26 vol. 17500 pp.

Graham John B., 1972. Genetic effects of family planning, *Family Planning*, vol. 21(2), p. 27–29.

Graham John B., Cohen B. H., and Schull William J., 1973. Genetic effects of family planning, *American Journal of Human Genetics*, vol. 24, p. 350–351.

Graham Robert K., 1970. *The future of man.* North Quincy, Christopher Publishing House, 200 p.

Hamilton William D., 1975. Innate social aptitudes of man: an approach from evolutionary genetics, *in:* Robin Fox (ed.), *Biosocial anthropology*, p. 133–155. New York, J. Wiley & Sons, 169 p.

Hardy Garrett H., 1908. Mendelian proportions in a mixed population, *Science*, vol. 28, p. 49–50.

Hay David A., 1985. *Essentials of behaviour genetics.* Oxford, Blackwell Scientific, 218 p.

Hay David A., 1999. Ideology matters more than science in the discussion of the genetics of intelligence, Cahiers de psychologie cognitive. *Current Psychology of Cognition*, vol. 18(2), p. 188–199.

Hed Helen M. E., 1987. Trends in opportunity for natural selection in the Swedish population during the period 1650–1980, *Human Biology*, vol. 59(5), p. 785–797.

Herrnstein Richard J. and Murray Charles, 1994. *The bell curve: intelligence and class structure in American life.* New York, The Free Press, Simon & Schuster, 872 p.

Hertig Arthur T., 1967. The overall problem in man, *in:* Kurt Benirschke (ed.), *Comparative aspects of reproductive failure: an international conference at Darmouth Medical School, Hanover (NH), July 25–29, 1966*, p. 11–41. New York, Berlin and Heidelberg, Springer, 486 p.

Higgins J. V., Reed Elisabeth W., and Reed Sheldon C., 1962. Intelligence and family size: a paradox resolved, *Eugenics Quarterly*, vol. 9(2), p. 84–90.

Hiorns R. W., Harrison G. Ainsworth, Boyce Anthony J., and Kuchemann C. F., 1968. A mathematical analysis of the effects of movement on the relatedness between populations, *Annals of Human Genetics*, vol. 32, part 1, p. 237–250.

Huynh T., Mollard R., and Trouson A., 2002. Selected genetic factors associated with male infertility. *Human Reproduction Update*, vol. 8, p. 183–198.

Illsley Raymond, Finlayson Angela, and Thompsom Barbara, 1963. The motivation and characteristics of internal migrants: a socio-medical study of young migrants in Scotland, *Milbank Memorial Fund Quarterly*, vol. 41, p. 114–144 and p. 217–248.

Imaizumi Yoko, 1992. Factors influencing the frequency of consanguineous marriages in Japan, *in:* Derek F. Roberts, Norio Fujik, and Kansi Torizuka (eds.), *Isolation, migration and health, 33rd Symposium of the Society for the Study of Human Biology*, p. 29–40. Cambridge, Cambridge University Press, 268 p.

Jacquard Albert, 1974. *The genetic structure of populations.* New York, Springer-Verlag, 569 p.

Jacquard Albert (ed.), 1976. *L'étude des isolats: espoirs et limites.* Paris, Institut national d'études démographiques, 344 p.

Jacquard Albert, 1979. La transition démographique: le capital génétique est-il affecté? *in:* IUSSP (ed.), *La science de la population au service de l'homme. Conférence sur la science au service de la* *vie/Population science in the service of mankind. Conference on science in the service of life*, p. 191–210. Liège, Institute of Life and International Union for the Scientific Study of Population, 246 p.

Jacquard Albert, 1983. Heritability: one word, three concepts, *Biometrics*, vol. 39, p. 465–477.

Jacquard Albert, 1984. Concepts of genetics and concepts of demography: specificities and analogies, *in:* Nathan Keyfitz (ed.), *Population and biology*, p. 29–39. Liège, Ordina, 318 p.

Jensen Arthur R., 1981. *Straight talk about mental tests.* New York, Free Press, 269 p.

Jorde Lynn B., Pitkänen Kari J., O'Brien Elisabeth, and Eriksson Aldur W., 1992. Consanguinity and genetic disease in Finland's Swedish-speaking minority, *in:* Alan H. Bittles and Derek F. Roberts (eds.), *Minority populations: genetics, demography and health. Proceedings of the 27th Annual Symposium of the Galton Institute*, p. 14–32. London, Macmillan, 276 p.

Kamin Leon, 1974. *The science and politics of IQ.* Potomac, Lawrence Erlbaum Associates, 183 p.

Kanisto Vaino, Lauritsen Jens, Thatcher A. Roger, and Vaupel James W., 1994. Reductions in mortality at advanced ages: several decades of evidence from 27 countries. *Population and Development Review*, vol. 20, p. 793–810, 921–924.

Keyfitz Nathan (ed.), 1984. *Population and biology: bridges between disciplines. Proceedings of a conference.* Liège, Ordina, 318 p.

Khlat Myriam, 1988. Consanguineous marriage and reproduction in Beirut, Lebanon, *American Journal of Human Genetics*, vol. 43(2), p. 188–196.

Khlat Myriam and Khoury Muin J., 1991. Inbreeding and diseases: demographic, genetic, and epidemiologic perspectives, *Epidemiologic Review*, vol. 13, p. 28–41.

Khoury Muin J., Cohen Bernice H., and Chase Gary A., 1987. An epidemiologic approach to the evaluation of the effect of inbreeding on prereproductive mortality, *American Journal of Epidemiology*, vol. 125(2), p. 251–262.

Kirk Dudley, 1969. *The genetic implications of family planning, biology and the family. Biology and the Family*, vol. 44, p. 80–83.

Kirk Dudley, 1996. Demographic transition theory, *Population Studies*, vol. 50(3) p. 361–387.

Kirkwood Thomas B. L. and Holliday Robin, 1986. Ageing as a consequence of natural selection, *in:* Alan H. BittLes and Kenneth J. Collins (eds.), *The biology of human ageing*, p. 1–16. Cambridge, Cambridge University Press, 260 p.

Kozinszky Z., Zadori J., Orvos H., Katona M., Pal A., and Kovacs L., 2003. Obstetric and neonatal risk of pregnancies after assisted reproductive technology: a matched control study. *Acta Obstetricia et Gynecologica Scandinavica*, vol. 82(9), p. 850–856.

Landry Adolphe, 1934. *La révolution démographique.* Paris, Librairie du Recueil Sirey, 231 p.

Lasker Gabriel Ward, 1963. *Human evolution, physical anthropology and the origin of man. Adapted by Frank N. Paparello from The evolution of man.* New York, Holt, Rinehart and Winston, 118 p.

Lasker Gabriel W. and Kaplan Bernice A., 1995. Demography in biological anthropology: human population structure and evolution, *American Journal of Human Biology*, vol. 7(4), p. 425–430.

Lebel Robert R., 1983. Consanguinity studies in Wisconsin, I: secular trends in consanguineous marriages, 1843–1981, *American Journal of Medical Genetics*, vol. 15(4), p. 543–560.

Leonetti Lockwood Donna and Newell-Morris Laura, 1982. Exogamy and change in the biosocial structure of a modern urban population, *American Anthropologist*, vol. 84(1), p. 19–36.

Leridon, H., 1973. *Aspects biométriques de la fécondité humaine. INED, Travaux et Documents 65.* Paris, Presses Universitaires de France.

Leslie Paul W., 1990. Demographic behavior, mating patterns, and the distribution of inbreeding, *in:* Julian Adams, Albert I. Hermalin, David Lam, and Peter E. Smouse (eds.), *Convergent issues in genetics and demography*, p. 63–79. Oxford, Oxford University Press, 361 p.

Lewontin Richard C., 1972. The apportionment of human diversity, *in:* Max K. Hecht and William S. Steere (eds.), *Evolutionary biology*, vol. 6, p. 381–398. New York, Plenum, 446 p.

Livshits Gregary and Nei Masatoshi, 1990. Relationships between intrapopulational and interpopulational genetic diversity in man, *Annals of Human Biology*, vol. 17(6), p. 501–513.

Loehlin John C., 1997. Dysgenesis and IQ, American Psychologist, vol. 52, p. 1236–1239.

Lynn Richard, 1996. *Dysgenics: genetic deterioration in modern populations.* Westport, Praeger, 237 p.

Lynn Richard, 1999. New evidence for dysgenic fertility for intelligence in the United States. *Social Biology*, vol. 46(1–2), p. 146–153.

Lynn Richard and Hampson Susan, 1986. The rise of national intelligence: evidence from Britain, Japan and the USA, *Personality and Individual Differences*, vol. 7, p. 323–332.

MacKay Trudy F. C., 2001. The genetic architecture of quantitative traits, *Annual Review of Genetics*, vol. 35, p. 309–339.

MacKlon Nick S., Geraedts Joep M., and Fauser Bart C. J. M., 2002. Conception to ongoing pregnancy: the black box of early pregnancy loss, *Human Reproduction Update*, vol. 8(4), p. 333–343.

Malécot Georges, 1948. *Les mathématiques de l'hérédité.* Paris, Masson, 63 p.

Manton Kenneth G. and Stallard Eric, 1994. Medical demography: interaction of disability dynamics and mortality, *in:* Linda G. Martin and Samuel H. Preston (eds.), *Demography of aging*, p. 217–278. Washington D.C., National Academy Press, 412 p.

Manton Kenneth G., Vertrees James C., and Wrigley J. Michael, 1990. Changes in health service use and mortality among U.S. elderly in 1980–1986, *Journal of Aging and Health*, vol. 2(2), p. 131–156.

Martin W. J., 1949. *The physique of young adult males.* London, HMSO, 66 p.

Mascie-Taylor C. G. Nicholas, 1984. Biosocial correlates of IQ, *in:* Clive J. Turner and H. B. Miles (eds.), *The biology of human intelligence*, p. 99–127. Nafferton, Driffield, North Humberside, Nafferton Books, 167 p.

Mascie-Taylor C. G. Nicholas, 1988. Assortative mating from psychometric characters, *in:* C.G. Nicholas Mascie-Taylor and Anthony J. Boyce (eds.), *Human mating patterns*, p. 62–82. Cambridge, Cambridge University Press, 237 p.

Mascie-Taylor C. G. Nicholas and Lasker Gabriel W., 1988. *Biological aspects of human migration.* Cambridge, Cambridge University Press, 263 p.

Mather Kenneth and Jinks L., 1982. *Biometrical genetics.* London, Chapman and Hall, 382 p.

Matsunaga Ei, 1966. Possible genetic consequences of family planning, *Journal of the American Medical Association*, vol. 198(5), p. 533–540.

Medawar Peter B., 1952. *An unsolved problem of biology.* London, H. K. Lewis, 24 p.

Medawar Peter B., 1965. Do advances in medicine lead to genetic deterioration? *Mayo Clinic Proceedings*, vol. 40(1), p. 23–33.

Medawar Peter B., 1974. The biological consequences of family planning, *Family Planning*, vol. 22(4), p. 71–76.

Moroni Antonio, 1969. *Historical demography, human ecology and consanguinity.* Liège, International Union for the Scientific Study of Population.

Morton Newton E., Yee S., Harris D. E. and Lew R., 1972. Bioassay of kinship, *Theoretical Population Biology*, vol. 2(3) p. 507–524.

Muhsam Helmut V., 1979. The demographic transition: from wastage to conservation of human life, *in:* IUSSP (ed.), *Population science in the service of mankind*, p. 143–165. Liège, IUSSP, 246 p.

Muller Hermann J., 1967. *What genetic course will man steer? Proceedings of the Third International Congress of Human Genetics*, p. 521–543. Basel, S. Karger.

Neel James V., 1992. The distinction between primary and secondary isolates, *in:* Derek F. Roberts, Norio Fujiki, and Kansi Torizuka (eds.), *Isolation, migration and health: 33rd Symposium of the Society for the Study of Human Biology*, p. 17–22. Cambridge, Cambridge University Press, 268 p.

Ōdegård O., 1980. Fertility of psychiatric first admissions in Norway 1936–1975, *Acta Psychiatrica Scandinavica*, vol. 62(3) p. 212–220.

Osborn Frederick, 1952. The eugenic hypothesis, II: megative eugenics, *The Eugenics Review*, vol. 44, p. 97–100.

Osborn Frederick, 1968. *The future of human heredity. An introduction to eugenics in modern society.* New York, Weybright and Talley, 133 p.

Perusse Daniel, 1993. Cultural and reproductive success in industrial societies: testing the relationship at the proximate and ultimate levels, *Behavioral and Brain Sciences*, vol. 16(2), p. 267–322.

Piazza Alberto, 1990. Migration and genetic differentiation in Italy, *in:* Julian ADAMS, Albert I. Hermalin, David Lam, and Peter E. Smouse (eds.), *Convergent issues in genetics and demography*, p. 81–93. New York/Oxford, Oxford University Press, 362 p.

Plomin Robert, 1997. Identifying genes for cognitive abilities and disabilities, *in:* Robert J. Sternberg and Elena Grigorenko (eds.), *Intelligence, heredity and environment*, p. 89–104. New York, Cambridge University Press, 628.

Plomin Robert, Defries John C., McClearn Gerard E., and McGuffin Peter, 2000. *Behavioral genetics: a primer.* New York, W.H. Freeman and Company, 449 p.

Plomin Robert, McClearn Gerald E., Smith Deirdre L., Skuder Patricia, Vignetti Sylvia, Chorney Michael J., Chorney Kathryn, Venditti C. P., Kasarda Steven, Thompson Lee A., Detterman Douglas K., Petrill Stephen A., Daniels Johanna, Owen Michael J., and McGuffin Peter, 1995. Allelic associations between 100 markers and high versus low IQ, *Intelligence*, vol. 21(1), p. 31–48.

Pollitzer William S., 1964. Analysis of a tri-racial isolate, *Human Biology*, vol. 36(4), p. 362–373.

Reddy Chengal, 1984. Distribution, endogamy, and isolation of Malas of Chittoor District, Andhra Pradesh, India, *Social Biology*, vol. 31(1–2), p. 108–113.

Reed T. Edward, 1969. Caucasian genes in American Negroes, *Science*, vol. 165(3895), p. 762–768.

Relethford John H. and Lees Francis C., 1983. Genetic implications of return migration, *Social Biology*, vol. 30(2), p. 158–161.

Relethford John H. and Mielke James H., 1994. Marital exogamy in the Aaland Islands, Finland, 1750–1949, *Annals of Human Biology*, vol. 21(1), p. 13–21.

Retherford Robert D. and Swell William H., 1988. Intelligence and family size reconsidered, *Social Biology*, vol. 35(1–2), p. 1–40.

Reynolds Vernon, Falger Vincent S. E., and Vine Ian (eds.), 1987. *The sociobiology of ethnocentrism: evolutionary dimensions of xenophobia, discrimination, racism, and nationalism.* London, Croom Helm, 327 p.

Riley James C., 1990. Long-term morbidity and mortality trends: inverse health transitions, *in:* John Caldwell, Sally Findley, Pat Caldwell, Gigi Santow et al. (eds.), *What we know about Health Transition: the cultural, social and behavioural determinants of health: the proceedings of an international workshop, Canberra, May 1989*, p. 165–188. Canberra, Australian National University, 499 p.

Ritsner Michael, Sherina O., and Ginath Y., 1992. Genetic epidemiological study of schizophrenia: reproduction behaviour, *Acta Psychiatrica Scandinavica*, vol. 85(6), p. 423–429.

Roberts Derek F., 1955. The dynamics of racial intermixture in the American negro: some anthropological considerations, *American Journal of Human Genetics*, vol. 7(1), p. 361–367.

Roberts Derek F., 1968. Genetic effects of population size reduction, *Nature*, vol. 220(1), p. 1084–1088.

Roberts Derek F., 1977. Assortative mating in man: husband/wife correlations in physical characteristics, *Bulletin of the Eugenics Society*, vol. 2(Suppl.).

Roberts Derek F., 1992. The Galton lecture for 1990: the price of isolation, *in:* Alan H. Bittles and Derek F. Roberts (eds.), *Minority populations: genetics, demography and health*, p. 35–67. London, The Macmillan Press, 186 p.

Roberts Derek F., Fujiki Norio, and Torizuka Kansi (eds.), 1992. *Isolation, migration and health: 33rd Symposium of the Society for the Study of Human Biology*. Cambridge, Cambridge University Press, 268 p.

Robine Jean Marie, Romieu Isabelle, and Michel Jean-Pierre, 2002. Trends in health expectancies, *in:* Jean-Marie Robine, Carol Jagger, Colin Mathers, Eileen Crimmins, and Richard Suzman (eds.), *Determining health expectancies*, p. 75–101. Chichester, Wiley, 428 p.

Rodgers Joseph Lee and Kohler Hans-Peter, 2003. *The biodemography of human reproduction and fertility.* Boston, Kluwer Academic Publishers, 258 p.

Rodgers Joseph Lee, Rowe David C., and Miller Warren B., 2000. *Genetic influences on human fertility and sexuality: theoretical and empirical contributions from the biological and behavioral sciences.* Boston, Kluwer Academic Publishers, 331 p.

Rose Michael R., 1991. *Evolutionary biology of aging.* Oxford, Oxford University Press, 221 p.

Rosenberg Noah A., Pritchard Jonathan K., Weber James L., Cann Howard M., Kidd Kenneth K., Zhivotovsky Lev A., and Feldman Marcus W., 2002. Genetic structure of human populations, *Science*, vol. 298(5602), p. 2381–2385.

Rubio Carmen, Simón Carlos, Vidal F., Rodrigo L., Pehlivan T., Remohí Jose, and Pellicer Antonio, 2003. Chromosomal abnormalities and embryo development in recurrent miscarriage couples, *Human Reproduction*, vol. 18(1), p. 182–188.

Rushton John Philippe and Nicholson Ian R, 1988. Genetic similarity theory, intelligence, and human mate choice, *Ethology and Sociobiology*, vol. 9(1), p. 45–57.

Schoenmaeckers Ronald, Callens Marc, Cliquet Robert, Corijn Martine, Deven Fred, and Lodewijckx Edith, 1997. *Gezinsvorming in Brussel-Hoofdstad.* Brussels, Centrum voor Bevolkings en Gezinsstudie.

Schwidetzky Ilse, 1950. *Grundzüge der Völkerbiologie.* Stuttgart, Ferdinand Enke Publisher, 312 p.

Shapiro Harry L., 1939. *Migration and environment: a study of the physical characteristics of the Japonese immigrants to Hawaï and the effects of environment on their descendants.* London, Oxford University Press, 593 p.

Sheets John W., 1980. Population structure of depreciated communities, I: the 1977 genetic demographies of Colonsay and Jura Islands, the Scottish Inner Hebrides, *Social Biology*, vol. 27(2), p. 114–129.

Sheps, M. C. and Menken J. A., 1973. *Mathematical models of conception and birth.* Chicago, University of Chicago Press.

Spuhler James N., 1962a. Empirical studies on quantitative human genetics, *in:* United Nations and World Health Organization (eds.), *The use of vital and health statistics for genetic and radiation studies, Proceedings of the seminar sponsored by the United Nations* *and the World Health Organisation held in Geneva, 5–9 September 1960*, p. 241–252. New York, Untied Nations, 259 p.

Spuhler James N., 1962b. The scope for natural selection in man, *in:* United Nations and World Health Organization (eds.), *The use of vital and health statistics for genetic and radiation studies, Proceedings of the seminar sponsored by the United Nations and the World Health Organisation held in Geneva, 5–9 September 1960*, New York, Untied Nations, 259 p.

Spuhler James N., 1963. Physical anthropology and demography, *in:* Philip M. Hauser and Otis D. Duncan (eds.), *The study of population: an inventory and appraisal*, p. 728–758. Chicago, University of Chicago Press, 864 p.

Spuhler James N., 1968. Assortative mating with respect to physical characteristics, *Eugenics Quarterly*, vol. 15(2), p. 128–140.

Spuhler James N., 1976. The maximum opportunity for natural selection in some human populations, *in:* Ezra B. W. Zubrow (ed.), *Demographic anthropology*, p. 185–226. Albuquerque, University of New Mexico Press, 299 p.

Steegmann Albert T. Jr., 1985. Eighteenth century British military stature: growth cessation, selective recruiting, secular trends, nutrition at birth, cold and occupation, *Human Biology*, vol. 57(1), p. 77–95.

Susanne Charles, 1967. Contribution à l'étude de l'assortiment matrimonial dans un échantillon de la population Belge, *Bulletin de la Société Royale d'Anthropologie et Préhistoire*, vol. 78, p. 147–196.

Sutter Jean, 1969. The effect of birth limitation on genetic composition of populations, *in:* Samuel J. Behrman, Leslie Corsa, and Ronald Freedman (eds.), *Fertility and family planning: a world view*, p. 213–251. Ann Arbor, The University of Michigan Press, 503 p.

Sutter Jean and Tabah Léon, 1948. Fréquence et répartition des mariages consanguins en France. *Population*, vol. 3(4), p. 607–630.

Sutter Jean and Tabah Léon, 1955. L'évolution des isolats de deux départements français: Loir-et-Cher, Finistère, *Population*, vol. 10(4), p. 645–674.

Termote Marc, 1984. Migration as a selection factor in a population. Some considerations on the relation between migration and genetics, *in:* Nathan Keyfitz (ed.), *Population and biology*, p. 133–146. Liège, Ordina, 318 p.

Thienpont Kriastiaan and Cliquet Robert L. (eds.), 1999. *In-group/out-group behaviour in modern societies: an evolutionary perspective*. Brussels, NIDI and CBGS Publications, 301 p.

Thiessen Delbert and Gregg B. Morin, 1980. Human assortative mating and genetic equilibrium: an evolutionary perspective, *Ethology and Sociobiology*, vol. 1(1), p. 111–140.

Twisselmann Fr., Moureau P. and François J., 1962. Évolution du taux de consanguinité en Belgique de 1918 à 1959, *Population*, vol. 17(2), p. 241–266.

United Nations Educational, Scientific, and Cultural Organization [UNESCO], 1984. Brain drain or the migration of talent and skills, *in:* United Nations (ed.), *Population distribution, migration and development*, p. 427–441. New York, United Nations, 505 p.

Vacher de Lapouge George, 1896. *Les sélections sociales*. Paris, Fontemoing, 503 p.

van den Berghe Pierre L., 1980. Incest and exogamy: a sociobiological reconsideration, *Ethology and Sociobiology*, vol. 1, p. 151–162.

van der Haegen Herman, Juchtmans G., and Kesteloot Christian, 1995. *Multicultureel Brussel*. Brussels, De Vlaamse Staatssecretaris bevoegd voor het niet-economisch wetenschappelijk ondrzoek van het Brussels Hoofdstedelijk Gewest, 48 p.

Vaupel James W., 1990. Kindred lifetimes: frailty models in population genetics, *in:* Julian Adams, Albert I. Hermalin, David Lam, and Peter E. Smouse (eds.), *Convergent issues in genetics and demography*, p. 155–170. Oxford, Oxford University Press, 361 p.

Verbrugge Lois M., 1984. Longer life but worsening health? Trends in health and mortality of middle-aged and older persons, *Milbank Memorial Fund Quarterly Health and Society*, vol. 62(3) p. 475–519.

Vetta, A. and Courgeau D., 2003. Comportements démographiques et génétique du comportement. *Population*, vol. 58(4–5), p. 457–488.

Villeneuve-Gokalp Catherine, 1997. Vivre en couple chacun chez soi, *Population*, vol. 52(5), p. 1059–1082.

Vincent Paul, 1961. *Recherches sur la fécondité biologique*. Paris, INED, PUF, 272 p.

Vining Daniel R., 1982. On the possibility of the re-emergence of a dysgenic trend with respect to intelligence in American fertility differentials, *Intelligence*, vol. 6(3), p. 241–264.

Vining Daniel R., 1986. Social versus reproductive success: the central theoretical problem of human sociobiology, *Behavioral and Brain Sciences*, vol. 9(1), p. 167–216.

Vogel Friedrich, 1992. Break-up of isolates, *in:* Derek F. Roberts, Norio Fujiki, and Kanji Torizuka (eds.), *Isolation, migration and health*, p. 23–28. Cambridge, Cambridge University Press, 268 p.

Vogel Friedrich, 1999. Alternative approaches to identify genes influencing intelligence. *Current Psychology of Cognition*, vol. 18(2), p. 252–262.

Vogel Friedrich and Motulsky Arno G., 1986. *Human genetics: problems and approaches*. Berlin, Springer-Verlag, 807 p.

Wachter Kenneth W. and Bulatao Rodolfo A., 2003. Biodemography of fertility and family formation, *in:* Kenneth W. Wachter and Rodolfo A. Bulatao (eds.), *Offspring: fuman fertility behavior in biodemographic perspective*, p. 1–18. Washington, The National Academies Press, 379 p.

Wachter, K. W. and Finch C. E., 1997. *Between Zeus and the Salmon: the biodemography of longevity*. Washington, National Academy Press.

Wahlund Sten, 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet, *Hereditas*, vol. 11(1), p. 65.

Waller Jerome H., 1971. Differential reproduction: its relation to I.Q. test score, education, and occupation, *Social Biology*, vol. 18(2), p. 122–136.

Walter Hubert, 1956. Räumliche Heiratsbeziehungen der Bevölkerung einer niedersachsischen Kleinstadt. (5. Tagung der Deutschen Gesellschaft für Anthropologie in Freiburg i. Br., 5 bis 7 April 1965).

Weinberg Wilhelm, 1908. Über den Nachweis der Vererbung beim Menschen, Separat-Abdruck aus Jahreshefte des Vereins für vaterländische Naturkunde in Wurttemberg, vol. 64, p. 368–382.

Weiss Kenneth M., 1990. The biodemography of variation in human frailty, *Demography*, vol. 27(2), p. 185–206.

Westermarck Edward A., 1922. *The history of human marriage*, 5th ed., rev. 3 vols. New York, Alberton, 1753 p.

White Karl R., 1982. The relation between socioeconomic status and academic achievement, *Psychological Bulletin*, vol. 81(3), p. 461–481.

Wijsman E. M. and Cavalli-Sforza Luigi L., 1984. Migration and genetic population structure with special reference to humans, *Annual Reviews of Ecological Systems*, vol. 15, p. 279–301.

Wilcox Allen J., Weinberg Clarice R., O'Connor John F., Baird Donna D., Schlatterer John, Canfield Robert E., Armstrong Glenn, and Nisula Bruce, 1988. Incidence of early loss of pregnancy, *New England Journal of Medicine*, vol. 319(4), p. 189–194.

Williams George C., 1957. Pleiotropy, natural selection and the evolution of senescence, *Evolution*, vol. 11, p. 398–411.

Williams George C., 1966. *Adaptation and natural selection. A critic of some current evolutionary thought*. Princeton, Princeton University Press, 307 p.

Willoughby Raymond R. and Coogan Marguerite, 1940. The correlation between intelligence and fertility, *Human Biology*, vol. 12(1), p. 114–119.

Witkop Carl J. Jr., Maclean C. J., Schmidt P. J., and Henry J. L., 1966. Medical and dental findings in the Brandywine isolate, *Alabama Journal of Medical Science*, vol. 3, p. 382–403.

Wolf Arthur, 1995. *Sexual attraction and childhood association: a Chinese brief for Edward Westermarck*. Stanford, Stanford University Press, 561 p.

Wood J., 1994. *Dynamics of human reproduction: biology, biometry, demography.* Hawthorne, Aldine.

Workman Peter L., 1973. Genetic analyses of hybrid populations, *in:* Michael H. Crawford and Peter L. Workman (eds.), *Methods and theories of anthropological genetics*, p. 117–150. Albuquerque, University of New Mexico Press, 509 p.

Wright Sewell, 1921. Systems of mating, I–V, *Genetics*, vol. 6, p. 111–178.

Wright Sewell, 1922. Coefficients of inbreeding and relationship, *American Naturalist*, vol. 56, p. 330–338.

Wright Sewell, 1931. Evolution in Mendelian populations, Genetics, vol. 16, p. 97–159.

Wright Sewell, 1943. Isolation by distance, *Genetics*, vol. 28, p. 114–138.

Wright Sewell, 1965. The interpretation of population structure by F-statistics with special regard to system of mating, *Evolution*, vol. 19, p. 395–420.

Wright Sewell, 1969. *Evolution and the genetics of populations, Volume II: the theory of gene frequencies.* Chicago, University of Chicago Press.

Wrong Dennis H., 1958. Trends in class fertility in Western nations, *Canadian Journal of Economics and Political Science*, vol. 24, p. 216–229.

Yanase T., 1992. Time trends in the break-up of isolates, *in:* Derek F. Roberts, Norio Fujiki, and Kansi Torizuka (eds.), *Isolation, migration and health: 33rd Symposium of the Society for the Study of Human Biology*, p. 23–28. Cambridge, Cambridge University Press, 268 p.

C H A P T E R

80

Lengthening of Life and the Population Health Status

JEAN-MARIE ROBINE AND CAROL JAGGER

Démographie et Santé, Institut national de la sante et de la recherché medicale (INSERM), Montpellier, France

Department of Epidemiology and Public Health, University of Leicester, United Kingdom

INTRODUCTION

Over the past 2 decades, in many developed and developing countries, life expectancy has greatly increased owing to the decrease in mortality. However, this positive change has raised new questions concerning the quality of the years of life gained, particularly for older people. Indeed, the individuals who escape death may still accumulate disabilities, after effects of accidents, or consequences of chronic diseases. So is there reason to fear that current and future progress in life expectancy will bring a deterioration in the mean health status of the population? Several theories have been put forward to answer this question. We will begin by presenting the general outlines of these theories.

These theories need to be confirmed with facts, and an accurate appreciation of the direction in which the facts are evolving is essential to adapt public health policies accordingly. This is why the concepts of health state expectancy, a composite measure of morbidity and mortality, and disability-free life expectancy (DFLE; or active life expectancy [ALE]) have been developed to address this question. We will discuss the development and definition of health state expectancies, defining the concepts underpinning the measurement of health states and giving some results of differentials in health state expectancies to demonstrate the relevance and utility of these indicators.

I. EXPANSION OR COMPRESSION OF MORBIDITY?

For a long time, the increase in life expectancy meant an improvement in the health of populations. Today, this is no longer enough because chronic diseases have replaced, or are progressively replacing, infectious diseases and the risk of becoming ill is not solely linked to the risk of dying (Riley, 1990). Thus, with a constant recovery rate, if the risk of dying diminishes more than the risk of becoming ill, the risk of being ill increases. In other words, the prevalence of chronic disease in the population can increase as a result of an increase in the duration of survival if the decrease in fatality is not offset by an equivalent decrease in incidence.

In the absence of pertinent data on the evolution of morbidity, the relationships that can exist between the patterns of evolution of these risks have been debated, gradually focusing on three theories: (1) an improvement in the state of health or a *compression of morbidity* (Fries, 1980, 1989); (2) a decline in the state of health leading to an *expansion of morbidity* (Gruenberg, 1977; Kramer, 1980; Olshansky *et al.*, 1991); and (3) *dynamic equilibrium*, a kind of status quo (Manton, 1982).

The theory of compression of morbidity states that (1) if the morbid period is defined as the period from appearance of the chronic disability until death, and (2) if the moment when such a morbid event appears can be postponed, and (3) if adult life expectancy is relatively constant, then (4) morbidity will be compressed into a shorter time (Fries, 1989).

According to Ernest Gruenberg and Michael Kramer, the decline in mortality is owing to the decrease in the fatality rate of chronic diseases and not to a reduction in their incidence or a slowing down in their rate of progression. Postponement of the moment of death causes more severe states of chronic disease to appear. This is the theory of the *pandemic of mental disorders and disabilities* or, more generally, expansion of morbidity.

Kenneth Manton (1982), who introduced the concept of dynamic equilibrium, proposed that the increase in life expectancy could be explained partly by a slowing down of the rate of progression of chronic diseases. Thus, the prevalence increases as mortality falls, but the prevalent states are, on average, less severe.

These three theories require the development of additional concepts such as the severity of prevalent states or that of disability. Indeed, the consequences of chronic disease are varied, ranging from death to the absence of any discomfort. They have been the object of international classifications (WHO, 1980, 2001), which place disability at their center. Disability is, at the same time, an indicator of the severity of morbid states and an indicator of the quality of years lived. Its introduction led to a considerable improvement in models of health (Fig. 80–1); the risks of disability can

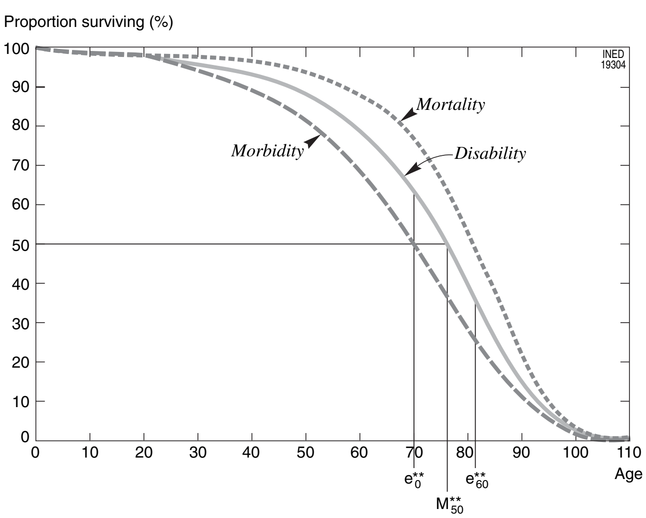


FIGURE 80–1 General model of health transitions according to the World Health Organization. The observed mortality and hypothetical morbidity and disability survival curves for females in the United States of America in 1980. $**e_0$ and $**e_{60}$ are the number of years of autonomous life expected at birth and at age 60, respectively; M_{50} is the age to which 50% of females could expect to survive without loss of autonomy. (From WHO, 1984.)

thus be added to the risks of disease. By distinguishing levels of severity, the risks of decline or improvement in the functional state of an individual can be measured.

First, the concept of disability allows us to rigorously define the progress of different scenarios of a populations' health corresponding to the three theories summarized above. Second, the breakdown of life expectancy into the years lived with or without disability provides the necessary tools to observe empirically which of these scenarios is confirmed. Thus, the theory of expansion of morbidity corresponds to the scenario in which the proportion of years lived with disability increases within life expectancy. The theory of compression of morbidity translates into the scenario in which the proportion of years lived with disability decreases within life expectancy. It thus becomes clear that the theory of compression of morbidity is not inevitably linked to the rectangularization of the survival curve (see Chapter 48). It is simply a verifiable fact that the survival rates without disability increase relatively more than do standard survival rates. Finally, taking into account the severity level of the disability, the theory of dynamic equilibrium corresponds to the scenario in which the proportion of years lived with disability—all levels combined—increases while the proportion of years lived with severe disability remains the same or even decreases within life expectancy. These theories correspond to different displacements of the survival curves of Fig. 80–1 to the right. One can imagine that they correspond to different phases in the health transition (see Chapter 57).

The first indicators of health state expectancy to be proposed were disability-free life expectancy (Sullivan, 1971) and, more particularly for the elderly, ALE (Katz *et al.*, 1983). A first calculation of health state expectancy, mostly DFLE, has been made for 49 countries, and repeated calculations exist for 14 of them (Robine and Romieu, 1998; Robine *et al.*, 1999). These calculations pertain to 20 of the 25 developed market economy countries; 4 of the least developed countries; 22 of 97 other developing countries; and 3 of the 22 economies in transition (Table 80–1). At present, it is impossible to compare these health state expectancies directly as the methods and the data used are very different from one country to another (UN, 1990). Similarly, for lack of comparable data, the World Health Organization (WHO) is currently producing crude estimates of disability-adjusted life expectancy for the 191 countries belonging to the organization (WHO, 2000).

An international research network, REVES (Réseau Espérance de Vie en Santé/Network on

TABLE 80–1 Health State Expectancy in the World According to Level of Economic Development

Level of development	Number of countries	Countries with calculations	
		n	%
DMEC	25	20	80.0%
LDC	47	4	8.5%
ODC	97	22	22.7%
ET	22	3	13.6%
Total	191	49	25.7%

DMEC, developed market economy countries; LDC, least developed countries; ODC, other developing countries; ET, economies in transition.

From Robine and Romieu, 1998.

Health Expectancy and the Disability Process) was set up in 1989 to facilitate international comparisons by examining the requirements for harmonization of health state expectancies (Bone, 1992). The network includes nearly 200 researchers from more than 100 research institutes or universities worldwide. The growth in the numbers of calculations shows there are as many possible health state expectancies as there are health concepts.

II. DEFINING AND CALCULATING HEALTH STATE EXPECTANCIES

An important first step toward valid international comparisons of health state expectancies is the harmonization of the conceptual approach to health state measurement. In the following section, we outline the functional approach to health, this being the main category of health state expectancies calculated to date. We also present the calculation methods actually used, because their convergence and harmonization constitute the second step in permitting international comparison of health state expectancies. However, it is important to note that other factors also require harmonization—such as the severity of the states of health taken into account, the protocol of surveys, and the formulation of questionnaires—before the calculations are strictly comparable.

In contrast to mortality, notions such as health or morbidity are complex and thus difficult to define. Health is “a composite of current state and prognosis (i.e., the probability of movement to other states) that occurs throughout an individual’s life” (Fanshel and Bush, 1970). In this way, health is not only “the ability to function now, but the outlook for future functional ability” (Fanshel, 1972). This life-course definition of health is the justification for the use of health state expectancies, as fundamental health indicators for populations, because they measure the lifetime spent in different health states.

During the 20th century, the epidemiological framework changed from communicable and acute diseases to long-term chronic diseases, with a consequent change in the approach to diseases and health. Beyond the presence or absence of disease, the classical biomedical approach, health status may be assessed through perceptual, functional, or adaptive approaches. With the functional approach, good health relates to ideas of effective achievement of roles and tasks, the fulfillment, without difficulty, of different activities. With the perceptual approach, good health relates to ideas such as well-being; a happy attitude to life; or a full, fruitful, and creative life. With the adaptive approach to health, good adaptation testifies to a harmonious relationship with one’s environment. The functional approach and the underlying conceptual framework of the disablement process are key factors in the health assessment of older people, making health state expectancies in combination with this approach and the resulting DFLE, of the utmost importance. Because the majority of health state expectancies calculated can be classified under the functional approach, we will look at this in more detail.

1. The Functional Approach

The functional approach to health is mainly addressed by four models, with some overlap: the model of disablement process (Nagi, 1965, 1976, 1991); the model of the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) (WHO, 1980); the model proposed for the revision of the classification, the International Classification of Functioning, Disability and Health (ICF, WHO, 2001); and the model of handicap creation process (Fougeyrollas *et al.,* 1998). The majority of these models distinguish between a maximum of five states: disease, impairments, functional limitations, activity restriction, and handicap.

Saad Nagi (1991) has the first three states (disease, impairments, functional limitations) followed by disability, which is defined as the inability or limitation in the fulfillment of activities and social roles in relation to work, the family, and an independent life (Nagi, 1991).

The ICIDH combines functional limitations and activity restriction under disability (Fig. 80–2), corresponding to any restriction or lack (resulting from an impairment) of ability to perform an activity in the

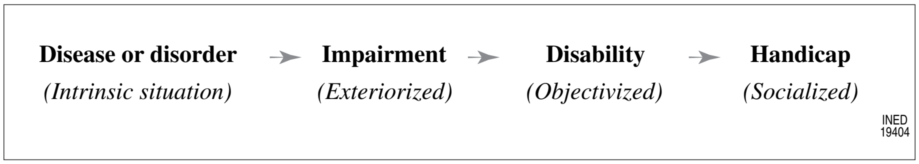


FIGURE 80–2 International Classification of Impairments, Disabilities and Handicaps (ICIDH): integration of concepts. (From WHO, 1980.)

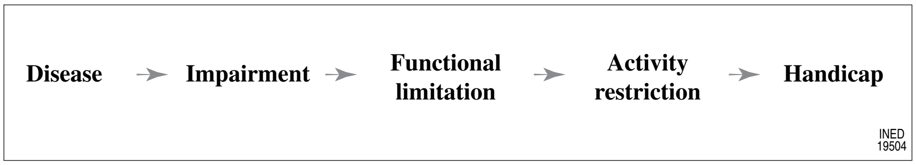


FIGURE 80–3 Disablement process, initial diagram by Philip Wood, 1975. (From Wood, 1975.)

manner or within the range considered normal for a human being. Under this framework, handicap is characterized by discordance between the activity and status of the individual and the expectations of his social environment (WHO, 1980). For fuller definitions see Appendix 80–1.

The boundary difficulties encountered, together with a need to further specify the area of mental health in the framework, made a revision of the ICIDH necessary. This revision, the ICF (WHO 2001), defines "functioning" and "disability" as umbrella terms covering the three new dimensions: (1) body functions and structure; (2) activities at the individual level, from simple to complex activities; and (3) participation in society. Functioning and disability are conceived as dynamic interactions between health conditions and contextual factors, which include both personal and environmental factors.

At the same time as the revision, an alternative model was devised that established the importance of the environment in the creation of handicap (Fougeyrollas *et al.*, 1998). This model of handicap creation puts the interaction between personal factors (organic systems and abilities) and environmental factors (facilitators or obstacles) at the heart of the model, the result of handicap situations (life habits). Risk factors (causes) are associated with personal factors, organic systems range from integrity to impairment, and abilities range from ability to disability.

Several attempts have been made to combine all or a part of the models described above through a disablement process (Verbrugge and Jette, 1994) or a general model (Robine *et al.*, 1997). The distinction between functional limitations and activity restrictions is central to these attempts, and the scheme with five levels (Fig. 80–3) (Wood, 1975) is preferred by those concerned with international comparisons as it clarifies the limits between impairment and disability and between disability and handicap. Thus, a difficulty concerning actions such as bending or reaching an object is a functional limitation, whereas difficulties concerning activities of daily life (dressing or domestic chores) are activity restrictions. The distinction between functional limitations and activity restrictions has been further blurred by existing measurement instruments that often combine these two levels and rely on apparently similar questions. Moreover, it has been suggested that functional limitations may be used as early indicators of activity restriction, thus avoiding potential heavier care burdens (Fried *et al.*, 1996; Lawrence and Jette, 1996). This two-step approach opens the way for research into new methods of measuring the functional health of populations by using functional limitation and activity restriction as two levels at which public health action could intervene.

2. The Health Expectancy Classification System

Initially, the term *disability-free life expectancy* was used without any reference to specific disability concepts. It was sometimes even used as a generic term to designate all health state expectancies. The concepts of the ICIDH (WHO, 1980) makes an initial distinction between life expectancies with or without impairment, life expectancies with or without disability, and life expectancies with or without handicap (social disadvantage). The clarification of concepts and terminology is very important; it facilitates communication between scientists at national and international levels and provides a means to promote these concepts more effectively among those responsible for health policy and the general public. It is the first step toward international harmonization of health state expectancies.

Table 80–2 summarizes the classification system currently used (Mathers, 2002; Robine 2002), which generalizes a first system developed by REVES for the WHO on the occasion of the first World Health Report (Robine *et al.*, 1995; WHO, 1995) and is based on previous work by The Netherlands Organization for Applied Scientific Research (Boshuizen and van de Water, 1994) and REVES (Mathers *et al.*, 1994).

Health expectancy is a generic term for all the health-related indicators expressed in terms of expectancy. It includes two main classes of indicators, the health state expectancies and the health-adjusted life expectancies.

Health state expectancy is a generic term for all the health-related indicators that express a defined state of health in terms of life expectancy. These indicators apply equally to positive and negative states of health. Thus, the sum of a complementary series of health state expectancies must always be equal to total life expectancy (Mathers *et al.*, 1994). The calculations of health state expectancy aim to provide indications on the current conditions of mortality and health. The repetition of calculations would make it possible to assess the progress of these conditions.

According to the concepts of the International Classification of Diseases (WHO, 1992), *life expectancy with or without disease* is the average number of years an individual can expect to live with or without disease if the current conditions of mortality and morbidity persist. A well-known example of this indicator is that of *life expectancy with or without senile dementia* (Ritchie *et al.*, 1993).

The International Classification of Handicaps (WHO, 1980) distinguishes among the concepts of (1) *life expectancy with or without impairment*, the average number of years an individual can expect to live with

TABLE 80–2 Classification System of Health Expectancies

Concepts	Indicators
Expectancy	Health expectancies (generic term for all)
Health state	**Health state expectancies**
Biomedical approach Disease (ICD)	Life expectancy with or without disease e.g., life expectancy with or without dementia
Functional approach	
Current system (REVES, 1994)[a]	
—Impairment	—Life expectancy with or without impairment
—Disability	
—Functional limitation	—Life expectancy with or without functional limitation
—Activity restriction	—Life expectancy with or without activity restriction
—Handicap	—Life expectancy with or without handicap
—Reduction in physical independence	—Life expectancy with or without reduction in physical independence
—ADL measures	—Active (independent) life expectancy[b]
—Mobility	—Life expectancy with or without mobility handicap
—Occupation	—Life expectancy with or without occupational handicap
—Other ICF (2001)	—Life expectancy with or without other handicap
—Impairment	
—Activity limitation	—Life expectancy with or without impairment
—Participation restriction	—Life expectancy with or without activity limitation
	—Life expectancy with or without participation restriction
Perceived health approach	Life expectancy in good perceived health / life expectancy in bad perceived health[c]
Health-adjusted	**Health-adjusted life expectancies**
Functional approach	Disability adjusted life expectancy (DALE)[d] (WHO, 2000)
Disability	

[a] The current REVES system for the functional approach combines the International Classification of Impairments, Disabilities, and Handicaps concepts, (WHO, 1980), the activities of daily living measures (Katz, 1983) and the REVES recommendations (1990).

[b] The use of the term independent life expectancy in place of active life expectancy is recommended to facilitate translation and comparability between countries.

[c] The terms healthy life expectancy or unhealthy life expectancy are often used for life expectancy in good perceived health or life expectancy in bad perceived health respectively. This was previously recommended by REVES, but given the current use of healthy life expectancy as a synonym of health expectancy by the WHO, this terminology should no longer be used.

[d] As proposed by the WHO (2000) to summarize the health of a population.

or without impairment; (2) *life expectancy with or without disability,* the average number of years an individual can expect to live with or without disability; and (3) *life expectancy with or without handicap*, the average number of years an individual can expect to live with or without handicap if the current conditions of mortality and impairment, disability, or handicap persist (cf. definitions in Appendix 80–1). The ICF (WHO, 2001), for its part, distinguishes among (1) *life expectancy with or without impairment,* the average number of years an individual can expect to live with or without impairment (i.e., with functional and structural integrity); (2) *life expectancy with or without activity limitation,* the average number of years an individual can expect to live with or without activity limitation; and (3) *life expectancy with or without participation restriction,* the average number of years an individual can expect to live with or without participation restriction if the current conditions of mortality and impairment, activity, or participation persist (cf. definitions in Appendix 80–1). The correspondence between the two classifications has yet to be defined, but life expectancy with a handicap in the original ICIDH meaning (WHO, 1980) is the nearest indicator to consequences of accidents and chronic diseases in daily life. Fortunately, all deficiencies (impairments) do not lead to a disability (or an activity limitation), and all disabilities do not lead to a handicap in daily life (or a participation restriction). The original ICIDH defines seven major dimensions of handicap (social disadvantage): handicaps of orientation, physical independence, mobility, occupation, social integration, economic self-sufficiency, and the other handicaps. In the same way, definitions are given for *life expectancies with or without handicap of physical independence, of mobility, of occupation, etc.* However, when the handicap is measured in a global manner, the corresponding indicator is a *life expectancy with or without handicap* in a general sense.

According to the recommendations of the committee of REVES on conceptual harmonization (Chamie, 1990) and to the attempts to combine the different conceptual models (Verbrugge and Jette, 1994; Robine *et al.,* 1997), life expectancy without disability sensu original ICIDH, encompasses the following concepts: (1) *life expectancy with or without functional limitation*, the average number of years an individual can expect to live with or without functional limitation; and (2) *life expectancy with or without activity restriction*, the average number of years an individual can expect to live with or without activity restriction if the present conditions of mortality and functional limitation or activity restriction, respectively, persist (cf. definitions in Appendix 80–1).

According to Katz *et al.* (1983), the ALE measures the average number of years an individual can expect to live without needing help to accomplish a whole series of basic activities of daily life or activities of daily living (ADL), if the current conditions of mortality and disability persist (cf. definitions in Appendix 80–1). ALE is, therefore, a life expectancy without handicap of physical independence in the sense of the original ICIDH. For this reason, it has been suggested that the term *active life expectancy* be replaced by *independent life expectancy*. This calculation can be extended to the need for help in realizing a whole series of complementary activities or instrumental activities of daily living.[1]

According to the perceived health approach, it is necessary to distinguish between perceived health expectancies. *Perceived health expectancy* is a generic term for all health state expectancies based on data on the perceived state of health. Thus, *life expectancy in good perceived health* is the average number of years an individual can expect to live in the favorable part of the distribution of perceived states of health (generally according to a scale of the form *very good, good, fair, bad, or very bad*).

Health-adjusted life expectancy is a generic term for all the life expectancies weighted by the social value given to different states of health in which the years are lived. The weighting used generally ranges from zero for death to one for perfect health (Mathers *et al.,* 1994; Mathers, 2002). This class of indicators is much less developed and contains mainly the disability-adjusted life expectancy, the new WHO indicator for the World Health Report (WHO, 2000), partly drawn from a Canadian indicator (Wolfson, 1996).

3. Calculation Methods

There are three calculation methods for health state expectancy: (1) the method of observed prevalence tables (Sullivan, 1971), (2) the method of double extinction tables (Katz *et al.,* 1983), (3) and the method of multistate tables (Rogers *et al.,* 1989). In this section the historic DFLE is taken as an example, but the different methods discussed here apply equally well to any other health state.

The main advantage of the *method of observed prevalence tables* (or Sullivan's method) stems from separate

[1] In French *espérance de vie active* means working life expectancy, with *actif* meaning in the labor force (occupied or unemployed). In English active life expectancy means life expectancy without physical dependence for personal care activities. As demographers have always paid attention to ease of translation, we strongly recommend the common use of independent and dependent life expectancy.

data collection on mortality and disability. The data on mortality come from standard mortality tables, and the data on prevalence of disability come from independent cross-sectional surveys.[2] However, the indicator obtained is not really a period indicator as the *period* prevalence is approximated by the prevalence observed in the survey.

The *method of double extinction tables* is based on the observation of two events during the study corresponding to two possible departures from the state of good health: death and entry into disability. In practice, it is sufficient to observe directly the probabilities of survival without disability. The main advantage of this method is that it provides a true period indicator from data that are not too difficult to collect. However, the method assumes that the disability studied is irreversible, which is the case with dementia, and that the recovery of lost functions is negligible.

The *method of multistate tables* was proposed to take recovery to good health into account. The main advantage of this method, based on transitions between different states of health, is that it provides a period indicator that takes the reversibility of states of disability into account. However, the data necessary for the calculation are scarce as they are difficult and expensive to collect. Finally, note that the multistate methods, as the method of double extinction, imply that the data on mortality and disability are collected in the course of the same study. The precision of mortality data depends, as a consequence, on the sample being of a sufficient size.

Most of the calculations of health state expectancy have been obtained with the method of observed prevalence tables (Sullivan's method, cf. Appendix 80–1). This method provides a very useful indicator as long as its limits are well understood (Robine and Mathers, 1993). The difference between the Sullivan and the multistate methods is not the use of prevalence versus the use of incidence and recovery. In both methods the prevalence is used in the calculation and comes from incidence and duration (recovery and lethality). But in the Sullivan method, the cross-sectionality observed (or current) prevalence comes mainly from past flows of incidence and recovery, whereas in multistate methods, the *equilibrium* prevalence (or table prevalence) comes from current flows only.[3] So the use of observed prevalences is not a suitable method for making instantaneous comparisons of health conditions in two countries or for establishing projections concerning the future health of populations. For all these reasons, it is preferable to observe and to use the period transitions between the different states of health. (For more details on the standard and advanced methods, respectively, see Cambois *et al.*, 1999; Laditka and Wolf, 1998.) EuroREVES, the European REVES committee has produced a first manual on the Sullivan method (Jagger, 1997), and a software application on the multistate approach has been finalized (Lièvre *et al.,* 2003).[4]

The health state expectancies aim to answer practical questions. Is the increase in life expectancy accompanied or not by an increase in the time lived with disability? Is DFLE increasing faster than is life expectancy, leading eventually to a compression of the time lived with disability? With the current concepts of disability, health state expectancies now aim to assess eventual differentials in the evolution of life expectancy without functional limitation and life expectancy without activity restriction. The disablement process—leading from accidents or chronic diseases to activity restrictions and social participation limitations—is complex, and measures to halt the process may be taken at different levels. Thus, an increase in the time lived with functional limitation may not necessarily be followed by an increase in the time lived with activity restriction or social participation limitation. Another question is whether or not an increase in the time lived without disability is accompanied by an increase in the time lived in good perceived health, a closer indication of individual satisfaction.

The goal of health state expectancies is clearly to monitor health transitions at work through the different components of health: mortality/longevity, measured or reported morbidity, functional limitation at the body level, activity restriction in daily life (i.e., in the daily environment), and perceived health. The

[2] In epidemiology, cross-sectional surveys study a population at one point in time: they serve to collect information on stocks (prevalence) or on past events. Longitudinal surveys follow a group of subjects over a period of time and allow the registration of new events (incidence). Demographers would prefer the term *point survey* to cross-sectional survey and the term *prospective survey* to "longitudinal survey", but the terms cross-sectional and longitudinal are standard throughout the literature.

[3] In this regard, the Sullivan method is much more robust than is the multistate approach. The methodology to assess the accuracy and quality of data collected through cross-sectional surveys is well developed, known, and used. It constitutes the topic of several chapters in statistical textbooks about the accuracy of estimations (e.g., confidence interval) and in epidemiology textbooks about the validity of indicators (e.g., sensitivity, specificity). Conversely, the methodology to assess the quality of data on individual transitions through longitudinal surveys is underdeveloped: in particular, how should expected misclassifications be separated from true transitions (according to the sensitivity-specificity of the indicator used) in the successive waves?

[4] Imach Version 0.8 (Lièvre *et al.,* 2003) available via the Internet (http://euroreves.ined.fr/imach).

period life table offers a strong conceptual and statistical framework for this, making it possible to design a complete family of integrated indicators. Health state expectancies, together with life expectancy, thus form a coherent set of health indicators to monitor health transitions at work in different countries. All the indicators are expressed in years so that basic combinations, such as summation or ratios, are easy to calculate and are immediately meaningful. The aim of these indicators is to summarize the health and mortality conditions over a period of time. This period characteristic is the essential precondition for monitoring change over time by repetition of the calculations.

III. TRENDS AND DIFFERENTIALS IN HEALTH STATE EXPECTANCIES

In addition to change over time, the estimates of health state expectancies already reported have brought information on sex, socioeconomic, and geographic differentials. They have also supplied information on the causes of morbidity and disability. The references of the studies summarized here can be found in Robine and Romieu (1998) or Robine *et al.* (1999), otherwise they are given in notes.

1. Sex Differentials

The wide differential between sexes found in life expectancies is not reproduced as strongly with the health state expectancies. Most studies indicate that life expectancy and positive health state expectancy (e.g., handicap-free, disability-free) are longer for women, and that the proportion of positive health state expectancy to total life expectancy is slightly lower for women in developed (Fig. 80–4) and in developing countries (Lamb, 1999).

Results from studies using data from repeated waves of surveys have suggested that the greater proportion of years lived with disability or handicap by women may be explained by the relatively higher survival of women after the development of these disabilities or handicaps (Robine and Ritchie, 1991; Mor *et al.*, 1994).

2. Socioeconomic and Sociodemographic Differentials

To date, socioeconomic variables have been included in studies from nine developed countries: Austria, Belgium, Canada, Finland, France, The Netherlands, Sweden, the United Kingdom (London), and the United States. All studies except one from the United States (Guralnik *et al.*, 1993) have demonstrated that social inequalities in health are much greater than has been shown by differential mortality: not only do the poorest and the least educated not live as long, but they also experience a greater part of their life with disability or handicap.

This was first observed in Canada, according to income levels (Wilkins and Adams, 1983). The gap in life expectancy between the richest and poorest sections of the community increased from 6.3 years for overall life expectancy to 14.3 years for occupational handicap-free life expectancy (Table 80–3).

Finnish and Dutch studies point to socioeconomic inequalities by level of education. The conclusions are

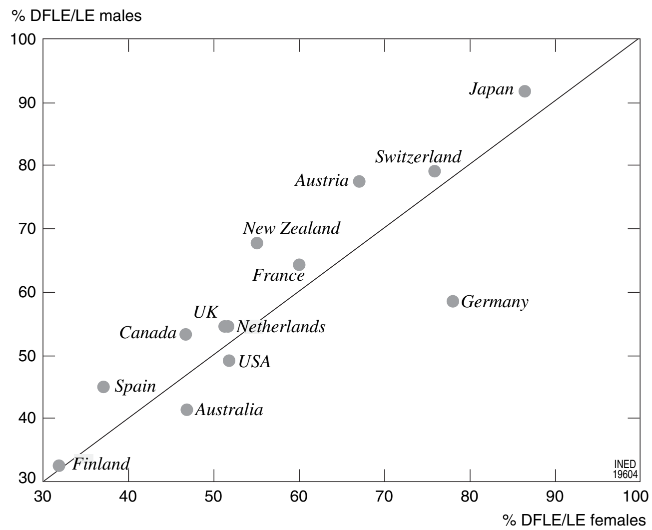


FIGURE 80–4 Proportion of disability-free life expectancy (% DFLE/LE): males versus females at age 65 in developed countries. (From Robine and Romieu, 1998.)

TABLE 80–3 Occupational Handicap-Free Life Expectancy in Canada, by Sex and Income Level, 1978

Income levels	At birth			
	Male		Female	
	LE	HE	LE	HE
Lowest	67.1	50.0	76.6	59.9
Second	70.1	57.9	77.6	61.8
Third	70.9	61.1	78.5	64.3
Fourth	72.0	62.6	79.0	63.5
Highest	73.4	64.3	79.4	67.5
Total	70.8	59.5	78.3	63.6
Differences highest/lowest	6.3	14.3	2.8	7.6

LE, life expectancy; HE, health expectancy.
From Wilkins and Adams, 1983.

similar: the higher the educational level, the higher the life expectancy and positive health expectancy. The calculations comparing different ethnic groups in the United States (Hayward and Heron, 1999) show dramatic differences to the detriment of the black population and to the benefit of the Asian population.

In France, repetition of the calculations at 10-year intervals shows that social inequalities in health remained constant for men over age 35 over the 1980s: gaps between the managerial staff group and the manual workers group in terms of life expectancy and DFLE, at ages 35 and 60, have not changed. In 1980 and 1991, at age 35, the managers group could expect an additional 5.4 years of life and an additional 7 years of life without disability compared with levels for the manual workers group. At age 60, the gaps reached 3 expected years of life and 4 expected years of life without disability. Nevertheless, this upholding of inequalities over the period is the result of an equivalent increase in life expectancy and DFLE for the socio-occupational groups (Cambois *et al.*, 2001). See Table 80–4.

3. National Geographic Comparisons

Several countries have computed estimates in order to make geographic comparisons across large geographic areas (Australia, Canada, France, Italy, Spain, and United Kingdom) (Robine and Romieu I, 1998) or between rural and urban areas (China and India). As with differences in life expectancy, differences in health expectancy across different geographic areas in the same country vary widely, although no general trends can be inferred.

In India, for example, comparisons in physical DFLE have been made between rural and urban males and females. Both life expectancy and physical DFLE are longer in more developed urban areas than in the less developed rural areas (see Table 80–5).

Figure 80–5 illustrates the area variations by using data for the United Kingdom. British researchers have computed expected years of ill health for numerous local areas, using data from the 1991 census. These calculations show that the local area with the shortest life expectancies also have the largest expected number of years in ill health. Whatever the causes of these area variations, it is more obvious that there may be a strong relationship between short life expectancy and amount of morbidity.

In all the situations studied and presented here based on sex, socioeconomic status, and geographic differentials, life expectancy and DFLE are positively associated with little evidence of the hypothetical trade-off between quantity and quality of life.

TABLE 80–4 Life Expectancy and Disability-Free Life Expectancy for Two Occupational Groups and for the Total French Male Population, 1980 and 1991

Occupational groups	Life expectancy				Disability-free life expectancy			
	35 years		60 years		35 years		60 years	
	1980	1991	1980	1991	1980	1991	1980	1991
Managers	41.3	43.5	19.1	21.1	35.9	38.2	14.3	17.0
Manual workers	35.9	38.1	15.8	18.0	28.6	31.4	10.5	12.8
Total male population	37.9	40.2	17.3	19.2	31.2	33.9	12.0	14.3

From Cambois *et al.*, 2001.

TABLE 80–5 Physical Disability-Free Life Expectancy in India by Sex and Residence, 1991

	At birth				At age 65			
	Male		Female		Male		Female	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Life expectancy	58.4	63.7	59.1	67.1	12.7	12.8	14.3	15.2
Physical disability-free life expectancy	56.6	60.7	56.6	62.6	6.6	7.4	7.1	7.8

From Guha, 1995.

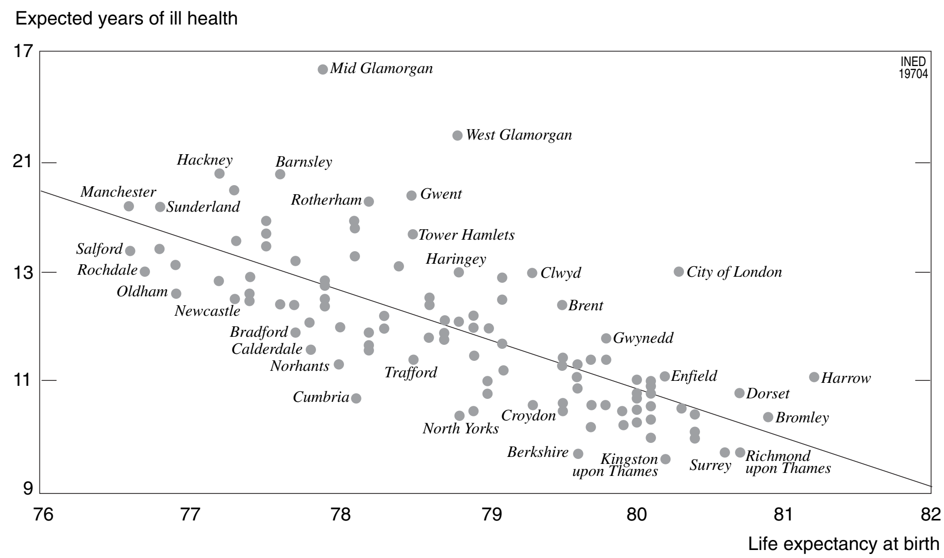


FIGURE 80–5 Life expectancy at birth and expected years of ill health for women, local authorities in England and Wales, 1991. (From Bone *et al.*, 1995: fig. 7, p. 37.)

4. Causes of Handicap, Disability, and Mortality

Potential gains in DFLE can be calculated after elimination of various pathologies, with a resulting ranking of the causes contributing to mortality and prevalence of morbidity (disability or handicap). Studies of this type have been undertaken to date for five developed countries (Australia, Canada, The Netherlands, United Kingdom, and United States). These studies have demonstrated an important effect produced by the elimination of osteo-articular diseases and accidents; in developed countries, these causes are of major importance after cardiovascular disease.

The results for The Netherlands in 1987–1988 show that although the elimination of fatal diseases leads to an increase in life expectancy and DFLE, it may also lead to an increase in life expectancy with disability, thus increasing the burden of disability to society (Nusselder *et al.*, 1996). However, the elimination of disabling nonfatal diseases results in a decline in life expectancy with disability (Table 80–6). Thus, the suppression of certain cancers would strongly increase life expectancy without increasing DFLE in the same proportion and therefore would strongly increase life expectancy with disability. Conversely, the suppression of nonfatal diseases, such as arthritis, would strongly increase DFLE without changing total life expectancy and therefore would strongly decrease life expectancy with disability. Between these two extremes, the suppression of fatal disabling diseases, such as heart diseases, would increase life expectancy and DFLE in various proportions.

5. Time Trends

Several time series of handicap-free life expectancy or DFLE have now been produced for 15 developed countries. A time series consists of at least two cross-sectional health surveys using the same measure of disability and handicap and comparable samples allowing comparisons over time. When the series are juxtaposed, they cover a period that extends over more than 30 years. Most authors distinguish between life expectancy without severe disability and life expectancy without disability, all levels of severity combined. DFLE most often means all levels of disabilities combined.

Figure 80–6 presents total life expectancy and life expectancy without disability—all disability levels combined—at age 65 in men in eight countries (United States, United Kingdom, Finland, Australia, France, New Zealand, Germany, and Canada) from 1970 to 1995. The analysis of the oldest series demonstrates that life expectancy without disability—all levels combined—is stagnating. However, the series across countries differentiate over time. For instance, the Canadian and Finnish series still suggest that DFLE is leveling off, whereas the American, French, British, and German series show that DFLE is now increasing, and the Australian series is decreasing. Hence, no general conclusion can be firmly drawn for this level of disability even if the general perception—when looking at Figure 80–7—is that although life expectancy is increasing whatever the country, DFLE—all severity levels combined—is apparently leveling.

TABLE 80–6 Change in Total Life Expectancy, Disability-Free Life Expectancy, Life Expectancy with Disability, and Percentage of Life Free of Disability Owing to the Elimination of the Specific Disease, The Netherlands, 1987–1988

Disease	Male at age 65				Female at age 65			
	LE	DFLE	LED	% DFLE/LE	LE	DFLE	LED	% DFLE/LE
At baseline	14.2	6.9	7.3	48.9	18.8	6.2	12.6	33.1
Chronic nonspecific lung disease	0.3	0.5	–0.2	2.2	0.1	0.2	–0.1	1.0
Heart disease	3.1	1.5	1.6	0.0	2.7	0.9	1.8	0.0
Cancer	2.7	0.9	1.8	–2.3	1.9	0.4	1.5	–1.2
Diabetes mellitus	0.1	0.0	0.1	–0.1	0.3	0.3	0.0	1.0
Arthritis/back complaints	0.0	0.7	–0.7	5.0	0.1	1.0	–1.0	5.3
Migraine/severe headache	0.0	0.1	–0.1	0.4	0.0	0.1	–0.1	0.4
Other neurological diseases	0.1	0.1	0.0	0.3	0.1	0.1	0.0	0.3

From Nusselder *et al.*, 1996.

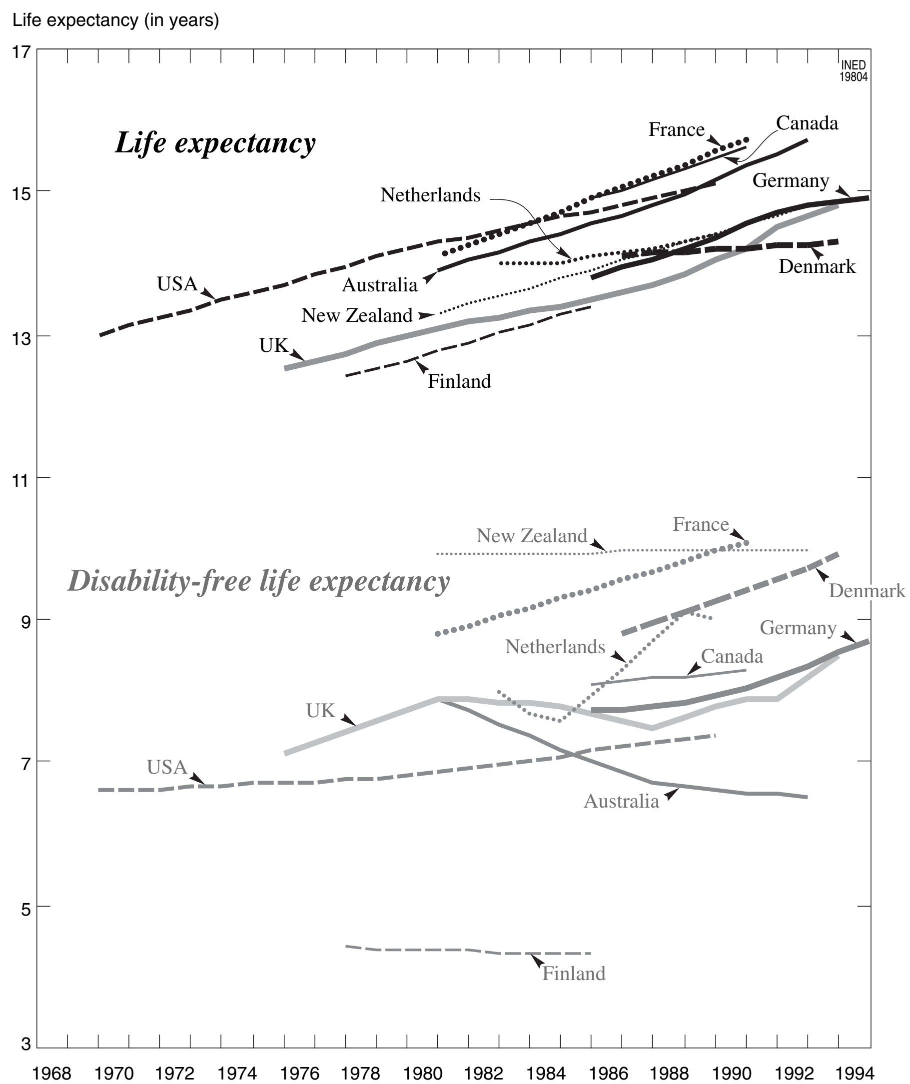


FIGURE 80–6 Evolution of life expectancy and disability-free life expectancy—all disability levels combined—in various countries at age 65. (From Robine and Romieu, 1998.)

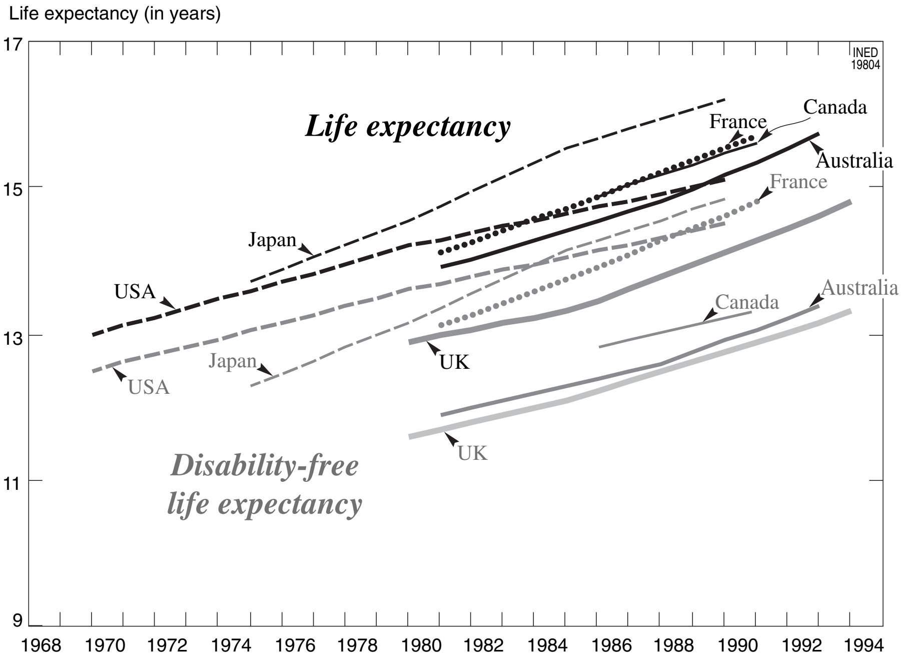


FIGURE 80–7 Evolution of life expectancy and life expectancy without *severe* disability in various countries, at age 65. (From Robine and Romieu, 1998.)

Total life expectancy and life expectancy without severe disability at age 65 is shown for men in six countries (United States, Japan, United Kingdom, Australia, France, and Canada) from 1970 to 1994 (Fig. 80–7). Life expectancy without severe disability roughly parallels total life expectancy, meaning that the number of years lived with severe disability is leveling off. This result can be verified whatever the developed country or the period of time studied or the indicator of severe disability considered (e.g., institutionalization in the United States; institutionalization or confinement to house for France; personal help or supervision required or impossibility to perform basic tasks in relation with self care, mobility, and verbal communication in Australia).

It is probable that the difference between the results obtained for severe disability and disability (all levels combined) is attributable to the less subjective declaration of severe disability compared with declaration of light disability. As time goes on, health requirements are increasing, and what was once considered normal for an age, sex, or social position is now considered disabling. This may well explain an apparent decrease in DFLE as well as incoherent trends between countries, with cultural differences explaining different perceptions of health. Nevertheless, this type of subjective trend has a greater effect on light disabilities than on severe disabilities, as the latter have always been perceived as disabling. The trends in severe DFLE are probably much more representative of the health status of the population.

In summary, it is apparent from the data available, that the increase in life expectancy is not accompanied by an increase in the time spent with severe disability. The results indicate at worst a pandemic of light and moderate, but not severe, disabilities. These results therefore tend to confirm the theory of *dynamic equilibrium,* which partly explains the increase in life expectancy by a slowing down in the rate of progression of chronic diseases (Manton, 1982). Thus, although the decline in mortality can lead to an increase in the prevalence of disabilities, these disabilities are less severe.

In France the general model of health transitions (WHO, 1984) has been used to compare the evolutions of life expectancy, DFLE, and life expectancy without chronic disease. The results show that the increase in life expectancy between 1981 and 1991 has been accompanied by a parallel increase in DFLE, and life expectancy without chronic disease has remained constant (Fig. 80–8). This apparent contradiction in the evolutions of morbidity and disability again illustrates the theory of dynamic equilibrium proposed by Manton in 1982: with the decline in mortality, the prevalence of chronic diseases increases but the diseases are less severe (Robine *et al.*, 1996).

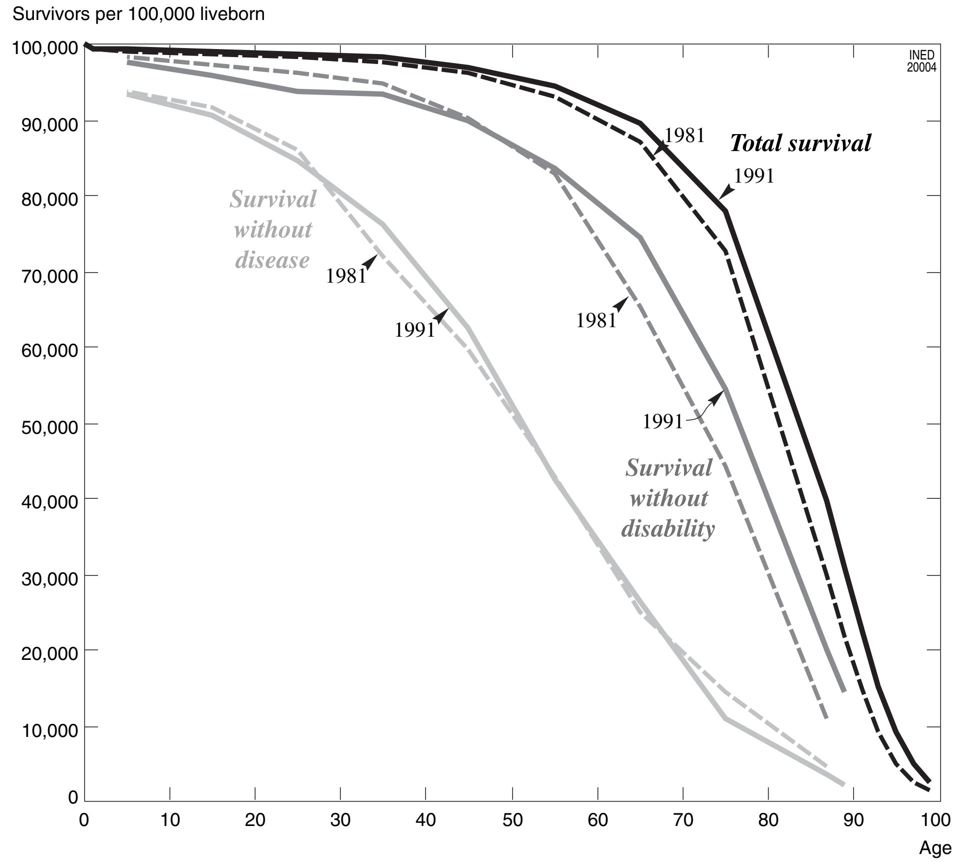


FIGURE 80–8 Survival without disease and survival without disability (World Health Organization model, 1984) France, 1981–1991, females. (From Robine *et al.*, 1996.)

IV. WHAT CONDITIONS CONSTITUTE THE YEARS LIVED IN BAD HEALTH OR WITH DISABILITY?

Finally, Tables 80–7 and 80–8 are given to illustrate the nature of years lived in bad health or in disability by people over 65 years.

In France, the health survey of 1991–1992 showed that a problem of disability or of handicap affected 39.8% of years lived outside institutions by men over 65 in France and 51.2% of years in women (Table 80–7). In men, 28.7% of years lived were with functional limitations of mobility and agility (unable to walk 200 m without stopping and without severe discomfort, great difficulty in going up and down stairs, or picking up an object from the floor), whereas 14.7% of years were with difficulties of communication (great difficulty in reading or following a conversation on the telephone), 12.1% of years were confined to the home, and 9.0% of years were lived with handicaps of physical independence (confined to a bed or an armchair or great difficulty getting dressed, washing, going to the toilet,

TABLE 80–7 Different Health State Expectancies at Age 65, France, 1991

Quality of life	Male		Female	
	Years	%	Years	%
Life expectancy	15.7		20.1	
Institutionalized	0.5		1.2	
Noninstitutionalized of which	15.2	100	18.9	100
With functional limitations of mobility and agility	4.4	28.7	7.9	41.6
With a diff. of communication	2.2	14.7	3.7	19.5
Home confined	1.8	12.1	3.2	17.2
Physically dependent (ADL)	1.4	9.0	2.1	11.0
Dependent for eating	0.5	2.9	0.7	3.6
Confined to bed or armchair	0.4	2.8	0.6	2.9
At least one of these problems	6.0	39.8	9.7	51.2
Without difficulty and performing	5.7	37.8	8.3	43.8
Without dementia[a]	15.1		19.2	
With dementia	0.6		0.9	

[a] Life expectancy with or without dementia (Ritchie *et al.*, 1994).
From Robine *et al.*, 1995.

TABLE 80–8 Different Health State Expectancies at Age 65, United Kingdom, Nottingham, 1985

Quality of life	Male		Female	
	Years	%	Years	%
Life expectancy	13.4	100	17.4	100
With mental impairment	1.1	8.3	2.0	11.3
Institutionalized, or with a				
Visual impairment	2.8	20.9	5.1	29.4
Hearing impairment	5.2	38.5	5.8	33.2
Poor or fair perceived health	5.0	37.3	9.0	52.0
With urinary incontinence	2.0	14.6	4.1	23.8
With mobility impairment	0.6	4.5	1.4	8.1
With depression	1.2	8.9	2.9	16.6
At least one of these problems	9.5	70.6	13.7	78.7

From Jagger *et al*., 1994; Bone *et al*., 1995.

or feeding oneself). For 2.9% of the time lived, men had great difficulty in feeding. It can also be seen that each of the most severe levels (time lived with great difficulty in feeding, in institutions, or with a dementia or time confined to a bed or an armchair) concerns only a small part of the years lived. The table also shows that although women declare many more functional limitations of mobility and agility, there are no great differences in the handicaps of physical independence. On the contrary, with regard to the proportion of years lived, not only without disability but also with good performance in the instrumental activities of daily life such as preparing a meal or filling in care expenditure claim forms, the proportion is higher for women (43.8%) than for men (37.8%).

Table 80–8 shows similar information from longitudinal surveys in Nottingham and Leicestershire, England, in 1985: 37.3% of years lived by men over 65 years are not perceived as years of good health (52.0% for women), 14.6% of years are lived with urinary incontinence, 8.9% with depression, 8.3% with mental impairments, and 4.5% with impairments of mobility (idea close to confinement to bed or an armchair).

The development of dementia-free life expectancies initiated by Ritchie (1991) has been conducted in 10 countries (Australia, Belgium, Denmark, France, Japan, The Netherlands, Spain [Catalonia], Switzerland, United Kingdom, and United States), although only 4 of the calculations are presented at a national level. In parallel with dementia-free life expectancies, other types of mental health expectancies are now being developed such as depression-free life expectancy and life expectancy in good mental health (Jagger *et al*., 1998).

CONCLUSION: POLICY RELEVANCE OF HEALTH STATE EXPECTANCY INDICATORS

Health state expectancies can be considered as a family of indicators in which each can be analyzed independently or according to each other. They can be added to each other: for example, the sum of DFLE plus life expectancy with disability is equal to total life expectancy. They can also be presented as a ratio of DFLE to total life expectancy—generally expressed as a percentage—indicating the proportion of life expectancy lived without disability. Health state expectancies offer a positive assessment of the time spent in different health states, similar to the way that life expectancy provides a positive indication of human longevity. They provide positive indications on the health of populations, on their vitality, or on their quality of life.

Health state expectancies can also assist in establishing public health priorities when potential gains are calculated. Gains in health state expectancies make it possible to classify priorities according to the survival or health criteria used in the calculation. This leaves a degree of choice to public health authorities by providing them with all the elements necessary to decide between longevity and, for example, duration of life with or without disability.

Health state expectancies can be used for direct comparison of the different groups that make up the population, whatever the distribution criterion used—e.g., sex, socio-occupational category, region—because, for their calculation, the years lived are divided by the number of survivors, which make them independent of the size and age structure of the populations from which the data are obtained.

As a rule, the calculations of health state expectancies are based on observed data: period life tables, results of population censuses, and results of various surveys (e.g., living conditions, health, disability, labour force).[5] This explains why estimations of health state expectancies are currently available for only about 50 countries, and why these estimations are not directly comparable from one country to another. In fact, the national characteristics of the different surveys, in terms of protocol, questionnaire, or question formulation, make international comparisons difficult. This is the major weakness of the present calculations of health state expectancy.

[5] However, when the data necessary to the methodological works devoted to health expectancies and its calculation methods were not available, they have been simulated (Mathers and Robine, 1997).

The profusion of possible indicators, as illustrated by Tables 80–7 and 8, makes it necessary to make a choice, consistent with the basic conceptual framework, because too many indicators may divert attention and too few indicators may hide the possible trade-off between the different facets of health as well as the effects.

Standardization of disability data should become a priority. A first objective could be the harmonization of the underlying concepts followed by the measurement of disability according to severity levels in the general population. For the most severe states, a consensus should not be too difficult to find, as almost all the countries use measures of personal care activities limitations to calculate severe disability (e.g., eating, dressing, washing). Moreover, these elementary activities are universal. The extension of this approach to other types of activity is worth exploring to standardize the measurement of less severe disability levels (e.g., mobility, household care activities, professional or school activities).

In conclusion, health state expectancy indicators are relevant and meaningful indicators for policymaking. Health state expectancies are not summary measures of population health aimed at replacing life expectancy or the most basic indicators, such as the prevalence or the raw figures. However, they do make it possible to assess whether the increase in life expectancy is accompanied by a compression of morbidity or by an expansion of disability. Very recent studies show that people with the lower risks (defined on the basis of smoking, body mass index, and exercise patterns) not only live longer but experience fewer years of disability before death (Vita *et al.*, 1998; Ferrucci *et al.*, 1999; Nusselder *et al.*, 2000). None of these three studies suggest that there is a trade-off between quantity and quality of life but rather, at least for two of them, that there is a possible compression of disability, with quantity and quality going hand in hand. Thus, if future increases in life expectancy are owing to better behaviors, they could be accompanied by a larger increase in DFLE, leading to a compression of morbidity. In fact, in the United States, certain forecasts of disability in elderly populations are very optimistic (Manton *et al.*, 1997).

Nevertheless, health state expectancies have not become a routine indicator of the health status of the population as they are not currently comparable across countries. Standardization of the concept and of the disability questionnaire used to calculate health state expectancies would resolve this problem. Direct comparisons across countries would then be as easy as the comparisons of life expectancies are today, and this task has already begun in the European Union (Robine *et al.*, 2001).

References

Bone M., 1992. International efforts to measures health expectancy. *Journal of Epidemiology and Community Medicine* vol. 46, p. 555–558.

Bone M., Bebbington A. C., Jagger C., Morgan K., and Nicolaas G., 1995. *Health expectancy and its uses*. London, HMSO. 90 p.

Boshuizen H. C. and van de Water H. P. A., 1994. *An international comparison of health expectancies*. Delft, The Netherlands, TNO Health Research.

Cambois E., Robine J. M., and Brouard N., 1999. Life expectancies applied to specific statuses: a history of the indicator and the methods of calculation. *Population: An English Selection* vol. 11, p. 7–34.

Cambois Emmanuelle, Robine Jean-Marie, and Hayward Mark D., 2001. Social inequalities in disability-free life expectancy in the French male population, 1980–1991, *Demography*, vol. 38(4), p. 513–524.

Chamie M., 1990. *Report of the Sub-committee on Conceptual Harmonization*. Brussels, REVES paper no. 41.

de Bruin A., Picavet H. S. J., and Nossikov A., 1996. *Health interview surveys: towards international harmonization of methods and instruments*. Copenhagen, World Health Organization, Regional Office for Europe.

Fanshel S., 1972. A meaningful measure of health for epidemiology. *International Journal of Epidemiology* vol. 1(4), p. 319–337.

Fanshel S. and Bush J. W., 1970. A health-status index and its application to health-services outcomes. *Operational Research* vol. 18(6), p. 1021–1066.

Ferrucci L., Izmirliam G., Leveille S. G., Phillips C. L., Corti M. C., Brock D. B., and Guralnick J. M., 1999. Smoking, physical activity and active life expectancy. *American Journal of Epidemiology* vol. 149(7), p. 645–653.

Fougeyrollas P., Cloutier R., Bergeron H., Côté J., and St. Michel G., 1998. *Classification québécoise: processus de production du handicap*. Lac Saint Charles, Canada, RIPPH/SCCIDIH.

Fried L. P., Bandeen-Roche K., Williamson J. D., Prasada-Rao P., Chee E., Tepper S., and Rubin G. S., 1996. Functional decline in older adults: expanding methods of ascertainment. *Journal of Gerontology: Medical Sciences* vol. 51A(5), p. M206–M214.

Fries J. F., 1980. Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* vol. 303, p. 130–135.

Fries J. F., 1989. The compression of morbidity: near or far? *Milbank Memorial Fund Quarterly* vol. 67, p. 208–232.

Grimley Evans J., 1983. Prevention of age-associated loss of autonomy: epidemiological approaches. *Journal of Chronic Diseases* vol. 37(5), p. 353–363.

Gruenberg E. M., 1977. The failures of success. *Milbank Memorial Fund Quarterly / Health and Society* vol. 55, 3–24.

Guha Roy S., 1995. Sub-national comparisons of disability indicators: the Indian example. *in: International Statistical Institute, Eds. 50th Session of the International Statistical Institute; 21–29 August; Beijing,* p. 1113–1130. Beijing, International Statistical Institute.

Guralnik J. M., Land K. C., Blazer D., Fillenbaum G. G., and Branch L. G., 1993. Educational status and active life expectancy among older blacks and whites. *New England Journal of Medicine,* 329(2), p. 110–116.

Hayward Mark D. and Heron Melonie, 1999. Racial inequality in active life among adult Americans, *Demography,* vol. 36(1), p. 77–91.

Jagger C., 1997. *Health expectancy calculation by the Sullivan Method: a practical guide.* Montpellier, Euro-REVES.

Jagger C., Morgan K., Clarke M., and Raymond N., 1994. *Pilot study (2) on the use of healthy active life expectancy measures: report* *to the Department of Health*. Leicester, University of Leicester and University of Nottingham.

Jagger C., Ritchie K., Bronnum-Hansen H., Deeg D., Gispert R., Grimley Evans J., Hibbett M., Lawlor B., MRC CFAS Group, Perenboom R., Polge C., and Van Oyen H., 1998. Mental health expectancy: the European perspective. *Acta Psychiatrica Scandinavia,* vol. 98, p. 85–91.

Katz S., Branch L. G., Branson M. H., Papsidero J. A., Beck J. C., and Greer D. S., 1983. Active life expectancy. *New England Journal of Medicine,* vol. 309, p. 1218–1224.

Katz S., Ford A. B., Moskowitz R. W., Jackson B. A., and Jaffe M. W., 1963. Studies of illness in the aged. The Index of ADL: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association,* vol. 185(12), p. 914–919.

Kramer M., 1980. The rising pandemic of mental disorders and associated chronic diseases and disabilities *Acta Psychiatrica Scandinavica,* vol. 62 (Suppl. 285), p. 282–297.

Laditka S. and Wolf Douglas A., 1998. New methods for analysing active life expectancy. *Journal of Aging and Health,* vol. 10(2), p. 214–241.

Lamb V. L., 1999. *Active life expectancy of the elderly in selected Asian countries.* Tokyo, NUPRI, 22 p.

Lawrence R. H. and Jette A. M., 1996. Disentangling the disablement process. *Journal of Gerontology: Social Sciences,* vol. 51B(4), p. S173–S182.

Lawton M. P. and Brody E. M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist,* vol. 9, p. 179–186.

Lièvre Agnès, Brouard Nicolas, and Heathcote Christopher R., 2003. The estimation of health expectancies from cross-longitudinal surveys, *Mathematical Population Studies,* vol. 10(4), p. 211–248.

Manton K. G., 1982. Changing concepts of morbidity and mortality in the elderly population. *Milbank Memorial Fund Quarterly/Health and Society,* vol. 60, p. 183–244.

Manton K. G., Corder L., and Stallard E., 1997. Chronic disability trends in elderly United States populations: 1982–1994. *Proceedings of the National Academy of Sciences USA,* vol. 94, p. 2593–2598.

Mathers Colin D., 2002. Health expectancies: an overview and critical appraisal, *in:* Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers, and Alan D. Lopez (eds.), *Summary measures of population health,* p. 177–204. Geneva, World Health Organization, 765 p.

Mathers C. D. and Robine J. M., 1997. How good is Sullivan's method for monitoring changes in population health expectancies. *Journal of Epidemiology and Community Health,* vol. 51, p. 80–86.

Mathers C., Robine J. M., and Wilkins R., 1994. Health expectancy indicators: recommendations for terminology. *in:* C. D. Mathers, J. McCallum, and J. M. Robine (eds.), *Advances in health expectancies*, p. 34–41 Canberra, Australian Institute of Health and Welfare, 450 p.

Mor V., Wilcox V., Rakowski W., and Hiris J., 1994. Functional transitions among the elderly: patterns, predictors, and related hospital use. *American Journal of Public Health,* vol. 84(8), p. 1274–1280.

Nagi S. Z., 1965. Some conceptual issues in disability and rehabilitation. *in:* M. B. Sussman (ed.), *Sociology and rehabilitation,* 100–113. Washington, American Sociological Association.

Nagi S. Z., 1976. An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly / Health and Society,* vol. 54, p. 439–467.

Nagi S. Z., 1991. Disability concepts revisited: implication for prevention. *in:* A. M. Pope and A. R. Tarlov (eds.), *Disability in America: toward a national agenda for prevention.* Washington, National Academy Press.

Nusselder W. J., Looman C. W. N., Marang-van de Mheen, van de Mheen H., and Mackenbach J. P., 2000. Smoking and the compression of morbidity. *Journal of Epidemiology and Community Health,* vol. 54, p. 566–574.

Nusselder W. J., Van Der Velden K., Van Sonsbeek J. L. A., Lenior M., and Van Den Bos G. A. M., 1996. The elimination of selected chronic diseases in a population: the compression and expansion of morbidity. *American Journal of Public Health,* vol. 86, p. 187–194.

Olshansky S. J., Rudberg M. A., Carnes B. A., Cassel C. K., and Brody J. A., 1991. Trading off longer life for worsening health: the expansion of morbidity hypothesis. *Journal of Aging and Health,* vol. 3, p. 194–216.

Riley J. C., 1990. The risk of being sick: morbidity trends in four countries. *Population and Development Review* vol. 16 (3), 403–432.

Ritchie K., 1991. La mesure de l'espérance de vie sans démence: étude préliminaire de faisabilité. *Cahiers Québécois de Démographie* vol. 20, 253–268.

Ritchie K., Jagger C., Brayne C., and Letenneur L., 1993. Dementia-free life expectancy: preliminary calculations for France and the United Kingdom, *in:* J. M. Robine, C. D. Mathers, M. R. Bone, and I. Romieu (eds.), *Calculation of health expectancies; harmonization, consensus achieved and future perspectives,* p. 233–244. Paris, John Libbey Eurotext.

Ritchie K., Robine J. M., Letenneur L., and Dartigues J. F., 1994. Dementia-free life expectancy in France. *American Journal of Public Health* vol. 84(2), 232–236.

Robine Jean-Marie, 2002. A new health expectancy classification system. *in:* Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers, and Alan D. Lopez (eds.), *Summary measures of population health,* p. 205–211. Geneva, World Health Organization, 765 p.

Robine Jean-Marie, Jagger Carol, and Romieu Isabelle, 2001. Disability-free life expectancies in the European countries: calculation and comparaison, *Genus,* vol. LVII(2), p. 89–101.

Robine Jean-Marie and Mathers Colin D., 1993. Measuring the compression or expansion of morbidity through changes in health expectancy, *in:* Jean-Marie Robine, Collin D. Mathers, Margaret R. Bone, and Isabelle Romieu (eds.), *Calculation of health expectancies: harmonization, consensus achieved and future perspective,* p. 269–286. Montrouge, John Libbey Eurotext, 388 p.

Robine J. M., Mormiche P., 1994. Estimation de la valeur de l'espérance de vie sans incapacité en France en 1991. *Solidarité Santé* vol. 1, 17–36.

Robine J. M., Mormiche P., and Cambois E., 1996. Evolution des courbes de survie totale, sans maladie chronique et sans incapacité en France de 1981 à 1991: application d'un modèle de l'OMS. *Annales de Démographie Historique* 99–115.

Robine J. M., Ravaud J. F., and Cambois E., 1997. *General concepts of disablement*. *in:* D. Hamerman (ed.), *Osteoarthritis*. Baltimore, Johns Hopkins University Press, 251 p.

Robine J. M. and Ritchie K., 1991. Healthy life expectancy: evaluation of a new global indicator of change in population health. *British Medical Journal* vol. 302, 457–460.

Robine J. M. and Romieu I., 1998. *Healthy active ageing: Health expectancies at age 65 in the different parts of the world*. Montpellier, REVES, REVES paper no. 318.

Robine J. M., Romieu I., and Cambois E., 1995. *Rapport final de la convention MIRE: Espérances de santé pour la population âgée de 65 ans et plus: Application des concepts de la classification internationale des* *handicaps aux calculs d'espérance de santé: France, début des années 1990*. Montpellier, Laboratoire universitaire d'épidémiologie et d'économie de la santé.

Robine J. M., Romieu I., and Cambois E., 1999. Health expectancy indicators. *Bulletin of the World Health Organization* vol. 77(2), 181–185.

Robine J. M., Romieu I., and Cambois E., van de Water H. P. A., Boshuizen H. C., and Jagger C., 1994. *Global assesment in positive health: contribution of the Network on Health Expectancy and the Disability Process to World Health Report 1995.* Montpellier, REVES/INSERM.

Rogers A., Rogers R. G., and Branch L., 1989. A multistate analysis of active life expectancy. *Public Health Report* vol. 104, 222–225.

Sullivan D. F., 1971. A single index of mortality and morbidity. HSMHA *Health Report* vol. 86, 347–354.

United Nations, 1990. *Disability Statistics Compendium* ST/ESA/STAT/Ser. Y/4, UN.

Verbrugge L. M. and Jette A. M., 1994. The disablement process. *Social Science and Medicine* vol. 38(1), 1–14.

Vita A. L., Terry R. B., Hubert H. B., and Fries J. F., 1998) Aging, health risks, and cumulative disability. *New England Journal of Medicine* vol. 338(15), 1035–1041.

Wilkins R. and Adams O. B., 1983. Health expectancy in Canada, late 1970s: demographic, regional and social dimensions. *American Journal of Public Health* vol. 73(9), 1073–1080.

Wolfson M. C., 1996. Health-adjusted life expectancy. *Rapports sur la santé* vol. 8(1), 41–46.

Wood P. H. N., 1975. *Classification of impairments and handicaps*. Geneva, WHO.

World Health Organization [WHO], 1980. *International Classification of Impairments, Disabilities, and Handicaps: a manual of classification relating to the consequences of diseases*. Geneva, WHO.

World Health Organization [WHO], 1984. *The uses of epidemiology in the study of the elderly: report of a WHO scientific group on epidemiology of aging*. Geneva, WHO.

World Health Organization [WHO], 1992. *International statistical classification of diseases and related health problems: 10th revision*. Geneva, WHO.

World Health Organization [WHO], 1995. *The world health report 1995: bridging the gaps*. Geneva, WHO, 118 p.

World Health Organization [WHO], 2000. *The World Health Report 2000: health systems: improving performance*. Geneva, WHO.

World Health Organization [WHO], 2001. *International classification of functioning, disability and health (ICF)*. Geneva, WHO, 299 p.

APPENDIX 80–1

1. International Classification of Impairments, Disabilities, and Handicaps

In the International Classification of Impairments, Disabilities, and Handicaps (ICIDH), the World Health Organization distinguishes among *impairments, disabilities and handicaps* (WHO, 1980). Basically, one finds an organic or functional anomaly corresponding to the intrinsic situation, which can be exteriorized by the *impairment*, objectivized by the *disability,* and socialized by the *handicap.*

Impairments

In the field of health, impairment corresponds to all loss of substance or alteration in a physiological or anatomical structure or function (WHO, 1980). Impairments are linked to pathologies that can be organic as well as functional.

Note that impairment is wrongly interpreted as an organic or functional lessening as it also means all forms of organic or functional surplus (having six fingers on one hand is a impairment) likely to hinder the normal function of an individual's faculties.

Disabilities

In the field of health, a disability corresponds to any partial or total reduction (resulting from a impairment) in the capacity to carry out an activity in the manner or in the limits considered normal for a human being (WHO, 1980).

Within the subject of disability, a distinction can be made between *functional limitations* and *activity restrictions* if one chooses to distinguish between the realization of *actions* and the realization of *activities*, which the WHO does not do.

Note that disability is wrongly interpreted only as an "impossibility or a difficulty in realizing an *activity*," whereas all forms of inappropriate, exaggerated, or untimely realization of an *activity* outside the normal must be considered.

Handicaps

"In the field of health, the handicap (social disadvantage) for a given individual results from an impairment or a disability which limits or prevents the accomplishment of a normal role (consistent with age, sex, social and cultural factors)" (WHO, 1980). Such a *handicap* is assessed on the basis of a judgment, which depends on the *performance* of *activities* making up the role studied.

2. Functional Limitation and Activity Restriction, REVES Classification System

Functional Limitation

Depending on the *degree of severity,* functional limitation is impossibility or difficulty in performing an *action.* The measurement of functional limitation is complicated, as individuals generally have daily *activities* in common (eating, personal hygiene, housework, etc.), but they are not always challenged by the same *actions* in the course of these *activities* (e.g., housework does not necessarily mean climbing stairs, reaching up, bending down, etc., for everyone). It thus becomes difficult to obtain information on the *performances* of the subject; to reveal a specific functional limitation, it is preferable to obtain information either on the *aptitude* in performing a defined *action* or the cause of the difficulty encountered when faced with a particular *activity,* generally including the realization of the *action* being studied.

Activity Restriction

This represents impossibility or difficulty, depending on the *degree of severity* in performing an *activity.* It is measured in terms of *performances* when the *activity* is considered to be an unavoidable component of a *role* studied. For example, all subjects whose *social role* is housekeeping will be questioned on their ability to wash the dishes, do the housework, and do the laundry.

Finally, a distinction can be made between the difficulty or impossibility, depending on the *degree of severity,* of carrying out a gesture, an action, an activity, or a role. When limited to the field of locomotion, which is more illustrative, a correspondence table grading these different types of difficulty can be drawn up. A difficulty in performing a *gesture* is an *impairment.* A difficulty in performing an *action* a *functional limitation.* A difficulty in performing an *activity* corresponds to an *activity restriction.* A difficulty in performing a *role* corresponds to a *handicap.*

3. Activities of Daily Living, Basic or Instrumental, Proposed by Katz *et al.* (1963) or Lawton and Brody (1969)

Personal Care Activities

These are a set of *activities* considered to be indispensable for ensuring *minimal independency* such as washing and eating. They correspond to *activities* that the subject must do for himself, with or without *help*, and that nobody else can do for him. From this definition, Sydney Katz *et al.* (1963) elaborated a model bringing together six activities of daily living considered to be common to all with the aim of *minimal independency*: feeding, bathing, going to toilet, dressing, transfer (moving in and out of bed or chair), continence. This forms the model of reference for many studies.

Household Care Activities

These are all the *activities* representing the *ability to live alone in a private home.* They include *household care activities* known as instrumental activities of daily living such as filling in forms, shopping, or cooking. They correspond to *activities* normally carried out by the subject in daily life, with or without help, but that may possibly be carried out by someone else. The model of reference was proposed by Powell Lawton and Eliane Brody (1969).

4. International Classification of Functioning, Disability, and Health (ICF, WHO, 2001)

In the context of a health condition:

Body functions are the physiological functions of body systems (including psychological functions).
Body structures are anatomic parts of the body such as organs, limbs, and their components.
Impairments are problems in body function or structure such as a significant deviation or loss.
Activity is the execution of a task or action by an individual.
Activity limitations are difficulties an individual may have in the performance of activities.
Participation is involvement in a life situation.
Participation restrictions are problems an individual may experience in involvement in life situations.
Environmental factors make up the physical, social, and attitudinal environment in which people live and conduct their lives

Sullivan's Method

Principle

The years lived by the population in a life table are separated into years lived with or without disability. The rates of institutionalization (which generally come from population censuses) and the short-term or long-term disability prevalence rates (which generally come from health or disability surveys) can be used for this purpose. Once the table is modified, life expectancies with or without disability are calculated in the usual way. Thus, by including only the years lived without disability, a life expectancy without disability is obtained, or conversely, by including only the years lived with disability, a life expectancy with disability is obtained.

Calculation

Using the survivors (b) from a life table (cf. table below), the number of years lived between each age (c) is calculated. The prevalence rates of disability (d) can then be used to calculate the number of years lived with disability. By simple subtraction, the number of years lived without disability (e) is obtained. These years are then added together (f) from any age x and divided by the total number of survivors at this age (b) to obtain the life expectancy without disability (g).

Example

The total number of years lived without disability from 65 years of age is 1,153,013.2 in the table. This total is divided by the number of survivors at 65 years old to estimate life expectancy without disability at this age: 1,153,013.2 (f) divided by 89,347 (b), i.e., 12.9 years (g).

For more information, see Jagger (1997).

TABLE 80–A Disability-Free Life Expectancy According to Sullivan's Method, France, 1991, Female (Simplified Estimation Calculated from Long-Term Disability)

Age (x) (a)	Survivors (Sx) (b)	Years lived between x and $x + a$ (c)	Prevalence of disability between x and $x + a$ (d)	Years lived without disability between x and $x + a$ (e)	Years lived without disability from x (f)	Long term DFLE from x (g)
0	100,000	496,176.5	0.0097	491,366.7	7,075,234.3	70.8
5	99,242	496,287.5	0.0242	484,295.5	6,583,867.6	66.3
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
65	89,347	436,686.7	0.1885	354,389.6	1,153,013.2	12.9
70	84,952	408,481.5	0.2740	296,546.1	798,623.6	9.4
75	78,000	363,545.5	0.3455	237,955.9	502,077.5	6.4
80	66,522	290,185.4	0.4675	154,519.8	264,121.6	4.0
85	48,434	297,869.1	0.6320	109,601.8	109,601.8	2.3

DFLE, disability-free life expectancy.
Data from Robine and Mormiche, 1994.

CHAPTER

81

Demographic Trends and Relationships between Generations

ANTONIO GOLINI AND RAFFAELLA IACOUCCI

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION: A FEW DEFINITIONS

Almost every discipline has its own definition of the word "generation." Jurists, economists, sociologists, historians, teachers, politicians, and even technocrats have their own understanding of the word, which depends on each discipline's interpretation of how individuals, events, or objects succeed each other in time. Thus, in political science, a group of persons who went through major experiences together are referred to as a generation, such as postwar generations or the 1968 generation. In the field of computer science, a recent discipline, the word generation refers to each of the successive stages of technological development corresponding to the creation of ever more sophisticated products; computer scientists speak of second-generation or third-generation computers.

Demographers use two different words, depending on whether a person's birth is considered in the context of the general population (cohort) or in the family context (generation). The two notions, however, are closely linked both structurally and conceptually, and for this reason, they cannot be dissociated.

Indeed, on one hand, when dealing with all the individuals that make up a population, demographers consider that all persons born in the same year represent a birth cohort (sometimes the group includes persons belonging to several cohorts, for instance over a 5-year span) and study their demographic behavior; in this case, for example, they speak of the fertility and mortality of the cohort born in the year t. The notion of cohort is very frequently referred to in demographic analyses (see Chapters 6 to 13).

On the other hand, when population specialists deal with individuals within a family and study how each person is related to the rest of the family at the time of his or her birth, they use the word generation, which comes from the Latin *generare*, to give birth to, to describe all the members of a family who are equally related to a common ancestor. Thus, from generation to generation, a newborn baby can be linked to its most distant ancestor, and conversely, the chain can be linked going back the other way, from the ancestor down to the most recent generation: upward from the children to the parents, grandparents, great-great-grandparents, great-great-great-grandparents and downward to the grandchildren great-grandchildren, etc.

In this sense, it is interesting to determine how many generations coexist at a given moment. The number of coexisting generations obviously depends on (1) biological or voluntary infertility (the more childless women there are, the more cases in which the generation chain is broken); (2) the distance between generations at the time of birth (the older the average age of the mother and the father at the time of childbirth, the lower the probability, all other things being equal, of observing a large number of coexisting generations); and (3) mortality conditions (the higher the life expectancy, the higher, all things being equal, the probability of having a large number of coexisting generations).

In most European countries today, we observe a situation of low fertility, an increase in the distance between generations and low mortality. Developing countries may soon follow, even though such a prospect seems quite remote for most. European countries, however, may be considered as precursors in dealing with problems tied to the aging of the population, in particular those that will be discussed in this chapter. Even though our description of these problems is limited to Europe, it can be used as reference by "young" countries when their turn comes to face the same type of demographic trends and determine the number of coexisting generations.

However, from a social, familial, and individual point of view, the important point is not only the *number* of coexisting generations but also, and especially, the *percentage* of the population that finds itself in a situation of coexistence with several generations and the *duration* of the period of coexistence; we may also consider the number of *years of life* spent by each individual in the position[1] of child, grandchild, parent, grandparent, or other positions. In the past, families in which four generations coexisted for a short period of time were rare, and cases of five coexisting generations were extremely rare, limited to families in which people had children young and lived very long. Today, cases of five coexisting generations remain quite rare, but cases of four coexisting generations have become relatively common. However, the proportion of persons coexisting at a given age with living relatives going back one, two, and three generations and the duration of this coexistence have radically changed and will continue to change because elderly persons, especially women, are living longer. Indeed, at the age of 65 or 70 years, some women are still the child of a living parent, something that recently was hardly conceivable.

This chapter deals essentially with the notion of generation. We will describe mainly the dramatic changes in the quantitative balance (within a population or a family) among living children, parents, grandparents and great-grandparents, as well as between adult children and elderly parents. We will then describe a few typical traits pertaining to psychological, cultural, social, or economic relationships between the different generations. Last, we will address the question of transmission, as well as the various aspects linked to changes in the distance between generations: these changes have had an impact well beyond simple material inheritance and age difference. In each case, our angle of analysis will be essentially demographic.

I. SEEING THE GENERATIONS

Figure 81–1 shows the situation of four successive generations: their life courses, the system of relations linking them within the general population (in relationship to their cohorts) or within the family and linking them to periods when demographic events are likely to occur (first union, marriage, childbirth, divorce, migration, chronic degenerative diseases or disabilities, death). The four diagonal stripes correspond to both cohorts and generations, according to the definitions given in the Introduction, whereas the number of diagonal stripes crossing the vertical columns corresponds to the number of generations coexisting during a given calendar year or a given period of time.

This diachronic and synchronic study shows the complexity of the links among generations, age, and period, in all respects: economic (both micro- and macroeconomic), educational, professional (aspects tied to social advancement), or even those concerning legislation, norms, etc.

Indeed, the number of events in each parallelogram of the Lexis diagram (i.e., the number of marriages, births, divorces, migrations, and deaths occurring within a given cohort during a calendar year and thus at a given age) depends not only on the conditions under which a given cohort was born and grew up—on the social, economic, political, and health context of that year or brief period (and especially, as concerns migrations and divorces, on changes in the current legislation)—but also on the person's age at the time and the type of support (family or social) she or he received. One example of the link between context and age, in Italy, is that of the generation of women born in 1950, whose formative years (18 to 20) took place in 1968 to 1970 in a very specific political and ideological climate; this combination occurred once again in 1974, when they turned 24, at the time of the first large-scale referendum on divorce. The events of 1968 and 1974 obviously had a different impact on women born in 1964, who were only 4 in 1968 and 10 in 1974.

These examples show the complexity—highlighted several times in this treatise[2] and widely researched in the literature—not only of the links among age, generation, and period but also of interrelations between generations; it seems that the instruments developed

[1] This position exists when both persons concerned in the family relationship (parent and child, grandparent and grandchild) are alive.

[2] In addition to Chapters 6 to 13 of Volume I, which have already been referred to, see also Chapters 17 and 18 (also in Volume I).

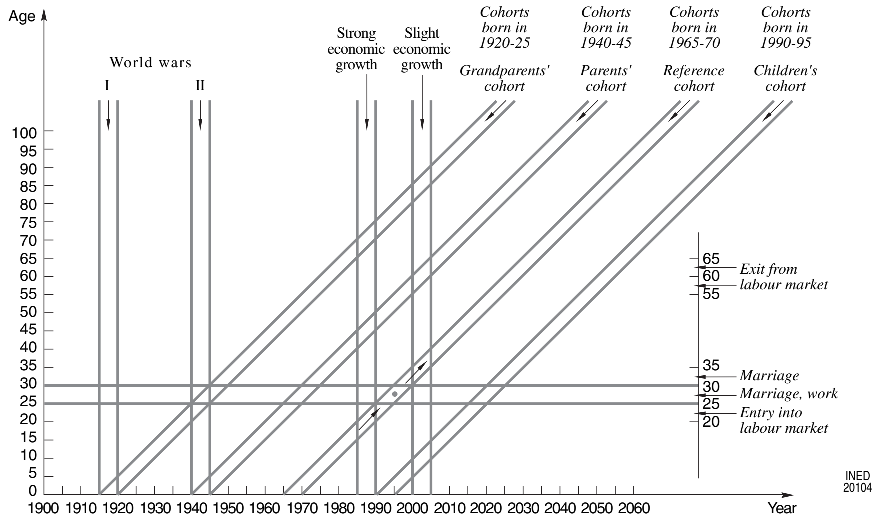


FIGURE 81–1 Relationships between age, cohort, and period in terms of population or family. (Adapted from Golini, 1992:779.)

through the APC (age/period/cohort) method of analysis are not sufficient to fully describe and explain this complexity. For example, the point drawn in Fig. 81–1 represents a woman belonging to the group of cohorts born in 1965–1970, and who reached the age of 25 to 30 years around 1995. The demographic behavior of this woman depends of course on her age but also on the period when she was 25 to 30 years old and on her own past as well as that of her cohort. These factors will all influence her future life in the next 10 to 20 years. But there are other factors as well, such as the coexistence of her parents or grandparents, along with the intense influence of the *generation of remembrance*.[3] This underscores the importance of how this woman experienced her formative years in her family of origin, how she perceived and internalized her remembrances of her parents' relationship as a fertile couple, and how she as an adult perceives her parents' marital life. A kind of parental or grandparental legacy thus emerges, a nonmaterial, emotional legacy taht carries its own specific risks, such as divorce or illness.

II. THE COEXISTENCE OF GENERATIONS IN THE FAMILY

It is not easy to determine the number of persons who, at different ages, still have at least one living grandparent, and how many grandparents, on average, are still alive. The same is true for the number of adults or elderly people who have grandchildren. Different studies have yielded estimates on these situations, which are the basis of significant (and widely researched) relationships between grandparents and grandchildren. Our main, though not exclusive reference is an Italian survey conducted by National Institute of Statistics (ISTAT).

1. Grandparents and Grandchildren

Figure 81–2 shows that the number of living grandparents varies considerably according to the age of the person and very little according to the respondent's sex. It can be seen that 97.8% of Italian children under 15 years have at least 1 living grandparent, and on average, they have 3 living grandparents; in other words, for 100 children, there are 293 grandparents; compared with the maximum allowed of 400 grandparents, this figure is very high. For children aged 15 to 25 years, the proportion and the average number remain high, respectively 82.8% and 2.2, representing a total of 182 grandparents for 100 children. For those aged 25 to 34 years, the number drops to 68 grandparents for 100 persons, and only 6 grandparents for 100 persons aged 35 to 69 years.[4]

If we change our perspective, we can see in Fig. 81–3, for Italy, how many adults or elderly persons

[3] I would like to thank Chiara Saraceno for having suggested this evocative phrase.

[4] It is difficult to understand why, in order to analyze a phenomenon such as relationships between grandparents and grandchildren, the National Institute of Statistics constructed its tables on the basis of such a large and inadequate age group (35 to 69 years).

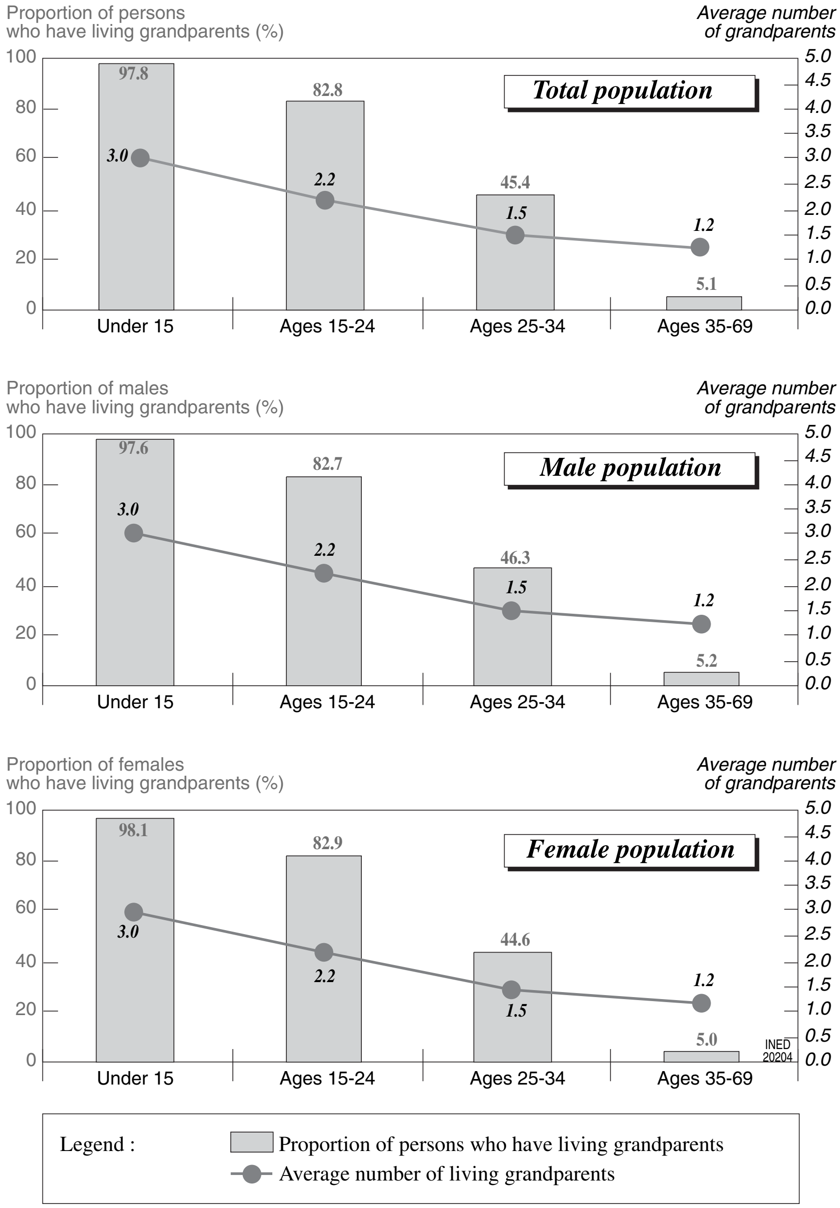


FIGURE 81–2 Proportions (%) of persons who have living grandparents and average number of grandparents, according to sex and age group up to 69 years in Italy, 1998. The average number is calculated on the basis of persons who have at least one living grandparent. (Data from ISTAT, 2001.)

have grandchildren and how many they have on average. One hundred years ago, the shortness of life spans made it quite unlikely that a person would live long enough to see all his/her grandchildren and especially long enough to see them grow up, whereas today it has become quite frequent. However, in the past, grandparents who were fortunate enough to survive long enough to see their grandchildren had a large number of them, whereas today grandparents are numerous but the number of grandchildren has diminished, with many already teenagers or adults. As a result, the grandchild/grandparent relationship is often more intense, and grandparents often find themselves playing an important role in the daily lives of not only their grandchildren when they are young children but their grown-up grandchildren as well.

Among persons aged 65 years and over, only 70.8% have at least one grandchild, and those who do have grandchildren have an average number of four grand-

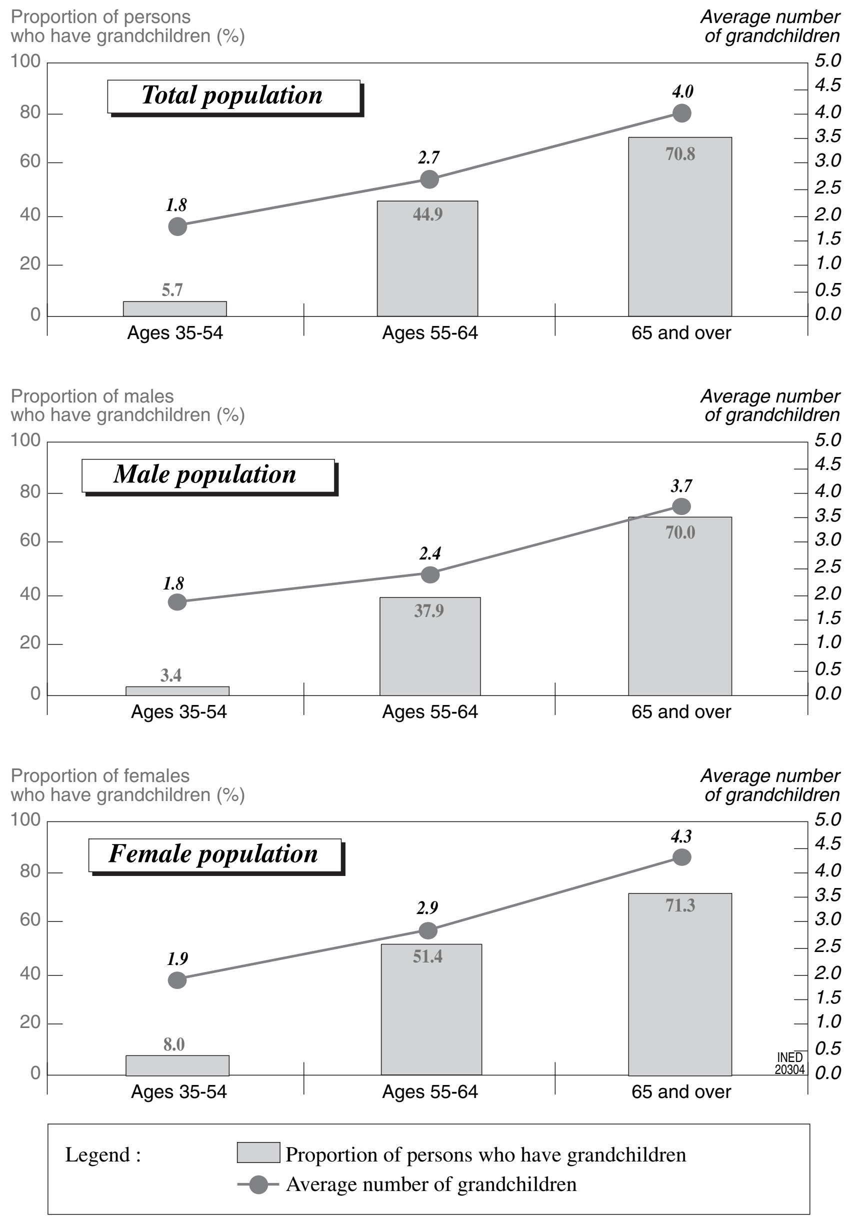


FIGURE 81–3 Proportions (%) of persons aged 35 years and over who have grandchildren and average number of grandchildren according to sex and age group: Italy, 1998. The average number is calculated on the basis of persons who have at least one living grandparent. (Data from ISTAT, 2001.)

children. This number varies considerably according to sex.[5] For 100 men aged over 65 years, there are 259 grandchildren (70.0% of the men of that age group are grandfathers, and each has on average 3.7 grandchildren), whereas for 100 women there are 307 grandchildren, an additional 19% (71.3% of the women of that age group are grandmothers, and they have on average 4.3 grandchildren). The fact that the average age of women over 69 years is much higher than that of men of the same age group increases the probabil-

[5] It is hardly necessary to mention that there is no perfect symmetry between the relationships linking grandchildren to grandparents, or young children to elderly adults, and those linking grandparents to grandchildren or elderly adults to young children. Indeed, although a child necessarily has two parents and four grandparents (not necessarily living), an elderly adult does not necessarily have children, and if she or he does, these in turn do not necessarily have children either. This explains the difference between the very high proportion of children under 15 years who have a living grandparent (97.8%) and the lower proportion of persons over 65 years who have grandchildren (70.8%). Of course, this reasoning only concerns biological relationships between generations.

ity that the average number of grandchildren will be higher than that observed for grandfathers. But one must also underscore the fact that about 30% of persons over 65 years do not have living grandchildren.[6] In these persons' families, the generational chain has been broken off and the elderly have been deprived of an entire category of descendents (grandchildren) to whom they could have transmitted an inheritance and on whom they could have relied for support.[7] The proportion of persons who do not have grandchildren is naturally much higher among persons aged 55 to 64 years, an age at which nearly 50% of the women do not yet have any grandchildren: this high proportion is owing in part to the age of the person concerned and to the increase in the average distance between generations or in the average number of children who decide not to have children themselves, an increasingly frequent phenomenon.[8]

According to a French survey conducted in 1999, the number of grandparents has reached 12.6 million, of whom 2 million are great-grandparents (Cassan *et al.*, 2001). At the age of 56 years, one out of two persons has at least one grandchild; at 66 years, three out of four. Among persons aged 70 years, 20% have no grandchildren, a lower proportion than that observed in Italy: in France, intergenerational relationships with descendants, down to grandchildren, are more frequent, because fertility is much higher in France than in Italy. In France, the age at which people become grandparents has increased over the years: the proportion of grandparents aged 50 years is about 35% for the cohorts born in 1930–1935 and 25% for those born in 1940–1945. At the time of the survey, grandparents had on average a little over four grandchildren, but the number of grandchildren depends on the number of children: persons who have one child have barely two grandchildren on average, whereas those who have three children have on average seven grandchildren. In addition, 85% of the women aged 85 years who have more than three children have great-grandchildren.

Most often grandparents and grandchildren enjoy close contacts and relationships. Today, grandparents usually do not live in the same house as their grandchildren; nonetheless, about 10% of grandparents share the home of at least one of their grandchildren, and the frequency increases with the age of the grandparents, especially grandmothers (ISTAT, 1999). According to a 1999 American survey, 5.5% of grandchildren were living with their grandparents; this proportion has been increasing constantly. Indeed, in 1970 the proportion was only 3.2% (Bryson and Camper, 1999). From 1992 to 1997, the most sizeable increase has been that of grandchildren living with their grandparents without either of their parents: this increase has been attributed to growing drug abuse among parents, to the increase in teenage pregnancies, to the increase in divorce rates and in the number of single-parent families, and to mental and physical diseases, AIDS, delinquency, and incarcerated parents, sexual abuse of minors, etc. (Fuller-Thomson *et al.*, 1997)

In any case, in Italy, physical proximity is quite common. Over 44% of grandparents have at least one of their grandchildren and his/her parents living within a radius of less than 1 km. This nearness encourages close relationships: 84.2% of grandparents take care of their grandchildren (81.2 of grandfathers and 87.0 of grandmothers) in one way or another, and in 29.8 of cases, they do so when the parents are at work (ISTAT, 1999). The educational potential of grandparents who can positively contribute to the development of their grandchildren's personal identity should also be mentioned. In addition, grandparents participate financially: in Italy, in 1999, their contribution to the small and larger expenses of their adult children, sons and daughters-in law, reached the equivalent of the impressive amount of 82.6 billion euros, or 15% of the total annual expenses of Italian families (Più and Fenacom, 2001).

2. The Increasing Probability of Having Living Parents and Grandparents

Of course, the probability of having living parents and grandparents varies according to the age of the person of reference and changes over time and depends on the degree of relationship (parents, grandparents, great-grandparents). In the absence of direct observations yielding information on the subject, we may try to evaluate variations through life tables per-

[6] It is difficult to assess how many grandparents have lost all their grandchildren and thus on a given survey date have ceased to be grandparents. The probability is likely to be quite low, unless they were grandparents of an only grandchild. The observations made in this article nevertheless remain valid, even in these cases.

[7] One must take into account the fact that in the 35- to 69-year age group, the proportion of persons who do not yet have grandchildren can be quite high, and all the more so if there is among them a high proportion of women who had a child after the age of 35. An National Institute of Statistics publication (1999) provides additional details, in particular the fact that the proportion of persons aged at least 75 years who do not have grandchildren is about 25%, a proportion that is not likely to decrease as the respondents grow older.

[8] If, from one generation to the next, the average age at first childbirth for women were constant (27 years old) and if the proportion of women without children also were constant (at about 15%), it is obvious that by the age of 55 years, one would expect the proportion of women who have grandchildren to reach 70% instead of the 51% resulting from the data provided in Fig. 81–3.

TABLE 81–1 Proportions (%) of Females Who Still Have a Living Mother, Grandmother, or Great-Grandmother, According to Italian Life Tables of 1911, 1951, 1990, and 1998

Age of (in years)		Life table			
The female	The relative	1911	1951	1990	1998
Mother					
0	27	100.0	100.0	100.0	100.0
5	32	96.6	99.2	99.8	99.8
10	37	93.2	98.3	99.5	99.6
20	47	86.1	95.5	98.4	98.6
30	57	76.8	89.7	95.7	96.3
40	67	59.1	76.6	89.2	90.9
50	77	27.4	46.5	72.2	76.6
60	87	3.3	10.7	33.5	41.2
70	97	0.0	0.0	3.1	5.6
Grandmother					
0	54	80.1	91.1	96.8	97.2
5	59	74.1	87.9	94.8	95.6
10	64	65.8	81.8	91.8	93.1
20	74	37.6	57.6	79.1	82.5
30	84	7.6	19.4	46.9	54.1
40	94	0.0	1.3	8.3	13.0
Great-grandmother					
0	81	14.7	30.4	59.4	66.0
5	86	4.5	13.1	38.0	45.5
10	91	0.0	3.7	17.2	23.8

Data from the following: until 1990, Golini and Silvestrini, 1995; for 1998–1999: ISTAT, 2001.

taining to different periods for a given average age at childbirth. The results of this computation are provided in Table 81–1 and illustrated in Fig. 81–4.[9]

Even though the evaluations of Table 81–1 are based on approximations, they are very similar in terms of quantities to those of the graphs of Fig. 81–2, which are based on real data. Indeed, these show that in 1998, 98.1% of children aged 0 to 15 years had at least one living grandparent, whereas the proportion for children aged 5 to 10 years calculated on the basis of the life tables of 1998, is of 93 to 96%.

The results of Table 81–1, which are thus relatively reliable, may be commented as follows.

1. *A phenomenal increase in the duration of the period spent in the position of child.* According to the 1911 life table, at the age of 50 years, only one in four persons still had a living mother, whereas in the 1998 life table, three out of four were in that situation. The proportions are even more spectacular if one considers women aged 60 years: only one out of 30 women still had her mother in 1911, whereas according to the most recent life table, one out of two at that age still have their mother. This means that in 50% of the cases, the person caring for a very elderly person (the mother of a 60-year-old is about 87 years) is him/herself elderly.[10] Last, in 1998, one in five Italian women died at over 90 years old, and it is most likely that they were taken care of by daughters over 65 years.

2. *A very high increase in the number of living great-grandmothers and, as a result, frequent coexistence of four generations.* Whereas in 1911, one in 20 little girls aged 5 years had a living great-grandmother, in 1998 the proportion was about one in two. Regardless of emotional and relational considerations, the coexistence of four generations, with the simultaneous presence in the family of a child of 5 years and great-grandmother aged 86 years, can clearly pose problems of organization: if the child's mother, age 32, is working and needs the grandmother's help, the latter will have to take care not only for her own mother but also for her daughter's daughter. Thus, women aged 55 to 59 years are often described as the "sandwich generation" because they are torn between the need to care for their mothers and to help their daughters by caring for their grandchildren. This situation probably partly explains the low fertility of Italians, because it is difficult for daughters to ask their mothers for extra help with a second child.

3. *The disappearance of orphanages.* In 1911, about 3.5% of girls aged 5 years had lost their mothers; in 1998, the proportion was only 0.2%. For girls aged 10 years, the proportions were, respectively, 6.8% (one out of 15 girls) and 0.4% (one out of 250). Biological orphans practically disappeared during the last century. It is true that they have been replaced by "social" orphans, that is, children of divorced or separated parents, but these children are often brought up by their grandparents, as shown in the above-mentioned American survey. Thus, because of the decline in the number of orphans and the simultaneous increase in the number of living grandparents, orphanages have practically disappeared from today's social landscape.

In countries where fertility and mortality are low, the age structure of multigenerational families, in which four generations often coexist, looks like a column with a swelling in the middle: very few young

[9] The results given in Table 81–1 and Fig. 81–4 were obtained by applying a method suggested in 1986 by Noreen Goldman, which has already been used in previous research (Golini and Silvestrini, 1995) and has been updated. The exercise consists in applying period life tables with a constant interval between generations.

[10] Indeed, one must take into account the fact that a 60-year-old woman has a husband aged 63 years on average. If we look at the health situation of a couple of that age, we find, according to the last National Institute of Statistics health survey, that 66% of the women and 62% of the men report suffering from at least one chronic disease.

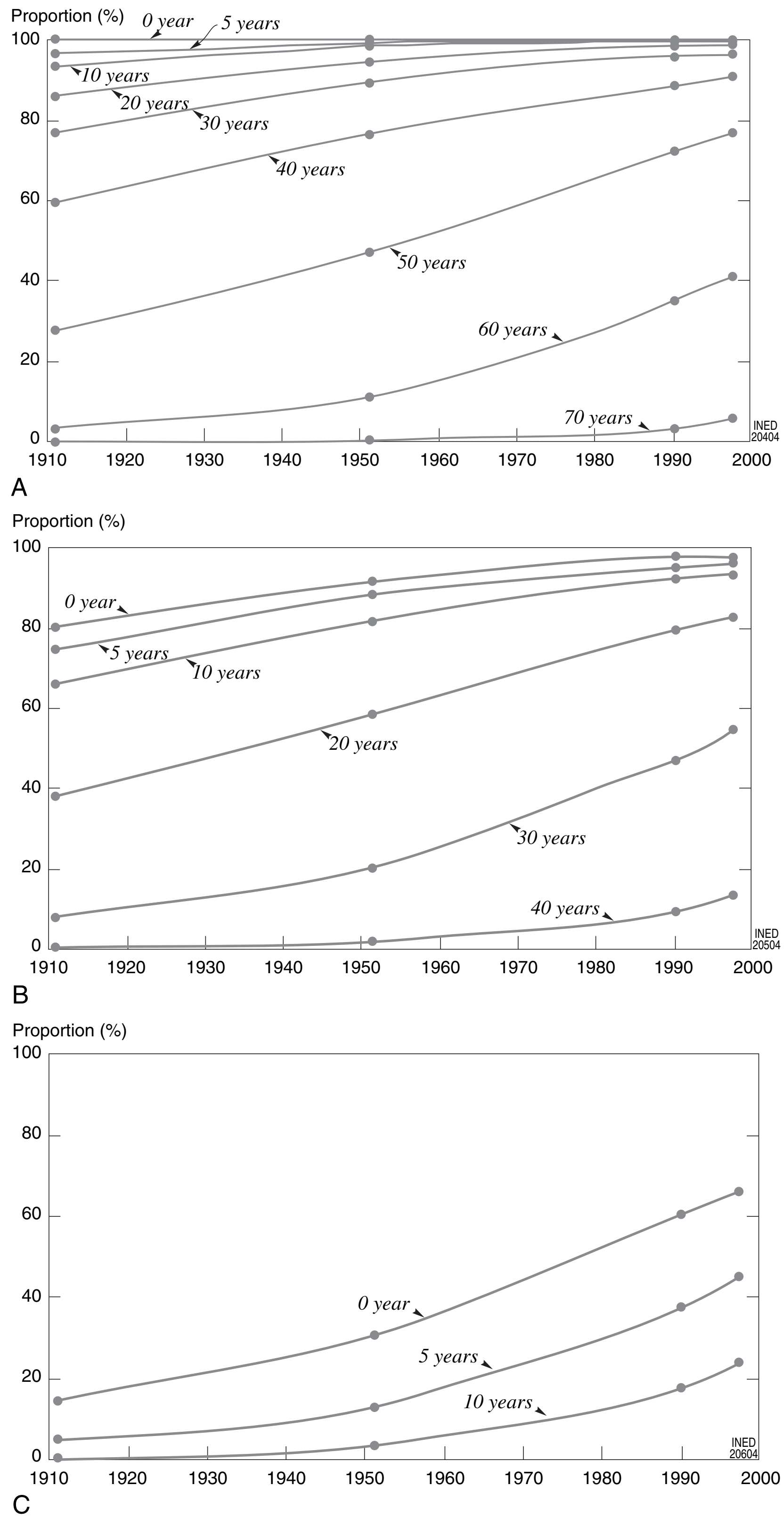


FIGURE 81–4 **(A)** Proportions (%) of females aged 0 to 70 years whose mother, who is 27 years older, is still alive, based on Italian life tables of 1911, 1951, 1990, and 1998–1999. **(B)** Proportions (%) of females aged 0 to 40 years whose grandmother, who is 54 years older, is still alive, based on Italian life tables of 1911, 1951, 1990, and 1998–1999. **(C)** Proportions (%) of children aged 0 to 5 years whose great-grandmother, who is 81 years older, is still alive, based on Italian life tables of 1911, 1951, 1990, and 1998–1999. (Data from the following: until 1990, Golini and Silvestrini, 1995; for 1998–1999 ISTAT, 2001.)

TABLE 81–2A Estimate of the Average Number of Living Relatives of an Italian Woman Aged 80 Years Old Who Has Been Married, According to Date of Birth

Year of birth (year of observation)	Husband, brothers, and brothers or sisters-in-law	Children and sons and daughters-in law	Nephews and their partners	Grandchildren
1900 (1980)	2.3	5.4	25.8	6.5
1930 (2010)	3.4	4.4	16.8	3.9
1960 (2040)	3.7	3.4	9.1	3.1

TABLE 81–2B Estimate of the Average Number of Living Relatives of a Yound Italian Aged 20 Years, According to Date of Birth

Year of birth (year of observation)	Brothers	Cousins	Parents	Uncles and aunts (including those by marriage)	Grandparents
1930 (1950)	1.8	7.1	1.7	9.8	0.8
1960 (1980)	1.4	5.8	1.8	8.5	1.2
1990 (2010)	0.8	3.7	1.9	7.2	1.9

Data from Barbagli *et al.*, 2003:201.

children at the bottom, very few great-grandparents at the top, and a large number of adults, parents, and grandparents (biological or adoptive) in the middle. The result are *family networks* that are smaller and more complex, both because of the increase in the number of coexisting generations and because of the increasing frequency of divorces, separations, cohabitations, and second marriages. This has led to an increase in the number and diversity of family relationships, within networks that have become more extended and less direct.

3. Transformations in Kinship Systems

More generally speaking, the entire kinship system has been deeply transformed by the increase in life expectancy and the fertility decline. Within the space of only 15 years, from 1983 to 1998, the proportion of families with at least one child dropped from 39% to 25%, whereas the proportion of families with at least one very elderly relative rose from 28% to 35% (Barbagli *et al*., 2003). The data provided in Table 81–2 reflect the profound changes that have occurred in the size and structure of kinship systems.

The ratio of elderly parents to adult children has drastically changed in the past 60 years. In 1980, for 100 elderly persons, there were 163 adult children; in 2010, the ratio may drop to 100 for 100; in 2040, for 100 elderly persons, there may be only 72 adult relatives (Barbagli *et al*., 2003). Over the medium and long term, a very difficult phase seems to be in the making, because there will be many elderly persons without adult children or without enough adult children they can rely on for support (Golini, 1999).

Economically advanced countries will thus soon have to deal with the fact that because of this disproportionate ratio between young people and adults, families will no longer be able to function as a support system for their elderly members and will not be able to care for them consistently. It is true that technology may be able to offset part of this problem, because new systems are being developed, for example, the use of microscopic cordless sensors to watch over elderly persons and young children. These systems also have applications in the field of the environment and traffic regulation, and these new developments are opening new avenues for exploration that may have many social repercussions.[11]

Aging and the vertical extension of the family are an object of concern for not only the elderly, and the issue of who will care for them, but also young children and young people. In countries with very low fertility, because of the couple's wishes, as in Italy, or because of authoritarian policies, as in China, the number of only children will rise and each child, or nearly, will be surrounded by numerous adults or elderly persons full of concern for him/her and eager to satisfy every one of

[11] This view was provided by Alberto Sangiovanni-Vicentelli, technologist and professor at the University of California at Berkeley, in an interview published in the June 13, 2003, issue of *Il Sole-24 ore* (p. 11).

her or his physical and psychological needs. One may wonder about the personality structure of these "little emperors" growing up in such an emotional and psychological environment. Only children often receive more attention from their parents, but this can sometimes have a negative impact: when this special attention is measured, it can help the child develop self-confidence, a sense of security, and wide-ranging interests, thus resulting in deep bonds between parents and child. If this attention is excessive and not well directed, the child may feel oppressed, become too dependent and incapable of facing individual responsibilities and difficulties later on in life, and end up rejecting his or her parents (Golini, 2003:141).

4. The Pattern of the Succession of Generations Has Changed

The growing imbalance between generations can be seen in an analysis that compares persons aged 20 to 24 years, 45 to 49 years, and 70 to 74 years (Figs. 81–5—81–8), over the span of a century (1950–2050), in Italy and France.[12] The aim of this research was to study how combined trends of births, deaths, and migrations have modified and will modify the relationship between three generations separated by an average of 25 years, as well as relationships between grandchildren and grandparents in a country with long-standing very low fertility, such as Italy, and in a country with low fertility, such as France. Naturally, the total population of a certain age group, separated from the previous and the next age groups by 25 years each way, cannot be considered as a generation in the strict sense of term as we have used it so far (in the sense of grandparents in relation to children and grandchildren). As concerns family categories, however, the historical series illustrated in Figs. 81–5 and 81–6 may be considered as approximations; to make things simpler, we will refer in the following comments to children, parents, and grandparents.

1. The long-term trend shows the successive birth peaks separated each time by 25 years. For example, in Italy, in 1990 we observe a population peak at ages 20 to 24 years, which shifts to 45 to 49 years in 2015 and 70 to 74 years in 2040. Naturally, this succession changes the relationship between the three age groups under consideration, not only from a numerical point of view but also from an economic and social point of view. The 45- to 49-year age group, which in 1950 represented 2.9 million persons, should revert to 2.6 million in 2050 after peaking at 4.7 million in 2015 (15 years later, this peak will cause a crisis in the pension system). In France, where births follow a more regular pattern, no such peaks are observed.

2. The ratio of adult children to elderly parents (45 to 49 years/70 to 74 years) dropped from 2.6 in 1950–1955 to 1.4 in 2000–2005 and will fall to 0.7 in 2040–2045. At that time, the present system of aid and care based on family support will be completely ruined (also because the life span of elderly persons is expected to increase considerably). In France these three ratios are, respectively, 2.2, 1.7, and 1.0; the system will be under much less pressure than that in Italy.

3. The children/parent ratio is even less favorable if one looks at the following generations (persons aged 20 to 24 years and 45 to 49 years). For the three periods under consideration in comment 2 above, the three values are, respectively, for Italy: 1.4, 0.8, and 0.7; for France, the values are 1.0, 0.9, and 1.0.

4. Looking at the information given in comments 2 and 3 above, we observe that certain generations are presently or potentially in a position of disadvantage: in 2005, in Italy, for example, the ratio of the 45- to 49-year age group to its parents will be 1.4 to 1, whereas that of their children will be 0.8 to 1.

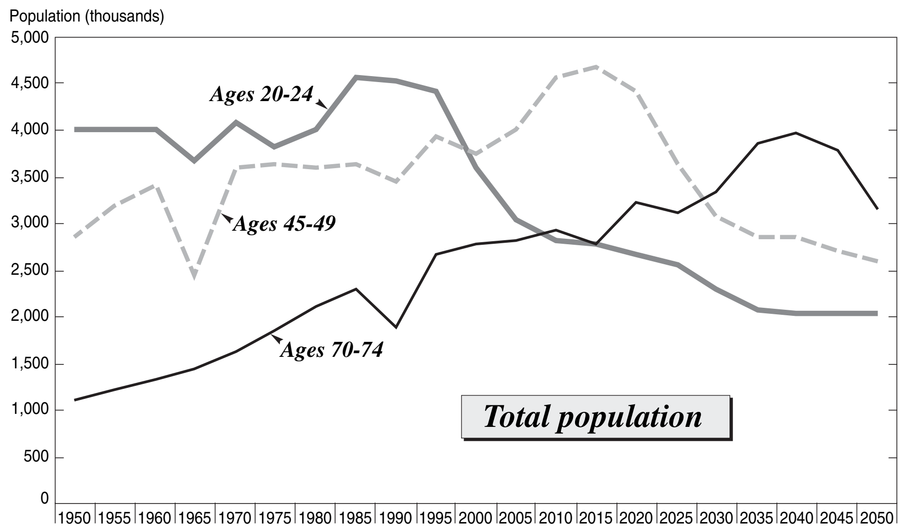
5. If the ratio of children aged 0 to 4 years to elderly persons aged 70 to 74 years is taken as a proxy of the ratio of grandchildren to grandparents, the number, in Italy, of grandchildren per grandparents would shift from 3.9 to 0.9 and 0.5. In France, the number of children and grandchildren has declined less sharply, and the ratio is expected to drop from 2.9 to 1.5 and 0.9.

III. FAMILY INHERITANCE AND RELATIONSHIPS BETWEEN GENERATIONS

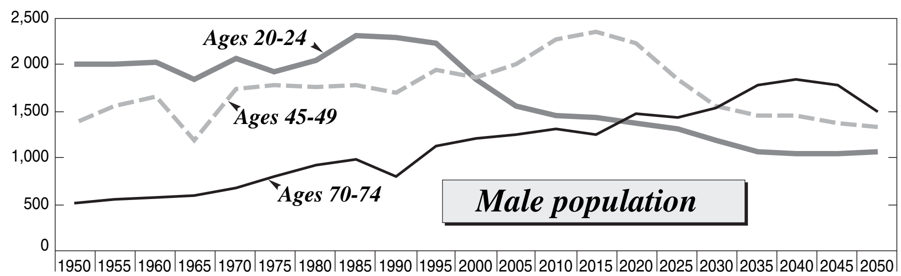
When speaking of inheritance and hereditary patrimony, we usually think in terms of either property, personal or real estate, accumulated by parents and grandparents during their life, or of physical or biological, genetically transmitted heredity. These subjects, which play an extremely important role in the relationship between generations, have already been extensively written about, and for this reason, we will only briefly mention them in this chapter. Moreover, in the field of genetic transmission, new evidence is being brought to light every day (see Chapters 79 and 96).

In addition to these types of heritage, there is another kind that is less direct and less obvious and is

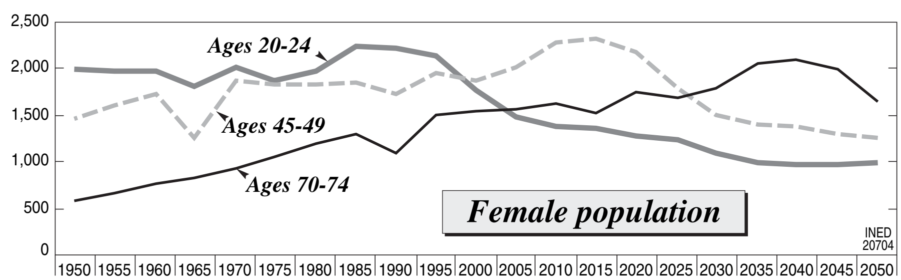
[12] All the data are taken from the United Nations Population Division (United Nations, 2003).



20-24	4027	3996	4025	3668	4097	3818	4043	4574	4530	4409	3633	3063	2839	2817	2676	2566	2294	2094	2033	2042	2075
45-49	2862	3195	3425	2471	3625	3643	3612	3657	3432	3940	3766	4020	4567	4693	4447	3661	3092	2868	2848	2711	2602
70-74	1119	1239	1364	1444	1625	1868	2133	2304	1908	2674	2775	2820	2950	2796	3233	3116	3356	3841	3979	3799	3155

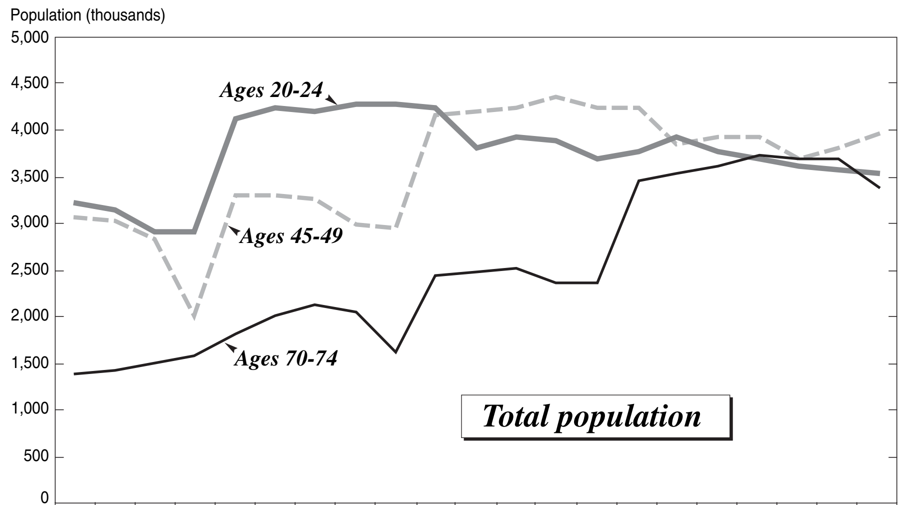


20-24	2011	2007	2030	1849	2077	1934	2044	2320	2301	2243	1851	1563	1451	1444	1377	1320	1180	1077	1046	1051	1068
45-49	1388	1568	1671	1187	1745	1788	1772	1796	1697	1954	1869	2001	2279	2355	2243	1850	1565	1455	1450	1385	1330
70-74	514.7	549	580.9	598.4	672.7	796.6	911.3	978.3	804	1137	1207	1238	1309	1252	1464	1420	1541	1775	1856	1787	1492

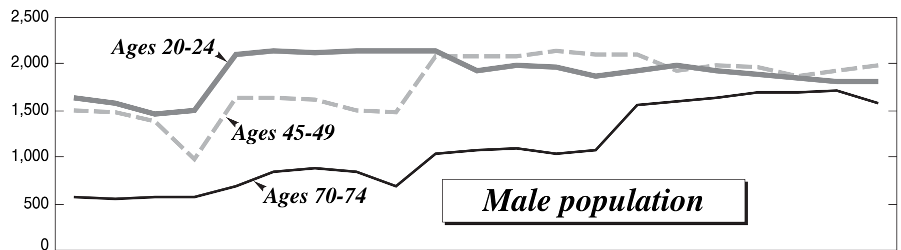


20-24	2 015	1 989	1 995	1 819	2 019	1 885	1 998	2 253	2 229	2 166	1 782	1 500	1 387	1 372	1 300	1 246	1 114	1 017	987	991	1 007
45-49	1 474	1 628	1 754	1 284	1 880	1 855	1 840	1 861	1 735	1 986	1 897	2 020	2 288	2 339	2 205	1 811	1 527	1 413	1 398	1 326	1 272
70-74	604	690	783	845	952	1 071	1 221	1 326	1 104	1 538	1 567	1 582	1 642	1 543	1 769	1 697	1 815	2 066	2 123	2 011	1 663

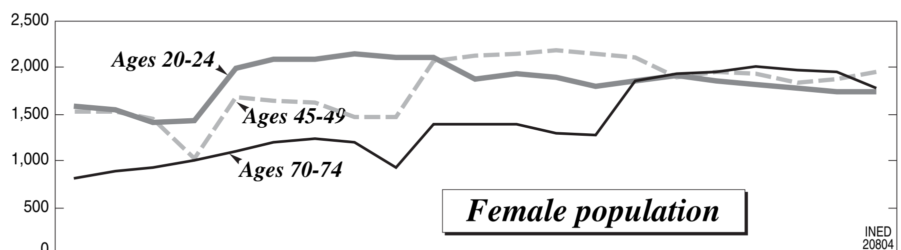
FIGURE 81–5 Sex-specific and age-specific population (in thousands), 1950–2050, Italy. (Data from United Nations, 2003.)



	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
20-24	3219	3133	2883	2912	4117	4247	4228	4305	4285	4240	3811	3923	3879	3673	3767	3909	3782	3689	3617	3565	3548
45-49	3041	3009	2842	1996	3310	3288	3243	2976	2949	4175	4223	4255	4343	4252	4232	3831	3945	3906	3707	3805	3951
70-74	1376	1439	1496	1578	1798	2016	2131	2034	1614	2440	2469	2488	2343	2356	3429	3536	3604	3712	3667	3676	3359

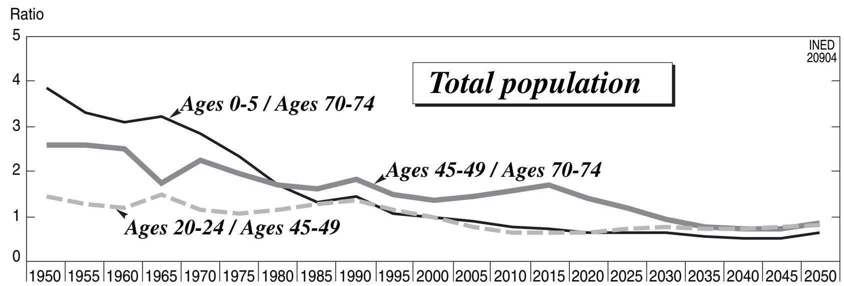


	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
20-24	1640	1593	1478	1500	2113	2160	2140	2164	2168	2145	1943	2004	1981	1876	1929	2002	1937	1890	1853	1826	1818
45-49	1514	1492	1399	984	1640	1645	1626	1502	1492	2101	2095	2097	2146	2116	2118	1934	1997	1978	1878	1933	2009
70-74	564.2	554	568	580	705.6	836.6	887.9	844.3	684.7	1042	1072	1096	1052	1078	1572	1609	1639	1702	1700	1722	1593

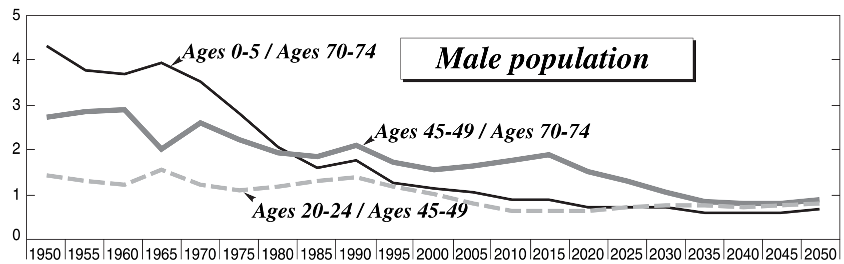


	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
20-24	1580	1540	1405	1412	2004	2087	2088	2141	2116	2095	1868	1919	1898	1797	1838	1907	1844	1799	1764	1738	1730
45-49	1526	1517	1443	1012	1670	1643	1617	1474	1457	2073	2128	2157	2196	2137	2114	1897	1948	1928	1829	1872	1942
70-74	811.8	885	928	998	1092	1180	1243	1190	929	1397	1397	1392	1291	1278	1857	1927	1965	2010	1967	1954	1765

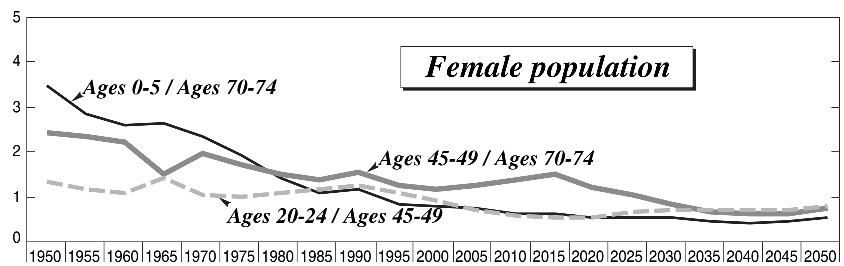
FIGURE 81–6 Sex-specific and age-specific population (in thousands), 1950–2050, France. (Data from United Nations, 2003.)



	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	2.558	2.58	2.51	1.712	2.231	1.95	1.693	1.587	1.799	1.473	1.357	1.426	1.548	1.679	1.376	1.175	0.921	0.747	0.716	0.714	0.825
0-5/70-74	3.867	3.277	3.072	3.206	2.824	2.312	1.696	1.313	1.439	1.03	0.947	0.893	0.762	0.733	0.616	0.641	0.604	0.527	0.498	0.509	0.599
20-24/45-49	1.407	1.251	1.175	1.484	1.13	1.048	1.119	1.251	1.32	1.119	0.965	0.762	0.621	0.6	0.602	0.701	0.742	0.73	0.714	0.753	0.797

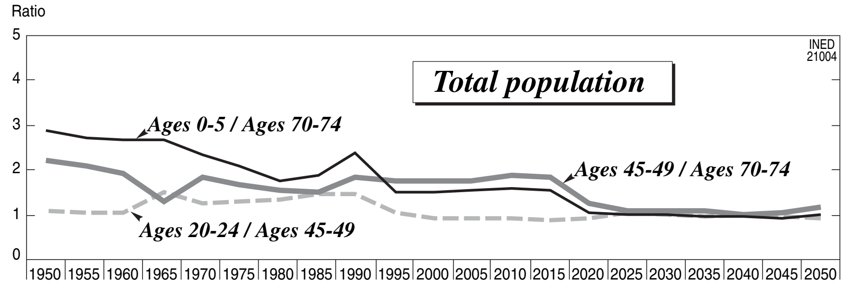


	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	2.697	2.856	2.877	1.984	2.594	2.244	1.944	1.836	2.11	1.719	1.548	1.616	1.742	1.88	1.532	1.303	1.016	0.82	0.781	0.775	0.891
0-5/70-74	4.305	3.777	3.689	3.958	3.495	2.784	2.038	1.588	1.757	1.245	1.121	1.048	0.885	0.843	0.7	0.725	0.678	0.588	0.55	0.557	0.652
20-24/45-49	1.449	1.28	1.215	1.557	1.19	1.082	1.154	1.292	1.356	1.148	0.99	0.781	0.637	0.613	0.614	0.714	0.754	0.74	0.721	0.759	0.803

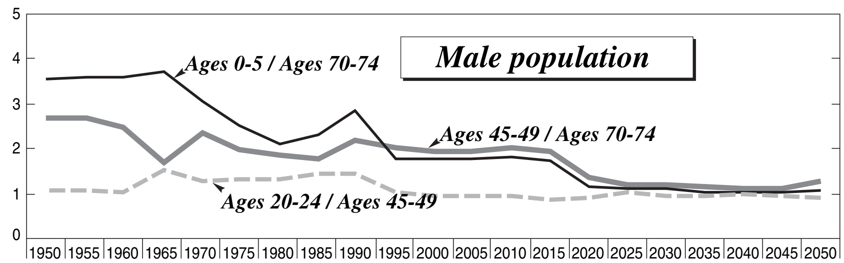


	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	2.44	2.36	2.238	1.519	1.975	1.731	1.506	1.404	1.572	1.292	1.21	1.276	1.394	1.515	1.246	1.067	0.841	0.684	0.659	0.66	0.765
0-5/70-74	3.493	2.879	2.616	2.673	2.35	1.961	1.44	1.11	1.208	0.871	0.813	0.772	0.664	0.644	0.546	0.571	0.542	0.475	0.452	0.466	0.551
20-24/45-49	1.367	1.222	1.137	1.416	1.074	1.016	1.086	1.211	1.284	1.091	0.939	0.743	0.606	0.587	0.59	0.688	0.729	0.72	0.706	0.747	0.792

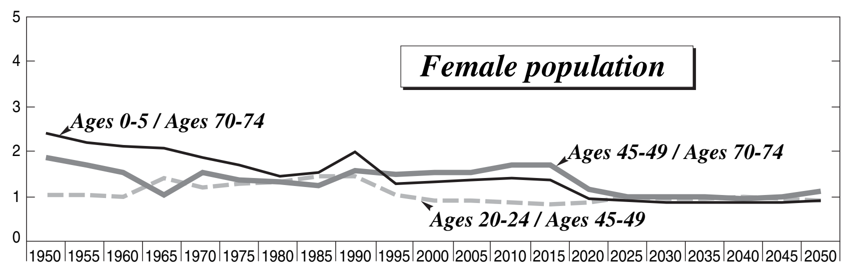
FIGURE 81–7 Ratios of the population aged 45 to 49 years to that aged 70 to 74 years, of population aged 0 to 4 years to that aged 70 to 74 years, and of the population aged 20 to 24 years to that aged 45 to 49 years, according to sex, 1950–2050, Italy. (Data from United Nations, 2003.)



	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	2.21	2.09	1.9	1.26	1.84	1.63	1.52	1.46	1.83	1.71	1.71	1.71	1.85	1.8	1.23	1.08	1.09	1.05	1.01	1.03	1.18
0-5/70-74	2.88	2.73	2.68	2.68	2.34	2.04	1.74	1.87	2.35	1.48	1.5	1.54	1.58	1.53	1.03	0.99	0.96	0.94	0.94	0.92	0.99
20-24/45-49	1.06	1.04	1.01	1.46	1.24	1.29	1.3	1.45	1.45	1.02	0.9	0.92	0.89	0.86	0.89	1.02	0.96	0.94	0.98	0.94	0.9



	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	2.68	2.69	2.46	1.7	2.32	1.97	1.83	1.78	2.18	2.02	1.95	1.91	2.04	1.96	1.35	1.2	1.22	1.16	1.1	1.12	1.26
0-5/70-74	3.58	3.61	3.59	3.71	3.04	2.52	2.13	2.3	2.84	1.77	1.76	1.79	1.81	1.72	1.16	1.11	1.09	1.05	1.04	1.01	1.06
20-24/45-49	1.08	1.07	1.06	1.52	1.29	1.31	1.32	1.44	1.45	1.02	0.93	0.96	0.92	0.89	0.91	1.04	0.97	0.96	0.99	0.94	0.91



	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
45-49/70-74	1.88	1.71	1.55	1.01	1.53	1.39	1.3	1.24	1.57	1.48	1.52	1.55	1.7	1.67	1.14	0.98	0.99	0.96	0.93	0.96	1.1
0-5/70-74	2.39	2.17	2.12	2.07	1.88	1.7	1.46	1.56	2	1.26	1.29	1.34	1.4	1.38	0.93	0.88	0.86	0.84	0.86	0.84	0.91
20-24/45-49	1.03	1.02	0.97	1.4	1.2	1.27	1.29	1.45	1.45	1.01	0.88	0.89	0.86	0.84	0.87	1.01	0.95	0.93	0.96	0.93	0.89

FIGURE 81–8 Ratios of the population aged 45 to 49 years to that aged 70 to 74 years, of population aged 0 to 4 years to that aged 70 to 74 years, and of the population aged 20 to 24 years to that aged 45 to 49 years, according to sex, 1950–2050, France. (Data from United Nations, 2003.)

linked to emotional relationships between generations, to these generations of remembrance mentioned above. These relationships (studied here from a bidirectional point of view, i.e., from parents to children and from children to parents) are often neglected despite their significance in the lives of individuals and in relationships between generations. Indeed, individuals cannot be separated from their pasts and from their relationships with persons of reference. We will discuss here three types of transmission: intergenerational transmission of disease risk, children who support widowed parents, and the intergenerational transmission of the divorce risk.

1. The Long-Term Consequences of Parental Conflicts

In 1992, an interesting and innovative French study was conducted on the health problems of adults seen in the light of family difficulties experienced during childhood (Menahem, 1992). The performances in school and in the professional world of children with divorced or separated parents have been extensively studied, especially in the United States, but there have been very few wide-ranging and statistically significant studies on the consequences of childhood traumas in terms of health. Table 81–3 and Fig. 81–9 clearly

TABLE 81–3 Average Number of Reported Illnesses, According to Several Variables Concerning the Adult Respondent's Youth

Variables and variants	No. of cases	Average no. of reported illnesses	Index
Serious conflict or separation or divorce of parents (before youth becomes 18)			
Neither separation nor conflict	4,199	2.99	98
Only separation or divorce	187	3.26	106
Separation or divorce and serious conflict	114	3.62	118
Serious conflict without separation	151	4.44	145
Death or serious illness of the respondent's mother			
Neither death nor serious illness	4,127	3.01	98
Death without illness, disability or serious accident	300	3.29	107
Illness, disability or serious accident	224	3.82	125
Serious financial problems (before youths becomes 18)			
Never or rarely	2,108	2.77	90
Occasionally	1,222	3.01	98
Often	1,321	3.59	117
Total	4,651	3.06	100

Data from Menahem, 1992:904.

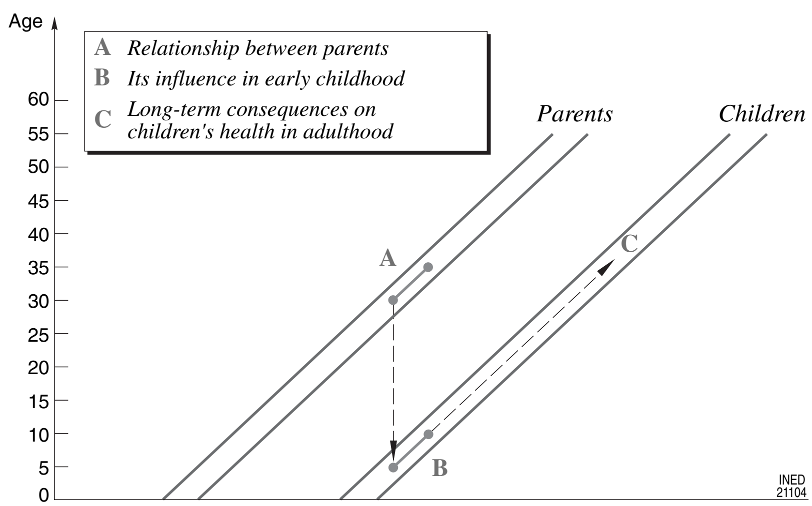


FIGURE 81–9 Pattern of the influence of the parents' relationship on the health of their children once they have become adults.

show the negative impact of childhood traumas; in particular, they reveal that even when the parents do not separate or divorce, serious conflicts between parents can have negative repercussions on their children once they have become adults.

The findings of George Menahem could be interpreted as a statistical confirmation of the Freudian hypothesis as to the significant impact of childhood traumas on the later life of a person. However, regardless of one's theoretical references, the survey clearly shows that there are some individuals whose physiological vulnerability is heightened in certain critical situations, causing them to fall ill, sometimes very seriously.

2. Children and their Parents' Health

Unlike what might have been expected, persons who have children do not have a lower mortality than persons who do not. An interesting Belgian survey covering 149,000 widows and widowers aged 30 to 84 years, interviewed in 1991 with a 5-year follow-up period, showed that—age, sex, duration of widowhood, and level of education being equal—neither having children nor their number seems to have any influence on how long a person survives after the death of his/her partner (Lusyne *et al*., 2002). A possible explanation can be the fact that childless widowers or widows can find alternative networks of support among their relatives, friends, and neighbors, which play the same supporting role as a child would for his or her widowed parent. One could also surmise that on the contrary, from an emotional point of view, the loss of a partner is so painful that neither the presence of children nor that of friends can help. However, these findings do not in any way contradict the large body of research showing the importance of children and relatives in the task of caring for the elderly, especially if they are widowed, ill, or disabled.

3. Transmission of the Divorce Risk

Children of divorced parents divorce much more often than do children of parents who stay together. The intergenerational transmission of the divorce risk has been observed in several industrialized countries: in the United States (Wolfinger, 1999), Australia (Dronkers and Hox, 1998), Germany (Engelhardt *et al*., 2003), and The Netherlands (Fisher and De Graaf, 2002). Overall, five main causal mechanisms have been identified: (1) the stress generated by the parents' divorce leads children to leave home early, get married, and have their first child quite soon, thus increasing the risk of divorce; (2) persons who, during childhood, experienced conflictual relationships at home may have difficulty creating and enjoying decent interpersonal relationships later on in life; (3) the loss of economic status after divorce has a negative influence on the children's lives (especially in terms of education and professional career), and this in turn increases their own risk of divorcing; (4) the stigma of the parents' divorce has a negative impact on the children's later opportunities in life; and (5) parents and children have in common a few hereditary character traits that can favor (or on the contrary prevent) divorce. All these mechanisms are not mutually exclusive may interact with one another (Engelhardt *et al*., 2003).

Less is known about the influence of family policies and family law on divorce risks. A recent survey, based on the *German Life History Study* and using a multivariate analysis of biographies shows that the interaction of family policies and laws on divorce with the transmission of divorce risk is complex but does exist (Engelhardt *et al*., 2003).

IV. TRANSFERS WITHIN THE POPULATION

Intergenerational transfers within a population and their potential inequity, especially when one observes or foresees strong, lasting, and significant numerical imbalances between generations (as described in the preceding chapter), are the subjects of increasingly wide-ranging and multidisciplinary debate. Not only does it raise the issue of the aging of the population (the increasing number of elderly persons, on one hand, and the decreasing number of children and young people, on the other), but also that of the changes in economic life cycles that will modify the duration of periods of economic dependence (e.g., the increase in the duration of studies and professional training); another issue is the modification of the legal context surrounding the transfers necessary to finance these dependent stages (the lengthening of the period of compulsory education, or changes in the legal retirement age, etc.). Thus, there are at least three spheres in which demographic changes interact with the system of social security and intergenerational transfers (Lee, 2020). Of course, aspects of collective psychology, or social or cultural aspects can contribute to bring about changes both in the economic cycle of life and in the legal framework.

These three spheres interact in turn with relationships between the generations of a given family, whose situation, good or bad, has a strong impact on periods of economic dependence, because this factor either shortens or lengthens the period during which a person will be receiving an education or working. Figure 81–10 shows quite clearly that the consequences of demographic changes for public finances depend not only on the age structure of the population (which as we know is the result of the history of successive cohorts) but also on the numerical size of the population.[13]

The redistribution of income between generations is a very complex problem. To quantify the phenomenon, several economists have elaborated a framework of analysis called "generational accounts." These accounts determine how much a typical member of the present generation and of future generations will pay in taxes during his or her life, and will deduct from that what the member has received in the

[13] Nicola Sartor (2003:33) underscores that "a fertility increase, and thus the demographic size of a population has a positive impact on public finance, at least on the long term, and this impact can be perceived when assessing the per capita tax reductions necessary to maintain public services at a constant level and ensure the long-term sustainability of the public debt. [*gli effetti che un aumento della fecondità, e quindi della dimensione demografica di una popolazione, sortirebbe sulla finanza pubblica sono positivi, sebbene nel lungo periodo, e possono essere espressi attraverso la stima delle minori imposte pro capite necessarie per mantenere invariati i servizi pubblici a tutti i cittadini e per assicurare la sostenibilità di lungo periodo del debito pubblico*]."

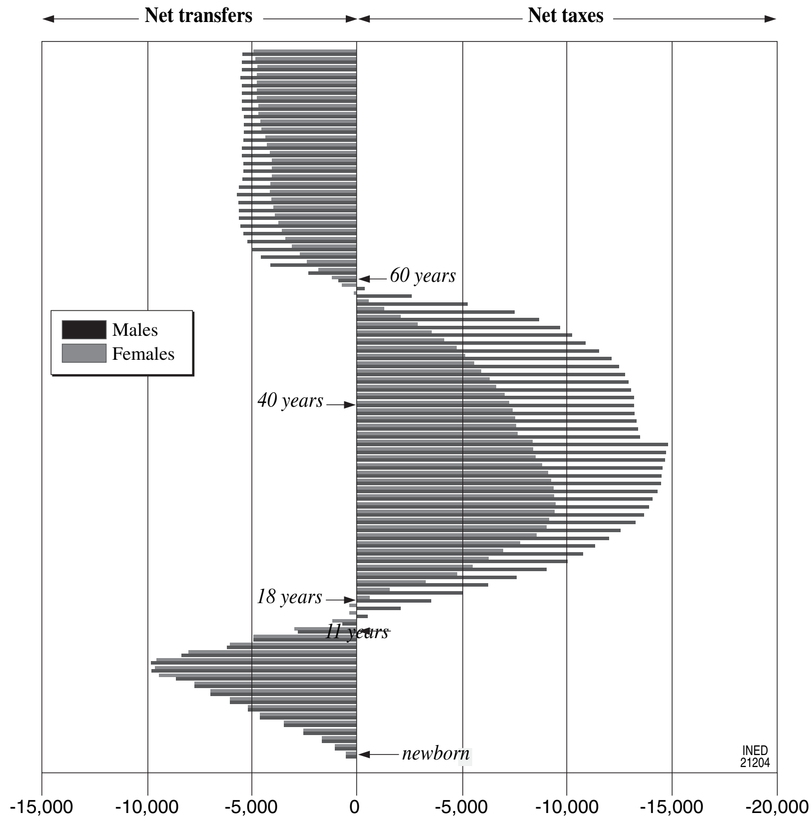


FIGURE 81–10 Net financial position of the different age groups/cohorts in public administration accounts, Italy. (Data from Sartor, 2003:32.)

form of transfers and services (an example is shown in Figure 81–11). The Organization for Security and Cooperation in Europe asked several experts (including Professor Nicola Sartor of Italy) to elaborate generational accounts according to a constant method. The results for Italy are shown in Table 81–4.

This theme is particularly sensitive, because the limited nature of resources means that one must take from one age group (that is a group of cohorts) or one social group to give to another. This entails serious technical and political difficulties, for many reasons; we will only mention the most important among them.

In a democracy, a political decision aimed at withdrawing resources or privileges from one age group or social group in favor of another in order to offset an imbalance must be based on a consensus. This consensus is often long and difficult to obtain, and in some cases, it is totally impossible owing to the opposition of the concerned group or lobby. When that happens, the problem is tied to the institutional system, with the risk that "given the way representative democracies function today, collective decisions may correspond only to the preferred options of a part of the group, those which have won over the short term, and not to those of the entire group of citizens." (Petroni, 2003:6).[14]

The population dynamics observed almost everywhere but especially in countries with very low fertility over a long period of time, such as Italy, Spain, or China—where one observes a strong decline in the number of children and a strong increase, both in absolute numbers and in proportion, of elderly persons—should necessarily result in a transfer of physical, human, and financial resources from the former to the latter. The best solution—which is very difficult to determine and implement, as this is a completely new phenomenon for which there is no precedent—would be to establish a horizontal transfer of resources to ensure that the per capita resources

[14] *"La forma assunta dalla democrazia rappresentativa ai giorni nostri, fa sì che le decisioni collettive non riescono a corrispondere all'intero insieme delle preferenze dei cittadini, ma soltanto a una parte di esse, quelle maggiormente legate al breve periodo."*

TABLE 81–4 Present Values of Taxes Paid during Life, after Deduction of Transfers. Italy (Billions of Dollars*)

Productivity increase:	1			1.5			2		
Interest rates:	3	5	7	3	5	7	3	5	7
Case A[a]									
Males									
Present generations	102	54	22	114	65	29	122	77	36
Future generations	433	340	316	465	354	306	508	374	306
Imbalance between generations[b]	326	533	1,336	310	446	970	315	385	741
Females									
Present generations	19	14	2	12	17	5	1	19	8
Future generations	79	88	26	51	93	50	5	94	65
Imbalance between generations[b]	327	532	1,333	310	446	976	325	385	737
Case B[a]									
Males									
Present generations	122	59	24	144	72	31	166	88	39
Future generations	258	206	192	273	213	185	290	224	185
Imbalance between generations[b]	111	249	709	90	195	500	74	155	369
Females									
Present generations	37	19	3	40	24	7	39	29	10
Future generations	79	65	27	76	70	40	68	74	49
Imbalance between generations[b]	111	248	703	92	195	499	74	155	368

Values are in billions of dollars as constant prices, adjusted according to the development of real incomes, converted according to 1993 average rates of development. The present generations correspond to cohorts born in 1993.

[a] Case A, World Bank population projections based on the hypothesis that fertility rates will return to replacement level in 2030; case B, a more rapid increase in fertility rates that reaches replacement level in 2010.

[b] The imbalance between generations is obtained through the difference between the discounted current value of the taxes paid by a typical member of present and future generations, expressed in percentages of the net payment of taxes by the present generation.

(which in many cases should be themselves increased) to be received by the children do not decrease.[15]

For the first time in the history of mankind, and especially in the West, we are in a situation in which human resources have become either stationary or are declining and in which there is an employment crisis. Such a situation requires great flexibility on the part of the work force and an extensive and significant system of mutual social support to organize this flexibility and weather the crisis. One of the most difficult problems is that social welfare measures for adults and the pension system for retirees are based on the prevailing structure, which means that the entire burden is borne by the working population; in times of demographic or economic crisis, this burden can become too heavy to bear.

For the pension system, the distribution pay-as-you-go model (which divides population and society horizontally between those who work and contribute and those who receive pensions) weakens the capacity to accumulate capital within the family and reduces the time span during which individuals can make their own decision. As a result, the responsibility of single persons over the long term is reduced, as is the responsibility of public decision makers, and their strategies and choices are increasingly considered over the short-term (Petroni, 2003). However, the capitalization advanced-funded model does not represent a fully convincing alternative, especially given the enormous long-term financial risk and the high management costs. Thus, over the middle term, the pension system seems to be in a dead-end situation; the way out seems to be a new efficient and applicable system based on two (or three) basic elements (distribution + capitalization + possibly private savings).

The close link between work and chances of survival, which has always existed in the history of mankind and was still very strong three or four generations ago, has to a large extent disappeared from collective, individual, and family psychology. Social welfare now often tends to be taken for granted and is no longer considered as something that must be gained through long and hard work, remunerated by the market or by public funds. Thus, aging genera-

[15] Concerning this issue, Nicola Rossi (1997) takes into account demographic trends, economic development and social solidarity measures and examines them also from a long-term point of view. Less support should be given to the fathers and more to the sons (*meno ai padri, più ai figli*).

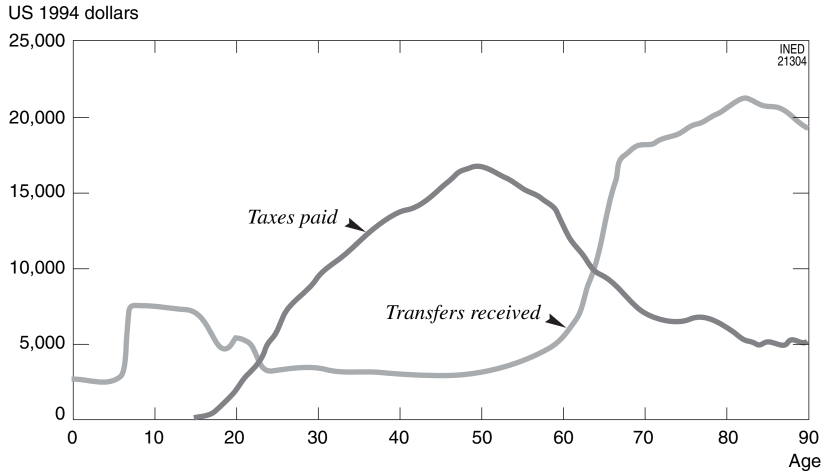


FIGURE 81–11 Age profile of the transfers received from the public sector and taxes paid, United States, 1994. (Data from Lee, 2002.)

tions tend to aspire more and more to the status of a paid *nonworker* rather than to that of pensioner, according to more varied forms of professional or social *occupation*.

V. SOCIAL, PROFESSIONAL, AND PSYCHOLOGICAL DISTANCE BETWEEN GENERATIONS

Demographic change cannot be separated from the modernization process as a whole. Thus, both within the family and among the general population, educational levels and all forms of mobility (professional, social, matrimonial, and territorial) have increased from one generation to the next.

In the realm of education, the differences among grandparents, children, and grandchildren is considerable, not only from a quantitative point of view (the number of years spent in school is much higher than it was in the past) but also from a qualitative point of view (in general, in the past, education had a cultural purpose, whereas today it is market-oriented). As concerns the professional realm, today's grandfathers and grandmothers generally had the same occupation all their lives, whereas their children and grandchildren often have many different occupations, either successively or simultaneously. As regards travel, if one compares today's generation of children with that of their parents (not to mention their grandparents), the number, frequency, and length of trips made by individuals aged 20 to 40 years has considerably increased. In addition, one must also take into account the fact that a rising proportion of young people will live with brothers and sisters, parents, and grandparents belonging to their father's or mother's second family. In contrast, many children of divorced couples only seldom see or will get to see their father.

All this increases the distance between generations. Situations and family networks today do not have much in common with those of the past and even less with those of the future. And at the level of the population itself, things have changed a great deal, because relationships between cohorts can be seen as the complex and collective consequences of the relationships between generations within the family.

Modernization comes along with new technologies; these are highly innovative and invasive and have also widened the gap between generations. The ability to react and adapt to modernization varies a great deal with age. The older generations rely on their past experience, and their ability to develop new competences is more limited. In addition, new technologies have a heavy impact on the quantity and quality of available employment as well as on the quality of life of the different generations and of society as a whole. Rapid modernization and the spread of Internet may widen the gap still further. The ability to adapt to new technologies and take advantage of their potential, as well as the ability to learn and master the language of technology and the possibilities they offer, vary with age. The paradox is that the older generations are providing the younger generations with the very technologies that contribute to the widening of the generation gap.

Present generations have much more time at their disposal than did their elders to accumulate property (because of economic and professional reasons and because the duration of life has increased). But the age at which their heirs inherit this patrimony keeps increasing. In Italy, the real mean age of a husband at the time of his death is 71 years; the children then inherit 50% of the inheritance and the widow the other half. On average, the widow is 68, and her remaining mean duration of life is 17 or 18 years. The children thus receive the first half of their inheritance at 42 years and the second half around 60 years, at the death of their mother. A few decades ago, people inherited at a much younger age.

Thus, the contributing factors to the increase in the social, professional, and psychological distance among generations are extremely varied; today this distance is probably the largest it has ever been. In addition, the increase goes along with a reversal in values, because there are few children and many elderly persons. In the past, old-timers were valued and not children, but today the opposite is true.

VI. A NEW SOCIAL CONTRACT BETWEEN GENERATIONS

From a global point of view, it must be pointed out that the conjunction of the demographic crisis with an economic crisis is mostly a European phenomenon. It is difficult to imagine what type of economic development will take place in Europe, especially in countries that are not very competitive, such as Italy. Will they be able to remain competitive compared with China or India, which have many young people and good educational systems? Will it be possible in Europe to maintain a social, cultural, psychological, and political balance among an aging population, traditional ethnic minorities (those linked to European history), and the young immigrant population that is absolutely necessary to these countries? It is difficult to imagine the consequences of new technologies, in particular robots, on the labor demand and more generally on the economy and society (e.g., one could imagine a system of provisional taxes on computers and robots). If the labor demand decreases, will it not be necessary to redesign (on schedule, on a regular basis, and as efficiently as possible) the social solidarity system and redefine the social contract between generations?

Indeed, strong measures aimed at redefining a social contract must be developed. One could, for example, imagine a pension system based on the three "foundations" (public pensions financed by contribution, pension funds financed by compulsory savings, and pensions financed through private insurance) with a complete revision of the system of retirement ages, and, in particular, the creation of a transitional period between full-time work and full-time retirement.

At the microeconomic level—still referring to the European experience—one observes a strong rise in individualism. This phenomenon is for the most part rooted in the rising level of education, which is based on personal investment and consequently on the necessity of taking advantage of one's "human capital" and taking good care of one's physical health. Individualism is also reinforced by the fact that during his or her life, a person may have an increasing number of families (in the past, people had only two families, that of their parents and their own, whereas today they can have three, four, or more families). The family is no longer experienced as a basic social structure in whose context a person will satisfactorily spend his or her adult life, but rather as a transitional stage experienced by an isolated adult moving from one family to another. If intergenerational links have weakened within families, one may expect them to weaken in society as well.

Is it possible to imagine a new social contract that would be increasingly intragenerational, without forbearing the traditional relationships and obligations that exist between generations? Such a project could be based on three main principles:

1. The creation of a *compulsory general civil service* devoted mainly to caring for the elderly and to the environment. This service could have several advantages: reducing individualism and promoting a spirit of self-sacrifice and team work, reinforcing or creating intergenerational spirit outside the family context in a social context, and reducing the cost of the social mutual support system.
2. Encouraging the development of the nonprofit sector, in particular volunteer work for elderly people in good health, who could help care for older persons with physical, psychological, or cognitive disabilities.[16] The senior citizens who participate in this volunteer work could earn credit, enabling them to receive aid if and when they need it at a later stage.
3. Measures to reduce the numerical imbalance between generations, whenever necessary. To achieve this, given that a mortality increase would not be desirable and that none has been foreseen (even if one cannot exclude the possibility), something must be done about the other two components of population dynamics: fertility and migration. As concerns fertility, an increase must be stimulated in those areas where it has been excessively low for a long time. This is a dif-

[16] This type of activity is already quite widespread in small Italian mountain towns or villages, where the population is very old and where there are absolutely no young people at all: efficient systems of mutual support have been set up enabling senior citizens to care for each other.

ficult task for many reasons; one of them is that it is difficult to imagine the goals and demographic behavior of today's youth, just as it is difficult to imagine what kind of parents they will be once they have decided to have children (as concerns industrialized countries, Chapters 103 and 104 deal with the question of fertility policies and their efficiency). As regards migration, the solution would be to favor immigration, but in that case, the difficulty remains of ensuring the peaceful and productive coexistence of native and immigrant populations.

As demographic changes accelerate and intensify (beyond the modernization and globalization processes that go along with these changes), the gap between generations grows wider and more conflict-filled. In many ways, this gap may become a potentially explosive issue for societies, if the latter are unable to establish a dynamic process of adjustment.

Acknowledgment

This chapter was translated from the French version by Zoe Andreyev.

References

Barbagli Marzio, Castiglioni Maria, and Dalla Zuanna Gianpiero, 2003. *Fare famiglia in Italia. Un secolo di cambiamenti.* Bologna, Il Mulino, 308 p.

Bryson Ken and Camper Lynne M., 1999. *Coresident grandparents and grandchildren.* Washington, Census Bureau, 198 p.

Cassan Francine, Mazuy Magali, and Toulemon Laurent, 2001. *INSEE première*, no. 776, p. 1–4.

Dronkers Jaap and Hox Joop, 1998. *The importance of family and education for divorce risks of siblings: an Australian example.* Florence, European University Institute, 16 p.

Engelhardt Henriette, Trappe Heike, and Dronkers Jaap, 2003. Differences in family policy and the intergenerational transmission of divorce: a comparison between the former East and West Germany, *Vienna Yearbook of Population Research*, p. 121–142.

50&PIÙ and Fenacom, 2001. *Essere anziano oggi. Liberare la forza dell'età matura.* Rome, Confcommercio.

Fischer Tamar F. C. and De Graaf Paul M., 2002. *Pre-divorce family characteristics and the impact of parental divorce on children.* Florence, European University Institute, 24 p.

Fuller-Thomson Esme, Minkler Meredith, and Driver Diane, 1997. A profile of grandparents raising grandchildren in the United States, *The Gerontologist*, vol. 37(3), p. 406–411.

Goldman Noreen, 1986. Effects of mortality levels on kinship, *in*: United Nations (ed.), *Consequences of mortality trends and differentials*, p. 79–87. New York, United Nations, 191 p.

Golini Antonio, 1992. Demografia, *in:* IEI (ed.), *Enciclopedia delle Scienze sociali. Vol. 2*, p. 774–791. Rome, Istituto della Enciclopedia Italiana, 816 p.

Golini Antonio, 1999. L'invecchiamento della popolazione in Italia, *Rivista italiana di economia, demografia e statistica*, vol. 53(4), p. 11–59.

Golini Antonio, 2003. *La popolazione del pianeta.* Bologna, Il Mulino, 141 p.

Golini Antonio and Silvestrini Angela, 1995. Cambiamenti familiari e relazioni generazionali: una lettura demografica, *in:* Pierpaolo Donati (ed.), *Quarto Rapporto CISF sulla Famiglia in Italia*, p. 89–126. Cinisello Balsamo, Edizioni San Paolo, 425 p.

Lee Ronald D., 2002. *Demographic change, welfare and intergenerational transfers: a global overview.* Berkeley, University of California, 25 p.

Lusyne Patrick, Vernaillen Neelke, Page Hilary, and Lievens John, 2002. *Excess mortality following conjugal bereavement: does having children buffer the effects?* Brussels and Ghent, Vrije Universiteit Brussel and Universiteit Gent, 20 p.

Menahem Georges, 1992. Troubles de santé à l'age adulte et difficultés familiales durant l'enfance, *Population*, vol. 47(4), p. 893–932.

National Institute of Statistics [ISTAT], 1999. Nonni e nipoti: le principali caratteristiche. Anno 1998, *Statistiche in breve*, 19 November, p. 1–4.

National Institute of Statistics [ISTAT], 2001. *Parentela e reti di solidarietà, Indagine Multiscopo sulle famiglie, Aspetti della vita quotidiana, Anno 1998*. Rome, ISTAT, 129 p.

Petroni Angelo M., 2003. Sulle pensioni pesano troppi alibi elettorali, *Sole-24 ore*, 25 mai, p. 1–6.

Rossi Nicola, 1997. *Meno ai padri, più ai figli*. Bologna, Il Mulino, 94 p.

Sartor Nicola, 2003. La bassa fecondità: *free-riding* demografico? (ancora sui rapporti economici e finanziari tra Stato e famiglia), *in*: Marco Breschi and Massimo Livi Bacci (eds.), *La bassa fecondità italiana tra costrizioni economiche e cambio di valori. Presentazione delle indagini e dei risultati*, p. 23–40. Udine, Forum, 263 p.

United Nations, 2003. *World population prospects: the 2002 revision.* New York, United Nations, Population Division, 1720 p. + CD-rom.

Wolfinger Nicholas H., 1999. Trends in the intergenerational transmission of divorce, *Demography*, vol. 36(3), p. 415–420.

C H A P T E R

82

Economic, Social, and Cultural Consequences of the Aging of the Population

JACQUES LÉGARÉ

Department of Demography, University of Montreal, Montreal, Canada

INTRODUCTION

The 21st century will increasingly be characterized by the proliferation of elderly people. Not only are the raw figures for older people increasing in all the national populations, but the proportion of the old in these populations is growing considerably. Such results were easily foreseeable as soon as the demographic transition reached its conclusion; that is, the control of mortality and fertility enabled these two phenomena to reach equilibrium. In 1956, Jean Bourgeois-Pichat revealed by means of population models the economic and social implications of aging (UN, 1956), but unfortunately, this famous document—according to the rules for United Nations (UN) publications—did not mention his name.

Such changes in the age structure of populations are not without consequences for the societies that undergo them. They have already aroused interest in the major international organizations: the UN (1988, UN-EEC 1992, 1994), the Organization for Economic Cooperation and Development (OECD, 1996, 1998, 2000), and the World Bank (1994). But, as with all theories, the theory of demographic transition has its limits when confronted with reality (Chapter 69), and the economic and social consequences of aging for actual populations differ from the forecasts for theoretical populations, sometimes to a marked extent. And it is this point that I shall endeavor to explain in this chapter.

The demographic history of the 20th century has shown that, for both fertility and mortality, the passage from a relatively high to a low level does not occur in a regular way. There have been variations both in an upward and downward direction during the secular decline in fertility, principally allied to civil wars and genocides in certain countries and to the world wars and baby boom in the West. The baby boom, frequently succeeded by a baby bust, has caused important variations in the size of the successive generations, and these changes have effects throughout the life cycle of these generations (Golini, 2003). However, the recent increases observed in certain countries in the former communist empire and those related to AIDS in Africa were unpredictable.

Another dimension completely unenvisaged by the demographic transition theory is that the replacement of the generations has no longer been attained in a good number of national populations, and the life expectancy of many children born in the early 21st century will probably reach 100 years. Among the major effects on the life of individuals and populations can be mentioned the reduction in the sizes of the total population of influential countries on the world stage and the dazzling increase of the very aged, including a large number of 100-year-olds.

I. DEMOGRAPHIC CONTEXT OF AGING SOCIETIES

Up to the present, the amazing decline in mortality and its implications for the survival of children and the growth of populations have been studied more fully than have its relations to demographic aging.

The elderly have always existed in human societies, but there were considerably fewer when mortality (excluding epidemics) was high. If indeed a model life table is interpreted correctly, it will be found that when life expectancy at birth is around 25 years, there will nevertheless be 15% of the initial cohort that survives until 60, and these survivors can hope on average to live a further 10 years. With the impressive progress in preventing premature death and with life expectancies at birth approaching 80 years over a large part of the planet, more than 90% of the initial cohort will reach advanced ages and experience an increasingly long old age. We are therefore witnessing what I have called a democratization of old age (Légaré, 1991), in the sense that it is no longer reserved to a few but is available to a very large majority of the population. One of the principal outcomes of all democratization is to destroy (for better or for worse) the homogeneity of the elite, formerly fawned upon but now part of a more heterogeneous group in which individuals are less fawned upon. That certainly applies to aging societies in which the subpopulation of retired persons increasingly resembles that of the whole population in socioeconomic and cultural terms.

Although demographic projections are quite often hazardous, especially in the long term, the same cannot be said for the projections concerning the numbers of older persons. Effectively, all elderly persons of 50 years or over in 2050 are already born, and updating these projections at national level depends on the precise evaluation of their behavior in terms of mortality and migration. Predicting the sizes of yet unborn generations is much more uncertain.

The past development of fertility and mortality that mechanically provokes the aging of populations not only disturbs the population pyramid at its summit but also affects the whole pyramid. Throughout the whole aging process, it is not only the median age of the whole population that increases but the median age of both the subpopulation of retired persons and the subpopulation of those of working age—those effectively to be found on the labor market. It follows that the population of 85-year-olds and over will grow more quickly than will that of the 65-year-olds and over (Suzman *et al.*, 1992) and, in most Western countries, an important decline in both the raw and relative numbers of persons of working age can be anticipated (Légaré, 2001a). In light moreover of past (and also future) differential life expectancy according to sex, there are, and will continue to be, more elderly (and especially very old) women than men, as is clearly characteristic of contemporary aging societies (de Jong Gierveld *et al.*, 2003).

Even though the aging of populations is regarded less as a negative phenomenon but just a logical consequence of one of the great advances in human progress—the increasingly effective control of mortality and fertility—it still gives rise to considerable apprehension. Future societies will certainly be older, but will they be more aged and engender a pandemic of sick or disabled persons? Are we capable of evaluating the implications of administering an aging society according to whether the total population is increasing or decreasing demographically? In this era of increasing globalization, such demographic mutations affecting all populations on the planet—although occurring at different times and to different degrees—will certainly have geopolitical repercussions.

To affirm that future societies will clearly be more aged than those of today is to assume that the threshold of old age is unalterable in time and space. Such an affirmation is being increasingly questioned, at least by certain demographers and historians (Desjardins and Légaré, 1984; Bourdelais, 1993). Will the physical and economic dependence that characterize aging be transferred to much more advanced ages, so that future societies will certainly be more elderly but without necessarily being more aged?

In fact, can aging be defined? According to Furetière (1690), "Old refers to a certain age at which, upon attainment, one is decrepit: men are old at 60, horses at 20." The qualifier "decrepit" refers to dependence, as evoked earlier. The difference between human beings and horses refers here to the important difference in their average lifespan. *Ceteris Paribus*, when we observe in human societies an important difference in life expectancy at birth, the threshold of aging must logically increase. To take account of this development, Norman Ryder (1975) suggested that the threshold of aging should not be defined as a certain time passed since birth, such as 50 or 60 years, but rather a certain number of years remaining to be lived. Employing the terminology of the life table, this would be the age x for which the life expectancy e_x equals, for example, 10 years. That is admittedly as arbitrary in one case as in another, but Norman Ryder's proposal admits the possibility of taking better account of reality. In addition, this period before death can be made to vary according to the assumed compression of morbidity (Fries, 1980, 1989; Myers and Manton, 1984). With the help of model life tables (Table 82–1), it can be seen that when

TABLE 82–1 Aging Threshold Defined as 10 Years Remaining To Be Lived, According to the Indices of Model Life Tables

Aging indicators	Life expectancy at birth e_0 (in years)				
	25	40	60	80	91
Old age threshold (x) for $e_x = 10$ years	60	65	70	75	84
Surviving ‰ to the age (x) (S_x)	150	300	500	750	840

Data from Coale *et al.*, 1983 (West model, feminine sex), except for $e_0 = 91$: Duchêne and Wunsch, 1990.

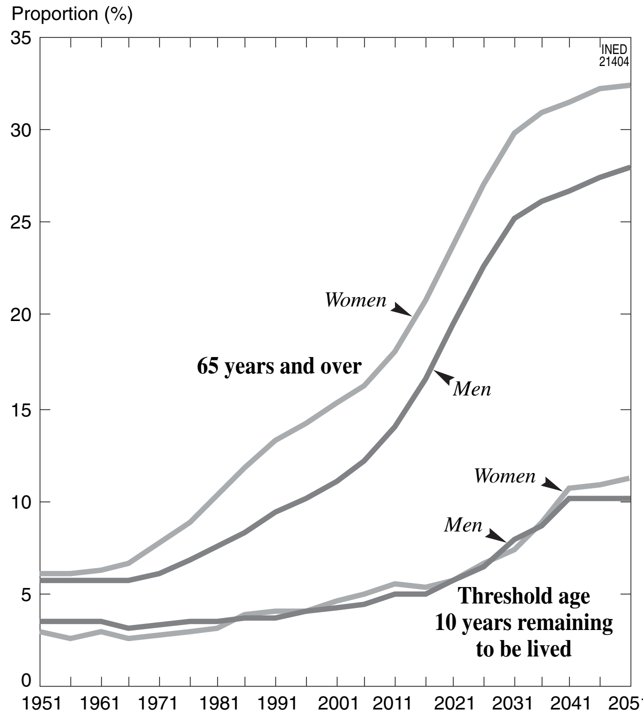


FIGURE 82–1 Evolution of the proportion of persons aged 65 years and over and proportion of persons surviving beyond the age threshold "10 years remaining to be lived," Quebec, 1951–2051. (Data from Statistics Canada, Canadian censuses of 1951–1996, and the Quebec Statistical Institute, *Perspectives démographiques du Québec de 2001–2051*.)

the life expectancy at birth varies between 25 and more than 90 years, the threshold age for 10 years remaining to be lived could itself vary from 60 to nearly 85 years. With such a definition, the proportion of the *aged* in an actual population (in this case that of Quebec) (Fig. 82–1) scarcely triples over a period of 100 years, whereas the proportion of 65-year-olds and over is multiplied by nearly 6. Such a development doubtlessly applies to all human populations and dismisses theories of hardened pessimists and smug optimists concerning the coming aging of our civilizations (Henripin and Loriaux, 1995; Gee and Gutman, 2000).

Considerable confusion has arisen in this area through assimilating the threshold of old age to that of retirement (Légaré and Desjardins, 1987). Although in the past, especially in farming societies, working only ceased when precluded by declining strength, it is completely different in industrial and postindustrial societies (Guillemard *et al*., 1995). The right to retirement has become a welfare entitlement that permits a period within the life cycle when living is sustained more from benefits than from earnings, but without incurring poverty or ill health. This is the new "third age" of Peter Laslett (1989), which extends from retirement to old age and consigns to a "fourth age" the period in the life cycle when an individual becomes dependent on other members of the population, especially for reasons of declining health. Because this new step in the life cycle has never previously existed in any civilization, our societies have to invent everything in order to ensure that this stage in the life of individuals shall be as beneficial as possible both for themselves and for the societies in which they have evolved. This reality is so new that it is not yet known how to designate those who form part of it. Because the term "old people" is no longer appropriate, one should perhaps use the Latin word *seniores*, or *aînés*' in French and "seniors" in English. These denominations correspond to the roles of mentors that these persons now play, or at least should play, in our industrialized societies. These roles fit well into the context of "active aging," an expression, when applied to young retired people, that is often understood as an extension of the opportunities that they may use for improving their social and functional independence through the enrichment of various aspects of their social life (as full- or part-time work), as well as their physical and mental health.

The demographic perspective illuminates in an original way the phenomenon of aging populations and has given birth to a well-defined niche in our discipline: the demography of aging (Loriaux *et al*., 1990; Martin and Preston, 1994; NRC, 2001). This is a field of study that clearly differs, for example, from the disciplines of gerontology and geriatrics, despite being limited to the study of the subpopulation of elderly persons (Myers and Légaré, 2000). Such developments will obviously have a repercussion in the area of transmitting knowledge about the aging of populations (Golini, 2002).

After these clarifications that shed light on certain little known demographic aspects concerning aging societies, it is now appropriate to verify the extent to which the aging of populations is really an economic weight and a delicate social problem for countries (Légaré, 2001b). Must the social contract be reviewed (Véron *et al.*, 2004)?

II. SOME ECONOMIC CONSEQUENCES OF DEMOGRAPHIC AGING

A consideration of the economic consequences of demographic aging leads us to enquire about the costs of the social programs allied to aging and to their financing of them (Gillion *et al.*, 2000; Queisser, 2000).

1. Estimated Costs of Dependence

It is customary in demography to estimate the burden represented by the dependence of the young and retired people relative to the economically active persons by calculating the dependency ratios. The sum of those from 0 to 19 years of age and from 65 and over, for example, is related to the total number of those of 20 to 64 years of age. Strictly speaking, this is only an indicator of the structure by age of a population and is consequently frequently now designated in English by the expression "demographic dependency ratio." For a better grasp of the development of the dependency costs of older people, it is preferable to follow the course of more significant indicators, such as those that relate the inactive persons of 55 years and over to all those who are active (Légaré *et al.*, 1991).

This dichotomy between dependants and non-dependants becomes completely understandable within a family or household when responsibility for dependency is entirely assumed by the active members among them, as is habitually the case in traditional societies. The situation becomes complicated in the case in which this responsibility falls to the community. Greater reference is then made to social legislation in which the aid given is allied to obligations based on acquired entitlements, rather than assistance on a voluntary basis. The main areas of expense involved by this legislation are education, health, and retirement, the social costs of the community. As these costs vary enormously according to age, educational expenses are important in a young society, whereas for an aging society, it is those allied to health and retirement (Fig. 82–2).

Although children in industrial societies generally depend on and are supported by the economically active population, especially for their professional training, the same does not apply to elderly people in relation to retirement and health costs.

The first subject of preoccupation is therefore the health expenses for elderly persons. It would be better, in fact, to refer to the cost of illnesses and disablement, because the costs relating to prevention in the area of

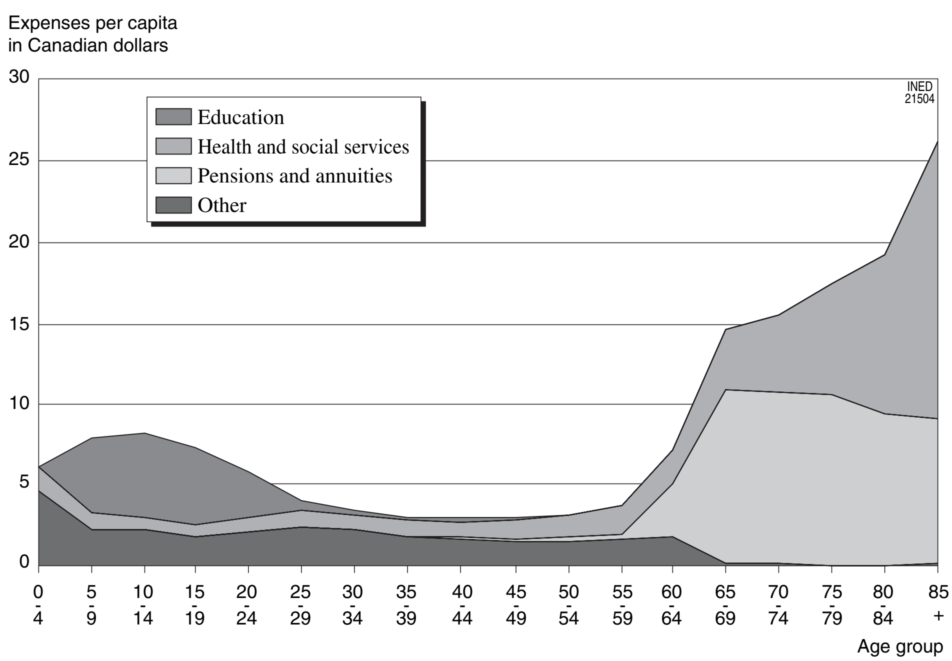


FIGURE 82–2 Public expenditures per capita according to sector and age, Quebec, 1998. (Data from the Quebec Statistical Institute.)

health apply especially to the younger ages. The past has clearly shown that the aging of populations correlated little with the astronomical increases in health expenses (Barer *et al*., 1995). These were first and foremost related to the employment of new technologies and to a greater supply than the demand for elderly persons. In future, it will be the expenses for social services allied to *caring* that will have precedence in relation to medico-hospital expenditures—budgetary appropriations will have to take account of the powerful medical and pharmaceutical lobbies allied to *curing*. An encouraging point for the future is that medico-hospital costs will not increase at the same rate as the growth in the elderly population, but according to the number of deaths of aged persons, which will be less when the baby boomers enter into the third or fourth ages. This will be the case because these expenses for elderly persons are concentrated around the time of their death, irrespective of their age (Zweifel *et al*., 1999).

Concerning the cost of retired persons, it must not be forgotten that in general they pay taxes on their incomes as well as the consumption taxes. The real costs for them should be net and not gross. In fact, as was stressed some 20 years ago by Henri Mendras and his colleagues of the *Observatoire français des conjonctures économiques* (Dirn and Mendras, 1984), the members of the new third age are frequently not only holders of the national heritage and ensured of their revenues but also good clients, especially for food, accommodations, and services. We may ask ourselves whether we are not experiencing a certain aristocratization of the new third age, as within many societies the elderly people possess a large part of the wealth and power of their nation and greatly influence its social and economic policies.

2. Financing of the Social Costs of Aging

According to the principle of solidarity, dependent persons are taken cared for by those who are not dependent. In more traditional societies, it is family solidarity that has precedence. In social solidarity systems (welfare societies), the state assumes a large part of the costs allied to dependence. To ensure the young people receive suitable training for professional life and the elderly have a time of well-deserved rest after the labors of an active life, an appeal is made to intergenerational solidarity. The adult generations support those in retirement in the hope that when they themselves reach retirement age, they will be supported by the succeeding generations.

Before the democratization of aging as previously defined, the impact of old people within society was not at all dramatic. There were very few old people and a large reservoir of adults of active age—a situation well illustrated by the population pyramid. The aging societies of tomorrow will, however, be better illustrated by spinning tops or age cylinders rather than by pyramids. The size relationships will be very different, and certain thresholds of tolerance relating to the burden of social charges in the incomes of an individual or household will have to be reasonably determined and respected.

It follows that the development of the present aging societies will generally be accompanied by structural changes that call into question certain operational methods in the functioning of social security within a society (Légaré, 2001a). Among the most important will be the development of education (especially by way of compulsory schooling) among the young. This is why the age of 15 years can no longer be considered the lowest limit of the economically active age group, because it does not accord with reality. In addition, the life expectancy of retired people never ceases to grow. Such changes create problems for the financing of the social expenses and induce governments to review certain methods of financing.

For retired persons, because there must be a certain equilibrium between payments and benefits and because the global amounts depend on the numbers of contributors and beneficiaries, certain parameters of the equation have to be changed (OECD, 2001). There is therefore a choice among increasing the contributions, reducing the benefits or raising the age of retirement, or employing a combination of these three elements. To maintain a reasonable living standard on retirement and to avoid possible poverty arising from the demographic structural changes, retired people will have to diversify their sources of income (Smeeding, 2003). As suggested by the "Four Pillars" model (AIEEA, 1996), these diversified sources could be the following:

State retirement benefits, via social security;
Retirement benefits coming from public or private systems, allied to employment;
Individual income from savings; and
Secondary employment incomes, for the younger.

This fourth source of income is directly allied to the calling into question of the low rates of employment for persons of 55 years and over (Guillemard, 2003). At the individual level, it is not imperative that the age at which the principal employment ceases should necessarily coincide with that of the access to social retirement benefits, because there may be several sources of income on retirement. Clearly, however, independent of age, social assistance should always be concerned with the most vulnerable and the most impoverished.

The baby boom generations decided to have fewer children; they had to accept their responsibilities regarding their retirement and avoid encumbering the less numerous cohorts that they had created with too heavy social charges. Even in a system of financing by distribution (pay as you go), certain reserve funds have to be set up to provide for important variations between the sizes of the younger and older generations. There are many ways of paying for these reserves, one of them being that of the user-payer. The same reasoning applies to the financing of health costs (Légaré *et al*., 2005). Inaction in this area risks the disastrous consequences in terms of intergenerational iniquity. With the help of demographers, economists have devoted themselves to measuring methods for limiting the damage, known as "generational accounting" (Auerbach *et al*., 1994, 1999; Gauthier, 1995). While continuing to improve the long projections needed for this approach, the use of these methods, despite their precariousness, should not be avoided.

III. SOME SOCIAL CONSEQUENCES OF DEMOGRAPHIC AGING

As the social consequences of aging are numerous, we shall limit ourselves here to addressing only those that are directly linked to demography.

1. Succession of Very Different Cohorts: Toward an Increasing Greater Autonomy for Old People

As was said in the preceding chapter, the accelerated demographic aging experienced by our societies is usually accompanied by important social mutations. The cohorts of old people follow but do not resemble one another (Stone, 1999). Not only are they different in terms of size (Golini, 2003; preceding chapter), but also they differ especially according to their socioeconomic characteristics (Marcil-Gratton and Légaré, 1987). The more important role of education among the young has already been mentioned, without losing sight of its repercussions throughout the various stages of life. In particular, a better level of education for women has the effect of increasing their participation in the remunerated labor market, of enabling them to acquire a certain degree of both economic and social independence, and of ensuring them an old age very different from that of their mothers and grandmothers. It follows, that when we wish to project the future by observing old people of today, we must always clearly distinguish between the effect of age and that of the birth cohort. The best example is the observation that relatively few old women drive a car, and those who do so have taken driving lessons at an advanced age. Most of tomorrow's old women, however, will have learned to drive in their youth and will be able to continue driving in their old age without any particular stress. Similar conclusions would be reached (though doubtlessly more negative) by studying the effects of smoking.

2. Ways of Living: Living Alone No Longer Necessarily Implies Isolation

Among adults, the most usual way of living is as a couple. In light of the important effect of mortality in producing widows and widowers, the diversity is sometimes greater among old people (Légaré and Martel, 2003). Even more so, the fact of having lost a spouse through divorce, a comparatively rare situation nowadays but one that can be envisaged as being more frequent in the future, places the individual entering old age in a quite different context from that of having lost a spouse by death (Martel and Carrière, 2003). Living without a spouse among other people generally reflects a certain physical, affective, or economic dependence (UN, 2001). In Third World countries, however, it is more often the case that elderly people with or without a spouse go to live in an extended family (Légaré *et al*., 2002), even though, with populations becoming increasingly urbanized and living spaces frequently more limited, accommodating a dependant relative at home is not easy.

The proliferation of single-person households in the Western world is doubtlessly allied to the high degree of economic independence enjoyed by adults without spouses. It must be stressed, however, that this applies equally to the young and old people. Among old people, it is principally women who live alone, particularly if they are in good health. Whereas, in the past people lived alone out of necessity, it is nowadays, a reasoned choice that does not necessarily imply isolation for the old person (Marcil-Gratton and Légaré, 1992). In Europe, if these people living alone are very old, this may mean that there is an extreme precarious state of accommodation (Gaymu, 2003). This is another good example of a cohort effect rather than one of age.

Everything, of course, is brought into question when the old person becomes disabled. Living in an institution often then becomes a necessity, even though this arrangement frequently entails being among very old people with social rather than health problems (Trottier *et al*., 2000). One must, however, be very careful when analyzing the way of life in an institution. The fact of there being more old people in institutions does not necessarily imply an increase in the population of old people suffering from bad health. This way of life is first and foremost much more a function of supply and demand. Nevertheless, the weaker the supply, the more the persons accommodated will be old and disabled.

3. Quality of Life Versus Quantity of Life: A Choice for the Aging Society

Although it is observable that old people live increasingly longer, can it be assumed that they are healthier (Légaré and Carrière, 1999)? It has already been observed that more members of a cohort are reaching retirement and old age, and that once this threshold is attained, they have an increased life expectancy. Effectively, in modern and therefore aging societies, premature deaths from disease have to a large extent been eliminated—most of the deaths among young people are allied to our ways of living, such as death by suicide or road accidents.

But even though establishing a death is simple, determining the health of an elderly person is much more complex (see Chapter 41). There are grounds, moreover, for clearly distinguishing the observed state of health from that as discerned by the person concerned. Among old people, the latter aspect often varies considerably from country to country, especially for cultural reasons (Egidi, 2003). Nevertheless, thanks to data arising from health surveys, demographers and epidemiologists have created indicators, such as life expectancy in good health, that enable a better understanding of the real situation (see Chapter 80). By comparing the evolution of life expectancy in all states of health with life expectancy in good health, it can be seen whether the years gained are those in the latter state. In short, has there been an increase in the years of life or of life in good health? It is easier to ask this question than it is to answer it. It can be affirmed, however, with considerable reserve, that although disability-free life expectancy (at all levels) has scarcely improved over the past 30 years, life expectancy without severe disability has tended to follow the improvement in life expectancy for all states of health (Robine and Romieu, 1998). Such an assertion does not automatically imply a future increase in disability among old persons (Jacobzone *et al*., 1998).

There is increasing insistence, moreover, that individuals deliberately lead a healthy life throughout their life cycle to preserve their physical and mental health better in old age. Campaigns against bad eating habits and nicotine addiction are concrete examples that are attaining a certain success. When asked, moreover, the large majority of *healthy* old people declare their preference for living less long but in good health, rather than much longer but disabled and ill (Somers, 1988). Respecting such a choice on the quality of life that transcends the apprehended increase in health costs (Bonneux *et al*., 1998) implies the abandonment of prolonging life by medical means and reorienting the current priorities of biomedical research on mortal diseases, such as cancer, toward such chronic diseases as Alzheimer's (van de Water *et al*., 1995; van de Water, 1997). Even more, it implies that in aging societies, priorities for the health of old people should move farther toward *caring* than toward *curing*, and that health personnel should learn to listen as much as to cure.

For some people, living for 120 years (which biogenetic engineering, having already proved itself in the veterinary domain, appears to make possible) may seem an objective. Yet there is still so much left to do in order to make all individuals equal in relation to the risk of death, even in our advanced societies but especially in the Third World countries, where the differences in mortality and morbidity are enormous compared with those that are industrialized.

4. Aid to Old People Who Are Losing their Autonomy: The Role of Informal Support Networks

Loss of autonomy is an evolving process for old people. Most people of 65 or over are not impaired—something that is too frequently forgotten. Subsequently, the need for assistance is gradually felt both for activities of daily living as defined by Sidney Katz *et al*. (1963):

Attending to personal care (feeding, washing, dressing, and going to the toilet unaided);
Getting into and out of bed;
Moving about the house.

Also important are the activities of domestic life, called instrumental activities of daily living as defined by Lawton and Brody (1969), such as,

Preparing meals;
Doing the shopping;
Carrying out domestic chores;
Managing personal finances;
Moving about outside the home.

Also included are activities of everyday life (Carrière *et al*., 1996), a mixture of certain activities from the two preceding categories.

Because, in light of the implied costs, institutionalization has increasingly become a solution of last resort, the leaders of aging societies wish to reinsert aid services for old people into the community and, in particular, to develop home care. They hope, moreover, that the informal networks (inferring principally families) will make up for the formal networks whether private or public, and ideally readopt the role that was allotted to them in former or more traditional societies. Now, in the course of the aging process, populations and families have themselves undergone considerable mutations; conjugal and parental life have been transformed in a radical way. The old people of tomorrow will have seen their married life more frequently ended by divorce. As for the parental dimension, even though in the succession of cohorts the decline in fertility was less brutal than in the period perspective, the size of the families fell by at least 50% (Carrière *et al.*, 2003). Lastly, given the extreme mobility of modern populations, children frequently do not live in the same region as that of their old disabled parents. In consequence, health service managers will have to apply themselves to developing the organization of formal aid networks for old people.

IV. SOME CULTURAL CONSEQUENCES OF DEMOGRAPHIC AGING

Two principles among the social characteristics that distinguish societies across the world and within multicultural countries are ethnic groups and religions. Given their demographic behavior during the second demographic transition (van de Kaa, 1987), several industrialized countries have had, are having and will have to resort to important international immigration—especially from southern countries—in order to mitigate partly their accelerated aging and apprehended negative growth. Societies, often monolithic from the point of view of ethnic groups and religions, will have to adapt to a new status of multicultural societies.

For rulers, the management of an aging society has not the same implications in Pakistan as in Canada or in China, both in the management of the aging phenomenon itself and for the services offered to old people. The role of the state in relation to the demographic problem is often ambiguous, especially for cultural reasons. In many Third World regions, an aging society will have to be administered even though the young population is still exploding, which is obviously not the case with industrial countries. The same will apply to the questions of the rights and duties of old people (UN, 2002).

It is first necessary to distinguish old people who are healthy from those who are not. The social role of the first will vary tremendously from one culture to another, in particular according to the sex and the matrimonial status of the individual. The independence observed for aged women in most industrialized countries, whether economic or social, has still a long way to go before being currently accepted in societies in which the emancipation of women is still in question. The role of the close or extended family will vary considerably from one culture to another.

This latter point becomes fully significant at the moment when an old person becomes disabled or ill, especially in multicultural societies. Contacts among old people, their family, and the medico-hospital system will vary enormously according to the ethnic group and religion. In addition, the services offered by immigrants in industrialized countries will often be more sympathetic than are those sometimes by the nationals. In contrast, the probable incompatibility between the cultural baggage of the person treated and that of the person treating greatly endangers the effectiveness of such a measure. Lastly, the decision to place a person in an institution will raise many problems according to the cultural environment of the old person.

As to death and the circumstances surrounding it, this then touches the foundation of the cultures, whether it concerns a premature decease such as with a child or the death of an elderly person. Religion and destiny are clearly opposed to the more individualistic positions of modern societies. Over time, it is individuals and not the collective households and their standards that have chosen the partner to the union, the number of children they will have, and the divorce in case of a failed marriage. It is difficult to see how society will be able to stop individuals in the near future from choosing the moment of their death, if they so desire (Légaré and Marcil-Gratton, 1990). The medical means and palliative care (its counterpart) appeal to values that differ enormously according to the cultures, so that those concerned with the health and social services must be very aware of these problems in multicultural societies.

CONCLUSION

The aging of populations is a great step forward and a benefit for humanity. And aging societies are here to stay. One of the main effects of this development is that several generations are now present at the same time and quite often for a fairly long period, whereas, in the past, two generations at most rubbed shoulders for a good part of their life cycle. This is not without a major implication at the level of intergenerational relationships. Can intergenerational conflicts be avoided (Kessler, 1996)? The best way to achieve this is to manage these aging societies in a climate of intergenerational ethics, which will only be achieved by a reasonable equilibrium between intergenerational solidarity and equity.

Acknowledgment

This chapter was translated from the French version by Ronald Wilkens.

References

Association internationale pour l'étude de l'économie de l'assurance [AIEEA], 1996. Studies on the Four Pillars, *The Geneva Papers*, vol. 21(81), p. 435–580.

Auerbach Alan J., Gokhale Jagadeesh, and Kotlikoff Larence J., 1994. Generational accounting: a meaningful way to evaluate fiscal policy, *Journal of Economic Perspectives*, vol. 8(1), p. 73–94.

Auerbach Alan J., Kotlikoff Larence J., and Leibfritz Willi (eds.), 1999. *Generational accounting around the world*. Chicago, University of Chicago Press, 534 p.

Barer Morris L., Evans Robert G., and Hertzman Clyde, 1995. Avalanche or glacier? Health care and the demographic rhetoric, *Canadian Journal on Aging*, vol. 14(2), p. 193–224.

Bonneux Luc, Barendregt Jan J., Nusselder Wilma J., and van der Maas Paul J., 1998. Preventing fatal diseases increases healthcare costs: cause elimination life table approach, *British Medical Journal*, vol. 316(1), p. 26–29.

Bourdelais Patrice, 1993. *Le nouvel âge de la vieillesse: histoire du vieillissement de la population*. Paris, Odile Jacob, 441 p.

Carrière Yves, Jenkins Eric, Gupta Neeru, and Légaré Jacques, 1996. The needs for home care services for the elderly in a context of population aging, *in: Facts, research and intervention in gerontology: home care*, p. 31–52. Paris, Serdi Publisher Company.

Carrière Yves, Légaré Jacques, Keefe Janice, Rowe Geoff, Martel Laurent, Lin Xiaofen, and Rajbhandary Sameer, 2003. *Microsimulations and the effect of changing family structure on the needs for formal home care services*. Ottawa, Statistics Canada.

Coale Ansley J., Demeny Paul, and Vaughan Barbara, 1983. *Regional model life tables and stable populations, second edition*. New York, Academic Press, 496 p.

De Jong Gierveld Jenny, Blommesteijn Marieke, and de Valk Helga, 2003. Social distinctions between older persons, the effects of gender and the gender system, *Genus*, vol. 59(1), p. 105–136.

Desjardins Bertrand and Légaré Jacques, 1984. Le seuil de la vieillesse: quelques réflexions de démographes, *Sociologie et sociétés*, vol. 16(2), p. 37–48.

Dirn Louis and Mendras Henri, 1984. Le 3e âge animera la société française, *Futuribles*, no. 80, p. 3–29.

Duchêne Josianne and Wunsch Guillaume, 1990. Les tables de mortalité limite: quand la biologie vient au secours du démographe, *in:* Michel Loriaux, Dominique Rémy, and Éric Vilquin (eds.), *Populations âgées et révolution grise. Les hommes et les sociétés face à leurs vieillissements*, p. 321–332. Louvain-la-Neuve, Éditions CIACO, 1118 p.

Egidi Viviana, 2003. Health status of older people, *Genus*, vol. 59(1), p. 169–200.

Fries James F., 1980. Aging, natural death, and the compression of morbidity, *New England Journal of Medicine*, vol. 303(3), p. 130–135.

Fries James F., 1989. The compression of morbidity: near or far?, *Milbank Memorial Fund Quarterly*, vol. 67(2), p. 208–232.

Furetière Antoine, 1690. *Dictionnaire universel contenant généralement tous les mots françois tant vieux que modernes, et les termes de toutes les sciences et des arts: divisé en trois tomes / recueilli et compile par feu Messire Antoine Furetière, abbé de Chalivoy, de l'Académie Françoise*. La Haye, Arnout & Reinier Leers, 3 vol., s.p.

Gauthier Hervé, 1995. Variables démographiques et charges sociales: comparaisons annuelles et intergénérationnelles, *Cahiers québécois de démographie*, vol. 24(2), p. 285–321.

Gaymu Joëlle, 2003. The housing conditions of elderly people, *Genus*, vol. 59(1), p. 201–226.

Gee Ellen M. and Gutman Gloria M. (eds.), 2000. *The overselling of population aging: apocalyptic demography, intergenerational challenges, and social policy*. Oxford, Oxford University Press, 241 p.

Gillion Colin, Turner John, Bailey Clive, and Latulippe Denis (eds.), 2000. *Social security pensions: development and reform*. Geneva, ILO, 769 p.

Golini Antonio, 2002. Teaching demography of aging, *Genus*, vol. 58(3–4), p. 135–163.

Golini Antonio, 2003. Current demographic setting and the future of aging: the experience of some European countries, *Genus*, vol. 59(1), p. 15–49.

Guillemard Anne-Marie, 2003. *L'âge de l'emploi: les sociétés à l'épreuve du vieillissement*. Paris, Armand Colin, 286 p.

Guillemard Anne-Marie, Légaré Jacques, and Ansart Pierre (eds.), 1995. *Entre travail, retraite et vieillesse: le grand écart*. Paris, l'Harmattan, 326 p.

Henripin Jacques and Loriaux Michel, 1995. Le vieillissement: discours à deux voix, *Population*, vol. 50(6), p. 1591–1638.

Jacobzone Stéphane, Cambois Emmanuelle, Chaplain Emmanuel, and Robine Jean-Marie, 1998. *The health of older persons in OECD countries: is it improving fast enough to compensate for population ageing?* Paris, OECD, Employment, Labour and Social Affairs Committee, 65 p.

Katz Sidney, Ford Amasa B., Moskowitz Roland W., Jackson Beverly A., and Jaffe Marjorie W., 1963. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function, *Journal of the American Medical Association*, vol. 185(12), p. 914–919.

Kessler Denis, 1996. Preventing conflicts between generations: for an active management of the allocation of resources between generations, *The Geneva Papers*, vol. 21(81), October, p. 435–468.

Laslett Peter, 1989. *A fresh map of life: the emergence of the third age*. London, Weidenfeld & Nicolson, 213 p.

Lawton Powell M. and Brody Elaine M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living, *The Gerontologist*, vol. 9(3), p. 179–186.

Légaré Jacques, 1991. Vers une démocratisation de plus en plus grande de la vieillesse: y a-t-il raison de tant s'inquiéter? *in:* Jacques Henripin and Yves Martin (eds.), *La population du Québec d'hier à aujourd'hui*, p. 139–143. Montreal, Les Presses de l'Université de Montréal, 216 p.

Légaré Jacques, 2001a. Ageing and social security program reforms: Canada in international perspective, *Canadian Journal of Policy Research (ISUMA)*, vol. 2(2), p. 110–118.

Légaré Jacques, 2001b. Population aging and its economic and social consequences, general, *in:* Neil J. Smelser and Paul B. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences*, vol. 17, p. 11,737–11,741. Oxford, Pergamon, 16,695 p.

Légaré Jacques, Bourbeau Robert, Desjardins Bertrand, and Deblois Chad, 2005. Variation in cohort size and lower mortality in the elderly: implications for pay-as-you-go health care systems, *in:* Yi Zeng *et al*. (eds.), *Longevity and Health*. New York, Kluwer p. 305–318.

Légaré Jacques and Carrière Yves, 1999. Dying healthy or living longer: a society's choice, *in:* Jean-Marie Robine, Bernard Forette, Claudio Franceschi, and Michel Allard (eds.), *Research and perspectives in longevity: the paradoxes of longevity*, p. 123–133. Berlin, Springer Verlag, 137 p.

Légaré Jacques and Desjardins Bertrand, 1987. Pour une remise en question de l'universalité de l'âge normal de la retraite, *European Journal of Population*, vol. 3(2), p. 123–129.

Légaré Jacques, Ibrahima Mahamane, Sossa Olivier, and Smuga Mélanie, 2002. *Le vieillissement démographique: un défi pour les pays en développement*. Montreal, International Federation on Ageing, 100 p.

Légaré Jacques and Marcil-Gratton Nicole, 1990. Individual programming of life events: a challenge for demographers in the 21st century, *Annals of the New York Academy of Sciences*, vol. 610, p. 98–105.

Légaré Jacques, Marcil-Gratton Nicole, and Carrière Yves, 1991. Vieillir en emploi: un choix inscrit dans l'avenir démographique du Québec, *in:* Rodrigue Blouin, Gilles Ferland, Alain Larocque, Claude Rondeau, and Lise Poulin-Simon (eds), *Vieillir en emploi*, p. 11–29. Sainte-Foy (Quebec), Les Presses de l'Université Laval, 199 p.

Légaré Jacques and Martel Laurent, 2003. Living arrangements of older persons in the early nineties: an international comparison, *Genus*, vol. 59(1), p. 85–103.

Loriaux Michel, Remy Dominique, and Vilquin Éric (eds.), 1990. *Populations âgées et révolution grise: les hommes et les sociétés face à leurs vieillissements*. Louvain-la-Neuve, Éditions CIACO, 1118 p.

Marcil-Gratton Nicole and Légaré Jacques, 1987. Vieillesse d'aujourd'hui et de demain. Un même âge, une autre réalité? *Futuribles*, no. 110, p. 3–21.

Marcil-Gratton Nicole and Légaré Jacques, 1992. Will reduced fertility lead to greater isolation in old age for tomorrow's elderly?, *Canadian Journal on Aging*, vol. 11(1), p. 54–71.

Martel Laurent and Carrière Yves, 2003. Descriptive analysis of the effects of marital dissolution on living arrangements of older persons, the Canadian case and an international comparison, *Genus*, vol. 59(1), p. 137–167.

Martin Linda G. and Preston Samuel H. (eds.), 1994. *Demography of aging*. Washington, National Academy Press, 411 p.

Myers George C. and Légaré Jacques, 2000. The behaviour and status of older persons, *in:* United Nations (ed.), *Generations and gender programme: exploring future research and data collection options*, p. 105–124. New York, United Nations, 166 p.

Myers George C. and Manton Kenneth G., 1984. Compression of mortality: myth or reality?, *The Gerontologist*, vol. 24(4), p. 346–353.

National Research Council [NRC], 2001. *Preparing for an aging world: the case for cross-national research*. Washington, National Academy Press, 308 p.

Organization for Economic Cooperation and Development [OECD], 1996. *Aging in OECD countries: a critical policy challenge*. Paris, OECD, 111 p.

Organization for Economic Cooperation and Development [OECD], 1998. *Maintaining prosperity in an aging society*. Paris, OECD, 141 p.

Organization for Economic Cooperation and Development [OECD], 2000. *Reforms for an aging society*. Paris, OECD, 217 p.

Organization for Economic Cooperation and Development [OECD], 2001. *Aging and income: financieal resources and retirement in 9 OECD countries*. Paris, OECD, 187 p.

Queisser Monika, 2000. Pension reform and international organizations: from conflict to convergence, *International Social Security Review*, vol. 53(2), p. 31–45.

Robine Jean-Marie and Romieu Isabelle, 1998. *Healthy active ageing: health expectancies at age 65 in the different parts of the world*. Montpellier, INSERM and REVES, 58 p.

Ryder Norman B., 1975. Notes on stationary populations, *Population Index*, vol. 41(1), p. 3–28.

Smeeding Timothy M., 2003. Income maintenance in old age: current status and future prospects for rich countries, *Genus*, vol. 59(1), p. 51–83.

Somers Anne Ramsey, 1988. Aging in the 21st century: projections, personal preferences, policies: a consumer view, *Health Policy*, vol. 9(1), p. 49–58.

Stone Leroy O., 1999. *Cohort Flow and the Consequences of Population Aging, An International Analysis and Review*, Ottawa, Statistics Canada, CD-ROM.

Suzman Richard M., Willis David P., and Manton Kenneth G. (eds.), 1992. *The oldest old*. New York, Oxford University Press, 444 p.

Trottier Helen, Martel Laurent, Houle Christian, Berthelot Jean-Marie, and Légaré Jacques, 2000. Living at Home or in an Institution: What Makes the Difference for Seniors?, *Health Reports*, Statistics Canada, Spring, vol. 11(4), p. 49–61.

United Nations Economic Commission for Europe [UN-EEC], 1992. *Demographic causes and economic consequences of population aging: Europe and North America*. New York, UN-EEC, 483 p.

United Nations Economic Commission for Europe [UN-EEC], 1994. *Social aspects and countries reviews of population aging: Europe and North America*. New York, UN-EEC, 447 p.

United Nations [UN], 1956. *The aging of Population and its Economic and Social Implications*. New York, UN, Department of Economic and Social Affairs, 168 p.

United Nations [UN], 1988. *Economic and social implications of population aging*. New York, UN, Department of International Economic and Social Affairs, 430 p.

United Nations [UN], 2001. Living Arrangements of Older Persons, *Population Bulletin*, no. 42/43 (special issue), 1–392 p.

United Nations [UN], 2002, Report *of the Second World Assembly on Aging Madrid, 8–12 April 2002*. New York, UN, 79 p.

van de Kaa Dirk J., 1987. Europe's second demographic transition, *Population Bulletin*, vol. 42(1), p. 1–59.

van de Water Harry P. A., 1997. Health expectancy and the problem of substitute morbidity, *Philosophical Transactions of the Royal Society of London. Series B Biological Sciences*, vol. 352(1363), p. 1819–1827.

van de Water Harry P.A., Van Vliet H.A., and Boshuizen Hendriek C., 1995. *The impact of substitute mortality and morbidity on public health policy*. Leiden, TNO Prevention and Health, 85 p.

Véron Jacques, Pennec Sophie, and Légaré Jacques (eds.), 2004. *Âge, générations et contrat social*. Paris, INED, 312 p.

World Bank, 1994. *Averting the old age crisis: policies to protect the old and promote growth, A World Bank policy research report*. New York, Oxford University Press, Inc., 402 p.

Zweifel Peter, Felder Stefan, and Meiers Markus, 1999. Ageing of population and health care expenditures: a red herring? *Health Economics*, vol. 8(6), p. 485–498.

CHAPTER

83

The Economic Consequences of Migration

MANON DOMINGUES DOS SANTOS

Centre de recherché en economie et statistique (CREST), Université de Paris, Paris, France

INTRODUCTION

In 2002, 175 million people were living outside of their country of birth or citizenship, representing only 3% of world population (United Nations, 2002). Should we conclude from the relative weakness of recorded migration that international migration does not constitute (or no longer constitutes) an important economic phenomenon? Certainly not.

As a first objection, it is useful to note that this global evaluation conceals quite pronounced regional differences: although foreign-born persons represent only 1.5% of the population of southeast Asia or Latin America, they constitute 5% of the European population, 8.6% of population of North America and 18% of the population of Oceania. International migration thus constitutes a major contemporary phenomenon for a good number of countries.

Moreover, observed migration very likely represents only a part of potential migration. Over the past quarter century, numerous traditional receiving and sending countries have established restrictive and selective migration policies that strongly limit migration. Indeed, as in he past, migration generates tensions that push legislators to want to regulate these flows. Natives in the receiving countries often fear the impact of immigration on their well-being and most especially on pay and employment opportunities. In contrast, in the sending countries emigration is often perceived by those who do not move as a loss of their most capable compatriots, which is able to compromise or slow their development.

Thus, the question of international migration constitutes a major political issue. To what degree can economic analysis allow us to understand the consequences of migration, in order to evaluate the pertinence of the migration policies that have been put in place? This chapter seeks to respond to this question.

The economic literature devoted to migration has experienced remarkable growth over the past 25 years.[1] There is, nonetheless, a marked dichotomy between, on the one hand, contributions focused on the determinants of the migration choice and the individual characteristics that may influence that choice and, on the other hand, more macroeconomic approaches analyzing the consequences of the observed migration flows for incomes, employment, or growth, without reference to an explicit microeconomic choice. The first set of contributions generally considers migration to be a rational economic choice, the individual *freely* choosing to locate in the region where his well-being is highest, taking into account possible migration costs. These works provide us with information on the qualitative and quantitative characteristics of unconstrained migration flows, that is, flows freed from the effects of all government inter-

[1] See *Handbook of Population and Family Economics* (1997), Volume 1B, Editions North-Holland.

ventions that could shut off or modify them. Substantial differences can be observed between this hypothetical migration, which would result directly from the implementation of individual choices, and actual migration, most notably when international migration is concerned. In fact, international migration is subject to migration policies seeking to facilitate entry of some people and to prevent the arrival of other people, depending on criteria that are sometimes economic but also political (right to exile) or social (family reunification). Migration policies thus constitute a selection prism between hypothetical migration and actual migration between nations.

It is useful to evaluate the pertinence and efficacy of this prism. More specifically, confronted with a relative deterioration of employment opportunities for unskilled workers, but also facing ongoing population aging, numerous developed countries, such as Australia, Canada, and the United States, have already adopted selective migration policies favoring immigration of skilled workers. Thus, the traditional welcoming countries are looking to limit immigration of less-skilled workers but stimulate that of elites, which the sending countries disapprove of. It therefore should be noted that, at present, the skills of migrants constitute the central focus of migration policies. A question thus emerges: to what degree do the relative skills of migrants *effectively* influence the economic performances of the countries involved? The objective of this contribution is precisely to spell out the responses to this question, theoretical and empirical, that economics can provide.

To understand the impact of migrant skills, a two-step procedure can be adopted. This procedure consists first in analyzing the determinants of salaries, of unemployment rates, and of national growth rates in order to be able to evaluate the degree to which the skills of migrants influenced these determinants. But worker skills are an established essential determinant of employment and growth. Training constitutes an effective bulwark against unemployment and a means to earn a higher salary. The accumulation of knowledge is likewise perceived as one of the principal engines of economic development. Thus, observing that worker skills are indeed a major determinant of economic prosperity is tantamount to acknowledging that the impact of migration very much depends in an essential way on the skills of migrants. Understanding the consequences of migration therefore requires that we study how skill level interacts with the migration choice in order to be able to define the relative skill level of migrants, with regard to their country of origin and of destination. It will then be possible to understand and evaluate the economic consequences of the migration flows on the countries involved.

The first part will thus deal with the lessons from theoretical and empirical work analyzing the skill level of migrants, and the second part will address the consequences of migration in the labor market. Finally, in the last part, the links between migration and economic growth will be made explicit.

I. IS MIGRATION SELECTIVE?

Worker skills being one of the keys to prosperity, it is necessary to know and evaluate the extent to which migration can be determined to be selective. That is, do the skill levels of emigrants and immigrants differ from those of their counterparts who remain or of the people in the country of destination? To illuminate this question, we begin by reviewing the lessons from work looking at the determinants of the migration choice before examining results of empirical studies.

1. Skills and Migration Choice

Most work bases the decision to migrate on a simple axiom: a worker will choose to locate in the region that assures him the highest well-being, taking into account possible migration costs. Thus, as Larry Sjaastad showed as early as 1962, migration may be considered as an investment, resulting from a weighing of the expected benefits and the costs that it entails. It is thus useful to analyze the impact of level of skill on the propensity to migrate by first analyzing the impact of skill on returns and then on the costs incurred by the move.

a. Skills and Expected Returns

Conforming to human capital theory, initiated by Gary Becker, individuals can choose to improve themselves in order to acquire the abilities that will enhance their productivity. Education facilitates their entry to the labor market and allows them to seek higher-paying jobs. The best-trained workers thus have an incentive to locate in an area where their skills are well remunerated. Consequently, the impact of skill level on the propensity to migrate is ambiguous, a priori: it depends on the relative earnings of human capital in the potential regions of residence (see Roy, 1951; Katz and Stark, 1987; Heckman and Honoré, 1990; Borjas *et al.*, 1992). In short, when the most prosperous regions provide a higher return to skill, it is the most skilled workers who have an incentive to migrate there, whereas in the opposite case, it is the least trained workers who are motivated to migrate.

b. Skills and the Costs of Migration

Emigration may also turn out to be costly. Migration engenders financial costs that one must be able to finance, such as transportation costs, liquidation of real estate, or the search for new housing or new employment. Migration also entails psychological costs attributable to being far from friends and family, to the necessary adaptation time, and to the difficulties inherent in integration to a new environment. But there are a number of arguments that suggest that migration costs diminish as the individual's skill level rises. First of all, if the migration entails substantial financial costs, the potential migrant may be subject to a liquidity constraint that prevents him from financing his move. Thus, as emphasized by John Vanderkamp (1971) or Gary Fields (1979), if current income, highly correlated with the initial skill level, is too low, this may prevent a worker from realizing a move that otherwise is optimal. Moreover, skilled workers generally have access to a more-developed network of acquaintances that, among other things, allows them to reduce their costs of job search or housing (Schwartz, 1973; Inoki and Suriban, 1981; Da Vanzo, 1983). Finally, knowledge of a foreign language can reduce migration costs and promote integration.

Theoretical work devoted to migration choices thus tells us that the selectivity of migration with regard to skill depends a priori on the relative returns to human capital in the regions of potential location. And this is so even if the costs of migration may be larger for less-skilled workers.

2. An Empirically Established Selectivity

To analyze the consequences of migration for the economies of sending and receiving countries, it is essential to estimate if the average skills of emigrants and immigrants differ from those of their counterparts who remain or those of natives living in the destination regions.

a. Relatively Skilled Emigrants

A considerable consensus emerges from the empirical work examining the impact of individual characteristics on the propensity to emigrate. The majority of studies show, in effect, a positive influence of the skill level of workers on their propensity to migrate (for the United States, see Nakosteen and Zimmer, 1980; for the United Kingdom, see Gordon and McCormick, 1981; for Japan, see Inoki and Surugan, 1981; for Cote d'Ivoire, see Vijverberg, 1993; for Kenya, see Hoddinott, 1994). William Carrington and Enrica Detragiache (1998, 1999), in a study of 61 developing countries, show that for some of these countries, the emigration rate of the most skilled workers is especially high, a phenomenon commonly referred to as a brain drain. A good number of countries, especially the small countries of Africa, the Caribbean, or South America, have lost more than 30% of their natives who have attained tertiary-level education. In Guyana, the country statistically most affected by this phenomenon, close to 70% of workers with a tertiary education have emigrated to the United States.

In total, emigrants appear to be relatively skilled compared with their those compatriots who remain. This selectivity of migration with regard to skill level can thus be explained by the importance of the phenomenon of diminishing migration costs as the level of skill rises.

b. Relatively Unskilled Immigrants

Evaluation of the relative capabilities of immigrants raises numerous methodological problems, as reflected by the numerous contributions on this theme.

A first approach consists of simply comparing the skill levels (diploma obtained or number of years of school completed) reported by immigrant workers and natives in specific surveys. Such a procedure usually shows that the natives are more skilled than are the immigrants (Table 83–1). The pertinence of this procedure is, however, limited. Indeed, one must think about the question to which an immigrant worker is responding, a worker who has done all or a part of his schooling in another country, when asking him about his level of schooling. Does he reply using a conversion scale that adequately compares diplomas given in his country of birth to those given in his present country? Can he evaluate the diploma that he would have been able to have given his capabilities? Moreover, can one compare the accumulated abilities from a year of study when the school systems are different? Are these abilities really transferable?

Mitigating the limits of this first approach, numerous authors, beginning with the original contribution of Barry Chiswick (1978), seek to estimate the initial relative skill level of immigrants and the evolution of their capabilities during their residence by evaluating the evolution of their performance in the labor market.

A number of these authors base their results on *cross-sectional analysis* (for Germany, see Dustmann, 1993; for Canada, see Baker and Benjamin, 1994). In simplifying, this approach consists, at a date t, of classifying immigrants according to the duration of their

TABLE 83–1 Percentage of Individuals Having Less than a Secondary-Level Diploma

Country	Foreigners	Natives
United States	35	15.7
Germany	48.5	13.2
France	63.3	33.4
Canada	22.2	23.1

Data from OECD, 2000.

stay, D. It rests on the central hypothesis that the earnings received after x years by an immigrant who arrived at date t are equivalent to the earnings received at date t by a worker having immigrated x years previously. The procedure consists, in short, of estimating the following equation:

$$\ln (w_i) = \alpha_1 X_i + \alpha_2 I_i + \alpha_3 D_i + \varepsilon_i,$$

where w_i represents the wage received by individual i, X_i is the vector of his socioeconomic characteristics such as education (in addition to age and sex), I_i is a dummy variable equal to one if the individual is an immigrant and zero otherwise, and D_i is the number of years since immigration if the individual is an immigrant and otherwise equal to zero.

Generally, α_2 is estimated to be negative and α_3 is estimated to be positive. In his study using American data for 1970, Barry Chiswick found that at their arrival, immigrants earned 17% less than did natives, but this gap diminished by 1% per year. Chiswick concluded that at their arrival immigrants experienced a loss of productivity, largely owing to the nontransferable nature of a part of their knowledge. Then, in the course of their stay, they accumulate human capital, which allows them to become assimilated with the native work force (Lalonde and Topel, 1991).

The interpretation of these results is, however, subject to caution. As George Borjas (1995) has emphasized, the sign of the coefficient α_3 may be owing to a cohort effect. That is, it may reveal a reduction in the relative skill level of cohorts, with the most recent immigrants being less qualified than the earlier ones. Longitudinal studies allow one to circumvent this constraint. They follow the same cohort of immigrants and natives over time, allowing one to observe individuals for a long period. For the United States, George Borjas (1999) shows that in 1960, immigrants earned, on average, 4% more than natives, whereas in 1990, they earned 16.3% less. According to the author, close to half of this reduction in the relative earnings of immigrants can be explained by changes in their observable characteristics, and especially their level of education. Further, George Borjas shows that in 1960 immigrants who had emigrated less than 6 years earlier earned 13.9% less than did natives compared with 38% less for immigrants who had arrived less than 6 years earlier in 1990. These results appear to confirm the lowering of the relative skill level of new cohorts of immigrants and the existence of a process of assimilation contingent on the duration of the immigrant's stay.

At the end of this first part, an observation is evident: although emigrants are more skilled than those compatriots who remain, immigrants are in general less qualified than are the native workers, even if they have a tendency to assimilate with the nonimmigrant population. Until recently, effective migration represented a departure of relatively skilled workers from the countries of emigration and arrival of comparatively less skilled workers in the receiving counties. Clearly, then, emigration had a tendency to reduce the size of the working population and the share of skilled workers, whereas immigration will have contributed to a larger labor force and growth in the share of less-skilled workers.

II. DOES MIGRATION INFLUENCE SALARIES AND EMPLOYMENT?

Evaluation of the influence of immigration on the well-being of native workers provokes much debate, often polemic, and made concrete by an abundant literature. Two hypotheses may, however, be envisaged. Certain authors advance the hypothesis of *replacement*, according to which immigrant workers substitute for natives. Immigration thus contributes to reducing earnings of natives and to increasing their likelihood of unemployment. Other authors favor instead the hypothesis of *segmentation*, according to which immigrants have jobs that natives do not want, or they would generate, in the long run, a demand for labor equivalent to the number of jobs they hold. To evaluate the pertinence of these two hypotheses, it is useful to review the lessons from economic theory in order to interpret the principal results of empirical work.

1. Lessons from Theory

The impact of migration on the conditions of remuneration and employment depend, of course, on the horizon that one is considering. Salaries and company hiring policies take time to adapt to a change in their environment. Taking account of these adjustment delays, it is useful to distinguish the short-term influence of migration from the long-term effects.

a. Short-Term Impact of Migration

The Lessons from Neo-Keynesian Models

With regard to the conditions prevailing in the labor market, an inflow or substantial departure of workers is like a shock that affects the size and the structure of the supply of labor. But whatever the manner in which the labor market being considered is regulated, earnings always take a certain time to react to a change in the economic environment. In the same manner, enterprises observe a certain delay before changing their production methods. Most traditional macroeconomic models, seeking to represent the short-run functioning of an economy, thus assume that prices, salaries, and productive capital are fixed. Their functioning may be summarized as such. In the short term, with prices being assumed fixed, it is the demand for goods and services coming from households, firms, and the government that determines output and employment. Unemployment thus results from the imbalance between the demand for labor needed to satisfy the demand for goods and services, given the state of technology, and the supply of labor, largely determined outside of the model. From a longer-term perspective, the *wage-price spiral* plays an important role. The unemployment rate affects the rate of growth of earnings, by virtue of the Phillips curve[2], and the evolution of wages and salaries has implications for prices. Inflation then feeds back on the demand for goods and services.

In the context of these neo-Keynesian–inspired models, immigration is analogous to an exogenous increase in the supply of labor and emigration is a decrease in supply. Thus, immigration[3] immediately increases the number of workers without affecting the level of employment, because prices and wages are fixed. The inflow of workers translates to an increase in unemployment, because the same number of jobs must be divided among a greater number of workers. However, this increase is only transitory. The increase in the unemployment rate moderates salary demands: wages fall and prices follow. This then stimulates demand for goods and services and hence for labor, this latter effect thereby contributing anew to reduce the unemployment rate. In the end, the reduction in prices is such that the initial effect on the unemployment rate is entirely offset. Immigration thus has only a transitory impact on the unemployment rate, and this effect persists less when prices and wages adjust rapidly.

By way of illustration, for the case of France, in relying on the estimates by Pierre Cahuc and André Zylberberg (2001) of the Phillips curve for the period 1970–1998, one can calculate that immigration resulting in an increase in the growth of the labor force by 1% would lead to a short-term increase in the unemployment rate by 2.4%. However, estimation of an average delay of adjustment of wages to prices of less than a semester (Artus and Muet, 1995) leads one to anticipate weak persistence of this effect.

The Lessons of Matching Models

Matching models (Pissarides, 1990) postulate that some time is required for a worker to find a job suited to his capabilities and for a firm to find a worker suited to the job vacancy that it seeks to fill. These delays imply that at any moment in time in the labor market there exist both unemployed workers searching for jobs and vacant positions. In this context, if one considers an inflow of immigrants, those who are newly arrived will devote some time to finding a job conforming to their expectations. Immigration[4] thus translates instantly into an equivalent increase in the number of unemployed, which raises the unemployment rate. Nonetheless, this increase from the recruitment pool increases the probability for an enterprise of rapidly filling a vacant position, which will diminish the cost of doing so. Thus, by increasing the profit associated with the creation of a job, immigration stimulates job creation. This process remains operative as long as the gain associated with listing a newly vacant employment remains positive, that is, until the unemployment rate returns to its initial level.

Hence, when wages and prices show a certain delay in adapting to changes in the economic environment and workers take some time to find a job suited to their skills, immigration can temporarily increase the unemployment rate and emigration can temporarily reduce it.

b. On the Long-Term Impact of Migration

The long-term impact of migration on wages and working conditions depends on how the labor market in question is regulated. There is a certain dichotomy between the lessons of Anglo-Saxon work, which is done in the context of a flexible labor market, and the lessons of European-inspired work, which favor the

[2] The Phillips curve shows an inverse relationship between the evolution of nominal earnings and the unemployment rate. More specifically, an increase in unemployment induces a decline in the rate of growth of earnings. One of the usual explanations for this causality is grounded in the process of salary negotiations. Basically, when the unemployment rate rises, employed workers are led to moderate their salary demands.

[3] For emigration, the direction of the effects is opposite.

[4] For emigration, the direction of the effects is opposite.

hypothesis of a rigid labor market characterized by the persistence of substantial unemployment. The former concentrate on the impact of migration on salaries, whereas the latter focus on the consequences of immigration for unemployment.

Flexible Labor Market, Migration, and Wages

To analyze qualitatively the impact of migration on wages, it is useful to consider an elementary model of the supply and demand for labor. This approach assumes that wages adjust to equilibrate supply and demand for each type of labor. The supply of a factor is constituted by the sum of the quantities of the factor held by natives and immigrants. Thus, immigration directly increases the supply of the kind of labor similar in productive characteristics to immigrant labor. The demand for labor coming from firms depends on the cost of labor, technological possibilities for substitution, and their situation in the market for goods and services. These are the properties of factor demands that condition the impact of immigration on the remuneration of all factors of production: the effect of an increase in the supply of a certain factor of production on the payments to all factors will depend principally on the elasticity of substitution between the different types of factors and their share in the total factor costs. Basically, labor immigration contributes to reducing the remuneration of factors of production that are perfectly substitutable for the labor of the immigrants, and it increases remuneration of factors of production that are complementary to it. Thus, in supposing that capital and skilled labor are complementary factors, and that unskilled labor is substitutable (Hamermesh, 1993), immigration of unskilled workers tends to reduce the wages of unskilled workers but increase remuneration of capital and of skilled labor. Furthermore, the emigration of skilled workers tends to increase remuneration of skilled workers and of capital, but it reduces compensation of unskilled natives.

Rigid Labor Market, Migration, and Unemployment

Numerous countries confront substantial unemployment. Wage rigidity and the structure of the working population are largely evoked to explain the persistence of a situation of underemployment. Basically, a country's unemployment rate may be considered as a weighted average of the unemployment rates of different categories of workers. Thus, migration can influence the national rate of unemployment via two mechanisms. Migration can influence category-specific unemployment rates: immigration of unskilled workers can affect the unemployment rate of unskilled native workers, whereas emigration of skilled labor can affect the unemployment of skilled labor who remain. Migration can also modify the respective shares of the different categories of labor in the total labor force. If migration increases or decreases the share of workers having structurally a higher rate of unemployment, this effect, called a *structural effect* or *compositional effect*, can likewise modify the average unemployment rate. Consider successively the mechanisms dealing with these two effects.

Migration, Equilibrium Category-Specific Unemployment Rates, and Unemployment

As previously noted, immigration (emigration) increases (reduces) the size of the labor force with productive characteristics similar to those of the immigrants (emigrants). From this perspective, evaluating whether migration influences the unemployment rate of the category of labor to which the immigrants belong entails examining the long-term links between the unemployment rate and the size of the labor force.

Recent findings in labor economics shed light on the sources of long-term category-specific unemployment. Endogenous mechanisms of salary determination are largely invoked. Thus, salary negotiations between employees and employers or salary policies adopted by firms, confronted with the turnover costs of labor or information asymmetries with regard to the characteristics of workers and the intensity of effort they put forth (for a review, see Cahuc and Zylberberg, 2001), can bring about a salary rigidity that gives rise to unemployment. In a general way, specifying the determinants of wages tends to limit the impact of migration on long-term unemployment rates by category (see Stiglitz, 1974; Calvo, 1978; Schmidt *et al.*, 1984; Bulow and Summers, 1986; Domingues Dos Santos, 1999b). Indeed, in equilibrium unemployment models, equilibrium category-specific unemployment rates turn out to be independent of the size of the labor force being considered. One result thus emerges: immigration (emigration) of a certain type of worker increases (reduces) the number of this type in the labor force, but this increase (reduction) will not have any impact on the long-term unemployment rate of the labor category being considered. Hence, immigration of unskilled workers will increase the number of unskilled workers, but this increase will not, in the long run, have any impact on the unemployment rate of unskilled workers. Correspondingly, emigration of skilled workers reduces the number of skilled workers but will have no long-term impact on the unemployment rate of skilled workers who do not migrate.

Migration, Compositional Effect, and Unemployment

Immigration and emigration can influence average unemployment rates if the migrants have, on average, a tendency to be unemployed that is different from the unemployment of natives or those who do not migrate.

Empirical work shows that immigrants have, on average, a higher unemployment rate than that of natives. For the case of France, the average unemployment rate of immigrants in 2000 was around 20.9%, compared to 9.9% for the entire labor force. This greater vulnerability to unemployment primarily can be explained by their lower skill level, with skill level clearly being one of the most effective bulwarks against the risk of unemployment.

Moreover, the propensity to emigrate seems to be greater when the individual is unemployed. Numerous arguments explain this selectivity of migration with regard to labor force status. Employed workers have, in effect, a higher opportunity cost of emigrating. With regard to speculative migration,[5] they are confronted with a risk of unemployment. They may also lose some of the advantages they had acquired, especially those linked to seniority.

Compositional effects would thus contribute to increasing the average unemployment rate in receiving nations and to reducing it in sending nations, because the tendency of migrants to be unemployed is greater than that of natives or of those who do not migrate.

The lessons from theoretical work lead us to think that migration has a negligible long-term effect on long-term category-specific unemployment rates. However, immigration could produce a transitory increase in unemployment in the receiving region, and emigration could result in a transitory decrease in unemployment in the sending region. It is useful now to evaluate the empirical relevance of these assertions.

2. Empirical Results

Empirical work on the impact of migration on wages and employment has two peculiarities. First, until recently, the majority of these studies looked at immigration to the United States. The specific features of American immigration and the functioning of the American labor market have thus limited the ability to generalize results to other countries. In addition, different econometric methods have been used, with approaches in terms of factor proportions, spatial correlations, and natural experiments being the most widely utilized.

[5] We refer to speculative migration when describing migration that takes place prior to securing employment in the destination.

a. Factor Proportions

The factor proportions approach, initiated by Jean-Baldwin Grossman (1982), entails three steps. The first step consists of determining, given the assumed technology of production, the theoretical labor demand functions. Then plausible values are determined for the elasticities of substitution among the different factors and the share of their remuneration in income as well as the relative skills of immigrant labor, in order to evaluate with which type of labor it is substitutable and the degree to which immigration modified the relative factor endowments. Finally, one can simulate the impact of immigration on the earnings of natives. George Borjas, Richard Freeman and Lawrence Katz (1997) estimated that between 27% and 55% of the reduction in relative earnings of unskilled workers in the United States between 1979 and 1995 could be attributed to immigration, which apparently increased by 15% to 20% the share of these workers in the labor force. In his own work George Borjas (1999) considers a technology with three factors of production: capital, unskilled labor, and skilled labor, with workers having less than a completed high school education being considered as unskilled. Using U.S. data, he estimates that in 1995, skilled workers represented 91% of the total labor force but only 68% of the immigrant population. He estimates that in the short run, that is, with the stock of capital fixed, a 10% increase in the labor force induced by immigration would reduce by 2.29% the earnings of skilled workers, and the earnings of unskilled workers would decline by 3.72%. In the long run, once the capital stock is adjusted by firms, earnings of skilled workers would rise by 0.46%, whereas earnings of unskilled workers would fall by 4.27%. However, the relative weakness of this effect has recently been called into question by George Borjas (2002). In refining the categorization of the work force by level of education and most notably in integrating work experience, he estimates that between 1960 and 2000, an increase in immigration that would have increased the supply of certain types of labor by 10% would have reduced by 2% to 3% the earnings of natives and by 2 weeks the number of weeks worked.

The approach in terms of factor proportions thus concludes that there is an important influence of immigration on the relative earnings of unskilled natives. However, this approach may prove to be unsatisfactory. The applications are all on the

American case, in which the labor market is presumed to be flexible. Further, this approach does not consist of *estimating* the actual influence of immigration but rather of *simulating* that influence by using a counter-factual method that should be viewed with caution.

b. Spatial Correlations

The approach in terms of spatial correlations consists of identifying the correlations that exist at a given moment between observed earnings or employment opportunities in a sample of regions and the density of their respective immigrant populations. In short, if, for example, it turns out that the regions having a high density of unskilled immigrant workers or substantial immigration of this type of worker are those in which earnings of unskilled native workers are the lowest or increase the least, this approach tends to conclude that there is a causal relationship. Work based on this approach typically finds a not very significant or insignificant effect of migration on local employment conditions, as well as a strong dependence on the period being considered (for United States: Lalonde and Topel, 1991; Altonji and Card, 1991; Borjas *et al.*, 1997; Card, 2001; for Germany: De New and Zimmerman, 1994; Pischke and Velling, 1997).

However, the spatial correlation approach presents two drawbacks that limit the significance of certain of the results.

The first is linked to the problem of simultaneity resulting from self-selection of migrants. In brief, if migrants are attracted by the regions where employment opportunities are favorable, observation of a positive correlation between immigration rate and employment opportunities may simply reflect the causes of the migration choice and not the consequences. Utilization of the *method of instrumental variables* permits one to relax this constraint somewhat (Altonji and Card, 1991; Pischke and Velling, 1997). This procedure consists of identifying certain variables that influence the location choice of immigrants but do not affect conditions in the labor market being considered. However, it appears that introduction of instrumental variables has little effect on empirical results: immigration does not seem to influence in any substantial way earnings and employment opportunities of natives.

The second drawback of this approach is to obscure the mobility of native populations that can vote with their feet. Clearly, if immigration pushes natives to emigrate, this will limit the impact of immigration on conditions of employment. However, David Card (2001), for the United States, and Jörn-Steffen Pischke and Johannes Velling (1997), for Germany, estimate that this phenomenon is not very consequential.

Evaluations using the method of spatial correlations thus show generally that the impact of migration on remuneration and employment opportunities of natives is relatively modest or even insignificant.

c. Natural Experiments

The authors of some empirical work base their estimates on what are called natural experiments. These studies exploit the consequences of policies that generated intense migration flows for a specific period in certain regions.

Thus, David Card (1990) evaluated the consequences of the "Mariel boatlift." In 1980, after a disagreement between the Cuban and Peruvian governments, Fidel Castro announced the opening of the port of Mariel for any Cuban desiring to leave Cuba. This announcement provoked an exodus of more than 125,000 Cubans to the United States between April and September 1980, with half of these migrants settling in Miami, where the labor force grew by nearly 7%. To evaluate the consequences of this massive immigration, David Card compared changes in unemployment rates and earnings in Miami with the changes observed between 1979 and 1985 in other cities having similar characteristics, such as Los Angeles, Houston, Atlanta, and Tampa/St. Petersburg. During the 1970s these cities had experienced changes in employment similar to those in Miami, and similar to Miami, they also had important black and Hispanic communities. The similarity of the labor markets of Miami, the experiment city, and the control cities, allows one to attribute to the Cuban immigration the differences in performance of the labor market subsequently observed. David Card concluded from his analysis that changes in unemployment rates and earnings in Miami were very close to those observed in the other cities. In the long term, then, the impact on unemployment and earnings of this immigration, even though it was massive, was apparently negligible.

Another study, done by Jennifer Hunt (1992), looked at the consequences of the flow of French repatriated from Algeria during 1962. After the signing of the Evian agreement granting independence to Algeria, 900,000 French returned to the mother country in 1962, principally in the southern regions. To evaluate the impact of this migration flow on employment opportunities of natives, Jennifer Hunt compared changes in unemployment rates and earnings in the 90 metropolitan departments, controlling for a set of external variables that could also influence the variables studied. She estimated that an increase of 1% in the labor force induced by this immigration would have increased the unemployment rate of nonmigrant natives by 0.2%. Moreover, their average salary in 1967 would have been at most 1.3% lower than what it otherwise would have been. The comparison of the results of David Card and Janet Hunt suggests that the American labor market has a greater absorption capacity than does the French labor market. The rigidities of the European labor market, characterized notably by a more marked protection of employment and higher replacement rates, can be evoked to explain the greater impact of immigration on unemployment rates.

In this regard, work by Joshua Angrist and Adriana Kugler (2003) seems to confirm the role of labor market rigidities on the capacity of concerned regions to absorb a flow of immigrant labor. These authors evaluate the impact of consecutive exoduses from the Bosnia and Kosovo wars on each of the receiving countries, using a database covering the period from 1983–1999. Their results suggest that an increase of 10% in the proportion of immigrants induces a reduction in the rate of employment of natives of between 0.2% and 0.7%. In addition, the authors show that this reduction is more pronounced if the country in question is characterized by a rigid labor market.

The work based on the analysis of natural experiments seems to confirm a moderate impact on the conditions of employment of natives. However, this apparent capacity of receiving economies to absorb the newly arrived appears to be weaker in the presence of institutions limiting the flexibility of the labor market.

At the end of this section, an observation is pertinent: empirical work seems to validate the lessons of theoretical work concerning the impact of migration on earnings and unemployment. The effect of migration on salaries and the propensity of natives to be unemployed appears to be relatively weak. In total, the replacement hypothesis seems to be invalidated: economies are not characterized by a given number of jobs that must be divided among the individuals present. Rather, they have a tendency to create jobs in proportion to the number of individuals who work and consume there.

III. MIGRATION AND GROWTH

Up until now, we have ignored any possible impact of migration on the prospects for growth of the countries involved. But the international mobility of workers, bringing with them their knowledge and a part of their belongings, can influence growth and technical progress in both the sending and receiving regions. Let us be explicit about the pertinent mechanisms here, so as to highlight the results of empirical work.

1. The Lessons of Empirical Work

The impact of migration on growth depends, of course, on its influence on the two principal engines of expansion: capital accumulation (physical and human) and technological progress.

Although the authors who base their analyses on the *Solow growth model*[6] exclude, a priori, any impact of migration on technological progress, because this variable is considered exogenous to the model, they nonetheless make explicit the effects of migration on the dynamics of overall capital accumulation. These works conclude that there is a positive influence of migration on the convergence of standards of living of sending and receiving regions (Barro and Sala-I-Martin, 1995). Two arguments support this conclusion. In the first place, workers have an incentive to migrate to the region where income per capita is highest. But because income per capita will be determined by the per capita stock of capital, migration flows will be from the region where the per capita stock of capital is lowest to the region where it is highest. In the second place, migrants bring with them relatively little capital. In total, then, emigration increases the capital stock per capita and hence income per capita, whereas immigration reduces these two variables.

However, these assertions are worth discussing in view of the selectivity of migration with regard to skill level. Indeed, it is assumed that migrants bring relatively little capital with them. Hence, evaluating the relative capital, physical and human, brought by migrants turns out to be essential.

We know in reality that on average, immigrants are less qualified than are natives. Moreover, if we consider the costs of mobility of physical capital held by individuals before their migration (costs of sale and purchase of personal property and real estate), it is reasonable to think that, on average, immigrants will have an overall stock of capital that is less than that of natives.

Nonetheless, we saw that emigrants on average are more skilled than are their compatriots who remain. Hence, the proposed assertion is only pertinent if emigrants experience substantial losses of physical capital or if they leave to their next of kin a large part of their

[6] The Solow growth model spells out links between growth of per capita income, accumulation of physical and human capital, and population growth. However, this approach considers technological progress as exogenous. By its nature, then, the Solow model excludes any possible influence of capital accumulation or population structure on technological progress.

goods. However, consideration of the financial transfers that emigrants send to their countries of origin can relax this condition. Numerous studies document the importance of the share of their income that emigrants remit to their country of origin (Blitz, 1977; Chandavarkar, 1980; Lebon, 1994; Cashin and Loayza, 1995). Thus, these substantial capital transfers, conditioned by prior emigration, can have a favorable impact on income of the sending region, a priori the less developed region.

However, work on migration in the context of *endogenous growth models* is primarily oriented toward the problem of brain drain. The principal objective is to explain the consequences of emigration of the most skilled workers from developing countries. According to endogenous growth theories, the average skill level of the labor force constitutes one of the principal engines of economic development. Thus, emigration of skilled workers would have a negative effect on both the income of workers who do not emigrate (Miyagawa, 1991) and the rate of growth of the sending region (Haque and Kim, 1995). Andrew Mountford (1997) shows, however, that this effect may be attenuated if the choice of training is taken into account. In brief, because remuneration of a skilled worker is greater in developed countries, a nonzero probability of being able to emigrate toward these regions increases the return to education and thus provides greater motivation for workers to acquire training. If some of these workers cannot emigrate, the effect of greater motivation to acquire training can compensate for the direct adverse effect of the brain drain. Moreover, I have shown (Domingues Dos Santos, 1999a) that it may be optimal for the region where innovation activity is less efficient to allow a part of its skilled manpower to emigrate if the region can benefit from technological advances realized in the receiving region, via a process of imitation. Finally, migration may favor the growth of developing countries if workers return home and diffuse some of the knowledge they acquired during their stay abroad (Domingues Dos Santos and Postel-Vinay, 2003).

Thus, when one excludes a priori any impact of migration on technological progress, migration will tend to favor the convergence of standards of living between the sending and receiving countries. Nevertheless, the brain drain may compromise the growth prospects of developing countries, which limits the generality of this assertion.

2. Results of Empirical Work

It is necessary to acknowledge that up until now, there is little work that tries to evaluate the impact of migration on the convergence and the growth prospects of the regions involved. The lack of homogeneous international data that would allow estimation of these phenomena is certainly one of the essential explanations for this deficiency. Let us nonetheless consider the principal results that have been found.

a. Migration and Convergence

Robert Barro and Xavier Sala-I-Martin (1995) studied the process of convergence for regions of the United States (for the period 1920–1990), for Japan (for the period 1955–1990), and for Europe (for the period 1950–1980), both with and without migration. They show that the differences in estimated convergence speeds are small. The work of Alan Taylor and Jeffrey Williamson (1994), however, suggests that these results are not robust with respect to modification of the period being considered. In carrying out a counterfactual analysis of 17 countries of the old and new world for the period 1870–1910, the authors conclude that 73% of the reduction in the dispersion of output per worker, and 50% of the dispersion of output per capita can be attributed to migration. Work by Moses Abramovitz (1986), Angus Maddison (1991), and Jeffrey Williamson (1995) tends to confirm this assertion. They decompose the period from the mid-19th century to the present into four subperiods based on the criterion of dispersion of per capita income: the initial period, when the dispersion is substantial; the period from 1870–1913, characterized by a convergence of per capita incomes; the interwar period, during which there was substantial divergence; and the period from 1950 to the present, one of strong convergence. But according to Jeffrey Williamson, it is precisely migration that differed across these periods: the reduction in transportation costs resulted in massive migration after 1870, whereas migration flows substantially declined owing to introduction of migration quotas and the increase in transportation costs and the uncertainty linked to the war.

Although work on migration in developed countries for recent periods suggests that the influence of migration flows on convergence is limited, it does seem that there was a considerable influence in earlier periods.

b. Brain Drain

The absence of work seeking to evaluate the consequences of brain drain is especially obvious. Michel Beine, Frédéric Docquier, and Hillel Rapoport (2001) nonetheless shed some interesting light on this question. Exploiting data constituted by William Carrington and Enrica Detragiache (1998), these authors seek to explain investment in human capital and the rate of growth for a sample of 50 developing countries. They confirm the assertion of a motivating effect on training. However, ex post the effect becomes negative once the migration rate exceeds 20% or when the percentage of educated workers is greater than 5%. The consequences with respect to growth remain generally limited, when the brain drain is not too intense.

Work on the links between migration and growth thus teaches us that international mobility of workers should tend to favor convergence when it occurs between relatively developed countries. At the same time, emigration of skilled workers from developing countries tends to limit the prospects for growth of these countries and thus their chance of reaching a level of prosperity comparable to that of better-off countries. The selective migration policies that have recently been implemented by these latter countries can only strengthen this observation.

Acknowledgment

This chapter was translated from the French version by David Shapiro.

References

Abramovitz Moses, 1986. Catching up, forging ahead, and falling behind, *Journal of Economic History*, vol. 46(2), p. 385–406.

Altonji Joseph G. and Card David, 1991. The effects of immigration on the labor market outcomes of less-skilled natives, *in:* M. de John Abowd and Richard B. Freeman (eds.), *Immigration, trade and the labor market*, p. 201–234. Chicago, University of Chicago Press.

Angrist Joshua and Kugler Adriana, 2003. Productive or counter-productive? Labor market institutions and the effect of immigration on EU natives, *The Economic Journal*, vol. 113(488), p. 302–331.

Artus Patrick and Muet Pierre-Alain, 1995. *Théories du chômage.* Paris, Economica, 109 p.

Association internationale pour l'étude de l'économie de l'assurance [AIEEA], 1996. Studies on the Four Pillars, *The Geneva Papers*, vol. 21(81), p. 435–580.

Baker Michael and Benjamin Dwayne, 1994. The performance of immigrants in the Canadian labor market, *Journal of Urban Economics*, vol. 12(2), p. 369–405.

Barro Robert J. and Sala-I-Martin Xavier, 1995. *Economic growth.* New York, McGraw Hill, 539 p.

Beine Michel, Docquier Frédéric, and Rapoport Hillel, 2001. Brain drain and economic growth: theory and evidence, *Journal of Development Economics*, vol. 64(1), p. 275–289.

Beine Michel, Docquier Frédéric, and Rapoport Hillel, 2002. *Brain drain and LDC's growth: winners and losers.* Lille, CADRE, Universitéde Lille, 48 p.

Blitz Rudolph, 1977. A benefit-cost analysis of foreign workers in West Germany, 1957–1973, *Kyklos*, vol. 30(3), p. 479–502.

Borjas George J., 1995. Assimilation and changes in the cohort quality revisited: what happened to immigrants' earning in the 1980s, *Journal of Labor Economics*, vol. 13(2), p. 201–245.

Borjas George J., 1999. The economic analysis of immigration, in: Orley Ashenfelter and David Card (eds.), *Handbook of labor economics*, vol. 3A p. 1697–1760. Amsterdam, North-Holland, 928 p.

Borjas George J., 2002. The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. Cambridge, National Bureau of Economic Research, 55 p.

Borjas George J., Bronars Steven G., and Trejo Stephen, 1992. Self-selection and internal migration in the United States, *Journal of Urban Economics*, vol. 32(2), p. 159–185.

Borjas George J., Freeman Richard B., and Katz Lawrence F., 1997. How much do immigration and trade affect labor market outcomes, *Brooking Papers on Economic Activity*, vol. 1(1), p. 1–67.

Bulow Jeremy I. and Summers Lawrence H., 1986. A theory of dual labor market with application to industrial policy, discrimination and Keynesian unemployment, *Journal of Labor Economics*, vol. 4(3), p. 316–414.

Cahuc Pierre and Zylberberg André, 2001. *Le marché du travail.* Bruxelles, De Boeck, 714 p.

Calvo Guillermo A., 1978. Urban unemployment and wage determination in LDC's: trade union in the Harris-Todaro model, *International Economic Review*, vol. 19(1), p. 65–81.

Card David, 1990. The impact of the Mariel Boatlift on the Miami labor market, *Industrial and Labor Relations Review*, vol. 43(2), p. 245–257.

Card David, 2001. Immigrant inflows, natives outflows, and the local market impacts of higher immigration, *Journal of Labor Economics*, vol. 19(1), p. 22–64.

Carrington William J. and Detragiache Enrica, 1998. *How big is the brain drain?* Washington, International Monetary Fund/Fonds monétaire international, no. WP/98/102.

Carrington William J. and Detragiache Enrica, 1999. How extensive is the brain drain? *Finance and Development*, vol. 36(2), p. 46–49.

Cashin Paul and Loayza Norman, 1995. *Paradise lost? Growth, Convergence and Migration in the South Pacific.* Washington, International Monetary Fund/Fonds monétaire international, 42 p.

Chandavarkar Anand, 1980. Use of migrant remittance in labor-exporting countries, *Finance and Development*, vol. 17, p. 36–39.

Chiswick Barry, 1978. The effect of Americanization on the earnings of foreign-born men, *Journal of Political Economy*, vol. 86, p. 897–921.

DaVanzo Julie, 1983. Repeat migration in the United States: who move back and who move on?, *Review of Economics and Statistics*, vol. 65(4), p. 552–559.

De New John P. and Zimmermann Klaus F., 1994. Native wage impacts of foreign labor: a random effects panel analysis, *Journal of Population Economics*, vol. 7(2), p. 177–192.

Domingues Dos Santos Manon, 1998. *Migrations, choix de qualification et croissance.* Paris, Université de Paris I, 21 p.

Domingues Dos Santos Manon, 1999a. Diffusion technologique et transferts: vers un drainage des cerveaux optimal, *Recherches économiques de Louvain*, vol. 65(3), p. 323–347.

Domingues Dos Santos Manon, 1999b. Migrations et dynamique du chômage, *Annales d'économie et de statistiques*, vol. 53(1), p. 43–68.

Domingues Dos Santos Manon and Postel-Vinay Fabien, 2003. Migration as a source of growth: the perspective of a developing country, *Journal of Population Economics*, vol. 16(1), p. 161–175.

Dustmann Christian, 1993. Earning adjustment of temporary immigrants, *Journal of Population Economics*, vol. 6(2), p. 153–168.

Field Gary, 1979. Place-to-place migration: some new evidence, *Review of Economics and Statistics*, vol. 61(1), p. 21–32.

Gordon Hughes and McCormick Barry, 1981. Do council housing policies reduce migration between regions?, *The Economic Journal*, vol. 61(364), p. 919–937.

Grossman Jean-Baldwin, 1982. The substitutability of natives and immigrants in production, *Review of Economics and Statistics*, vol. 54(4), p. 596–603.

Hamermesh Daniel, 1993. *Labor demand*. Princeton, Princeton University Press, 464 p.

Haque Nadeem U. and Kim Se-Jik, 1995. Human capital flight: impact of migration on income and growth, *IMF Staff Papers*, vol. 42(3), p. 577–607.

Heckman James J. and Honoré BoE., 1990. The empirical content of the Roy model, *Econometrica*, vol. 58(5), p. 1121–1149.

Hoddinott John, 1994. A model of migration and remittances applied to Western Kenya, *Oxford Economic Papers*, vol. 46(3), p. 459–476.

Hunt Jennifer, 1992. The impact of the 1962 repatriates from Algeria on the French labor market, *Industrial and Labor Relations Review*, vol. 45(3), p. 556–572.

Inoki Takenori and Suruga Terukasu, 1981. Migration, age and education, *Journal of Regional Science*, vol. 21(4), p. 507–517.

Katz Eliakim and Stark Oded, 1987. International migration under asymetric information, *The Economic Journal*, vol. 97(387), p. 718–726.

Lebon André, 1994. Situation de l'immigration et la présence étrangère en France, 93–94. Paris, La Documentation Française, 139 p.

Maddison Angus, 1991. Dynamic forces in capitalist development: a long-run comparative view. Oxford, Oxford University Press, 333 p.

Miyagiwa Kaz, 1991. Scale economies in education and the brain drain problem, *International Economic Review*, vol. 32(3), p. 743–759.

Mountford Andrew, 1997. Can brain drain be go for growth in the source country?, *Journal of Development Economics*, vol. 53(2), p. 287–303.

Nakosteen Robert A. and Zimmer Michael A., 1980. Migration and income: the question of self selection, *Southern Economic Journal*, vol. 46(3), p. 840–851.

Organización para la Cooperación y el Desarrollo Económicos [OCDE], 1999. Trends in international migration. Paris, OCDE Editions.

Pischke Jörn-Steffen and Velling Johannes, 1997. Employment effects of immigration to Germany: an analysis based on local labor markets, *Review of Economics and Statistics*, vol. 79(4), p. 594–604.

Pissarides Christopher A., 1990. *Equilibrium unemployment theory*. Oxford, Basil Blackwell, 173 p.

Roy Andrew D., 1951. Some thoughts on the distribution of earnings, *Oxford Economic Paper*, vol. 3, p. 135–146.

Schmidt Christoph, Stilz Annette, and Zimmermann Klaus, 1994. Mass migration, union, and government intervention, *Journal of Public Economics*, vol. 55(2), p. 185–202.

Schwartz Aba, 1973. Interpreting the effect of distance on migration, *Journal of Political Economy*, vol. 81(5), p. 1153–1169.

Sjaastad Larry A., 1962. The costs and returns of human migration, *Journal of Political Economy*, vol. 70(5), p. 80–93.

Stiglitz Joseph E., 1974. Alternative theory of wage determination and unemployment in LDC's: the labor turnover model, *Quarterly Journal of Economics*, vol. 88(2), p. 194–227.

Taylor Alan M. and Williamson Jeffrey G., 1994. Convergence in the age of mass migration, *NBER Working Paper 4711*.

United Nations, 2002. *International migration report 2002*. New York, United Nations, Department of Economic and Social Affairs, 7 p.

Vanderkamp John, 1971. Migration flows, their determinants and the effects of return migration, *Journal of Political Economy*, vol. 79(5), p. 1012–1031.

Vijverberg Wim P., 1993. Labour market performance as a determinant of migration, *Economica*, vol. 60(238), p. 143–160.

Williamson Jeffrey, 1995. The evolution of global labor market since 1830: background evidence and hypotheses, *Exploration in Economic History*, vol. 32(2), p. 141–196.

CHAPTER

84

Immigration and Integration in the Developed Countries: A Conceptual Framework

VICTOR PICHÉ

Inter-university Centre for Demographic Studies, Ethnicity and Society Research Group, University of Montreal, Montreal, Canada

INTRODUCTION

The pattern of international migration to and from developed countries has changed radically since the early 1970s. Immigration to North America and Europe has gradually originated mainly from the developing countries, producing significant shifts to ethnic and racial compositions (Kuijsten, 1994; Massey *et al*., 1998). Over time, Western countries have awoken to their multiethnicity, mostly through heated public debates on the need (or not) to shut the door on immigration. But at the same time, multiethnicity was not just pervasively present but an inescapable reality, requiring political and social responses to the many challenges posed by the new immigration. In countries such as the United States, Canada, and Australia, whose histories are inextricably bound to large-scale immigration, multiculturalism has long been a focus of debate and policy. Western European immigration countries, in contrast, have only recently been compelled to consider the relationships among ethnic diversity, national identity, and citizenship, and to consider multiculturalism-oriented policies (Castles, 1993:28; Bauböck and Rundell, 1998).

It is against this changing background that integration of immigrants has become a key policy and research issue. Demography has not stood aside from this resurgence of interest in integration research because "there is a political demand which demographers and statisticians cannot ignore" (Magaud, 1997:2). Over and above the purely descriptive studies of migration and indicators of integration that have long typified demography's contribution, a growing body of demographic research focuses on the process of integration and especially on factors of inclusion and/or exclusion of immigrants and minorities.[1]

Research into the determinants of integration is still lacking in many respects. First, the indicators of integration are too often one-dimensional. Studies on the measures of economic integration are a case in point: comparative studies that find immigrant incomes to be lower than native-born incomes are often held to be evidence of immigrants' disadvantage on the labor market. But, the validity of this type of exercise depends on factoring other variables into the equation, such as period of arrival, age, educational level, sex, and previous work experience. In short, integration is a multidimensional process that involves a great many factors that are linked to immigrants' social and economic status in a given society.

Second, the wide variations in the number and types of factors included in empirical studies

[1] Demography's contribution is undeniably quite recent and as yet still tentative. Two recent illustrations of how demographers have stepped into the ethnic demography arena are Rallu *et al.* (1997) and Halli and Dreidger (1999).

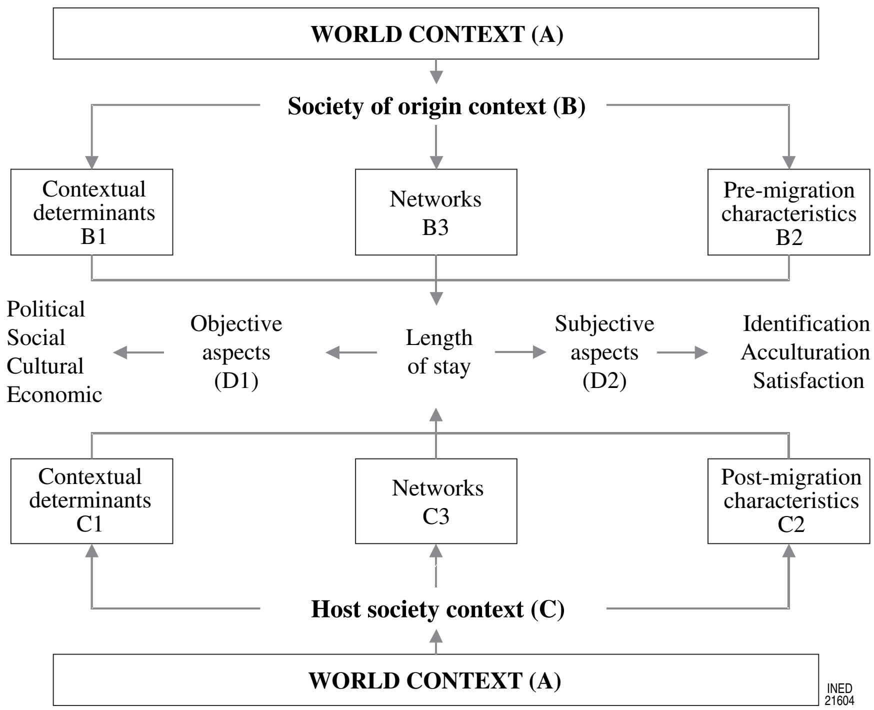


FIGURE 84–1 The integration of migrants: a conceptual framework.

constitute a serious obstacle to a comparative perspective. Factors of integration vary along theoretical and conceptual lines: hence, two schools of thought have developed in a parallel fashion. One approach insists that integration is mainly predicated by individual, human capital-related factors (e.g., educational level, age), whereas another perspective gives more importance to macrostructural and political factors. It is worth pointing out that the micro approach, heavily influenced by human capital theory (Borjas, 1994) and culturalism (D'Souza, 1995), has largely dominated North American research, whereas European (especially Western European) research on immigration has centered almost exclusively on historical and structural factors (Castles and Kosack, 1985; Simon, 1999) with a particular focus on racism and discrimination (Wrench and Solomos, 1993; Wieviorka, 1994). Arguably, this compartmentalization can yield partial or distorted results, because integration is a process that results from the complex interplay between immigrants' characteristics and the specific historical (economic and political) contexts of the host societies (Portes and Rumbaut, 1990).

What is proposed here in order to account for the complexity of the integration process, and especially its multidimensional character, is a conceptual framework that pinpoints what current research considers to be the key factors of integration. It aims to incorporate multiple types of factors and levels of analysis (Fig. 84–1).

I. DEFINING AND MEASURING INTEGRATION

There is no clear consensus on what is meant by the concept of "integration," which tends to be used in the literature to cover a wide range of situations. In some cases it is equated with assimilation and involves "the filtering out of migrants' cultural differences and migration characteristics" and "the convergence of behaviors in key areas" (Tribalat, 1996:254). Areas of convergence include indicators such as labor market performance, family and marriage practices, language, religion, education, housing, and citizenship. In this approach, immigrants' behaviors must converge, that is, resemble those of the native-born population. In most cases, however, integration refers to the multicultural model and involves both convergence in very narrow areas (e.g., relating to human and civic rights) and retention of certain cultural and other differences of migrant communities (Kymlicka, 1998). Inherent in this latter concept, therefore, is that immigrants should not entirely forsake their identity (Breton, 1994:240). In fact, there is an interplay of influence between immigrants and members of the host society in the process of integration, leading both groups to change over time.

Over and above theoretical concepts, what concerns us here are the operational and empirical definitions of integration. Broadly speaking, most authors conceptualize integration as a process that is measurable on a continuum ranging from nonintegration (e.g., leaving the country) to "successful" integration (e.g., exhibiting the same performance as the nonimmigrant or native-born populations). For many aspects of integration (e.g., economic, linguistic, residential), the indicators seem readily measurable and, as such, are the subject of consensus in so far as they are widely used and not routinely criticized. For example, the criterion of linguistic integration is knowledge of the language of the dominant group (or community). Whether in the home, at work, and during socializing and/or choosing media or a residential neighborhood, the basic questions are "which of the two languages (language of origin or language of the host country) is preferred?" and "what factors facilitate or hinder the adoption of the official language(s)?" As to economic (essentially the labor market) and educational integration, success is often measured against the average of the native-born population. The factors of economic and educational integration are therefore those that facilitate or hinder performance in these areas. Hence, successful economic, linguistic, and residential integration is achieved when differences in indicators between immigrant and native-born communities become nonsignificant.

By contrast, the indicators of social and cultural integration are much more problematic (Rallu, 1997). There is no ready consensus on a norm in these areas (Breton, 1994:243), and thus, the issue of successful integration poses huge problems of definition and remains normative and ideological. What are the criteria of success, and to whom do they apply: to immigrant groups or the host society? It is legitimate to suppose, for example, that indicators of success in the labor market are as valid for the economic integration of members of the host society as for immigrants. But the modes by which economic performance is achieved may be a critical issue. Ethnic enclaves are a case in point. Although some would argue that enclaves provide readily-available jobs and on-the-job training for immigrants (positive aspect), others considers enclaves as reflecting a measure of refusal to integrate into the supposedly nonethnic dominant economy (negative aspect).

Residential integration raises more problems still. However apparently neutral residential segregation measures may be, ethnic concentration, for instance, is seen as both positive and negative: positive because it is an expression of solidarities and the pooling of community resources enabling access to housing, and negative because the host society often perceives ethnic concentration as barriers to interaction between the immigrant and nonimmigrant populations. The same can be argued for educational integration. There seems to be a consensus on the measures of educational achievement (e.g., marks, drop-out rates, educational difficulties and low rates of education). However, where the values transmitted by the school and those of immigrant children and parents diverge, it is harder to define measures of integration or nonintegration. That, indeed, holds true for all indicators of social and cultural integration. The difficulty here, to quote Breton (1994:240–241), is "to identify the assemblage of attitudes and behaviors from which to conclude that an individual or category of individuals has become completely integrated." How are we to determine the cultural norm that should be the benchmark against which to measure the degree of integration? In a pluralistic society with such a diverse mix of values, defining that norm can be a highly controversial undertaking.

Ultimately, the definition of integration used here is very close to that of David Coleman (1994:43): integration involves the opportunity for immigrant populations and minority groups to participate fully in all aspects of society (social, economic, and political). And although equality of opportunity in society and the absence of discrimination on the grounds of national origin are certainly key factors in successful integration, so, too, are the characteristics of immigrants themselves, and the circumstances of their migration. That is what is conveyed by the conceptual framework presented below.

II. CONCEPTUAL FRAMEWORK FOR STUDYING THE FACTORS OF INTEGRATION

Figure 84–1 shows a simplified version of the conceptual framework proposed here to study the factors of integration.[2] Three broad groups of factors are identified. First are factors linked to the world context (A-type factors), which affect both societies of origin and host societies: generally, these are mainly but not exclusively macrostructural-type factors impinging on the economic and political spheres. This group of factors refers to such things as international relations, north–south relations, external/foreign policies, and globalization. They shape the conditions at origin,

[2] Adapted from the Goldlust and Richmond model, 1974.

which act as push factors, and the host country policies on selection and integration: in Fig. 84–1, this is called the society of origin context (B-type factors) and the host society context (C-type factors).

Figure 84–1 shows that the B- and C-type context variables are symmetrical. Three sets of factors are identified for origin (sending country) and destination (receiving country): (1) contextual determinants or structural factors (B1 and C1); (2) the pre- or postmigration micro-individual characteristics (B2 and C2); and (3) networks (B3 and C3). In the first case (B1), examples for the society of origin include factors such as economic and social development, industrialization, new urban development, and educational levels; political systems; state policies affecting migration; wars; and natural disasters. For the host society, (C1) factors would include economic and demographic conditions, the degree of pluralism, levels of discrimination, new urban development and stratification, and government policies, especially immigration policies.

Premigration characteristics (B2) refer to individuals: these may include education or vocational qualifications, demographic characteristics, previous experience, and motives. Likewise postmigration characteristics (C2) include immigration status, age, sex, occupational mobility, and social class. One characteristic—length of residence—is singled out in this model as a key factor in the integration process.

Implicit in the network variables are the circumstances in which the migration proper (B3) and the process of integration in the host society (C3) unfold. In both cases, these refer essentially to the role of the family, formal networks (e.g., churches, ethnic voluntary organizations, trade unions, employers' associations), and informal networks (e.g., intermediaries, friends) in facilitating emigration (origin) and integration (destination).

Finally, the conceptual framework identifies two types of dependent variables by which integration can be measured: the objective (D1) and subjective (D2) aspects of integration. The objective measures of integration tend to comprise behavioral or performance indicators such as type of participation in political and social life, type of media consumption, and economic activities (e.g., employment, unemployment, income). The subjective aspects are more perceptual and attitudinal and concern ethnic or linguistic identity and satisfaction with postmigration living conditions. The remaining part of this chapter will focus on the objective aspects of integration only.[3]

[3] For an application of this model to economic integration, see Piché *et al.* (2002a,b).

III. INFLUENCE OF THE WORLD CONTEXT

Contemporary international migration cannot be examined in a purely national, or even regional, setting. The formation of a world system (Wallerstein, 1979) requires international migration trends to be conceptualized within their global context, and a body of recent research has sought to demonstrate the relevance of such an analytical framework (Gungwu, 1997; Applebaum, 1999; Dieckhoff, 2000; Simmons, 2002). Stephen Castles (1993:17), for example, argues that the world is entering into a new phase of mass population shifts in which Europe-bound migration and the situation of ethnic minorities in Europe can only be approached in the global context. Likewise, Alan Simmons (1995:343) argues that the history of international migration cannot be separated from the history of the growth of world capitalism. He also hypothesizes three possible explanations for the changing directions and forms in the *global system of international migration* since the mid-1960s: they have acquired a more global scope (involving a growing number of hitherto comparatively remote countries), more flexible time-scales (a rise in short-term migration), and more decentralized control mechanisms (growing importance of families and networks in migration strategies increasingly outside state control).

Although it may appear obvious that the world economic context shapes not only the volume and direction of world migration patterns but also integration processes, research in this area is still mostly speculative and conjectural. Also, the effects of the world context are still unclear as to their positive or negative role in integration. As an illustration, five hypotheses can be suggested as ways in which the world context may be linked to integration.

1. The world economic situation—with its numerous crises and responses in the form of structural adjustment programs—has led to labor market tightening that more specifically affects the economic living conditions of recent immigrant groups (for the United States: Borjas, 1994; for Canada: Bloom *et al.*, 1994; for Europe: Coleman, 1994), which may block economic integration and fuel anti-immigration sentiments.

2. Trade globalization, reflected among other things in the formation of large-scale regional blocs, shifts discussions of immigration to the international level and makes immigration a strategic issue in treaty negotiations, be it North American Free Trade Agreement[4] in

[4] North American Free Trade Agreement (NAFTA) concluded in 1992 by the United States, Canada, and Mexico, applied from January 1994.

North America (Simmons, 1996) or the Maastricht Treaty in Europe[5] (Miles, 1992). These reorganizations may either favor integration through such things as freedom of movement for nationals in the new regional spaces and the adoption of universal antidiscriminatory measures, or impede it through the development of restrictive immigration and integration policies (e.g., exclusion of nonmember countries, tightening up of admissible categories, especially relating to families and refugees, restriction in welfare programs).

3. One frequently cited consequence of globalization is a trend toward the strengthening of local identities (Gungwu, 1997:8), which may reflect new solidarities or degenerate into ethnic conflicts (Weiner, 1997:101).

4. Globalization also affects ideologies and norms, thereby encouraging the emergence of a world public opinion (Badie, 1997). This arguably creates leverage in favor of human rights[6] (Withol de Winden, 1994) and antiracist, antidiscriminatory, and nonassimilationist policies (Gungwu, 1997:15–16). But at the same time, it can be responsible for the rise in anti-immigration sentiments (Wrench and Solomos, 1993:7) and an upsurge in nativism[7] (Weiner, 1997:100).

5. Finally, migration-driven transnationalization brings international relations into the forefront (Badie, 1993), often allowing immigrant populations to live in networks that may represent a threat to national sovereignty (Hollifield, 1994; Schnapper, 2001). New forms of international migration, the demands for recognition of minority identities, and the growing importance of transnational networks make national borders more permeable and lead to a radical rethinking of national identities and the very concept of citizenship (Castles, 1998; Soysal, 1998). For some, the forces of globalization are undermining the very foundations of the nation-state (Dieckoff, 2000; Seymour, 2002).

In short, the current context of globalization is producing radical changes to the way we approach integration. As was said earlier, research in this area is still conjectural and insufficiently comparative. Also, the effects of such global (and sometimes very abstract) phenomena are not readily measured by current methods. In any event, there is no denying the real and often opposing effects exerted by current world forces. Immigrant groups and their allies may be comforted by the portion of world public opinion that favors more tolerance and multicultural pluralistic systems that aim to facilitate the integration of immigrant populations. But at the same time, opposing forces are calling for more restrictive measures to stem the new migration flows that, they argue, pose a threat to national identities and political stability. Be that as it may, the study of the factors of integration can no longer afford to brush this kind of issue aside.

[5] The treaty signed in 1991 by the 15 European Union Member States (Germany, Austria, Belgium, Denmark, Spain, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, United Kingdom, Sweden).

[6] Myron Weiner (1997:114) goes so far as to talk of "global human rights."

[7] Nativism is a school of thought that aims to preserve the cultural character of a country as it has been shaped by the native-born population.

IV. CONTEXTUAL FACTORS IN THE COUNTRIES OF ORIGIN AND DESTINATION

However necessary the world system may be to the equation, it is not enough to address the practical modes of integration. In a sense, the world setting is often remote of the real life of immigrant populations, which unfolds in specific countries and societies in which the pressures of the world system are mediated through national histories (Tarrius, 2002). Integration research, therefore, must consider the contexts characterizing societies from which immigrant populations leave and into which they enter. There is a growing consensus among analysts of migration on the necessity to take into account factors related to the country of origin: international migration is therefore defined as a system comprising a set of countries interchanging a relatively large number of migrants (Kritz and Zlotnik, 1992:2). In this context, the countries are linked to one another by migration and other flows (e.g., capital, goods).

With respect to integration research, little consideration has been given to the macrostructural aspects of the country of origin. It is reasonable to think, however, that integration of immigrant groups will vary with the greater or lesser social, cultural, and economic distance between the country of origin and the country of immigration. Local social and economic indicators include, among other things, language, religion, types of family organization, and the level of economic development of the country of origin. In a recent study, for instance, it was demonstrated that the economic level of the country of origin of immigrant populations had a significant influence on their degree of economic integration measured by income and socioeconomic status: other factors being equal, immigrants from poor countries have more problems than do others integrating into the labor market (Renaud *et al.*, 1997).[8]

[8] The question of national origins will be considered further in the section on micro-individual factors.

Host country contextual factors have received far more extensive attention. Three factors are particularly relevant for immigrant populations: government policies regarding selection and integration, labor market conditions, and the characteristics of their ethnic community (Portes and Rumbaut, 1990:85). Political factors relate to how open or closed the country is to certain groups. To that extent, integration policies may become a key factor in the success or failure of integration. With growing pluralism, some countries are striving to develop integration policies either by encouraging assimilation and discouraging the perpetuation of minority groups, or by adopting a policy of multiculturalism designed to help immigrant minority groups to preserve their language, culture, and traditions (Coleman, 1994:41). Unfortunately, the present state of research in this field does not allow any conclusion to the effect that one model of integration is better than any other; that conclusion would require a comparative approach that does not currently exist. However, selection and integration policies may arguably have tangible effects within a country (Portes and Zhou, 1995), with the success of integration varying between groups of immigrants as a function of differential access to integration measures.

The second contextualizing factor for integration is the labor market situation. Labor market integration theories distinguish two levels of factors: macro and micro. Micro-individual factors will be dealt with in the following section. The macrostructural approach focuses on organizational aspects of the labor market that determine the rules of entry according to specific internal market rationales. Strictly speaking, what we are dealing with are multiple, differentiated markets rather than a single market. Hence, in a segmented labor market, the methods of integration will vary by segment, and some groups, regardless of their individual characteristics, will end up concentrated in specific segments (Portes, 1981). Arguably, therefore, there are institutional barriers to employment, at least in certain particular sectors, which in a way act as factors of exclusion of certain categories of workers, including immigrant groups. As a result, various groups of immigrants will either end up concentrated in insecure jobs (Coleman, 1994:55) or will develop strategies for integration based on the ethnic economy and form niches (Light, 1984) or even enclaves (Portes and Manning, 1986). In short, in this approach, some groups of immigrants, especially the lesser-skilled ones, are consigned to the bottom rung of the socio-occupational ladder (Castles and Kosack, 1985), mainly because of their vulnerable position on the market or institutional barriers that often involve discrimination against them.[9] This body of literature focuses on discrimination and racism that some categories of immigrants suffer, and that occur more or less across the developed world (for the United States: Poston, 1994; for Canada: Richmond, 1992; for France: Tribalat, 1996; for Western Europe: Castles, 1993). Authors such as John Wrench and John Solomos (1993:4) are in no doubt that immigration and racism are becoming critical issues for many European countries.

The third and final contextual factor affecting integration is linked with community and ethnic support structures (Musterd and Murie, 2001). Most immigrants move to societies in which there is already an established community, prompting a number of authors to argue for a strong family and ethnic network influence on integration. In many countries, massive immigration has resulted in the formation of strong ethnic identities (Featherstone, 1990; Gungwu, 1997:15). Networks operate at all levels: they are involved in the search for jobs and housing, as well as the organization of cultural and religious activities. Two types of integration have been particularly studied in connection with community networks: ethnic enclave labor markets and residential concentration. Although the ethnic factor is important in both cases, its impact remains moot. Does it facilitate integration, or does it rather contribute to the formation of ethnic ghettos? The same question has been asked about the influence of Islam in Europe (Coleman, 1994:67–69). In the case of North America, the ethnic diversity that has resulted from immigration is seen as an asset for society and plays a major role in the redefinition of national identities (Piché, 2002).

V. MICRO-INDIVIDUAL FACTORS

Contextual factors may be important in accounting for differential modes of integration between various groups of immigrants, but they do not explain all the inter- and intragroup differences. Intergroup differences may also be connected with the characteristics of the individuals that make up the different migration flows. A number of individual factors have been identified as potentially influencing the process of integration: length of residence, period of arrival, age at entry, language skills, education, sex, and immigration status (admission class and type of migration). If integration is seen as a long-term process (Bastenier and Dassetto, 1995), time variables such as length of residence and

[9] In this approach, immigrant populations are often seen as cheap labor (Burawoy, 1976).

period of arrival are key factors (Goldlust and Richmond, 1974; Tribalat, 1996). As economic integration is also the key to social integration, human capital variables will also be central factors: immigrants who are young and educated and who know the official language(s) find integration easier. Finally, migrant status plays a major part in the integration process: (1) illegal (undocumented) immigrants have many more problems integrating than do migrants entering through official categories; (2) among selected immigrants, refugees have more difficulties than do other categories of immigrants; and (3) among the refugee groups, refugee claimants are the most disadvantaged.

Research into the individual factors of integration is mostly based on cross-sectional data, mainly from censuses and, more occasionally, immigrant surveys. These studies usually establish comparisons between immigrant and native-born populations, with a strong emphasis on economic aspects. Income gaps between the two groups show generally that immigrants are on arrival disadvantaged in comparison to the native-born population. There is however a time-related rising trend in immigrant incomes after a period of adjustment to the new environment, which enables them to optimize their skills. American and Canadian research suggests that immigrants quite quickly rise up to the average income levels of the native-born population, except for recently arrived cohorts (Chiswick, 1986; Lalonde and Topel, 1992; Bloom *et al.*, 1994; Beaujot, 1997). Furthermore, between-group variations can be observed in immigrants' probability of economic success. In the United States, a social and economic hierarchy of immigrant groups by region of origin can be observed, with immigrants of European origin at the top of the hierarchy and non-Europeans—especially Latin Americans (Poston, 1994), recent Third World immigrants (Lalonde and Topel, 1991, 1992), and more specifically Mexicans (Chiswick, 1978) at the bottom. The same trend is found in Canada (Bloom *et al.*, 1994). In continental Europe, the use of ethnic categories is more problematic than in America or England, especially so in France (Simon, 1998). Nevertheless, recent surveys in France also show the existence of ethnic stratification (Tribalat, 1996).

These studies on the process of differential integration by national or ethnic origin have been considerably enriched by recently conducted longitudinal surveys. Three such surveys merit particular mention. The first is the establishment of new immigrants (ENI) survey carried out in Montreal, Quebec (Canada). This is a longitudinal 10-year follow-up survey of a sample cohort of newly arrived immigrants in Montreal in 1989. The study confirmed that the individual characteristics mentioned above were important for economic integration (measured by access to first job and income): age at entry, sex, migrant status, and language skills are all discriminant factors. But more importantly, even after all these factors are taken into account in multivariate analyses, national origin remains a significant factor in integration, suggesting that some form of discrimination is operating (Piché *et al.*, 2002a,b).

After 10 years, however, between-group income and employment status differentials have disappeared (Renaud, Piché and Godin, 2003). Three possible explanations may be advanced for this finding. First, the only way these groups of immigrants can counteract stigma and job mismatching entailed by initial discrimination is by developing "reskilling" strategies. What this amounts to is a strategy of overcompensation (these groups must do more to achieve the same results as the native-born population): this may account for the disappearance of differentials after 10 years. A second alternative explanation is that recruitment practices in Quebec may be very different from those found in some immigrants' countries of origin. The organizational culture, which includes interview practices, comes readily to mind. Over time, immigrant groups learn the host country's specific labor market culture, thereby experiencing a labor market socialization process. Hence, the idea of differentiated social capital on entry may offer insights into differential processes of economic integration. The third and final hypothesis is that host societies, especially employers, have changed in the past 10 years, becoming as it were more receptive and less exclusionary or discriminatory. The implementation of charters of rights, the demands and lobbying of immigrant groups, and the setting up of equal opportunity programs may have had a significant effect on discriminatory practices. Taken to the extreme, this third hypothesis would mean that job segregation would to all intents and purposes have vanished, which appears unlikely. Obviously, these are not either/or conjectures, and the three mechanisms may have operated simultaneously and in conjunction with one another.

A second example is provided by the National Institute for Statistics and Economic Studies (INSEE)'s Permanent Demographic Sample, from which a subsample of respondents, aged between 4 and 18 years in 1975, were followed-up in France from 1975–1990 (Héran, 1998). The multivariate analyses of unemployment-related factors for men and women aged 25 to 33 years in this population also show the importance of national origin for certain groups, especially young people from the Maghreb, who are significantly disadvantaged on the job market (Richard, 2000). The study concluded that "the likelihood is that some firms practice discriminatory policies towards young people of foreign origin, especially Muslims and those from the Maghreb" (Richard, 2000:72–73).

A third and final example is the joint INED/INSEE "geographical mobility and social integration survey"[10] carried out in France in 1992. This was another longitudinal survey, this time on three samples: immigrants, children of immigrants born in France, and a control population representative of the French population (Dayan *et al*., 1997:115). The findings on the role of national origin concur with those of the Quebec study: (1) human capital factors—especially age at entry, year of entry, educational level, and French language skills—have significant effects on economic integration; (2) sex is also significant, with female immigrants having more difficulties on the labor market than do males, particularly because of domestic responsibilities; and (3) the net effect of national origin (i.e., after control of other factors) is still significant, with a negative effect for immigrants of both sexes from Algeria, Turkey, Asia, and sub-saharan Africa, whereas no discernible disadvantage was found compared with the control population for male immigrants from Portugal and Spain (Dayan *et al*., 1997:123–131).

What explanations can be advanced for ethnic stratification found in all developed countries? First, recent research has discarded the idea of race or ethnic group as essential categories that brand certain groups as having innate characteristics that make them less fitted to the labor market (Zuberi, 2001). Also, differential integration cannot be attributed to human capital factors only, because even after controlling for these factors, some groups of immigrants still have greater difficulty than do the native-born population or other groups of immigrants in finding and keeping a job. Three possible explanations may be suggested. First, some immigrant groups may have more difficulty negotiating and securing the recognition of skills acquired in their country of origin (Borjas, 1994). Indirect discrimination would arguably be at play here, with training undergone in some countries being wrongly and unjustifiably undervalued. Second, access to ethnic networks may vary widely by immigrant group, which may explain the greater difficulty experienced by certain groups (Poston, 1994). Finally, differential labor market access by national origin may result from direct discrimination. This seems to be the most popular conclusion given countless references to discrimination in the literature on immigration (Richmond, 1992; Coleman, 1994; Wrench and Solomos, 1993; Richard, 2000).

[10] *"Mobilité géographique et insertion sociale."*

CONCLUSION

What is clear from this review of factors is that integration is a multidimensional process, and it is pointless to oppose macrostructural and micro-individual factors or economic and ethnic factors. Recent research shows that a multiplicity of factors are involved in the integration process and that they vary over time. Furthermore, this research also tends to suggest that integration processes in developed countries are sufficiently similar to warrant more comparative studies (Lloyd, 1993). The typically individualist tradition (United States, Canada, Australia) would benefit from being more greatly informed by the more historical and macrostructural European perspectives, and vice versa.

Various shortcomings in the research on integration need to be addressed in the planning of future research. First, the concept of integration itself raises problems of definition and measurement that are too infrequently ignored. Yet it is very difficult to lay down benchmark criteria of *successful* integration by which to measure the degree of integration, so that what is considered as an indicator of integration by some is not by others. Arguably, the measure of integration is always normative: the benchmark is usually set by reference to the control group, which is either the native-born population or the whole population (national norm). But, as was stated at the beginning, there is no consensus on the norm that should hold for cultural and social dimensions. In the economic sphere, by contrast, there is a consensus on the need for immigrant groups to have the same employment opportunities as those of the general population.

The second conclusion relates to methodology. The research data are drawn from censuses and mainly quantitative surveys. These are basically cross-sectional data and, as such, are unsuited to the study of an essentially longitudinal phenomenon. Simply put, the problem stems from the fact that integration is a process whereas the indicators are derived from cross-sectional data. Longitudinal studies such as those carried out in Canada and France are really the path of the future. Event history-type retrospective surveys and methodologies make this type of survey very useful if not essential.

Third, data collection methods remain highly divided. In particular, there is a sharp division between the quantitative and qualitative approaches. Arguably, this separation is futile, and what is required is more multimethod research. Also, there is too little multidimensional research on the factors of integration. The focus is too often on a very limited set of factors taken in isolation. Where multiple factors are considered, the net effects, or the relative weights of the factors, are rarely assessed.

Finally, the selectivity bias inherent in any study of integration cannot be underrated. The immigrant samples studied are affected by two types of selectivity. The first—selective immigration—stems from the selection criteria used. Here, the categories of immigration are in effect *synthetic* variables that are closely related to the modes of integration, more specifically on the labor market. Any comparative study must therefore always take into account (factor out) the sample composition by admission categories. The second type of selectivity—selective emigration—is more significant. It is a known fact that a significant percentage of immigrants—with wide intergroup variations—leave the host country sometime after entry. What is being studied, therefore, are immigrants who have *stayed*, or are still there at the time of census or survey. The fact that several factors of emigration are directly linked to the factors of integration may seriously contaminate the results based purely on immigrants who have stayed. If those that have most difficulties integrating have left, then studying those who remain will skew the results upward (overestimating success). The opposite may also happen (the most highly skilled leave, skewing the results downward). Furthermore, the direction of the biases may vary between periods. Only a longitudinal approach would, in theory at least, check this bias. But there is a risk of attrition and, hence a risk of bias, even there. Future research should make more allowance for these biases in interpreting the results.

Acknowledgment

This chapter was translated from the French version by Glenn D. Robertson.

References

Applebaum Richard P., 1999. Multiculturalism and flexibility: some new directions in global capitalism, *in:* Rodolfo D. Torres, Louis F. Miron, and Jonathan Xavier Inda, (eds.), *Race, identity, and citizenship*, p. 337–354. Malden, Blackwell, 452 p.

Badie Bertrand, 1993. Flux migratoires et relations transnationales, *in:* Bertrand Badie and Catherine Wihtol de Wenden (eds.), *Le défi migratoire: questions de relations internationales*, p. 27–39. Paris, Presses de la Fondation nationale des sciences politiques, 185 p.

Badie Bertrand, 1997. L'essor des politiques de ghetto devient une source d'instabilité internationale, *in:* Serge Cordellier and Beatrice Didiot (eds.), *L'état du monde 1998*, p. 40–42. Paris, La Découverte, 704 p.

Bastenier Albert and Dassetto Felice, 1995. L'intégration des immigrés en Europe: un cadre conceptuel, *in:* Hubert Gérard and Victor Piché (eds.), *La sociologie des populations*, p. 417–430. Montreal, Presses de l'Université de Montréal and AUPELF/UREF, 508 p.

Baubock Rainer and Rundell John (eds.), 1998. *Blurred boundaries: migration, ethnicity, citizenship.* Vienna, Ashgate, 356 p.

Beaujot Roderic, 1997. Comportements démographiques et statut socio-économique des immigrants canadiens, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 147–163. Paris, INED and John Libbey, 398 p.

Bloom David E., Grenier Gilles, and Gunderson Morley, 1994. *The changing labor market position of Canadian immigrants.* Cambridge, National Bureau of Economic Research, 36 p.

Borjas George J., 1994. The economics of immigration, *Journal of Economic Literature*, vol. 32(4), p. 1667–1717.

Breton Raymond, 1994. L'appartenance progressive à une société: perspectives sur l'intégration socioculturelle des immigrants, *in:* MAICC and CEETUM (eds.), *Actes du Séminaire sur les indicateurs d'intégration des immigrants*, p. 239–252. Montreal, Ministère des affaires internationales, de l'immigration et des communautés culturelles, 353 p.

Burawoy Michael, 1976. The functions and reproduction of migrant labor: comparative material from Southern Africa and the United States, *American Journal of Sociology*, vol. 81(5), p. 1050–1087.

Castles Stephen, 1993. Migrations and minorities in Europe. Perspectives for the 1990s: eleven hypotheses, *in:* John Wrench and John Solomos (eds.), *Racism and migration in western Europe*, p. 17–34. Oxford, Berg, 293 p.

Castles Stephen, 1998. Globalization and the ambiguities of national citizenship, *in:* Rainer Bauböck and John Rundell (eds.), *Blurred boundaries: migration, ethnicity, citizenship*, p. 223–244. Vienna, Ashgate, 356 p.

Castles Stephen and Kosack Godula, 1985. *Immigrant workers and class structure in western Europe.* London, Oxford University Press, 534 p.

Chiswick Barry, 1978. The effect of Americanization on the earnings of foreign-born men, *Journal of Political Economy*, vol. 86(5), p. 897–921.

Chiswick Barry, 1986. Is the new immigration less skilled than the old? *Journal of Labor Economics*, vol. 4(2), p. 168–192.

Coleman David, 1994. International migrants in europe: adjustment and integration processes and policies, *in:* Milos Macura and David Coleman (eds.), *International migration: regional processes and responses*, p. 41–76. New York, United Nations, 200 p.

Dayan Jean-Louis, Echardour Annick, and Glaude Michel, 1997. Le parcours professionnel des immigrés en France: une analyse longitudinale, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 113–146. Paris, INED and John Libbey, 398 p.

Dieckhoff Alain, 2000. *La nation dans tous ses États.* Paris, Flammarion, 356 p.

D'Souza Dinesh, 1995. *The end of racism.* New York, The Free Press, 724 p.

Featherstone Mike (ed.), 1990. *Global culture: nationalism, globalization and modernity.* London, Sage, 411 p.

Goldlust John and Richmond Anthony H., 1974. A multivariate model of immigrants adaptation, *International Migration Review*, vol. 8(2), p. 193–225.

Gungwu Wang (ed.), 1997. *Global history and migrations.* Boulder, Westview Press, 320 p.

Halli Shiva S. and Driedger Leo (eds.), 1999. *Immigrant Canada: demographic, economic, and social challenges* Toronto, University of Toronto Press, 358 p.

Héran François, 1998. La mobilité sociale et professionnelle à la lumière de l'échantillon démographique permanent: six approches longitudinales, *Économie et statistique*, no. 316–317, p. 63–76.

Hollifield James Frank, 1994. Entre droit et marché, *in:* Bertrand Badie and Catherine Wihtol de Wenden (eds.), *Le défi migratoire. Questions de relations internationales*, p. 59–87. Paris, Presses de la Fondation nationale des sciences politiques, 185 p.

Kritz Mary M. and Zlotnik Hania, 1992. Global interactions: migration systems, processes, and policies, *in:* Marie M. Kritz, Lim L. Lean, and Hania Zlotnik (eds.), *International migration systems: a global approach*, p. 1–16. Oxford, Clarendon Press, 354 p.

Kuijsten Anton, 1994. International migration in Europe: patterns and implications for receiving countries, *in:* Milos Macura and David Coleman (eds.), *International migration: regional processes and responses*, p. 21–39. New York, United Nations, 200 p.

Kymlicka Will, 1998. Multicultural citizenship, *in:* Gershon Shafir (ed.), *The citizenship debates*, p. 167–188. Minneapolis, University of Minnesota Press, 316 p.

Lalonde Robert J. and Topel Robert H., 1991. Immigrants in the American labor market: quality, assimilation and distributional effects, *American Economic Review*, vol. 81(2), p. 297–302.

Lalonde Robert J. and Topel Robert H., 1992. The assimilation of immigrants in the U.S. labor market, *in:* George J. Borjas and Richard B. Freeman (eds.), *Immigration and the work force: economic consequences for the United States and source areas*, p. 67–92. Chicago and London, The University of Chicago Press, 281 p.

Light Yvan, 1984. Immigrant and ethnic enterprise in North America, *Ethnic and Racial Studies*, vol. 4, p. 195–216.

Lloyd Cathie, 1993. Research and policy issues in a European perspective, *in:* John Wrench and John Solomos (eds.), *Racism and migration in western Europe*, p. 251–263. Oxford, Berg, 293 p.

Magaud Jacques, 1997. Préface, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 1–2. Paris, INED and John Libbey, 398 p.

Massey Douglas S., Arango Joaquin, Hugo Graeme, Kouaouci Ali, Pellegrino Adella, and Taylor J. Edward, 1998. *Worlds in motion: understanding international migration at the end of the millenium.* Oxford, Clarendon Press, 362 p.

McNicoll Claire, 1993. *Montréal une société multiculturelle.* Paris, Editions Belin, 317 p.

Miles Robert, 1992. Migration, racism and the nation-state in contemporary Europe, *in:* Vic Satzewich (ed.), *Deconstructing a nation: immigration, multiculturalism and racism in '90s Canada*, p. 21–45. Halifax, Fernwood Publishing, 538 p.

Musterd Sako and Murie Alan, 2001. *A comparative report on social exclusion and integration in neighbourhoods in 11 cities in six European countries: Amsterdam, Antwerp, Berlin, Birgmingham, Brussels, Hamburg, London, Milan, Naples, Paris and Rotterdam.* Amsterdam, URBEX and University of Amsterdam, 44 p.

Piché Victor, 2002. Immigration, diversity and ethnic relations in Québec, *Canadian Ethnic Studies*, vol. 34(3), p. 5–27.

Piché Victor and Bélanger Liane, 1995. *Une revue des études québécoises sur les facteurs d'intégration des immigrants.* Montreal, Ministère des affaires internationales, de l'immigration et des Communautés culturelles, 52 p.

Piché Victor, Renaud Jean, and Gingras Lucie, 2002a. Economic integration of new immigrants in the Montreal labor market: a longitudinal approach, *Population E*, vol. 57(1), p. 67–82.

Piché Victor, Renaud Jean, and Gingras Lucie, 2002b. L'insertion économique des nouveaux immigrants dans le marché du travail à Montréal: une approche longitudinale, *Population F*, vol. 57(1), p. 63–90.

Portes Alejandro, 1981. Modes of structural incorporation and present theories of labor immigration, *in:* Mary M. Kritz, Charles B. Keely, and Silvano M. Tomasis(eds.), *Global trends in migration*, p. 279–297. New York, Center for Migration Studies, 465 p.

Portes Alejandro and Manning Robert D., 1986. The immigrant enclave: theory and empirical examples, *in:* Susan Olzak and Joane Nagel (eds.), *Competitive ethnic relations*, p. 47–68. Orlando, Academic Press, 252 p.

Portes Alejandro and Rumbaut Rubén G., 1990. *Immigrant America: a portrait.* Berkeley, University of California Press, 300 p.

Portes Alejandro and Zhou Min, 1995. Destins divergents: immigration, pauvreté et esprit d'entreprise aux États-Unis, *in:* Hubert Gérard and Victor Piché (eds.), *La sociologie des populations*, p. 365–390. Montreal, Presses de l'Université de Montréal and AUPELF/UREF, 508 p.

Poston Dudley L., 1994. Patterns of economic attainment of foreign-born male workers in the US, *International Migration Review*, vol. 28(3), p. 478–500.

Rallu Jean-Louis, 1997. L'intégration, à quel type de société? *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 379–397. Paris, INED and John Libbey, 398 p.

Rallu Jean-Louis, Courbage Youssef, and Piché Victor (eds.), 1997. *Old and new minorities/Anciennes et nouvelles minorités.* Paris, INED and John Libbey, 398 p.

Renaud Jean, Piché Victor, and Gingras Lucie, 1997. Immigration et intégration économique à Montréal, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 89–111. Paris, INED and John Libbey, 398 p.

Renaud Jean, Piché Victor, and Godin Jean-François, 2003. Discrimination ou ajustement? L'insertion économique des immigrants dans leur dix premières années au Québec, *Sociologie et société*, vol. 35(1), p. 165–184.

Richard Jean-Luc, 2000. Une approche de la discrimination sur le marché du travail. Les jeunes adultes issus de l'immigration étrangère en France, *Revue européenne des migrations internationales*, vol. 16(3), p. 53–81.

Richmond Anthony, 1992. Immigration and structural change: the Canadian experience 1971–1986, *International Migration Review*, vol. 26(4), p. 1200–1221.

Schnapper Dominique, 2001. De l'État-nation au monde transnational: du sens et de l'utilité du concept de diaspora, *Revue européenne des migrations internationales*, vol. 17(2), p. 9–34.

Seymour Michel (ed.), 2002. *États-nations, multinations et organisations supranationales.* Montreal, Liber, 499 p.

Simmons Alan, 1995. Migration internationale et capitalisme global: examen critique des théories, *in:* Hubert Gérard and Victor Piché (eds.), *La sociologie des populations*, p. 341–364. Montreal, Presses de l'Université de Montréal and AUPELF/UREF, 508 p.

Simmons Alan B. (ed.), 1996. *Internatinal migration, refugee flows and human rights in North America: the impact of trade and restructuring.* New York, Center for Migration Studies, 348 p.

Simmons Alan B., 2002. Mondialisation et migration internationale: tendances, interrogations et modèles théoriques, *Cahiers québécois de démographie*, vol. 31(1), p. 7–33.

Simon Patrick, 1998. Nationalité et origine dans la statistique française: les catégories ambiguës, *Population*, vol. 53(3), p. 541–568.

Simon Patrick, 1999. L'immigration et l'intégration dans les sciences sociales en France depuis 1945, *in:* Philippe Dewitte (ed.), *Immigration et intégration: l'état des savoirs*, p. 82–95. Paris, Éditions La Découverte, 444 p.

Soysal Yasemin N., 1998. Toward a postnational model of membership, *in:* Gershon Shafir (ed.), *The citizenship debates*, p. 189–217. Minneapolis, University of Minnesota Press, 316 p.

Tarrius Alain, 2002. La *mondialisation par le bas. Les nouveaux nomades de l'économie souterraine*. Paris, Balland, 169 p.

Tribalat Michèle, 1996. *De l'immigration à l'assimilation. Enquête sur les populations étrangères en France.* Paris, La Découverte/INED, 302 p.

Wallerstein Immanuel, 1979. *The capitalist-world economy.* Cambridge, Cambridge University Press, 305 p.

Weiner Myron, 1997. The global migration crisis, *in:* Wang Gungwu (ed.), *Global history and migrations*, p. 95–116. Boulder, Westview Press, 320 p.

Wieviorka Michel, 1994. *Racisme et xénophobie en Europe, une comparaison internationale*. Paris, Éditions La Découverte, 305 p.

Wihtol De Wenden Catherine, 1994. Migrations et droits de l'homme, *in:* Bertrand Badie and Catherine Wihtol de Wenden (eds.), *Le défi migratoire: questions de relations internationales*, p. 159–177. Paris, Presses de la Fondation nationale des sciences politiques, 185 p.

Wrench John and Solomos John (eds.), 1993. *Racism and migration in western Europe*. Oxford, Berg, 293 p.

Zuberi Tukufu, 2001. *Thicker than blood: how racial statistics lie.* Minneapolis, University of Minnesota Press, 193 p.

CHAPTER

85

Economic Growth and Population Growth

ARNAUD DELLIS AND PIERRE PESTIEAU

University of Hawaii, Manoa, Hawaii, United States

Université de Liège, Liège, Belgium, and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

INTRODUCTION

The purpose of this chapter is to explore the hypothesis of an influence of population on productivity and technological progress. The recent past has seen the various indicators of productivity and technological progress follow an uncertain evolution. Most recently, a convergent trend to a decline in productivity has been detected in many industrial countries. What does this evolution owe to demographic factors? Ready-made answers to these questions exist. For some, there is an unavoidable connection between falling productivity and an aging population; for others, a smaller work force necessarily implies higher productivity.

We attempt to provide an economist's answer to these questions, based on rigorous reasoning and confirmed by factual evidence. Anticipating on what follows, it will appear that the empirical relationships among population, productivity, and technology—however those three terms are defined—are both tenuous and ambiguous in the present period. Consequently, it would be misplaced, supposing it were possible, to adopt an active population policy with the aim of modifying the current evolution of productivity and technological progress. It would be more efficient to use other instruments that are more certain in their effect.

The theoretical relationship between population and production has been the subject of extensive discussion ever since Malthus and has recently received fresh attention in certain formulations of the theory of endogenous growth. The strength of this theory lies in incorporating the dynamic and interactive dimension of this relationship; its weakness in being too often at variance with empirical evidence and in producing disparate conclusions that depend closely on the initial assumptions.

Our purpose in this chapter is not to present a comprehensive synthesis of the theoretical and empirical works devoted to the influence of population on economic growth but to give a selective presentation of a small number of works and new estimates.

I. POPULAR WISDOM AND SCIENTIFIC UNCERTAINTY

Economics is often compared to medicine. Both can be understood at several levels, most specifically those of the man in the street and of the specialist. Imagine your child comes home from school with a cold. Even before you do anything about treating it, you start looking for the cause, hesitating little between considering a draft and the classmate with a cold. If you ask your family doctor (and *a fortiori* a specialist), he or she will be more concerned about treating the young patient than explaining the origin of the problem. And any answer that is given will be hedged with qualifications.

The same duality, the same contrast between the assurance of the man in the street and the hesitancy of the specialist, is found in economics. Ask an economist for his view about the influence of population growth on productivity, and the answer will be anything but simple and precise. The economist will explain that the question is badly formulated or that it is meaningless. Or the economist will suggest that the terms be reversed, acknowledging that variations in productivity perhaps have an influence on fertility and family structure. The economist may ask if you are interested in the levels of population and production or in their rates of growth. And when your question does receive an answer, chances are that it will take the form of some 20 pages of qualified and careful arguments. It is in this spirit that the present chapter is written.

Popular wisdom, however, does not trouble with these intellectual precautions. It opts without hesitation for a clear and unequivocal answer. Yet at this level there is no single answer. Two in particular are encountered, which we will refer to as Ricardo-Malthusian and Boserupian. It can be noted that both are usually presented in a dramatic context, that of an aging and declining population or, conversely, that of an uncontrolled population explosion.

The Ricardo-Malthusian thesis (see Malthus, 1803; Dupâquier, 1980; Fauve-Chamoux, 1984), for which precursors can be found in the work of Giovanni Botero and of David Hume, is based on two fundamental principles: diminishing returns and the pressure of population on the means of subsistence. It leads to one recommendation: zero growth. We can summarize this thesis by using a diagram (Fig. 85–1). If the means of subsistence in a country are sufficient, the population will tend to grow faster than these means. According to Malthus, the means of subsistence increase at an arithmetic rate, and population increases at a geometric rate. The results are (1) a reduction in the subsistence goods available for each person, causing prices to rise; and (2) an expansion of the active population with a decrease in wages. This takes us to the threshold of subsistence. This is the Malthusian trap.

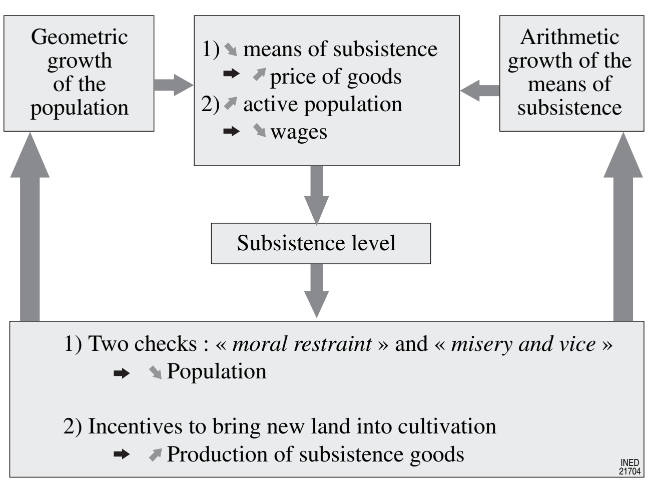


FIGURE 85–1 Illustration of the Ricardo-Malthusian thesis.

At this stage of the process, two types of check to population growth come into operation:

Moral restraint (or the preventive check)—in the form of delayed marriage, which causes fertility to fall;
Misery and vice (or the positive check)—in the form of wars, epidemics, etc., which cause mortality to rise.

The combined effect of these two phenomena is a reduction in population.

In contrast, the higher price of subsistence goods plus the fall in wages is an incentive for farmers to bring new land into cultivation. The resulting rise in the output of the means of subsistence, combined with the decline in population, means that we cross back over the threshold of subsistence, and the Malthusian cycle begins again.

The main criticism of the Malthusian theory formulated by the Boserupian theory concerns the assumption that technological progress is exogenous. More fundamentally, these two theories differ in their approach to the causality of the relationship. For Malthus, the means of subsistence is the key variable in the process of population regulation. For Boserup, population, by exerting a pressure on the economy, is the engine of growth.

The Boserupian thesis—which is also sometimes described as populationist—argues forcefully that a population that is declining, and hence aging, creates a mentality, a value system, and a social balance that are unfavorable to growth and technological progress. Conversely, a population that is growing and young is a guarantee of social and economic dynamism. The symbolism of this thesis is obvious. The idea is simple and powerful and has been incorporated into the programs of politicians sympathetic to this view. It has also found many supporters among the most famous names of economics: Stein Hansen, John Maynard Keynes, Colin Clark, and Simon Kuznets, for example.

The theory rests on several arguments that, although intuitively attractive, often turn out to be unsatisfactory when applied to the developed countries.

Simon Kuznets (1960) and Julian Simon (1981) have developed an argument that holds that the larger the population the greater the likelihood of it containing a genius who will make a major discovery.[1] The difficulty comes from the fact that this argument is not empirically verified. How could we explain that levels of economic growth and innovation are far higher in countries such as Luxembourg or Germany compared with very populated countries of Asia or Africa? Correlating large numbers of major discoveries,[2] as recorded by historians of science, with population size in ancient Greece and Rome is not really a convincing method of proof (Simon, 1981).

Creative pressure: the pressure exerted by population growth is a stimulus to improve production techniques and innovation. This thesis was keenly defended by Eugène Dupreel (1928). A few decades later, Alfred Sauvy (1966, 1979) noted that "*La pression démographique fait jaillir le progrès*" ("Population pressure brings forth progress").

A third type of argument, presented notably by Esther Boserup (1981, 1992), who drew her inspiration from the history of agriculture and urbanization, rests on the hypothesis that certain technological innovations require a critical mass.[3]

Finally, there is an argument whereby a rise in the relative price of capital, after an increase in the active population, could induce capital-augmenting technical changes. But it can equally be observed that in the case of a fall in population, wages will increase, and as a consequence, we would expect to see labor-augmenting technical changes (Arthur, 1989).[4]

II. SOME DEFINITIONS

The debate over the relations among population, productivity, and technology often suffers from ambiguities and misunderstandings that are easily avoided if care is taken over definition of the terms. Let us begin with the demographic component. The expression "population trends" can denote a variety of phenomena. First and foremost, it includes the population's size and rate of growth, two different concepts. There are also the variations of demographic structure, including the ratio between the economically active and nonactive populations and, within the active population, the age structure of the labor force for we know that the population size varies much more slowly and less extensively than does its structure (Coale, 1986). Many countries have thus experienced a net growth in their active population although their total population was stagnating.

The expression "demographic variable" is often applied to phenomena that are more cultural and sociological in nature: the growing labor force participation of women, urbanization, new modes of family life. In general, these phenomena—sometimes incorrectly described as sociodemographic—have a much greater incidence than do population growth and structure on economic growth and technology. We will not discuss this here (see Volumes I and II devoted to the determinants of fertility, mortality, and migration). In addition, through the next section we will consider only variations in the size of the population.

Similar ambiguities are found in the economic component. How many times, for example, is technological progress confused with productivity? In the terms proposed by André Vincent (1968:1), "productivity is the ratio between an output and the factors of production with which it was obtained, or indeed between an output and some of the factors with which it was obtained."

If only one factor is considered, this is partial productivity. A common approach is to divide the volume of output by the amount of labor used; this gives the partial (or average) productivity of labor, a concept widely employed in the works discussed here. There are several problems associated with this concept. First, if population is taken as an indicator of labor, labor productivity and average income are one and the same although their economic significance is completely distinct. Average income is an aim of economic policy, a measurement of its performance; labor productivity, in contrast, is held to represent the state of technology. Furthermore, any variation in the partial productivity of labor may well reflect interaction effects with the other factors as much as the specific action of the labor factor.

In any analysis of productivity, therefore, we need to consider simultaneously all the factors employed to establish a diagnosis of the behavior of the national economy. By proceeding thus, we obtain an indicator of total factor productivity. In a sense this indicator could be taken to reflect the state of technology and its variation, the rate of technical change. Technological progress is usually associated with the introduction and application of new production procedures that

[1] For example, Simon Kuznets (1960:328) writes: "Population growth . . . would, therefore, produce an absolutely larger number of geniuses."

[2] On this point, it can be noted that it is not so much discoveries as their technological applications which interest us here.

[3] See also John Maynard Keynes (1937) and William Reddaway (1939), who stress the importance of demand and investment in technological innovation and see a direct relationship between population and demand.

[4] On the Ricardo-Malthusian and Boserupian theses, see also Ahlburg (1987), Blanchet (1990), Cigno(1984), Hansen (1994), Keyfitz (1996), Kuznets (1973), Lee (1986, 1992), and Srinivasan (1988).

enable existing factors of production to produce more. It can be noted that the equivalence of technological progress and variation in total productivity only holds if the technology is always used efficiently. If this is not the case—in other words, if for any given level of technology, it is possible to produce more by making more efficient use of the factors of production—we speak of technical inefficiency. It can then be shown that the variation in total factor productivity can be categorized into two terms: technical progress and gains in technical efficiency.

In many respects, the populationist arguments based on the idea of population pressure have much more to do with technical efficiency than with technical progress. The correlation between discovery and technological advance, on the one hand, and population growth, on the other, appears doubtful as a universal law, even if it has been empirically verified in certain countries and at certain periods. However, it is plausible to think that population growth and its corollary, the relative scarcity of the other factors of production, will cause a society to use its resources more efficiently, to avoid any kind of waste, and, in short, to move closer to the efficiency frontier.

Thus, it appears that the gap between labor productivity and technical progress can be both conceptually and empirically very wide. It follows that the effect of population changes on one or other of these variables could be quite different. As an example, let us accept the idea of a positive correlation between population growth and technical progress. For population growth also to have a positive effect on labor productivity, we need to be certain that the technical efficiency and the amounts of the other factors can be held constant.

At the empirical level, it has to be admitted that it is relatively straightforward to calculate the partial productivity of labor. The difficulty lies in measuring the labor force that diverges from the total population as a function of age structure, male and female participation rates, retirement and schooling, and the length of the working week—multiple variables that vary greatly across time and space. In contrast, measuring technical progress is an impossible undertaking. The closest approximation that can be achieved is the total factor productivity, estimation of which requires many assumptions. It may be added that measuring technical progress by the amount of investment in research and development or by an enumeration of major inventions or patents is even more questionable. Clear, then, is that the concept of technical progress, at least at the macroeconomic level, remains a black box into which economists hesitate, more than others, before plunging.

III. EMPIRICAL STUDY

In the most aggregated models, production involves only two commodities: one that is produced and can be used either for consumption or for investment; the other figures only as a factor of production and represents labor. The transformation of the inputs, capital *(K)* and labor *(L)* into output *(Y)* can be represented by a *production function* such as

$$Y_t = F(K_t, L_t, t)$$

where t denotes the period considered. This function reflects the presence of technological constraints and depends on the experience and knowledge prevailing at the time. It varies between periods; hence the inclusion of t in the function. A simple form of production function is the Cobb-Douglas function, which is written

$$Y_t = K_t^{\alpha} L_t^{1-\alpha} e^{\gamma t}$$

where α stands for the share of national product going to capital and γ denotes the rate of technical progress.[5] We can also write $y_t = k_t^{\alpha} e^{\gamma t}$, where $y = Y/L$ and $k = K/L$. Suppose that the rate of technical progress is a function of n, the rate of growth of labor, which in the present case is equated with the population. We write

$$\hat{y}_t = \alpha(\hat{K}_t - n_t) + \gamma \qquad (1)$$

where $\hat{y}_t$ is the rate of growth of average labor productivity and $\hat{K}_t$ is the rate of growth of the capital stock. What impact does the rate of population growth have on the growth of productivity? Deriving $\hat{y}_t$ relative to n_t, we obtain

$$\frac{d\hat{y}_t}{dn_t} = -\alpha + \frac{\partial \gamma}{\partial n_t}$$

Put differently, for population growth not to depress the growth of productivity, an effect generated by diminishing returns, it will have to influence significantly technical progress.

Empirical studies of the relationship between population growth and economic growth are numerous. For a long time, no conclusion could be reached one

[5] We can illustrate the distinction between technical progress and efficiency gains by rewriting this function:

$$Y_t = K_t^{\alpha} L_t^{1-\alpha} e^{\gamma t} E_t$$

where E denotes the rate of efficiency (between zero and one). Differentiating, we obtain

$$\hat{y}_t = \alpha \hat{k}_t + \gamma + \hat{E}_t$$

In other words, the growth in income per head depends on growth in capital per head, technical progress and the variation in technical efficiency ($\hat{E}_t$). Naturally, where this efficiency does not vary $\hat{E}_t = 0$.

way or the other. In the past decade, an increasing number of authors seem to obtain a significant negative relationship.

Didier Blanchet (1996), James Brander and Steve Dowrick (1994), and Allen Kelley and Robert Schmidt (1994) represent this new type of research, which has access to more complete retrospective data and more sophisticated statistical techniques compared with earlier studies (see also Kormendi and Meguire, 1985; Barro, 1989; Ehrlich and Lui, 1991; Yip and Zhang, 1996). It should also be noted that recent years have been marked by new attempts to model the interaction between fertility and economic growth in a context of endogenous growth. The majority of these models usually end up with a negative relationship between fertility rate and economic growth rate. Rather than produce a synthesis of these recent works, we preferred to conduct our own verification of this new and relatively recent consensus. We decided to verify the relationship by using two approaches: a cross-sectional approach and a time-series approach.

In the cross-sectional approach, we used the data compiled by Robert Summers and Alan Heston (1991), known as the Penn World Tables (for a description of the different version of the tables, see Summers and Heston, 1991). These authors formed an economic and demographic data set covering the period 1950–1990. We divided these 40 years into seven subperiods of 10 years with 5-year overlaps. The results of estimating the simple model $\hat{y} = a + bn$ (where $\hat{y}$ is the rate of growth of real gross national product per person, n the rate of population growth, and a and b the parameters to be estimated) are given in Table 85–1. They show a negative relationship between these two variables.

Ideally, this relationship would be tested over a long period, not only 40 years. This is what we do in the time-series approach. Here we used the data compiled by Angus Maddison (1995). These cover 56 countries over a period from 1870–1994. This 124-year period has been divided into subperiods of 15 years, with 5-year overlaps. Three cases are studied:

One in which all the data are used as if we were dealing with a single entity.
One in which the constant a is country specific. In this way the national particularities are taken into account.
One in which the constant a is specific to each period, which captures the temporal differences.

The regression results are presented in Table 85–2. As in the cross-sectional approach, we obtain a negative relationship that is significant except when the constant is country specific. In the latter case, the time-series approach produces a weaker association between population growth and economic growth than does the cross-sectional approach; in the latter, parameter b, which measures the relationship, is almost always over one-half in absolute terms, whereas it is always below this value in the time-series approach.

Tables 85–3 and 85–4 present the correlation coefficients by country and by period. A difference is observed depending on the group of countries. For most of the "industrialized" and African countries, the relationship is negative; in contrast the majority of the countries of Asia and eastern Europe present a positive correlation. Regarding the evolution of the relationship over time, a stable negative correlation does not emerge until the second half of the 20th century.

IV. THE GROWTH THEORY PERSPECTIVE

Traditional growth theory (see Barro and Sala-I-Martin, 1995) devotes little attention to the demographic dimension nor *a fortiori* to its links with

TABLE 85–1 Estimation of the Simple Model (Cross-Sectional Approach)

Period	No. of observations	Constant (a)	Population growth (b)	R squared
1950–1960	60	3.78 (0.4785)[a]	–0.6576 (0.2058)[a]	0.1497
1955–1965	74	4.196 (0.4858)[a]	–0.6140 (0.1976)[a]	0.1182
1960–1970	125	4.1687 (0.5102)[a]	–0.5190 (0.2163)[b]	0.0447
1965–1975	126	3.7052 (0.479)[a]	–0.3942 (0.2119)[c]	0.0272
1970–1980	133	3.7928 (0.4581)[a]	–0.6549 (0.1911)[a]	0.0823
1975–1985	134	2.8058 (0.4386)[a]	–0.7987 (0.1784)[a]	0.1318
1980–1990	111	2.5418 (0.4055)[a]	–1.0441 (0.1815)[a]	0.2329
Total	763	3.5 (0.19)[a]	–0.6537 (0.0820)[a]	0.0771

Value in parentheses is the standard error.
[a]Significant at the 1% level; [b]significant at the 5% level; [c]significant at the 10% level.

TABLE 85–2 Results of the Time-Series Approach

	No. of observations	Constant (a)	Population growth (b)	R squared
1	944	2.1102 (0.1189)[a]	–0.1722 (0.0706)[b]	0.0063
2	944	individual for each country	0.0386 (0.1225)[c]	0.1380
3	944	individual for each period	–0.4377 (0.0666)[a]	0.2171

Value in parentheses is the standard error.
[a]Significant at the 1% level; [b]significant at the 5% level; [c]nonsignificant.

TABLE 85–3 Correlation Coefficients by Country

Country	Correlation coefficient	Country	Correlation coefficient
Industrialized countries		**Latin America**	
Australia	–0.1723	Argentina	0.4611
Austria	0.0039	Brazil	0.7058
Belgium	–0.1620	Chile	–0.2045
Canada	–0.0243	Colombia	–0.1618
Denmark	–0.2919	Mexico	0.4644
Finland	–0.4927	Peru	0.0763
France	0.7941	Venezuela	–0.3891
Germany	–0.0272		
Great Britain	–0.4539	**Asia**	
Italy	–0.0674	Bangladesh	0.5723
Japan	–0.2477	Burma	–0.6711
Netherlands	0.1284	China	0.6110
New Zealand	–0.1585	India	0.7162
Norway	–0.1131	Indonesia	0.5313
Sweden	0.0664	Pakistan	0.6918
Switzerland	0.1253	Philippines	0.0441
United States	–0.1292	South Korea	–0.3315
		Taiwan	0.0530
Other European countries		Thailand	0.5961
Greece	–0.4871		
Ireland	0.5184	**Africa**	
Portugal	–0.6728	Democratic Republic of Congo	–0.7366
Spain	0.1324	Ivory Coast	–0.4076
Turkey	–0.6498	Egypt	–0.1042
		Ethiopia	–0.8417
Eastern Europe		Ghana	0.6664
Bulgaria	0.4379	Kenya	–0.6176
Czechoslovakia	–0.3320	Morocco	–0.7990
Hungary	0.3696	Nigeria	–0.7458
Poland	–0.4243	South Africa	0.1080
Romania	0.3104	Tanzania	0.1453
Soviet Union	0.3716		
Yugoslavia	0.3044		

productivity and technical progress. Both population growth and the rate of technical progress are treated as exogenous. In the long run, growth in labor productivity or in average income depends solely on the rate of technical progress. From Eq. 1 above, in equilibrium, we get

$$\hat{y} = \gamma/(1 - \alpha) \tag{2}$$

where, it will be remembered, $\hat{y} = \hat{Y} - n$ is the difference between the rate of growth of national output and the rate of population growth, γ is the rate of technical progress, and $(1 - \alpha)$ is the share of national income accruing to labor.

A rise in population growth will have no effect on the growth of productivity; however, it depresses the level of productivity. Because increasing returns are excluded from this analysis, the effects of scale that population growth brings about, according to Colin Clark (1978), cannot occur. Julian Simon (1986) and Julian Simon and Gunter Steinmann (1992) have

TABLE 85–4 Correlation Coefficients by Period

Period	Correlation coefficient
1870–1885	0.3544
1880–1895	–0.6564
1890–1905	0.1902
1900–1915	0.0932
1910–1925	–0.1555
1920–1935	–0.1963
1930–1945	0.0888
1940–1955	–0.2420
1950–1965	–0.5505
1960–1975	–0.4823
1970–1985	–0.3324
1980–1994	–0.1426

developed the concept of induced technological progress, a concept first introduced by Kenneth Arrow (1962). Technological progress is viewed here as a cause and as a consequence of growth. Population growth would have a positive feedback effect on scientific knowledge, leading to an acceleration of growth and progress. With this hypothesis, we rewrite Eq. 2:

$$\hat{y}(n) = \gamma(n)/(1 - \alpha) \tag{2'}$$

Most of the works that have attempted to formalize the influence of population on productivity and economic growth consider the situation before long-term equilibrium and often resort to simulation techniques (see Steinmann and Komlos, 1988). It has to be confessed that the conclusions they produce usually follow from the initial assumptions about the relations between population and technology, which need to be checked for empirical validity.

In an empirical perspective, a whole series of works have attempted to quantify the role of particular factors in economic growth. Working with the production function used above, the exercise would start with the following equation:

$$\hat{Y} = \gamma + \alpha\hat{K} + (1 - \alpha)n \tag{3}$$

In this equation the growth of output thus appears as the sum of the contributions of technical progress, capital, and labor. Authors such as Edward Denison (1967) and Jean-Jacques Carré *et al.* (1972) have attempted to disaggregate the contribution of each factor. Thus, for population they distinguish its size, time spent working, educational level, activity rates, and the age–sex structure of the population.

Inclusion of age and sex reflects the hypothesis that productivity varies with these two characteristics and, more specifically, that young and old workers are less productive than are those in their middle years and that female workers have lower productivity than do male workers. In the case of France, the inclusion of age caused the index of labor quality to rise by 2% between 1949 and 1962; at the same time, the inclusion of a sex variable has a negligible effect. Overall, the contribution of the age and sex variables to growth in the various Organization for Economic Cooperation and Development countries is tiny. The contributions of capital and above all of technical progress predominate clearly.

These works are based on the paradigm of what is known today as exogenous growth, the rate of which is determined by the growth—quantitative but also perhaps qualitative—of the population. Two decades ago a new paradigm emerged, that of endogenous growth, under which the rate at which the economy grows can be influenced by the savings ratio, investment in education, or the size of the population.

This paradigm is founded on recognition of increasing returns to scale and of a technological externality. Whether for physical or human capital, economic agents fail to see that not only their own investments but the sum of all investments influence output. Hence they underestimate the real return of their actions, which makes necessary intervention by the public authorities.

In the original version developed by Paul Romer (1986), endogenous growth depends on population size, which can be seen as reflecting the scale of the economy.

Let there be a Cobb-Douglas production function:

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$$

where A_t is a technology function that depends on the capital stock $K_t = \sum K_t^i$, where K_t^i is the stock accumulated by agent i at period t. Each agent thinks that the return on his investment is $\dfrac{\partial Y}{\partial K_t^i} = \alpha A_t K_t^{\alpha-1} L_t^{1-\alpha}$ whereas in reality if $A_t = K_t^{1-\alpha}$, its real return is $\alpha L_t^{1-\alpha}$. If the rate of savings, s, is fixed in this economy, we write the following:

$$\hat{K}_t = s(1 - \alpha)L_t^{1-\alpha}$$

Thus, we see that both the rate of growth of the economy and the social return to capital depend on population size. Naturally this is an elementary version of an endogenous growth model. Yet it illustrates well the main message of this new paradigm: the rate of economic growth does not depend on population growth alone.

The problem with endogenous growth is that its message depends on assumptions made about technology and externalities. Depending on these assumptions, the relationship between the rate of economic growth and the size of the population, its rate of growth, and even the population structure can vary in opposite directions.

An interesting line of research is the one that combines the endogeneity of fertility specific to the new home economics inspired by the work of Becker, with the technological externality of the endogenous growth models. Two articles provided the initial impetus for this research (Barro and Becker, 1989; Becker *et al.*, 1990; see also Challier and Michel, 1996; Ehrlich and Lui, 1997); it has been validated by several empirical studies (Rosenzweig, 1990; Barro, 1991). These authors work from the idea that agents choose both the number of their children and the level of skills—known as the level of human capital—they wish to give them. Based on the choice—one might even say the trade-off—between these two variables that the authors describe as the quantity and quality of children, we can identify two stable equilibria for the economy: one has high fertility but low levels of human capital and economic growth, the other has low fertility but high levels of human capital and economic growth.

This theory of endogenous fertility sheds fresh light on the negative relationship that we obtained in the previous section.

V. DEMOGRAPHIC STRUCTURE AND PRODUCTIVITY

So far, emphasis has been placed on population increase or decrease without considering the question of age structure. Yet as the most pessimistic projections indicate, it is going to take time for the populations of the industrialized countries to fall significantly, whereas soon we will be witnessing radical changes in their age distributions. Changes in age structure can influence productivity in two ways. First, it is observed that profiles of activity and of productivity vary with age and that, in addition to the age effect, there is a double cohort effect: a size effect (the productivity of a cohort is a function of its size) and a technical change effect (the training of young people incorporates technical progress). Second, at any given date the contents of the basket of goods and services differ sharply when different age groups are considered. This difference influences the productivity of the economy considered in the aggregate.

1. Age and Productivity

Are the efficiency and productivity of labor differentiated by age? An affirmative answer would mean that as regards productivity, for an active population constant over time, some demographic structures are more favorable than are others. We immediately come against the question of how to measure individual productivity. Because the productivity of labor cannot be observed directly, the usual solution is to approach it through its earnings, that is, its cost. This convention amounts to accepting that wages are determined by the marginal productivity of workers. This is a questionable assumption, at least in some respects, although rejecting it is hard in the absence of an effective alternative (for a discussion of this question, see Chesnais, 1978).

Age-earning profiles have an asymmetric hump shape that varies with sex, type of employment, educational level, sector of activity, size, and nature of the firm. Incomes rise up to a certain age more or less close to retirement and then level off and fall back slowly.

It should be noted that what we usually observe are cross-sectional profiles and not the evolution of wages over the life cycle of a cohort. Thus, there is an age effect and a cohort effect, with the latter being a function of technical progress and of cohort size. A younger cohort will tend to have a higher productivity and hence a higher wage at a given age. But in addition to this technological factor, it would seem that both the level and the form of the age-earnings profile are a function of cohort size. A large cohort, such as that of the baby boom, would have a lower wage level and a less steeply rising profile than would a small cohort placed in the same circumstances. This hypothesis has been shown to be correct for many countries. In some European countries, the arrival in the labor force of large cohorts has also had the effect of raising relative unemployment levels (Lee *et al.*, 1988; Bloom *et al.*, 1987).

Thus, a change in the age structure of the economically active population, in the sense of aging, for example, will have contrasting effects on the average productivity of labor. We make the assumption of a constant active population whose membership is subject to aging. The age effect should lead to a productivity increase owing to the relative decline in the number of young workers and the concomitant increase in the proportion of older workers. The size effect mentioned above should also have an incidence of this kind, the young workers present in smaller numbers having a higher productivity. However, the "technical progress" effect should imply a decrease in productivity. There will be relatively fewer young people with technical skills that are more advanced than those of their elders.

Paul Levine and Olivia Mitchell (1988) have attempted to evaluate the influence that the American baby boom can be expected to have in 2020, when old workers will belong to this cohort and when young workers will be relatively few in number. Their conclusion is of a generalized increase in wages, particularly for the young workers. In a long-term simulation on the Netherlands, Evert van Imhoff (1988) highlights the crucial importance of the relative number of young workers whose skills encompass the most advanced technology. He assumes that it is more expensive and harder for the older workers to acquire equivalent skills. For this reason, the growth and productivity of the Dutch economy are extremely sensitive to population variations.

2. Age and Consumption

A distinction must be made between the level and structure of consumption. The theory of the life cycle assumes that individuals save while they are economically active to fund their consumption in old age. Two phases can be identified—that of saving, which concerns the period of activity; and that of dis-saving, which covers the period of retirement—of varying intensity depending on whether or not the individual in question leaves accumulated assets at death. This produces a typical age-specific assets profile that rises and then falls. At the macroeconomic level, the older the population, the lower the savings ratio. In the short-term, a strong consumption demand may be favorable to productivity, but in the medium-term, a low level of savings implies a reduction in the capital stock, which will contribute to depressing labor productivity.

An age effect also emerges clearly in the structure of consumption. The basket of goods and services at a given date is very different according to whether we consider young or retired couples. This is clear in the case of expenditure on the home and consumer durables, which is more concentrated at the beginning of the active period and is influenced by family behavior (Ermisch, 1988). The age effect is also observed in expenditure on food, leisure, and clothing, but the effect is most marked in health expenditure, which rises sharply over the life cycle.

If the emphasis is shifted from losses of expenditure to commodities in particular, we can speak of "old" commodities (e.g., televisions) and "young" commodities (e.g., motorcycles). For the present discussion, we need to know whether the basket of goods and services for elderly people incorporates a more advanced technology, a greater use of capital, than that of young households. Not much empirical work has been done on this question (see Luptacik and Schmoranz, 1980, 1986, 1989; Serow, 1984; Ekert-Jaffé, 1989). One widely held view is that elderly people have greater recourse to services than do younger people, and consequently, their basket of goods and services would imply a lower productivity. Technical progress, as measured by the growth in the total productivity of the factors of production, is known to be much lower in the service sector (zero for 1979–1983) than in manufacturing industry as a whole (1.8% per year for 1979–1983) (Englander and Mittelstädt, 1988).

Even if this idea was empirically correct for the present, it could be invalidated in the future. One of the consequences of population aging and longer life spans is the need for a gradual substitution of the traditional care services for elderly people by techniques based on smaller labor inputs.

VI. BACK TO SOCIOLOGY

The preceding sections sum up quite well the position of the majority of economists. Overall, demographic variations have no predictable influence on technical progress. They influence the partial productivity of labor to the extent that they are not accompanied by parallel changes in the other factors of production. Can nothing more be said? Are there not grounds for going beyond the economic standpoint? As is known, the strength—but also the weakness—of economic methodology lies in its scientific ambition. It aspires to rigor and universality and shuns particular and transitory considerations of a psychosociological origin.

Now if it is true that we cannot say that population growth necessarily has a positive influence on the state of technology and skills in a society, nor can we deny that such an influence has operated in some cases. Thus, to explain the rapid and sustained growth experienced by France in the period 1950–1970, some authors have invoked the recovery in the birth rate and the behavioral changes it is thought to have produced. According to Carré *et al.* (1972:252), "An attitude of narrow pessimism and economic conservatism ('malthusianism') may have given way to the desire to prepare a better future for the young people coming along; then, with the arrival of the new generations, the dynamism characteristic of the young would have spread to the rest of society. The changes on which growth depends thus came to be accepted where they had hitherto been rejected."

In their study of French economic growth published in 1972, Jean-Jacques Carré *et al.* challenge this explanation but do not actually reject it. The authors even admit that it could be of some importance. In his critique of the populationist thesis, which he judges unverifiable, Joseph Stassart (1965) acknowledges that there may be links between demographic change and a psychosociological structure of a nation, and that this linkage could provide some basis for the idea of creative pressure.

In other words, because there is no definitive theory of technical progress and because empirical studies on growth factors often conclude that for some explanations a historical and sociological analysis extending further into the past is needed, we cannot say that demographic variables never have an effect on growth and technology.

As Ron Lesthaeghe and Johan Surkyn (1988) have shown for fertility, economic factors do not explain everything; cohort effects also play an important role. Some birth cohorts are more motivated, more inventive, and more dynamic than are others. Why should this be? A desire for vengeance, an armed conflict, a political challenge, a recovery in fertility? To address these questions, we must proceed case by case, and even then, many areas of uncertainty remain. There is no better illustration of the present limitations of the social sciences, including economics.

There is another angle from which sociology can shed light on the relationship between population growth and economic growth. An aging population will tend to have a more static organization; levels of occupational mobility will be lower (Keyfitz, 1982). However, it will contain more small families, and members of smaller sibling groups are known to achieve greater educational and labor market success (Blau and Duncan, 1967). Lastly, the way in which fertility is reduced across income groups is important. An extreme case can be imagined: families experiencing chronic economic insecurity continue to have many children, whereas the middle class adopts a lower fertility. The consequences of such a change are obvious. Few studies have been conducted on this specific point (McNicoll, 1986).

Empirical works dealing with these social changes are not only rare but should be treated with caution. They are based on a period of higher population growth, and there is a possibility that social behavior, modes of family life, and the rules of social and spatial mobility become modified in an aging society.

CONCLUSION

The purpose of this chapter was to review the works dealing with the influence of population variations on economic growth. We have attempted to provide a synthesis and a critical analysis. Three lessons can be drawn.

First, there are no theoretical arguments or empirical studies that enable us to conclude unequivocally in favor of any kind of relationship between population growth and technical progress. It may be added that at present we possess no good explanatory theory of technical progress. Historians and economists are reduced to supplying interpretations and explanations on an ad hoc basis.

Next, the influence of population on productivity and, in particular, the partial productivity of labor is obvious. This is consistent with the law of diminishing returns. An aging population should experience rising productivity. This influence can, however, be countered by four factors: the variable gap between population and active population, the constraint of scarce resources, capital accumulation, and technical progress.

Lastly, the aging of a population has multiple consequences, both economic and noneconomic. These consequences can ultimately affect the culture and values of our societies. In what way and how? No one can say. By contrast, what is likely is that these mutations can have an effect on technical progress and efficiency in production. This effect is as unpredictable as the effect of aging on culture and social values.

Acknowledgments

We would like to thank Serge Feld and Sergio Perelman for their valuable suggestions. We also thank Francois Héran, Graziella Caselli, Jacques Vallin, and Guillaume Wunsch, as well as the members of the San Miniato seminar for their comments.

This chapter was translated from the French version by Godfrey I. Rogers.

References

Ahlburg Dennis A., 1987. The impact of population growth on economic growth in developing nations: the evidence from macroeconomic-demographic models, *in:* David Gale Johnson and Ronald Lee (eds.), *Population growth and economic development: issues and evidence*, p. 479–521. Madison, University of Wisconsin Press, 702 p.

Arrow Kenneth, 1962. The economic implications of learning by doing, *Review of Economic Studies*, Vol. 29(3), p. 155–173.

Arthur W. Brian, 1989. Competing technologies, increasing returns and lock-in by historical events, *Economic Journal*, vol. 99(394), p. 116–131.

Barro Robert J., 1989. A cross-country study of growth, saving and government, *in:* Douglas Bernheim and John Shoven (eds.), *National saving and economic performance*, p. 269–301. Chicago, University of Chicago Press, 385 p.

Barro Robert J., 1991. Economic growth in a cross-section of countries, *Quarterly Journal of Economics*, vol. 106(2), p. 407–444.

Barro Robert J. and Becker Gary S., 1989. Fertility choice in a model of economic growth, *Econometrica*, Vol. 57(2), p. 481–501.

Barro Robert J. and Sala-I-Martin Xavier, 1995. *Economic growth*. New York, McGraw Hill, 539 p.

Becker Gary S., Murphy Kevin M. and Tamura Robert, 1990. Human capital, fertility and economic growth, *Journal of Political Economy*, vol. 98(5), p. S12–S37.

Blanchet Didier, 1990. Population growth and income growth during the demographic transition: does a Malthusian model help explain their relationship? *Population: An English Selection*, vol. 2, p. 37–52.

Blanchet Didier, 1996. L'inversion des effets de la croissance démographique sur la croissance économique. Interprétations générales et spécificité africaine, *in:* Jean Coussy and Jacques Vallin (eds.), *Crise et population africaine*, p. 51–69. Paris, CEPED, 580 p.

Blau Peter and Duncan Otis D., 1967. *The American occupational structure*. New York, Wiley, 520 p.

Bloom David E., Freeman Richard and Korenman Sanders, 1987. The labour-market consequences of generational crowding, *European Journal of Population*, vol. 3(2), p. 131–176.

Boserup Esther, 1981. *Population and technology.* Oxford, Basil Blackwell, 255 p.

Boserup Esther, 1992. Croissance démographique et économique en économie ouverte, *Population*, vol. 47(6), p. 1505–1512.

Brander James A. and Dowrick Steve, 1994. The role of fertility and population growth in economic growth: empirical results from aggregate cross-national data, *Journal of Population Economics*, vol. 7(1), p. 1–27.

Carre Jean-Jacques, Dubois Paul and Malinvaud Edmond, 1972. *La croissance française.* Paris, Le Seuil, 709 p.

Challier Marie-Christine and Michel Philippe, 1996. *Analyse dynamique des populations. Les approches démographiques et économiques.* Paris, Economica, 215 p.

Chesnais Jean-Claude, 1978. Age, productivité et salaires, *Population*, vol. 33(6), p. 1155–1188.

Cigno Alessandro, 1984. Consumption versus procreation in economic growth, *in:* Gunther Steinmann (ed.), *Economic consequences of population change in industrialized countries*, p. 2–28. Berlin, Springer-Verlag, 415 p.

Clark Colin, 1978. Population growth and productivity, *in:* Simon Julian (ed.), *Research in population economics*, Vol. 1, p. 143–154. Greenwich, JAI Press, 347 p.

Coale Ansley J., 1986. Demographic effects of below replacement fertility and their social implications, *in:* Kinsley Davis, Mikhail S. Bernstam and Rita Ricardo-Campbell (eds.), "Below-replacement fertility in industrial societies: causes, consequences, policies," *Population and development review*, sup. to vol. 12, p. 203–216.

Denison Edward, 1967. *Why growth rates differ?* Washington DC, Brookings Institute, 494 p.

Dupaquier Jacques, 1980. Avez-vous lu Malthus? *Population*, vol. 35(2), p. 279–290.

Dupreel Eugène G., 1928. Population et progrès: essai sur les conséquences des variations démographiques, *in:* Eugène Dupréel (ed.), *Deux essais sur le progrès*, 2e partie. Brussels, Maurice Lamertin, 272 p.

Ehrlich Isaach and Lui Francis, 1991. Intergenerational trade, longevity and economic growth, *Journal of Political Economy*, vol. 99(5), p. 1029–1059.

Ehrlich Isaac and Lui Francis T., 1997. The problem of population and growth: a review of the literature from Malthus to contemporary models of endogenous population and endogenous growth, *Journal of Economic Dynamics and Control*, vol. 21(1), p. 205–242.

Ékert-Jaffé Olivia, 1989. Vieillissement et consommation: quelques résultats tirés des enquêtes françaises sur les budgets des ménages, *Population*, vol. 44(3), p. 561–579.

Englander A. Steven and Mittelstadt Axel, 1988. La productivité totale des facteurs: aspects macroéconomiques et structurels de son ralentissement, *Revue Economique de l'OCDE*, no. 10, p. 7–64.

Ermisch John F., 1988. Demographic patterns and the housing market, *in:* Ronald D. Lee, W. Brian Arthur and Gerry Rodgers (eds.), *Economics of changing age distributions in developed countries*, p. 155–182. Oxford, Clarendon Press, 221 p.

Fauve-Chamoux Antoinette (ed.), 1984. *Malthus hier et aujourd'hui.* Paris, Editions du CNRS, 503 p.

Hansen Stein, 1994. La population: un défi pour les spécialistes des sciences sociales et économiques, *Revue Internationale des Sciences Sociales*, no. 141, p. 395–408.

Kelley Allen C. and Schmidt Robert M., 1994. *Population and income change: recent evidence.* Washington, World Bank, 116 p.

Keyfitz Nathan, 1982. *Population change and social policy.* Boston, Abt, 265 p.

Keyfitz Nathan, 1996. Population growth, development and the environment, *Population Studies*, vol. 50(3), p. 335–359.

Keynes John M., 1937. Some economic consequences of a declining population, *Eugenics Review*, vol. 29(1), p. 13–17.

Kormendi Roger and Meguire Philip, 1985. Macroeconomic determinants of growth: cross-country evidence, *Journal of Monetary Economics*, vol. 16, p. 141–163.

Kuznets Simon, 1960. Population change and aggregate output, *in:* National Bureau Economic Research [NBER] (ed.), *Demographic and economic change in developed countries*, p. 324–340. Princeton, NBER, Princeton University Press, 536 p.

Kuznets Simon, 1973. *Population, capital and growth.* New York, W.W. Norton, 342 p.

Lee Ronald D., 1986. Malthus and Boserup: a dynamic synthesis, *in:* David Coleman and Roger Schofield (eds.), *The state of population theory: forward from Malthus*, p. 96–130. London, Basil Blackwell, 311 p.

Lee Ronald D., 1992. Croissance démographique, progrès et pauvreté, *Population*, vol. 47(6), p. 1533–1554.

Lee Ronald D., Arthur W. Brian and Rodgers Gerry (eds.), 1988. *Economics of changing age distributions in developed countries.* Oxford, Clarendon Press, 221 p.

Lesthaeghe Ron and Surkyn Johan, 1988. Cultural dynamics and economic theories of fertility change, *Population and Development Review*, vol. 14(1), p. 1–45.

Levine Paul B. and Mitchell Olivia S., 1988. The baby boom's legacy: relation wages in the twenty-first century, *American Economic Review. Papers and Proceedings*, vol. 78(2), p. 66–69.

Lupthacik Mikulas and Schmoranz Ingo, 1980. Demographic changes and economic consequences: demo-economic multiplier for Austria, *Empirical Economics*, vol. 5(1), p. 55–67.

Luptacik Mikulas and Schmoranz Ingo 1986. *The modelling of demographic-economic dependences within an input-output framework* (mimeograph).

Luptacik Mikulas and Schmoranz Ingo 1989. An extension of static input-output model for demographic-economic analysis, *in:* Alois Wenig and Klaus F. Zimmermann, *Demographic change and economic development*, p. 253–271. Berlin, Springer-Verlag, 324 p.

Maddison Angus, 1995. *L'Economie mondiale 1820–1992. Analyse et statistiques.* Paris, OECD, 333 p.

Malthus Thomas R., 1803. *An essay on the principle of population, or a view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*, 2e ed. New-York, The Modern Library, 352 p.

McNicoll Geoffrey, 1986. Economic growth with below-replacement fertility, *in:* Kinsley Davis, Mikhail S. Bernstam, and Rita Ricardo-Campbell (eds.), "Below-replacement fertility in industrial societies. Causes, consequences, policies," *Population and development review,* sup. to vol. 12, p. 217–239.

Reddaway William, 1939. *The economics of a declining population.* London, Allen and Unwin, 272 p.

Romer Paul, 1986. Increasing returns and long run growth, *Journal of Political Economy,* vol. 94(5), p. 1002–1037.

Rosenzweig Marc, 1990. Population growth and human capital investments: theory and evidence, *Journal of Political Economy,* vol. 95(5), p. S38–S70.

Sauvy Alfred, 1966. *Théorie générale de la population,* 2 vols. Paris, PUF, 776 p.

Sauvy Alfred, 1979. De la baisse de la fécondité en Europe occidentale, *in:* Milos Macura (ed.), *The effects of current demographic change in Europe on social structure: proceedings of the third European population seminar,* p. 81–92. Belgrade, Ekonomski Institut, 215 p.

Serow William J., 1984. The impact of population change on consumption, *in:* Gunter Steinmann (ed.), *The effects of current demographic change in Europe on social structure: proceedings of the third European population seminar,* p. 168–178. Berlin, Springer-Verlag, 415 p.

Simon Julian L., 1981. *The ultimate resource.* Princeton, Princeton University Press, 415 p.

Simon Julian L., 1986. *Theory of population and economic growth.* Oxford, Basil Blackwell, 232 p.

Simon Julian L. and Steinmann Gunter, 1992. Population, natural resources, and the long-run standard of living, *in:* Julian Simon (ed.), *Population and development in poor countries: selected essays,* p. 89–121. Princeton, Princeton University Press, 463 p.

Srinivasan Tirukodikaval N., 1988. Population growth and economic development, *Journal of Policy Modeling,* vol. 10(1), p. 7–28.

Stassart Joseph, 1965. *Les avantages et les inconvénients d'une population stationnaire.* Liège, Université de Liège, Faculté de droit, 256 p.

Steinmann Gunter and Komlos John, 1988. Population growth and economic development in the very long-run: a simulation model of three revolutions, *Mathematical Social Sciences,* vol. 16(1), p. 49–63.

Summers Robert and Heston Alan, 1991. The Penn World Tables (Mark 5): an expanded set of international comparisons, 1950–1988, *Quarterly Journal of Economics,* vol. 106(2), p. 327–368.

Van Imhoff Evert, 1988. *Optimal economic growth and non-stable population.* Rotterdam, Kripsreproneppel, 214 p.

Vincent André, 1968. *La mesure de la productivité.* Paris, Dunod, 305 p.

Yip Chong K. and Zhang Junsen, 1996. Population growth and economic growth: a reconsideration, *Economics Letters,* vol. 52, p. 319–324.

CHAPTER

86

Population and Development

JEAN COUSSY

Ecole des hautes etudes en sciences sociales (EHESS), Paris, France

INTRODUCTION

"Population and development" is a dated concept. Its introduction in the 1950s was inspired by the conjunction of the "discovery"[1] of underdevelopment and fears aroused by the first phase of the demographic transition. The formulation was intended to respond to the needs of emerging states and aid donors, and posed its questions in macroeconomic terms: what was the role of population growth in underdevelopment and its reproduction? Was there an optimal population for a given area of land (problem of density)? Was there an optimal rate of population growth? Did large countries have an advantage in the process of development (problem of dimension)? How does population growth increase the costs of development by requiring demographic investments and the accumulation of capital? What demoeconomic strategy should emerging nations adopt? Were international inequalities a function of differences in demographic regulation? Was economic growth a necessary condition for improvement in demographic indicators (reduction of mortality and morbidity)? Would increases in the standard of living lead to increased fertility (and a Malthusian trap) or fertility decline (and the second phase of the demographic transition)? Could international action (financial aid and information) help resolve the effects of underdevelopment and population growth?

In sum, traditional macroeconomic and macrodemographic questions were reformulated and quantified at the world level. Ideological debates and traditional politics exercised a strong influence on early research and the first statements of position.

There was a rapid re-emergence of Ricardo's classical ideas: a limited supply of land, decreasing returns to agriculture, and the increasing price of food products (wage goods) would lead to a decrease in both profits and the rate of accumulation, and eventually the halt of economic growth. The conclusions can be refined slightly by hypothesizing a phase of increasing returns preceding the phase of decreasing returns. This modification leads to a rediscovery of the concept of optimum population, the level of population that maximizes per capita production.

Ideas based on decreasing returns to agriculture were counterbalanced by a vision that was more optimistic and more favorable to agriculture. This school of thought borrowed several themes from the physiocrats: the capacity of agriculture to create a surplus, making reproduction possible; the complementarity of productive activities (illustrated by the zig-zag of François Quesnay[2]); the necessity of a boost in agricultural activity preceding the takeoff of an industrial economy (Bairoch, 1967); and the possibility that technical progress in agriculture would make it possible to avoid any obstacles imposed by food supply,

[1] In the sense used by Ignacy Sachs in *La découverte du Tiers Monde* ("The discovery of the Third World"), 1971.

[2] In the *Tableau économique* (Quesnay, 1758), a reprint of which can be found in *Quesnay et la physiocratie* (Sauvy and Hecht, 1958), among other sources.

even with a spectacular growth in world population (Klatzman, 1983).

Contemporary thinkers retained little of Malthus' highly elaborate analysis of development; his analysis of demand was notably neglected in the analyses of development of the 1950s. Instead, attention was focused on his analysis of the positive effects of increases in standard of living on birth rates and the negative effects of population growth on standards of living. The succession of these two relationships produces the Malthusian trap. The term "Malthusian" continued to denote a pessimistic vision of the effects of population growth, and it will be used in that sense in the following analysis.

In opposition to the Malthusian vision, both in scientific analysis and at the major world population conferences (notably the one in Bucharest), Marxist positions refuse to attribute poverty and underdevelopment to population growth. Instead, they argue that capitalist dynamics are at the root of these negative outcomes, as well as excesses or deficits of population themselves.[3] Countries disconnected from capitalism could use population as an instrument of development until rising standards of living led to a voluntary reduction in birth rates.

Mercantilist populationism, the quest for population growth as a means toward economic growth and power, resurfaced in numerous poor states. Though less explicitly cited than are Marxist views, this ideology was incontestably present in the outlook of many leaders. According to this perspective, all population growth was desirable, even when it lowered average revenue, because leaders hoped that it would result in an increase in the number of people available to join military forces, an opportunity for revenue from taxation, or an increase in rates of accumulation.[4]

I. THE GLOBAL ANALYSES OF THE PERIOD 1950–1960

During the years 1950–1960, the growing availability of statistics and the introduction of the first models transformed the traditional debates. The available statistics stimulated comparative studies on the growth of per capita gross domestic product (GDP).

A surprising number of tests simply demonstrated that subtracting the population growth rate from the rate of growth of the GDP led to a sharply reduced per capita GDP rate, and even occasionally a negative rate, and suggested that population growth was thus responsible for the slow pace of development. This conclusion confounded arithmetic operations with causal explanations and relied on the strong unstated hypothesis that population growth had not produced any increase in GDP.

Serious studies of the relationship between the growth rate of per capita income and the population growth rate repeatedly showed that there was no observable correlation (Chesnais and Sauvy, 1993). These findings led to doubt over the Malthusian interpretation of underdevelopment and criticism of the results of the macro models then under construction.

The authors of the first macro models worried about the influence of population growth on needs and the possibilities of accumulation. Prasanta Chandra Mahalanobis (1953–1954) using the model constructed by G. Feldman (1964) for soviet planning, Ansley Coale and Edgar Hoover (1958), Léon Tabah (1968, 1969), Hollis Chenery and Alan Strout (1966) using Roy Harrod (1939), and Elsey Domar (1957) assumed exogenous population growth, a macroeconomic production function (with complementary factors, returns on a constant scale, and no technological progress), and a Keynesian savings function.[5]

Under these hypotheses, population growth requires a proportional increase in investment if per capita capital, average income, and employment rates are to be maintained. The absence of this investment sets the stage for decreased per capita income, decreased savings rates, further decreases in investment, etc. In response to population growth, there

[3] These excesses and deficits evolve along with the different eras of capitalism. Capitalism, for instance, caused forced migrations (overpopulating the Antilles and underpopulating Africa), created Indian unemployment with the growth of the textile industry in Manchester, shifted the accumulation of capital at the world scale (Amin, 1976), and lowered the cost of reproduction of the workforce of poor countries by promoting the importation of food goods to the detriment of national producers.

[4] However, these leaders do not seem to have incorporated Alfred Sauvy's theory of the optimal demography of power. The power optimum is defined as the point at which decreasing returns lead to marginal productivity below the minimum wage rate that can be imposed on the producer. The politics of power seem to have been more complex. States probably surpassed not only the economic optimum, but also the power optimum (if only for the purpose of building military forces), even when they had the coercive means necessary to lower the wage rate (in turn occasionally reducing productivity).

[5] In postulating savings as a function of *average* per capita income, the models implicitly broke with the whole tradition of studies on accumulation, which had always been linked to inequality of income. They neglected the possibility that population growth could increase accumulation by increasing inequality (or could reduce both), that it might reduce accumulation by lowering profits (Ricardo, 1817), or that accumulation might, in a capitalist regime, require a certain level of unemployment in order to reduce salaries and increase the profits that generate accumulation (Marx, 1850–1859).

must be an increase in national savings (the solution of Feldman [1964] and Mahalanobis [1953]), a recourse to foreign aid (Chenery and Strout, 1966), a reduction of population growth (Enke, 1974), or the simultaneous application of these three methods (Tabah [1868, 1969] showed that demographic inertia rendered simple solutions implausible).

These models were first applied in simulations for national planners to measure the effects of policies to slow population growth. They were used to measure the aid needs created by population growth (which increases the gap between internal savings and investment needs). Everyone who requested aid used the models, and all of the international organizations made calculations inspired by them, which gave the models diplomatic importance and scientific credibility. They are constantly cited by partisans of population regulation and had, at one point, an ideological influence. Their structure was later integrated into more complex models, notably those of the Club of Rome (Meadows *et al.*, 1977). These latter added other Malthusian hypotheses: exhaustion of nonrenewable resources, decreasing returns to agriculture, and degradation of the environment by population.

But these models were quickly criticized for their Malthusian biases. They argued for a negative effect of population growth on per capita income at a time when, as stated earlier, no negative correlation had been demonstrated, and this conclusion was based on a very simplistic hypothesis (zero productivity of labor in case of a shortage of capital). There were attempts to improve the model by borrowing from Robert Solow's (1972) neoclassical model of growth, a macroeconomic production function with substitutable factors. With a substitutability of labor for capital, population growth no longer automatically produces unemployment and does not reduce the marginal productivity of labor to zero in case of a shortage of capital. In a market economy, development can take place via a reduction of capital intensity resulting from the reduction of salaries, international specialization, or development of the informal sector. In the case of public intervention, investments in human capital can play this role.

For seemingly minor reasons that nevertheless retain their importance today, the original version of the neoclassic model maintained its Malthusian biases: the marginal productivity of labor was assumed to be decreasing, the Keynesian savings function had been kept by Robert Solow (for circumstantial pedagogical reasons), and the neoclassic production function was used most often in its "Cobb-Douglas" form in which the sum of elasticities of production in relation to all factors is equal to one, thus implying that the elasticity of production with respect to population is less than one.

It would be hasty to conclude that no macro model has escaped Malthusian biases. Some models introduced production functions with a sum of elasticities greater than one, increasing returns to scale, and Boserup (1970) effects of stimulation of savings, investment, and productivity by demoeconomic tensions.[6] Because both positive and negative effects of population growth are added together, the results of macro models are no longer determined before the parameters are set. An interesting econometric calculation (Thirlwall, 1972) has even shown that it is possible to produce a period of equilibrium by balancing positive and negative effects. This is consistent with the period of apparent insensitivity of the rate of growth of the per capita GDP to population growth.

Despite this progress, macro models have been eclipsed owing to their narrow fit to the issues of the years 1950–1960. The media success of their Malthusian conclusions and their utilization by the Club of Rome has made them scientifically suspect (Cole *et al.*, 1974). The idea of assistance needs has been devalued by descriptions of the perverse effects of donor aid. The strategic role that development theory once assigned to accumulation has ceded its place to analyses of the progress of productivity, the importance of which appeared in the macro models, which themselves emphasized, following the theory of Edward Denison (1967), that the growth of production factors had less weight than the "third factor." Finally, as Peter Tamas Bauer (1976) has pointed out, the national scale of these models privileged collective norms, devaluing the preferences of families, and seemed superficial compared with disaggregate studies and the progress of micro analyses.

II. DISAGGREGATE ANALYSES AND STRUCTURAL STUDIES

Disaggregate studies of the demoeconomic evolution of poor countries[7] have revealed the great variation in economic and demographic parameters within countries, and the range of possible demoeconomic

[6] The Keynesian function is not Malthusian when growth is under way because the increase in revenues leads to a boost in savings. This virtuous circle is observed in Asia.

[7] This expression might be more appropriate than the expression "population and development," which has been accused of exaggerating the specificity of the problems of poor countries, ignoring periods of regression, and having political (unity of the Third World) and normative (the number of definitions of development introduced moral conditions) connotations.

dynamics. These studies divided national units into subunits that exchanged flows (of people, goods, revenue, capital, information, etc.) and carried out relationships (e.g., equalities and inequalities, complementarities, and conflicts) that sometimes had as much influence on the national demoeconomic dynamic as do national averages and aggregates.

1. Vertical Stratification by Income

Vertical stratification by income creates population dynamics that can differ from the expected effects of global variation in GDP. The persistence of high birth rates in Africa has often been attributed to the fact that only a high-income minority has changed its way of life, and that the majority of the population still lives in a setting in which increases in income have no effect or a positive effect on the number of children. However, incontestable inequalities in access to health services (public and private) have made less difference to life expectancy than some would have believed.[8] We have also seen that income distribution, rather than average income, best explains accumulation rates and the capacity to respond to the need for investment created by population growth. Family solidarities have changed with the level of inequality. The initial hostility of political leaders toward antinatalist campaigns has decreased with their growing fear of classes that are considered dangerous.

2. Regional Disparities

Regional disparities within underdeveloped countries are such that averages and national totals are often not informative. This does not necessarily mean that the strategic parameters identified by macro models are meaningless, as certain regional monographs have argued, but it does suggest that only geographic decomposition into relatively homogenous zones can produce numeric estimates that are realistic and verifiable.[9] Once these relatively homogenous regions have been identified, measures of usable land, demographic load capacity, available production factors, elasticity of food supply, growth or reduction in agricultural returns, opportunities for investment and capacities for accumulation, threats to the environment and conditions for sustainable development, the existence of an informal sector, and the proximity of family planning services and their audience, take meaning. The macro demoeconomic models described earlier may also be best tested at the regional level.

Defining homogenous regions makes it possible to identify both complementarities (comparative advantages) and conflicts (competition and fights over sharing profits). The outcome of national demoeconomic problems is primarily a function of these relationships, including internal migration, the geographic distribution of family members by age, flows of food toward populated zones, intrafamilial interregional financial transfers, and movements of capital toward previously disadvantaged zones or high-potential zones. The geographic redistribution of population and of production is an essential element in reducing demoeconomic tensions (in the case of cooperation) or reinforcing these tensions (in the case of interregional conflict). Economic policy as a function of population is thus often a policy of territory management.

3. Decomposition into Sectors

The process of decomposition into sectors (e.g., traditional and modern, with growing returns and with declining returns, rural and urban, formal and informal) was developed as part of the extension of Arthur Lewis' (1954) dual economy model. The dualist model early emphasized that the overpopulation of rural areas (defined by a marginal productivity of zero) made it possible, by internal migration, to develop a nontraditional sector toward labor with non-null productivity. The principal problem was financing, obtained either by internal taxation or from international capital (of which Arthur Lewis was one of the first to underline the abundance to come). Relations between sectors thus made it possible to loosen the constraints imposed both by Ricardian decrease in agricultural productivity and by the lack of capital described by the macro models of the 1950s. The model was enriched by the migratory models that followed, notably Todaro's (1969) model, and by increasingly detailed definitions of nonagricultural sectors as in the three-part models distinguishing urban sectors with high and low intensity of capital (Hugon, 1983).

[8] The Bariloche model even proposed life expectancy as a synthetic indicator of economic development (Herrera, 1977), but medical progress and the diffusion of information have extended life expectancy even in the absence of economic growth (Preston, 1975; Vallin, 1989, 1992).

[9] This observation is valid for both economic parameters (e.g., production factors, saving and investment rates, growth and shrinkage of returns, scarcity of the land factor, elasticity of food supply, vulnerability of the environment, presence of nonrenewable resources, pollution, level and distribution of income) and demographic parameters (fertility rates, forms of morbidity, life expectancy, mortality rates, propensity to migrate or to receive migrants). For each of these variables, regional statistics are too dispersed for their average to explain the internal dynamics of the nation.

4. Urban–Rural Relations

The conflicts and complementarities of urban–rural relations continue to inspire lively debate on their capacity to resolve the demoeconomic problems of poor countries. In abstract theories, this controversy is often expressed through an opposition between studies of flows, which emphasize complementarities, and studies of prices, which emphasize conflicts. In empirical analysis, urbanization and its consequences have been the object of urbo-pessimist descriptions: overpopulation of megalopolises; increased population growth owing to better conditions of sanitation (in contrast to the effects of urbanization in preceding centuries); degradation of the environment; exhaustion of resources in rural areas; rapid depopulation of rural areas unjustified by the ratio of urban to rural productivity; use of political power to obtain terms of exchange unfavorable to rural areas, discouraging them from food production; recourse to imports and food aid, competition with local agriculture; and the (disputed) existence of an urban bias in social policy (Lipton, 1976). Urbo-optimist descriptions, in contrast, have underlined the stimulating effects of the growth of effective demand for food products, the provision of services and of goods that stimulate agriculture, financial transfers, the growth of informal activities, the value added by urban services, etc. Analysis can only move forward if there are efforts to measure flows, prices, income levels and security, productivity, taxation, donor support, instabilities, etc., and, simultaneously, to take into account the capacity of economic fluctuations, interventions by the state, and exterior actions on incomes and internal politics (e.g., structural adjustment programs) to modify complementarities and conflicts.

5. Horizontal Stratification

Differences between families, between cultures, between religions, or between ethnic groups—sometimes called horizontal stratification—lead to differences in demoeconomic behavior. Behaviors can differ in the domains of marriage transfers, long-term investments, household structure, stability of unions, division of labor within the family, modes and rhythms of accumulation, etc. Horizontal stratification produces transfer zones, areas where financial engagements are respected, and collective procedures of accumulation that in turn create specific economic dynamics. But a nuanced comparative understanding of these dynamics and the interactions between them is often blocked by political and ideological resistance to research and to the production of differentiated statistics.

III. MICRO ANALYSES

Micro analyses of the demography of development have replaced development planners' macro investigations with observations of individual and family demoeconomic behavior in poor countries. They are oriented in two complementary but often conflicting directions: the microeconomic interpretation of demographic behavior and the social analysis of demoeconomic strategies.

The microeconomic interpretation of demographic behavior considers behavior in terms of economic rationality. These analyses examine both substantive rationality (utilitarian and materialist) and formal rationality (optimization strategies in accordance with microeconomic theory). In this interpretation, borrowed from Gary Becker (1964), population change in developing countries is a function of the costs and benefits of demographic decisions as assessed by the population and, in particular, is a function of the changes that development induces in these costs and benefits. For example, the number of births depends on parental predictions of the economic costs and returns of the marginal child; this number changes if development increases the costs of children (education, mode of living), lowers the immediate contributions to production (agricultural), or raises expected adult earnings and reduces the investment risk by increasing the life expectancy of infants and children. Analogous calculations are made for age at marriage, marriage choices, the duration of unions, the size of polygamous households, marriage payments, the amount of intergenerational transfers, the roles and status of women, the extent of networks in which gifts and counter-gifts take place, migrations to urban areas or abroad, etc. Each decision is influenced by long-standing microeconomic calculations, by changes produced by development-induced variation in economic parameters, and by the resulting feedback effects on the shape and pace of development.

This microeconomic interpretation is not only a theoretical construct. The framework is designed to be used in an applied context and to show that change has resulted from this type of calculation. Individual actors' resistance to macrolevel decisions to slow the pace of population growth was rational; conversely, couples have sometimes regulated births in countries where the state was still officially hostile to such practices.

The microeconomic framework also seeks to be predictive, arguing that the world is moving toward globalization of economic behavior and that methodological focus on the individual will be better suited to the study of poor countries as development succeeds in imposing economic rationalization of behavior.

Finally, microeconomic interpretation of behavior is normative, aiming to rehabilitate the behavior of individual actors. It is normative in its judgments on the past: unlike macro voluntarism, it claims to respect household preferences (Bauer, 1984). It is normative for the future, to the extent that it instructs planners to trust in families rather than leaders at the macrolevel to understand change. Finally, it is normative in arguing that if authorities let market signals work (leading to a rise in prices when population growth creates scarcity), households will adapt to new opportunities and new scarcities: in demography, as in economics, the market will promote the convergence of the micro optimum and the macro optimum.

Of course, these optima are Pareto optimal and do not take a position on the existing distribution or the demographic implications of this distribution. In addition, liberal conclusions must be nuanced to take into account the behavior of the "lone rider," for instance, if a household wants to increase its progeny in a country that is already overpopulated. Furthermore, the market only functions if there is perfect information with no delay, but experience shows that, for example, environmental degradation can be well under way, and already irreversible, before prices reflect an increase and households take it into account in their decisions. Even national authorities have very long delays in information and decisions (Meadows *et al.*, 1977). Finally it is possible that, contrary to the basic hypotheses of the theory, behaviors are not modified according to the costs and benefits of demoeconomic strategies because motivations are not uniquely economic.

Social analyses of demoeconomic microstrategies show that demographic decisions, and even microeconomic decisions, are shaped by social norms and by economic calculations. Development may expand the role of calculations, but households adapt to new opportunities in light of both the information available to them and pre-existing social norms.

Increased resources and diversified opportunities are sometimes put to traditional usages: modes of consumption are conserved while the volume of consumption grows; family size increases; the accumulation of social capital continues to be more important than the accumulation of economic capital, so social capital is maintained to the detriment of economic capital (ostentatious consumption, confusion of business stocks with family property); transfers increase; the number of wives increases; marriage payments rise; migration takes place in a family framework; and economic investments are made in the context of social investments. Whatever their origin (microeconomic calculation, rational convention, religious or symbolic group membership), social norms are not, at least in the short term, reassessed with each change in economic opportunities and constraints.

Reciprocally, social norms are used for economic ends and give birth to strategies largely unpredicted by outside analysts: the accumulation of economic capital, as much as the accumulation of social capital, relies on social controls. Social controls facilitate the use of tontines, the fosterage of children enrolled in school, and the collective financing of migration. The security provided by the exchange of transfers and counter-transfers makes it possible to take advantage of even fleeting economic opportunities, and strong social sanctions against those who forget their traditional obligations ensure economic "trickle down"[10] from urban residents to peasant families.

The economic application of social norms progressively transforms their content. Systems of gifts and counter-gifts become monetary, family relations are turned into economic relations, transfers are used for exploitative ends (notably of young people by older people), marriages become financial transactions, and economic benefits are extracted from political power (both traditional and new). This "economicization" provokes tensions over the level of transfers and recriminations against family structures.

These changes usher in a period of demographic and economic innovations that simultaneously modify both economic and social relationships. The family is particularly affected by these innovations, which influence the extent of the family, the division of labor, age at marriage, the roles and statuses of men and women, the position of younger family members, intergenerational relationships, attachment to age grades, the evaluation of rights and obligations, the rules of cohabitation, the location of the morning meal, the modes of preparation of family meals, etc. These innovations could all be described both as economic acts effecting a rise in a globalizing economic rationality and as adaptations to change by family structures (specific to different regions and different cultures). The result is variation in the levels of success in development and variation in the modes of successful development, such as the diversity of family adaptations to adjustment programs (Coussy and Vallin, 1996).

[10] The trickle down hypothesis argues that any increase in income in a country, even if it initially affects only the highest incomes, eventually reaches low incomes.

IV. THE INTERNATIONALIZATION OF RELATIONS BETWEEN POPULATION AND DEVELOPMENT

Since the origins of economic demography, it has been recognized that international economic relations affect the interactions between population and development. International migrations usually result from demographic disequilibrium (although there are forced migrations that are nonequilibrating). The importation of grain, according to David Ricardo (1817), was a means of preventing population growth from slowing development. The neoclassic economic school described international exchanges in conformity with comparative advantages as an imperfect substitute for migration. Alfred Sauvy (1966) argued that exports made it possible for populated countries to specialize in sectors with increasing returns. In the Hecksher-Ohlin-Samuelson model, external trade allowed countries that were poor in capital but rich in labor to specialize in sectors that were labor intensive. According to Arthur Lewis, we have seen, foreign investments made the growth of dualist economies possible. Finally, in the first macro demoeconomic models, outside aid takes on the task of meeting the needs created by population growth.

The capacity of external economic relations to reduce the tensions created by rapid population growth is balanced by their tendency to reduce the effect of growth in stimulating internal supply: decrease in food self-sufficiency, discouragement of saving and displacement of local investors, reduction of Boserup effects, abandonment of projects to develop equipment materials[11], etc.

Along with these possible detrimental effects, external economic relations can have negative effects on the international and internal distribution of income. Dependency theorists, including Marxists, conceive of these negative effects as part of a dynamic in which demoeconomic problems come not from population growth but rather from dependence on foreign countries. Dependency can reduce production and can slow the second phase of the demographic transition by maintaining poverty levels. Dependency-based analyses have not incorporated the possibility that demoeconomic tensions can be regulated at the international level, as in the cases of global improvements in health, the international origins of the green revolution, the use of agricultural exchanges to reduce food insecurity related to climactic instability, the jump-starting of certain national economies by migration, elevated growth rates in countries that specialize in exports, and the growth of employment at the global level. However, some of the relationships revealed by dependency analysis have become even more marked, notably dependency on foreign capital, the power of donors over internal policy, international competition that lowers the remuneration of labor, and barriers to immigration imposed by rich countries.

Currently, however, dependency-based analyses are overshadowed by the general acceptance of the new dynamic of the international economy: inequalities of development among under-developed countries, liberal globalization, erosion of borders, and financial globalization.

The inequality of rates of economic growth has displaced the image of a homogenous Third World. Instead, there are contrasts between, on the one hand, the Asian miracle and renewed growth in South America and, on the other hand, the specter of marginalization in sub-Saharan Africa and South Asia and economic regression in poor countries in transition.

Asian growth has changed views on the optimal size of nations. The "Little Dragons" were able to benefit from growing returns by turning to world markets, whereas large countries such as India and China, contrary to expectations, were braked in the short term by their continental expanse. Open small countries have used population density to their advantage; this situation was not unexpected but served as a counterexample to the textbook view of overpopulated islands. It has often been said, notably in the context of the theory of endogenous development, that productivity and demographic investments in health and education played a larger role in development than production factors. Today, however, the debate is being reopened.[12]

Asian countries have entered the second phase of the demographic transition, which has weakened belief in the inevitability of the Malthusian trap and led to a resurgence in the expectation that development will reduce birth rates. This expectation is con-

[11] In contrast, supply is stimulated when foreign relations increase exports (the case of balanced exchange), when they introduce factors complementary to labor (bringing capital to populated countries), when they make high-cost activities competitive (agricultural productivity), or when exterior aid creates an internal demand in countries where there is a lack of effective demand. Agriculture in west Africa, for example, experienced negative effects of food aid but benefited from urban demand financed in part by financial aid (and taxation on exports).

[12] On the one hand, backers of liberal reform in India cite works blaming Indian delay on the weakness of growth in global productivity (Ahluyallah, 1991; Bhagwati, 1993–1995). On the other hand, using the works of Alwyn Young (1995), Paul Krugman (1994) attributes the success of southeast Asia to the abundance of production factors, which announced the end of the Asian miracle, rather than growth in productivity.

firmed by the example of Africa, where the relatively rich countries of southern Africa have been the first to experience fertility decline. But other examples from Latin America (Cosio-Zavala, 1994) and Africa show that fertility decline can also be a result of poverty (the Malthusianism of poverty).

The contrast between rapid economic growth and declining fertility in Asia and countries (especially in Africa) with low economic growth where the demographic transition has not yet advanced significantly has led to a resurgence of 1950s-era hypotheses on the negative effects of population growth. The world econometric model of the "Asian miracle" (Banque mondiale, 1993; World Bank, 1993a,b) uses a Cobb-Douglas function that, for the reasons discussed above, shows an inverse relationship between population growth rates and growth of per capita income. It is likely not a coincidence that an earlier direct study of the correlation between the two rates (Blanchet, 1996) now seems to indicate a negative correlation rather than independence.

Liberal globalization has created new opportunities for development, first recognized by Arthur Lewis, through exports resting on foreign investments. It has also reinforced the preference for microlevel analyses, which are better suited for analysis of liberalized economies and more in line with current ideology. But globalization has liberalized international movements of capital more than movements of people, and there is a tendency to try to regulate population problems by relocation rather than immigration in developed countries, which is not the original classical optimum. This tendency has also reactivated the idea of development aid as a means of fighting migration, despite limited results obtained by this method.

The erosion of borders has accelerated the international diffusion of modes of consumption, health-related knowledge, changes in family structures, intergenerational relations, and ways of regulating fertility—so much so, that this diffusion sometimes precedes diffusion of income levels. It has increased transborder exchanges, both official and unofficial; family relations and ethnic group memberships have played an important role in these exchanges and have sometimes been reinforced by them in return. But the erosion has also reached more distant borders, allowing the globalization of family activities: migratory networks, financial networks, commercial activities, and the mastery of networks of production and exchange. Liberal globalization has often been cited as a reason to move beyond family-based capitalism in Europe. However, the economic efficacy of families, notably of Chinese families, has popularized the reliance on families in national strategies of development and raised hopes in countries benefiting from a diaspora, as evidenced by the Indian policy toward nonresident Indians. The role of networks based on identity, religious, or ethnic affiliations, which sometimes speak in terms of rupture with the global economy, is more ambiguous.

Financial globalization has modified the conception of capital movements driven by the disparity of population structures. The period 1950–1960 was dominated by models that established a relationship between the population growth of poor countries and their needs for external financing. These models were useful for projections and for justifying international aid, but the models proved to be less apt for describing and explaining the evolution of flows of capital over time. It quickly became clear that the movements of capital to poor countries were less a response to the needs of these countries than a component of the financial logic of the formation of savings and the allocation of capital. In the 1970s, the abundance of world liquidities increased the flow of capital toward poor countries above their capacity of productive absorption (notably in Africa). In the 1980s, the debt crisis reduced flows toward the south, sometimes to the extent that transfer balances became negative. Currently, capital movements are being renewed as an aspect of financial globalization: as a means of matching excess savings in developed countries, where there are numerous workers that have savings, and as a means of savings deficits in emerging countries. This process can be seen as an international extension of life-cycle trends in savings, as, for example, pension funds in Anglo-Saxon countries have a significant effect on loans to emerging countries. The reason for these loans is again demographic, but with a financial justification substituted for a justification based on needs. Capital flows are directed toward rapidly developing Asian countries, not toward African countries headed toward marginalization. And the volatility of these flows has had negative effects in Mexico and, more recently, the countries of the Association of Southeast Asian Nations.

CONCLUSION

The relationships between population and development in poor countries are now studied in the context of economic dynamics that can range from rapid growth to crisis or adjustment. International inequalities in growth have created new relationships between population and economics. These relations can only be studied, even at the national level, by dealing with subunits in which complementarities and conflicts have decisive effects on demoeconomic dynamics. In a period of liberalization, it is logical to orient analysis toward microanalyses (in addition to the simultaneous study of micro- and macrolevels in calculable models of general equilibrium). But microstrategies cannot be reduced to traditional microeconomics; they must be integrated with microsocial and familial relationships that lead actors to make innovations, including international innovations, that are simultaneously both economic and demographic.

Acknowledgment

This chapter was translated from the French version by Sarah R. Hayford.

References

Ahluwalia Isher J., 1991. *Productivity and growth in Indian manufacturing.* Oxford, Oxford University Press, 242 p.

Amin Samir, 1976. *L'accumulation à l'échelle mondiale.* Paris, Union générale d'édition, 2 vol., 502 p.

Bairoch Paul, 1967. *Diagnostic de l'évolution économique du tiers monde, 1900–1968.* Paris, Gauthier Villars, 261 p.

Bauer Peter Tamas, 1976. *Dissent on development.* London, Weidenfeld and Nicholson, 550 p.

Becker Gar, 1964. *Human capital: a theoretical and empirical analysis, with special reference to education.* New York, Columbia University Press, 187 p.

Bhagwati Jagdish Natwarlal, 1993. *Indian transition: freeing the economy.* Oxford, Clarendon Press, 108 p.

Blanchet Didier, 1996. L'inversion des effets de la croissance démographique sur la croissance économique. Interprétations générales et spécificité africaine, *in*: Jean Coussy and Jacques Vallin (eds.), *Crise et population en Afrique: crises économiques, politiques d'ajustement et dynamiques démographiques,* p. 51–69. Paris, CEPED, 580 p.

Boserup Esther and Métadier Jacques, 1970. *Évolution agraire et pression démographique.* Paris, Flammarion, 221 p.

Chenery Hollis and Strout Alan, 1966. Foreign assistance and economic development. *The American Economic Review,* vol. 56(4), p. 679–733.

Chesnais Jean-Claude and Sauvy Alfred, 1973. Progrès économique et accroissement de la population, *Population,* vol. 28(4–5), p. 843–858.

Coale Ansley and Hoover Edgar, 1958. *Population growth and economic development in low income countries: a case study on India's prospects.* Princeton, Princeton University Press, 389 p.

Cole Hugh S., Freeman Christopher, Jahoda Marie, and Pavitt Keith, 1974. *L'Anti-Malthus: une critique de "Halte à la croissance."* Paris, Le Seuil, 244 p.

Cosio-Zavala Maria-Eugenia, 1994. *Changements de la fécondité au Mexique et politiques de fécondité.* Paris, L'Harmattan, 256 p.

Coussy Jean and Vallin Jacques (eds.), 1996. *Crise et population en Afrique: crises économiques, politiques d'ajustement et dynamiques démographiques.* Paris, CEPED, 580 p.

Denison Edward F., 1967. *Why growth rates differ?* Washington, Brookings Institute, 494 p.

Domar Evsey D., 1957. *Essays in the theory of economic growth.* New York, Oxford University Press, 272 p.

Enke Stephen, 1974. Reducing fertility to accelerate development, *Economic Journal,* vol. 84(334), p. 349–366.

Feldman G. A., 1964. On the theory of growth rates of national income, *in*: Nicolas Spulber (ed.), *Foundations of Soviet strategy for economic growth: selected Soviet essays, 1924–1963.* Bloomington, Indiana University Press, 530 p.

Harrod Roy, 1939. An essay in dynamic theory, *Economic Journal,* vol. 49(193), p. 14–33.

Herrera Amilcar O., 1977. *Un monde pour tous: le modèle mondial latino-américain.* Paris, PUF, 271 p.

Hugon Philippe, 1983. Pression démographique, secteur informel et choix technologiques dans les pays du Tiers Monde, *Revue Tiers Monde,* vol. 24(94), p. 399–408.

Klatzman Joseph, 1983. *Nourrir dix milliards d'hommes?* Paris, PUF, 296 p.

Krugman Paul, 1994. The myth of Asia's miracle, *Foreign Affairs,* vol. 73(6), p. 62–78.

Lewis Arthur W., 1954. Economic development with unlimited supplies of labour, *Manchester School of Economic and Social Studies,* vol. 22(3), p. 139–191.

Lipton Michael, 1976. *Why poor people stay poor: urban bias in world development.* Cambridge, Harvard University Press, 467 p.

Mahalanobis Prasanta Chandra, 1953. Some observations on the process of growth on national income, *Sankhia,* vol. 12(4), p. 307–312.

Marx Karl, 1850. *Grundrisse der Kritik der politischen Ökonomie, Rohentwurf, 1857–1858 [und] Anhang, 1850–1859.* Berlin, Dietz, 2 vol., 1102 p.

Meadows Dennis L., Behrens William W. III, Meadows Donella, Nail Roger F., Randers Jorgen and Zahn Erich K. O., 1977. *Dynamique de la croissance dans un monde fini.* Paris, Economica, 610 p.

Meslé France and Vallin Jacques, 1996. *Mortality in the world: trends and prospects.* Paris, CEPED, 24 p.

Meslé France and Vallin Jacques, 1997. La mortalité dans le monde: tendances et perspectives, *in*: Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: enjeux et problèmes,* p. 461–479. Paris, INED-PUF, 630 p.

Preston Samuel H., 1975. The changing relation between mortality and economic development, *Population Studies,* vol. 29(2), p. 231–247.

Quesnay François, 1758. *Tableau économique. in*: INED (ed), *Francois Quesnay et la physiocratie, II: textes annotés,* p. 675–682. Paris, INED, 1005 p.

Ricardo David, 1817. *On the principles of political economy and taxation.* London, J. Murray, 589 p.

Sachs Ignacy, 1971. *La découverte du Tiers Monde.* Paris, Flammarion, 276 p.

Sauvy Alfred, 1966. *Théorie générale de la population.* 3rd ed. Paris, PUF, 2 vol., 776 p.

Sauvy Alfred and Hecht Jacqueline (eds.), 1958. *Quesnay et la physiocratie.* Paris, INED, 2 vol., 1008 p.

Solow Robert M., 1972. *Théorie de la croissance économique.* Paris, Dunod, 155 p.

Tabah Léon, 1968. Démographie et aide au Tiers monde, I: les modèles. *Population,* vol. 23(2), p. 509–534.

Tabah Léon, 1969. Démographie et aide au Tiers monde, II: calcul du déficit au cours de la période 1960–1980. *Population,* vol. 24(3), p. 501–540.

Thirlwall Anthony P., 1972. A cross sectional study of population growth and the growth of per capita income in a production function, *The Manchester School of Economic and Social Studies,* vol. 40(4), p. 339–356.

TODARO Michael P., 1969. A model of labor migration and urban unemployment in less developed countries. *American Economic Review,* vol. 59(1), p. 138–148.

VALLIN Jacques, 1989. Théorie(s) de la baisse de la mortalité et situation africaine, *in*: Gilles Pison, Mpembele Sala Diakanda, and Étienne van de Walle (eds.), *Mortalité et Société en Afrique au sud du Sahara,* p. 399–431. Paris, INED, PUF, 446 p.

VALLIN Jacques, 1992. Theorie(s) of the declining of mortality and African situation, *in*: Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in Sub-Saharan Africa,* p. 405–436. Oxford, Clarendon, 450 p.

WORLD BANK, 1993a. *The East Asian miracle: economic growth and public policy.* Oxford, Oxford University Press, 34 p.

WORLD BANK, 1993b. *Le miracle de l'Asie de l'Est, croissance économique et politiques publiques.* Washington, World Bank, 34 p.

YOUNG Alwyn, 1995. The tyranny of numbers: confronting the statistical realities of the East Asian growth experience, *Quarterly Journal of Economics*, vol. 110(3), p. 641–680.

CHAPTER

87

The Relationships between Population and Environment

FRANÇOISE BARTIAUX AND JEAN-PASCAL VAN YPERSELE

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Institut d'astronomie et de géophysique Georges Lemaître, Université catholique de Louvain, Louvain-la-Neuve, Belgium

INTRODUCTION

The relationships between population and environment are complex and numerous, and this area of research has been little frequented by demographers. In consequence, our contribution has a double objective: to take account of this complexity and diversity, and to suggest some lines of reflection in an attempt to understand why demographers have been so little interested in this branch of inquiry. For this reason we shall attempt to throw some light on the paradigms that underlie the studies carried out in demography and the natural sciences of the environment. By proceeding in this way, we hope to establish some guidelines for demographers who may wish to investigate the relationships between population and environment. For although some choices have been made (often implicitly), others remain to be determined and ambiguities must be removed, above all those concerning the relationship of man to nature. This problem appears with the formulation of definitions of environment and population that marks our entry into the debate.

The field of population environment is very vast and cannot be completely covered within the confines of this chapter. We would have liked to limit ourselves to the works published in the journals most frequently consulted by demographers, but it is not always apparent whether or not to label authors as demographers, and in any case, the problem would remain for those working in collaboration with scientists from other disciplines, especially because of the multidisciplinary character of the subject and the high specificity of demography.

What Definitions of Environment?

For Yvette Veyret and Pierre Pech (1993:2), it was in 1942 that the term *environnement* appeared for the first time in French geographical literature as a synonym for *milieu géographique*,[1] following the method of the Anglo-Saxon authors.

Nowadays, various surveys among the general public have shown that the word "environment" has various meanings, depending on the social coordinates of the respondents[2] (Guérin-Pace and Collomb,

[1] According to France Guérin-Pace and Philippe Collomb (1997:1), who trace the history of the word *environnement* and its presence in French dictionaries, it was in 1964 that this word appeared in the Robert dictionary for the first time, and was defined there "as the entire natural conditions (physical, chemical and biological) and cultural (sociological) in which living organisms develop."

[2] The *Populations, Espaces de vie, Environnements* (*Populations, Living Spaces, Environments*) survey carried out in 1992 by INED with a sample of 5000 persons representative of the French population, showed that "the less educated populations describe their environment in terms of its immediate proximity, the more favoured social classes speak of the quality of life and natural equilibrium, urban populations make reference to architecture, shops and noise, and country people invoke their rural surroundings, fresh air and nature on their very doorstep. For the oldest populations it is the social environment and quality of human relationships that predominates, and for the youngest the preservation of the natural heritage and accessibility of unspoiled nature, the bearer of freedom and doubtless of hope" (Guérin-Pace and Collomb, 1997:12).

1997:12) and the activities relating to the environment that the person concerned may be carrying out[3] (Dalla Valle *et al*., 2001:165).

According to the survey conducted by France Guérin-Pace and Philippe Collomb (1997:11), "problems allied to pollution are seldom mentioned in the responses." Yet it is this very aspect that has been favored by (the few) demographers who have investigated the question of links between environment and population, as will soon become apparent. We therefore refer here to the definition adopted by Pascal Baud *et al*. (1995:226), for whom environment is now the environment of living beings, which is both fragile and precious and which must be preserved. But the latter normative part of the definition (the environment must be preserved) already contains an action program that demographers are not inclined to follow and constitutes the ambiguity of the connections to nature and action that we shall discuss later.

And What Definitions of Population?

To tell the truth, the definition of population is hardly clearer than that of environment. In the words of Pierre Lannoy and Lionel Panafit (undated), it was in the 18th century that population was born as a concept that enabled the idea of a new link between individuals and the first structures of the nation states that (to employ the expression of Michel Foucault, 1976) controlled individuals through their "biopower." Demography was thus born, as were other sciences pursuing the same aim of public control, such as criminology and epidemiology.

Population is therefore a way of thinking that has become so natural for demographers that we sometimes forget its abstract character. The same surely applies to environment; we have already noted its varied significance among the general public.

The Field of Research within Population-Environment

As stressed by Nathan Keyfitz (1996, p. 335), demographers have shown little interest in the emergence of this new field, which for want of a better term we shall call *population and environment*, and have contributed little to its development.

This opinion is not shared by Wolfgang Lutz, Alexia Prskawetz, and Warren Sanderson (2002b:1) for whom "this literature [hundreds of scientific articles and reports] forms the groundwork for a new interdisciplinary field of studies that we call P-E [population environment] analysis". This field is constituted by a relatively simple study question: "What are the effects of changes in the human population on the natural environment (P-E)? This question has a complement [. . .]: What are the effects of changes in the natural environment on the human population (E-P)?" (Lutz *et al*., 2002b:225). The only recommended approach is therefore dynamic, and any study in the field should, in their opinion, include some demographic forecasts of the population under review (Lutz *et al*., 2002b:242).

According to these authors, the studies in this field are either positive, "the discussion of the way things are or will be, under certain circumstances"; or normative, "how things should be" (Lutz *et al*., 2002b:11).

The concept of the ecological footprint, introduced by Mathis Wackernagel and William Rees in 1996, aims at linking a population and environment in a synchronic and nondynamic way. For a given population living at a certain level, it involves calculating the equivalent surface for producing the resources and absorbing the waste required for this number of people at this standard of living. This concept, nevertheless, does not appear to have greatly interested demographers.

Since the United Nations Conference on Environment and Development held in Rio in 1922, environmental questions are more frequently treated in conjunction with economic and social aspects by integrating them within the framework of sustainable development.[4] To avoid overextending the field of our review, we shall limit our study to the relations between population and environment.

[3] According to an enquiry carried out in Belgium, where 6800 persons over the age of 16 years were questioned in 1998 within the seventh wave of the Panel Study on Belgian Households. The sample was representative at first of the population living in Belgium, but the more disadvantaged households and persons were more difficult to follow on a year-to-year basis, and the representivity tends, as a consequence, to suffer.

[4] The definition given to this concept in the Brundtland report (WCED, 1987; CNUED, 1988) is worth quoting in its entirety:

> Sustainable development is a development that meets the needs of the present without compromising the ability of future generation to meet their own needs. It contains within it two key concepts:
>
> The concept of "needs," in particular the essential needs of the world's poor, to which overriding priority should be given, and
>
> The idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs.
>
> Even in the narrowest sense of the term, sustainable development presupposes a regard for social equity between the generations, a regard that must extend, quite logically, within a same generation.

I. A REVIEW OF THE WORKS ON POPULATION AND ENVIRONMENT

In this review of the literature, we shall try to limit ourselves to works by (or related to) demographers that concern the interactions between population and environment, despite there having been fewer of them engaged in this topic than biologists, geographers, agronomists, climatologists, and even anthropologists and public health doctors.

As Aïcha Ouharon (1997) stresses, with some exaggeration,

> "the study of the population-environment relationship is marked by two views of the world derived from two disciplines: Economics and Biology. Dominant economic thinking presumes that all individuals are endowed with creativity and perfect reasoning. Confronted by choices under the constraints imposed by the environment, they are presumed to take the correct decision—they follow coherent objectives and employ adequate means for attaining them. Progress enables them to be masters of their destiny and to shape the environment as it pleases them. The other view, offered by biology, recalls that men live in close interdependence with their natural surroundings. Man does not exist for himself, but participates in the equilibrium of a whole system. His survival, as with that of other animal and vegetable species, depends on an equilibrium between his actions and the responses of the natural surroundings. Having succeeded in removing the obstacles imposed by the environment on its reproduction, the human species endangers the survival of other living species, and finally perhaps its own."

Is it because demography feels itself torn between these two paradigms—not easily reconcilable and divergent in the manner in which the (not perfectly rational) individual is represented—that demographic studies on population and environment are not numerous? We shall return to this discussion in the following section, after reviewing the demographic studies of population and environment by successively approaching the population doctrines, analytical methods, case studies, and quantitative global studies.

1. Population Doctrines

A non-negligible part of demographic work on the subject lies within the scope of population doctrines synthesized by Éric Vilquin in this treatise (see Chapter 97). We shall not therefore reopen these researches here, except to stress the proximity of "population and resource" problems and those concerning "population and environment," even though the second is (or could be) much wider.

The vast discussion opened by Nathan Keyfitz (1996) is in line with the history of population doctrines and concerns the past 50 years of demographic, development, and environmental growth. The subject is very wide, and the Keyfitz also includes some reflections on racism and, especially, on the rights of future generations.

Let us stress that it is not always easy to apportion the studies and classify them among the population doctrines or among the global quantitative studies. Many of them are simultaneously in both groups. John Bongaarts (1996), for example, in a quantitative work containing details of the demographic pressure on the supply and consumption of foodstuffs in developing countries, concluded that demand should also be studied, and he estimated that the best way of reducing the demand would be to reduce population growth in the Third World.

2. Syntheses

An issue of *Population Report* presents the state of research and offers numerous figures on the relation between population and environment (Green, 1992).

Other syntheses have been compiled on this topic for countries from the south by Évelyne Thiltgès and Dominique Tabutin (1992); Paloma Agrasot, Dominique Tabutin, and Évelyne Thiltgès (1994); and Dominique Tabutin (1995). According to the latter: "Knowledge about the relations between population and environment is still incomplete. There is a terrible lack of reliable data (especially on the environment) and detailed research integrating both phenomena. Work is based solely on the fragile and uncertain." (Tabutin, 1995:194–195).

Francis Gendreau, Patrick Gubry, and Jacques Véron (1996) have also devoted work to the relations between population and environment in developing countries by approaching the subject from four main angles: problems; production systems and environment; health and environment; and town, living space, and environment.

A brief synthesis of population and environmental problems is to be found in the article by Aïcha Ouharon (1997). Rajendra Pachauri and Lubina Qureshi (1997) report on a conference held on the theme of "Population, Environment and Development." The United Nations Fund for Population Activities (UNFPA) centered its 2001 annual report on the relationships between population and environmental changes (UNFPA, 2001). Mention should also be made of the *Atlas of Population and Environment* published by the American Association for the Advancement of Science (Harrison and Pearce, 2000), which contains numerous statistics on population and environment.

In preparation for the World Summit on Sustainable Development in Johannesburg (2002), the United

Nations Environment Program (UNEP) published the *Global Environment Outlook-3*, or *GEO-3*, in order to take stock of the recent development in the environment, its state, and prospects from now until 2030 (UNEP, 2002). The secretary general of the United Nations, moreover, ordered a series of reports on the five key areas: water, health, energy, agriculture, and biodiversity (WSSD, 2002a–e). Except where otherwise indicated, the data found in the Appendix were taken from the *GEO-3* report.

3. Analytical Methods

As regards methodology, we would especially mention the already cited work by Wolfgang Lutz *et al*. (2002a) on analytical methods in population and environment. The aim of the work is to "consolidate" the field of research from a methodological point of view while presenting a "multiplicity of methods" (Lutz *et al*., 2002b:2) ranging "from highly quantitative dynamic models to decomposition methods and various qualitative approaches" (Lutz *et al*., 2002b:1). The simulation studies presented are derived from the PEDA (*population, environment, development and agriculture*) model applied to various southern countries (see the case studies under the next section), whereas the decomposition methods are based on the equation $I = PAT$, which will be examined in a following section.

4. Case Studies

A number of studies on the interactions between population and environment are presented in the form of case studies. We shall cite some of these, but we acknowledge that our selection is both partial and unilateral.

Mary Tiffen *et al*. (1994), has studied the effect of population density on soil erosion in Kenya and concludes, contrary to widely held opinion, that high population density was helping to protect soils.

Erza Markos (1997) has described the demographic responses to the data on the degrading of the environment and insecurity of the food supply in the northern regions of Ethiopia subject to drought. These data come from a survey involving 2000 married women, and from in-depth interviews with local officials, old people, and development agents. He has also provided a bibliography on the demographic effects of famines.

Véronique Petit (1997) has also combined quantitative and qualitative data in her research into the reasoning behind the migrations of the Dogons of Sangha (Mali), as well as on the consequences for the sedentary populations, in terms especially of the relations between age categories and between men and women.

Wolgang Lutz (1994) has studied the interactions among population, development, and environment in Mauritius. He has mainly used simulation methods, on the basis of the PEDA model. Together with various collaborators, he has also carried out simulation studies that integrate population phenomena (sometimes including the AIDS epidemic), as well as food and environmental safety enquiries for other Third World regions—Cape Verde (Wils, 1996), Yucatan (Lutz *et al*., 2000), Botswana (Sanderson *et al*., 2001), Namibia (Sanderson, Fuller *et al*., 2001), and Mozambique (Wils *et al*., 2001).

Various case studies relating to the Maghreb have been carried out by Slimane Zamoun *et al*. (1995); others, by Francis Gendreau and Claude Meillassoux (1991).

Overall, it appears that the majority[5] of case studies in this area (if not all) are related to countries from the south—a characteristic that seems significant to us and to which we shall return later.

5. Global Quantitative Studies

A review of the works published from 1990–1997 by the *Population and Development Review* indicates that the few articles that come within our sphere of interest are quantitative global studies focusing on two issues: the capacity of the Earth to feed its inhabitants (Smil, 1991; Ehrlich *et al*., 1993; Dyson, 1994; Heilig, 1994; Smil, 1994; Bongaarts, 1996; Naylor *et al*., 1997) and climatic warming (Smil, 1990; Bongaarts, 1992; MacKellar *et al*., 1995; and, on energy: Gilland, 1995).

The first theme is linked to what biologists and certain geographers call *carrying capacity*. Kingsley Davis (1990) has clearly shown that this concept could not be introduced as such in the study of human populations, and Joel Cohen (1995) has also queried the pertinence of applying the concept of carrying capacity used in ecology. Analyzing 26 definitions of human carrying capacity, he observes that the concept is far from having a universal definition and that, despite the veneer of quantification, a good number of published estimates of human carrying capacity should be viewed as political instruments destined to influence public opinion rather than as unprejudiced analyses.

Most of the works published in the *Population and Development Review* are aimed at a more precise

[5] The interdisciplinary study: *The Earth as transformed by human action. Global and regional changes in the biosphere over the past 300 years* (Turner *et al*., 1990) nevertheless includes chapters on the Caucasus, the Russian plain, the great plains of the United States, Sweden, the Hudson basin in the United States, and Switzerland.

measurement of the relationship concerned, and sometimes they introduce concepts and variables that are perhaps unexpected but nevertheless important in the matter. Thus, Gerhard Heilig (1994) denounces the oversimplified underlying model (demographic growth implies an increased demand for food and, consequently, an intensification in agriculture and/or its extension, thus entailing deforestation and transformations of the soil properties) and shows that many changes in the type of soil covering can be attributed to factors other than demographic pressure.

This theme concerning the capacity of the Earth to feed its inhabitants also forms the basis of other studies, and we shall cite three of these, two of which are linked to the concept of density (maximum or not).

In a publication of the UNFPA, Norman Myers defines maximum density as "the number of inhabitants that the planet can accommodate without irreversibly reducing its capacity for accommodating others in the future" (FNUAP, 1991:71; UNFPA, 1991). This definition is usable not only on a planetary scale but also at national level. In the latter case, the UNFPA has remarked that account must be taken of certain reservations that may be made.

One of the areas in which this concept of maximum density has been applied is that of food production. The UNFPA (1991:72) cites the study by the *World Hunger Program* (Chen *et al*., 1990), according to which the planetary ecosystem could, in the present state of agricultural techniques and if the share of foodstuffs were equalized, accommodate 5500 million individuals under good conditions, but only if they were satisfied with a vegetarian diet. If they obtained 15% of their calories from animal products, as is generally the case in South America, the tolerable effective total would fall to 3700 million. If they derived 25% of their calories from animal products (as with the majority of inhabitants in North America), the Earth could only accommodate 2800 million human beings.

For foodstuffs, Hervé Le Bras (1994) concludes that food production is continuing to grow at a slightly faster rate than are populations, but that the abundance and global increase conceal formidable local inequalities and serious ecological problems that are also localized: "soils are deteriorating in many places, some fragile ecosystems are threatened, and poverty is increasing on the surface of the globe. These questions have simply nothing to do with overpopulation. They come from political and social disorders, and unequal distributions of the population at numerous parts of the planet."

Joel Cohen (1995) commented that the capacity of the Earth to feed its inhabitants is determined partly by processes that the natural sciences as well as the human sciences do not satisfactorily explain, and partly by the choices that we and our descendants still have to make. To illustrate this point, he has studied the problem of water. As with other natural constraints, the limitation associated with water interacts strongly with human choices and time limits. Cohen thinks, however, that estimates of human carrying capacity may, despite their limits, serve as useful indicators of the relations between humans and the Earth. He finally considers the suggestions for future action and concludes by offering four recommendations: developing institutions that consider the goals of efficiency and equality in a balanced way; improving the regard for social well-being, material flows and the consequences of each action; integrating our knowledge of populations, economies, environments and cultures; and creating a better understanding of mutual aid, including a quantitative documentation of the benefits derived by the rich from helping the poor live better lives.

With reference to climatic warming and its links with demographic growth, a critical synthesis of some studies carried out on the subject is to be found in our paper presented at the 22nd International Union of the Scientific Study of Population (IUSSP) Conference (Bartiaux and van Ypersele, 1993) and in the book by Brian O'Neill *et al*. (2001), *Population and Climate Change*. The studies reviewed (see Harrison, 1991; Bongaarts, 1992; Lutz, 1992; MacKellar *et al*., 1995) are all based on the celebrated equation ($I = PAT$) of Paul Ehrlich and John Holdren (1971). According to this equation, the influence (I) of the population (P) on the environment (or, in this case, on c1imatic warming) varies according to consumption (A for "affluence") and technology (T). The units of measurement vary according to the problem under consideration.

A presentation of this equation and its variants and a criticism of its bases are to be found in our own chapter, as well as several UNFPA studies that employ the incorrect results produced by an inadequate application of it. Although, indeed, the influence (I) of the population (P) on climatic warming depends not only on consumption (A) and technology (T) but also on the level of aggregation and therefore on the spatial scale employed (cf. the results of Lutz, 1992). This influence also varies according to whether the choice is to study consumption levels per inhabitant (choice more unfavorable for developed countries[6]) or growth,

[6] In particular, we have calculated that between 1950 and 1990, the increase in carbon dioxide emissions per inhabitant in the developed countries has played a more important part in the increase of carbon dioxide emissions worldwide than has the demographic growth of developing countries (Bartiaux and van Ypersele, 1993).

particularly demographic growth (more unfavorable for developing countries); whether it is postindustrial revolution (more unfavorable for developed countries) or post-decolonization (more unfavorable for developing countries; and whether or not to include methane data [rare] (more unfavorable for developing countries). And then, as Landis MacKellar *et al*. (1995) have added, it must be considered whether the analytical unit employed for the population is the individual (more unfavorable for developing countries) or the household (more unfavorable for developed countries, where the size of households is rapidly decreasing[7]), etc. Who still believes that quantified studies are neutral and objective?

Let us finally note the following remark by Hervé Le Bras (1994): "the classification of developed countries according to their participation in the greenhouse effect per inhabitant is in inverse order to the densities." Le Bras attributes this to a negligent use of working and living spaces in empty or new countries, where nature appears to be less restrictive, less congested, and more bountiful. Comparing the effect of the principle causes of wastage and that of demographic growth, he concludes that demographic growth does not play the leading role in the climatic crisis that people wish to attribute to it.

II. WHY ARE DEMOGRAPHERS SO LITTLE INTERESTED IN ENVIRONMENTAL QUESTIONS?

Asking this question also involves enquiring about the specificity of demography and about its paradigms, while comparing them with the paradigms of the sciences for which environment is the object of study.

For these disciplines, we shall examine successively the spatial and temporal scales, the envisioned links between nature and culture, the causal or systematic perspective, the place for research action and value judgments, and finally the customary interdisciplinarity. These aspects to be taken into consideration in comparing demography and the environmental sciences are partly inspired by José Prades (1996).

1. Spatial Scales

Spatial scales in the natural sciences of the environment are varied and range from the local to the global. They can be presented as concentric circles starting with the environmental elements with which human beings are in direct contact.

a. Microenvironment (Radius of about 100 m)

From his immediate environment, the human being takes the elements necessary for his metabolism, and these basic needs have been estimated in the following way, bearing in mind that these are average figures and therefore variable, modifiable and modified according to the societies: oxygen (about 0.5 kg/day), water (2 liters/day) and food (1 kg/day). In addition, every person needs to be surrounded for living by a microclimate defined by a certain range of temperature and humidity, as well as by acceptable levels of radiation (e.g., light, ultraviolet, radioactivity, microwaves). Toxic elements (chemical compounds and microorganisms) can only be present in very limited amounts. Noise must not exceed a certain level. The body must also be able to eliminate the gaseous, liquid, and solid waste from its metabolism without his immediate environment suffering. To guard against diseases, human beings have adopted many chemical elements from the vegetable, animal, and mineral world. The quality of this *microenvironment* must be adequate throughout life, but even more during pregnancy as well as the first and last years of life, when living beings are more fragile. The standards characterizing this quality are defined among others by the World Health Organization, on the basis of epidemiological studies (WHO, 2000).

The habitat and workplace are important components of the microenvironment. Many writers have studied the link between the health of the occupants and the temperature and humidity prevailing inside buildings (see Boardman, 1991), quality of the air (Leslie and Lunau, 1992), and electromagnetic radiation (Hester, 1992). Tobacco smoke is probably the main domestic pollutant (Leuenberger, 1993). In developing countries, smoke from domestic homes is also an important pollutant (WSSD, 2002b). Reference to this subject can also be seen in Chapter 50.

b. Local Environment (Radius of about 10 km)

In a society that is little industrialized, it is the local environment (within a radius of some 10 km around the dwelling place) that provides food, water, and energy (firewood), whereas in rich countries, environment is less directly involved and used more as a source of recreational areas.

In towns, road transport and heating cause pollution, which in certain cases has immediate and serious effects on human health. In December 1952, for example, 4000 victims were recorded in London after

[7] Decrease in the size of households reduces the economies of scale that are possible in larger households.

a stagnant polluting fog (smog) over the town (McNeill, 2000). Apart from industrial catastrophes, often deadly (e.g., 3000 deaths in Bhopal, India, in 1984), chronic atmospheric pollution through certain components, sulfur in particular, is responsible for many attacks on human health. Reference to this subject can also be seen in Chapter 50.

This list should also be completed by evoking the pollution and degradation of soils, deforestation, etc.

c. Regional Environment (Radius of about 10 to 1000 km)

In developed countries, a large part of the resources that we consume—food, water, energy, and material goods—come from far away. The book *Factor Four* (von Weiszäcker *et al*., 1997a,b) cites a study by the Wuppertal Institute for Climate, Environment and Energy on the peregrinations of pots of strawberry yoghurt consumed in Germany. According to this study, each yogurt, its ingredients, and its glass pot represents an accumulated journey totaling 3500 km, to which must be added 4500 km for the materials of the suppliers. Our solid waste is often transported on this regional scale—not near enough to be seen daily, but not too far on account of high transport costs.

Many environmental risks are present on a regional scale: air pollution, acid rain, photochemical pollution (ozone, smog), pollution of ground water, rivers, etc.

d. Global Environment (Encompassing the Whole Earth)

Globalization of the economy, with the effect that goods are exchanged between widely separated regions (exotic foods out of season; minerals, oil, gas, or uranium transported thousands of kilometers), is accompanied by a globalization of environmental problems.

One of the first global environmental problems was that of the thinning down of the stratospheric ozone layer, consequent upon the emissions of chlorofluorocarbons (CFCs) used in refrigeration and aerosol canisters. In fact, it was only in 1985—when British researchers published for the first time some observations that the ozone layer had lost 40% of its depth over the Antarctic between 1977 and 1984—that human beings became aware that in acting locally (by using a spray in their bathroom, for example) every Earth-dweller could affect the global environment. This awareness led to the Vienna Convention and the Montreal Protocol, which have resulted in a drastic reduction in CFC emissions. The stratospheric ozone layer should in principle regain its normal state toward the middle of the 21st century (UNEP, 2002).

The second global problem is that of the climatic changes resulting from the gas emissions, principally carbon dioxide,[8] resulting in the greenhouse effect. The most optimistic scenario for the emission of gas having a greenhouse effect[9] is world population rising to a maximum of 8700 million in the middle of the century and then declining; economy oriented toward economic, social, and environmental viability (including greater equity), evolving rapidly toward services and a reduction in the intensity of the consumption of energy and materials, thanks to important technological progress. The Special Report on Emissions Scenarios (SRES) B1 scenario of Nakićenović and Swart (2000) associated with the least sensitive of the climatic models[10]—the Intergovernmental Panel on Climate Change (IPCC), a group of intergovernmental experts on climatic issues—assembled on the initiative of the United Nations, projects an average air temperature of the global surface for the year 2100 that is 1.4°C higher than in 1900, as well as a rise in the average sea level (mainly by thermal expansion) of 9 cm. At the other extreme, with the most sensitive version of the climatic models and the SRES A1FI scenario, characterized by the same population development but with a very rapid economic growth and an energy system based on a high intensity of fossil fuel consumption (Nakićenović and Swart, 2000), the figures are +5.6°C and +86 cm. The intermediate SRES A1B scenario,

[8] Carbon dioxide has the property of being transparent to solar radiation but does not afford an easy passage to the invisible infrared rays emitted by the earth's surface in response to heating by the sun. It is a gas with a so-called greenhouse effect, which plays a similar role to that of the glazing in greenhouses. If the percentage of carbon dioxide in the atmosphere increases, the Earth can no longer cool itself as much by infrared emissions. The heat is trapped at surface level. The climate heats up, the seas send more water vapor into the atmosphere, rain becomes more abundant on a global average, and subtle equilibriums that control the planetary climates are disturbed. In consequence, therefore, of human activities, the percentage of carbon dioxide in the atmosphere (its concentration, which constituted 368 parts per million in volume in the year 2000) has augmented by 30% since the preindustrial era. The present rate of increase is 0.4% par annum.

[9] It should be noted that none of the SRES scenarios (Nakićenović and Swart, 2000) considered by the IPCC envisaged supplementary initiatives taken after 2000 with the specific aim of protecting the climate, not even the work of the Kyoto Protocol. The reduction of emissions envisaged for certain of these scenarios resulted from other factors than those of a desire to protect the climate, such as the concern to reduce air pollution or preserve non-renewable resources.

[10] The sensitivity of a climatic model to an increase in the concentration of gases with a greenhouse effect is conventionally measured by the increase in average world temperature simulated by a doubling of the preindustrial concentration of carbon dioxide, when the climate of the model is expected to be stabilised at a new equilibrium temperature. This sensitivity is typically between 1.5°C and 4.5°C.

which has the same demographic base and postulates a balance[11] between the energy sources, gives +3°C and +38 cm when associated with a model of average sensitivity (Houghton *et al*., 2001). The hydrological cycle will also be accelerated by increased evaporation, with increased droughts in some areas, and more floods in others. To put these figures into perspective, it must be borne in mind that the speed of warming will most probably be higher than it has been at any other period since the end of the last ice age, 10,000 years ago.

Such rapid changes in climate will have effects on different aspects with regard to the habitability of our planet. Thus, the rise in temperatures will cause an increase in cardiovascular diseases during heat waves, and an extension in the geographical area of tropical diseases is equally to be anticipated (e.g., the possibility of malaria in Europe). Agricultural zones and ecosystems will move, if they are able to do so. A spatial redistribution of agriculture will take place, and water resources will be affected. The consequences of the rise in the average sea level will be increased flooding in the coastal areas and an invasion of the underground water by salt water.

The IPCC points out that the vulnerability of human health, socioeconomic systems, and (to a lesser extent) some ecosystems depends on economic systems and the institutional infrastructure. This implies that vulnerability to climatic changes is generally higher in developing countries, where the economic and institutional conditions are less favorable (McCarthy *et al*., 2001).

In demography, the predominant spatial reference is that of the state, and this observation is in agreement with what was stated above concerning the birth of the population concept and the Foucault analysis. Administrations, in particular national statistical institutions, have set up national systems for the collection of demographic data—mainly the civil registry, census, and national enquiries. Most of the analytical tools developed by demographers refer to these sources of national data.

Environmental problems, however, know no frontiers. Ulrich Beck (1992), a German sociologist and pioneer in sociological and environmental studies, argued that "an older industrial society, whose axial principle was the distribution of 'goods,' was being displaced by an emergent 'risk society,' structured, so to speak, around the distribution of 'bads'" (Szerszynski *et al*., 1996:2).

[11] Balance means that there is no excessive reliance on one particular source of energy, assuming that similar rates of improvement apply to all technologies concerning energy supply and end use (Nakićenović and Swart, 2000).

Demography, with its predilection for national reference, is therefore not well placed for studying environmental questions, and it is not surprising that few quantitative studies on the subject carried out by demographers should be on the global scale (Bongaarts, 1992; Lutz, 1992; Bartiaux and van Ypersele, 1993) or on the local or regional scale (Markos, 1997; see also the numerous case studies collected in *Les spectres de Malthus* by Gendreau and Meillassoux, 1991). There are few local studies on population and environment concerning the northern regions, and this appears to demonstrate the reluctance of demographers to abandon the national scale for their studies on developed countries, or the fact that the local level has little pertinence for these countries. For John Caldwell (1996:333), "Demography will remain a distinct discipline because of its approach: [. . .] that the study should be representative of sizeable or significant and definable populations." Demographers should nevertheless recall that the pertinence of the population chosen for study does not depend solely on the data and available analytical techniques, for the latter frequently channel the problems.[12]

It appears to us, however, that the spatial scale is not the only reason for the relative absence in developed countries of regional studies on the links between population and environment, and that, as we shall see later on, the relationship that demographers have established (consciously or not) between nature and culture also plays a part.

2. Time Scales

Environment is characterized by a great variety of time scales. Stocks of resources provided by the environment have often been accumulated over long periods. Oil, for example, results from the transformation of organic matter, such as plankton, over millions of years. The ozone layer that protects the Earth from an excess of harmful ultraviolet rays took almost 3000 million years to accumulate (Graedel and Crutzen, 1993:220), but its depth diminished through the emissions of CFCs, gases only invented around 1930. Soils are the outcome of very slow erosion and a physicochemical transformation of Earth's crust, but it requires only a few years of ill-conceived irrigation, for example, to make the soil sterile through salinization.

The climate of our planet has experienced a succession of glacial and interglacial periods in the course of the past 2 million years. The characteristic periods of

[12] As, for example, the study of households in demography (Bartiaux, 1991).

these variations, associated with astronomic factors, are 23,000, 41,000, and 100,000 years. It required 10,000 years for the 2-km-thick ice cap covering Scandinavia, Russia, and North America to melt 20,000 years ago, but the warming that will result in the 21st century from the gas emissions with greenhouse effect arising from the burning of fossil fuels will be at least 20 times faster.

The time scale for certain factors is short (carbon monoxide coming from badly ventilated stoves can kill within a few hours), for others it takes tens of years (exposure to radon, a radioactive gas arising from the soil and from certain building materials, causes cancer in the long term) (Lindvall, 1992).

As remarked by Jean-Paul Deléage (1991:247), "what characterizes ecological periods is the combination of the infinitely short and extremely long, from which has arisen the impression of very great stability that has shaped the perception of all generations up to the present regarding their relations with the environment surrounding them."

In demography, the time reference for calculating period indices is the year, which seems too rigid for studying, for example, the influence of environmental factors on death and fertility. Retrospective analysis is not to be disregarded, but prospective analysis by cohort appears more suitable. Research into epidemiology and public health concerns the influence of environmental factors[13] on health and death, and it is doubtless useful for demographers to see the extent to which their concepts and methods are compatible with those of demography.

In their demographic perspectives, demographers extend their time reference up to 20, 30, or 50 years and even longer, but then the degree of uncertainty increases rapidly. Over recent years, long-term projections of world population have been revised downward several times. For example, for 2025, the high estimations of the United Nations projections carried out in 1996 (8580 million earth dwellers) were much closer the 1992 average estimate of 8500 million.

3. View of the Links between Nature and Culture

The question of links between environment and society and between nature and culture is fundamental for our problem, even though it is only rarely approached directly by demographers. Yet one can recognize, as Jean-Guy Vaillancourt (1996) stresses for sociology, that demography is largely anthropocentric and functions principally according to the paradigm of human exemptionalism, according to which "humans are not considered as beings that are dependent on nature in their physical activities, but rather as cultural entities" (Vaillancourt, 1996:29–30). This is especially valid for demography as practiced in the north, as we have noted above with regard to the time scales.

Jean-Guy Vaillancourt (1996) has reviewed some trends in American sociology that are opposed to this paradigm and that adopt a contrary view by insisting on the role of the environment in social organization (that is to say, human ecology [close to the Chicago school], then environmental sociology and, in the 1990s, ecosociology). These tendencies, however, are marginal in sociology and especially in environmental sociology, in which the "constructivist" paradigm is dominant—environmental "problems" are nonexistent, they are merely social constructions set by lobbies (especially scientific) that have an interest in so doing (see Hannigan, 1995). Although at least two invitations to follow the paradigm of human ecology have been made to demographers in two major reviews of the discipline—*Population and Development Review* (Law *et al*., 1992) and *Population Index* (Namboodiri, 1994)—it seems that it has not been followed up and that demographers, who have up to now hardly studied the consequences of environmental factors for populations (E–P), at least for developed countries, continue rather (albeit rarely) to consider the inverse relation (P–E).

4. Causal or Systemic Perspective

This subject, which has divided certain demographers in passionate disputes, is scarcely a preoccupation in the natural sciences of the environment. Some causal and linear relations established by certain of them have been introduced by others into more complex models, with retroactive effects and the intervention of new processes.

No comprehensive model exists for the relationship between population and environment, as is the case in climatology, but the simulations of the Club of Rome and those conducted by Wolgang Lutz at the International Institute for Applied Systems Analysis are attempts in this direction. The question remains whether the simulation models of population and environment will provide a means of reconciling the causalists and systematicians. Only the future will tell, but it does seem that, irrespective of the hypothesis, the choice of a causal or systemic perspective is not a prerequisite for studying population–environment relationships.

[13] See especially the chapter by Francis Sartor (Chapter 50) of this treatise.

5. Research Action and Value Judgments

This topic is more complex, because it raises the question of the relationship between science (in this case demography) and politics. Michel Foucault (1976) would contend that since its origin demography has been linked to politics and has played a role of scientific legitimation, as can be maintained by comparing the efforts (relatively weak) in theorizing mortality to the advances in fertility.

As far as population–environment relationships are concerned, are demographers expecting a (possible) demand for a policy? Do they wish to participate in the management of the environment (just as certain demographers contribute to the town and country planning territory)? Do demographers, for example, wish to participate in quantifying the responsibilities between the north and south concerning the increase in greenhouse gases? Judging by some of the publications issued by demographers on the subject, the answer appears to be in the affirmative. The political implications of this are evident.

Can demographers participate in the management of the environment in a neutral way, without value judgments or ideological prejudices? Certain studies cited above appear to indicate the negative, and it is necessary to remain vigilant.

This idea that science must clarify political choices is, as Pierre Hamel (1996) has shown, inherited from the "Enlightenment" tradition and a component of modernity. But human behavior is not always predictable by the human sciences (as demographers well know), and philosophers and sociologists are telling us that we have entered the postmodern era. What then will be the role of science (and of demography) for studying the environment?

6. Interdisciplinarity

Study of the relations between population and environment necessarily requires interdisciplinarity, associating researchers from the social sciences with those from the natural sciences. Marcel Jollivet (1992) has termed those engaged in the interdisciplinary study of the relationships between nature and society "frontier runners," and has analyzed the theoretical and methodological question that surround these efforts.

Brigitte Dumas (1996), has defined interdisciplinarity (in relation to environment) as being "a combined employment of some disciplines, entailing reciprocal transformations in each of them." If this definition is accepted, then demography, with its blurred outlines, is closer to interdisciplinarity than are other social sciences such as social psychology, which has a more rigid theoretical framework and a well-standardized method.

Another advantage of demography for working on the relations between population and environment is doubtless its quantification and measurement equipment, highly esteemed by the natural sciences of the environment. It was indeed a demographer, Wolfgang Lutz, who was responsible for showing the risk of errors inherent in the celebrated equation ($I = PAT$) in the population–environment field, and this is very important.

But beyond quantification, demographers know, especially through their experience in the area of family planning, that values and modes of social organization have to be taken into account, in order avoid failure.

CONCLUSION

When they are interested in the interactions between environment and societies, environmental specialists from the natural sciences very often have a simplistic view of demographic phenomena. Little inclined to question the parameters A (for affluence,, i.e., the level of consumption) and T (technology) of the equation $I = PAT$, or the fact of dynamic demographic transitions, they often have the tendency to attribute all the ills of the Earth to the "population explosion." In so doing, they follow the expression introduced by the American biologist Paul Ehrlich (a butterfly specialist), whose book *The Population Bomb*, which appeared in 1968, has had an enormous and lasting repercussion.

In collaborating more with these researchers, demographers could doubtless show how their own knowledge enables a better understanding of the complexity of the interactions between population and environment.

Although, up to the present, demographers have little studied the influence of population(s) on environment, and even less that of environment on population(s), it seems to us that they could give these questions more consideration were they to reconsider their spatial scales, especially if they were prepared to abandon the paradigm of human exemptionalism (and not only for southern countries). In so doing, demographers could adopt the tradition of case studies and other monographs, or of demographic doctrines, by clarifying their value judgments and by being vigilant about the normative presumptions that guide their work.

The relations and interactions between population and environment are many and often call for collaboration with other disciplines (e.g., medicine for studying the effects of an heat wave; economics for studying energy consumption). Research into more sustainable development models (WCED, 1987; CNUED, 1988) in both the north and the south widens the context in which these studies are made. In this area, as in others, demographers will increasingly call for interdisciplinary study.

They will also have to (continue to) keep their distance from the economic paradigm of action (the individual acts rationally and the market controls supply and demand, including nonrenewable resources), without adopting the biological paradigm uncritically (and a concept such as carrying capacity) or even the constructivist paradigm of the sociologists. Is this a difficult balance or an impossible challenge?

Acknowledgment

This chapter was translated from the French version by Ronald Wilkins.

References

Agrasot Paloma, Tabutin Dominique, and Thiltgès Évelyne, 1994. Les relations entre population et environnement dans les pays du Sud: faits et théories, *in:* Hubert Gérard (ed.), *Intégrer population et développement*, p. 383–420. Louvain-la-Neuve, UCL, CIDEP, Academia; Paris, L'Harmattan, CEPED, 824 p.

Bartiaux Françoise, 1991. *Formation et transformations des ménages des personnes âgées. Une analyse par ménage de différents recensements.* Louvain-la-Neuve, Academia, 256 p.

Bartiaux Françoise and van Ypersele Jean-Pascal, 1993. The role of population growth in global warming, *in:* International Union for the Scientific Study of Population [IUSSP], *International Population Conference, vol. 4*, p. 33–54. Liège, Belgium, IUSSP, 384 p.

Baud Pascal, Bourgeat Serge, and Bras Catherine, 1995. *Dictionnaire de géographie.* Paris, Hatier, 432 p.

Beck Ulrich, 1992. *Risk society: towards a new modernity.* London, Sage, 260 p.

Bongaarts John, 1992. Population growth and global warming, *Population and Development Review*, vol. 18(2), p. 299–319.

Boardman Brenda, 1991. *Fuel poverty: from cold homes to affordable warmth.* London, Belhaven Press, 267 p.

Bongaarts John, 1996. Population pressure and the food supply system in the developing world, *Population and Development Review*, vol. 22(3), p. 483–503.

Caldwell John C., 1996. Demography and social science, *Population Studies*, vol. 50(3), p. 305–333.

Chen Robert S., Bender W. H., Kates Robert W., Messer Ellen, and Millman Sara R., 1990. *The hunger report.* Providence, Brown University World Hunger Program.

Cohen Joel E., 1995. *How many people can the Earth support?* New York, Norton, 532 p.

Commission des Nations unies sur l'environnement et le développement [CNUED], 1988. *Notre avenir à tous.* Montréal, CNUED, Éditions du Fleuve, 424 p. (Known as the *Brundtland Report*).

Dalla Valle Christine, Mons Johanne, Bartiaux Françoise, and Yzerbyt Vincent, 2001. Les actions environnementales des consommateurs, *in:* Marie-Paule Kestemont, Françoise Bartiaux, Nadine Fraselle, and Vincent Yzerbyt (eds.), *Points d'ancrage pour une politique de développement durable: production et consommation*, p. 45–172. Brussels, Services fédéraux des Affaires scientifiques, techniques et culturelles, 208 p.

Davis Kingsley, 1990. Population and resources: fact and interpretation, *Population and Development Review*, vol. 16(Suppl.), p. 1–21.

Deléage Jean-Paul, 1991. *Histoire de l'écologie. Une science de l'homme et de la nature.* Paris, La Découverte, 330 p.

Dumas Brigitte, 1996. Science normale et sciences humaines: les voies de l'interdisciplinarité pour la resolution des problèmes environnementaux, *in:* Robert Tessier and Jean-Guy Vaillancourt (eds.), *La recherche sociale en environnement: nouveaux paradigmes*, p. 75–88. Montreal, Presses de l'Université de Montréal, 176 p.

Dyson Tim, 1994. Population growth and food production: recent global and regional trends, *Population and Development Review*, vol. 20(2), p. 397–411.

Ehrlich Paul R., 1968. *The Population Bomb.* New York, Ballantine, 201 p.

Ehrlich Paul R., Ehrlich Anne H., and Daily Gretchen C., 1993. Food security, population, and environment, *Population and Development Review*, vol. 19(1), p. 1–32.

Ehrlich Paul R. and Holdren John P., 1971. Impact of population growth, *Science*, vol. 171(3977), p. 1212–1217.

Foucault Michel, 1976. *Histoire de la sexualité, Vol. 1: La volonté de savoir.* Paris, Gallimard, 211 p.

Gendreau Francis, Gubry Patrick, and Véron Jacques, 1996. *Populations et environnement dans les pays du Sud.* Paris, Karthala and CEPED, 308 p.

Gendreau Francis and Meillassoux Claude (eds.), 1991. *Les spectres de Malthus: déséquilibres alimentaires, déséquilibres démographiques.* Paris, EDI, 442 p.

Gilland Bernard, 1995. World population, economic growth, and energy demand, 1990–2100: a review of projections, *Population and Development Review*, vol. 21(3), p. 507–539.

Graedel Thomas E. and Crutzen Paul J., 1993. *Atmospheric change: an earth system perspective.* New York, Freeman and Co., 446 p.

Green Cynthia P., 1992. *The environment and population growth: decade for action.* Baltimore, Johns Hopkins University, Population Information Program, 31 p.

Guérin-Pace France and Collomb Philippe, 1997. *Les Français et l'environnement. L'enquête "Populations, espaces de vie, environnements."* Paris, INED and Presses universitaires de France 255 p.

Hamel Pierre, 1996. Crise de la rationalité: le modèle de planification rationnelle et les rapports entre connaissance et action, *in:* Robert Tessier and Jean-Guy Vaillancourt (eds.), *La recherche sociale en environnement: nouveaux paradigmes*, p. 61–74. Montreal, Presses de l'Université de Montréal, 176 p.

Hannigan John A., 1995. *Environmental Sociology, A Social Constructionist Perspective.* London: Routledge, 236 p.

Harrison Paul, 1991. *Population–environment: quantifying the linkages.* London, British Society for Population Studies, 12 p.

Harrison Paul and Pearce Fred, 2000. *AAAS Atlas of Population and Environment.* Berkeley, American Association for the Advancement of Science and University of California Press, 204 p.

Hester Gordon L., 1992. Electric and magnetic fields: managing an uncertain risk, *Environment*, vol. 3(1), p. 6–11.

Heilig Gerhard K., 1994. Neglected dimensions of global land-use change: reflections and data, *Population and Development Review*, vol. 20(4), p. 831–859.

Houghton John T., Ding Yihui, Griggs David J., Noguer Maria, van der Linden Paul J., Dai Xiaosu, Maskell Kathy, and

Johnson Cathy A. (eds.), 2001. *Climate change 2001: the scientific basis.* Cambridge, Cambridge University Press, 881 p.

Jollivet Marcel (ed.), 1992. *Sciences de la nature, sciences de la société Les passeurs de frontières.* Paris, CNRS editions, 589 p.

Keyfitz, Nathan, 1996. Population growth, development and the environment, *Population Studies*, vol. 50(3), p. 335–359.

Lannoy Pierre and Panafit Lionel, s.d. *La population a-t-elle une date de naissance? La démographie selon M. Foucault.* Louvain-la-Neuve, Laboratoire d'anthropologie et de sociologie, Université catholique de Louvain, 7 p. (unpublished).

Law Babbi S., Clarke Alice L., and Lockridge Kenneth A., 1992. Toward an ecological demography, *Population and Development Review*, vol. 18(1), p. 1–33.

Le Bras Hervé, 1994. *Les limites de la planète: Mythes de la nature et de la population.* Paris, Flammarion, 351 p.

Leslie George B. and Lunau Frank W. (eds.), 1992. Indoor air pollution: problems and priorities. Cambridge, Cambridge University Press, 329 p.

Leuenberger Philippe, 1993. Le poumon de l'homme à l'interface avec l'environnement, *in:* Pierre Hainard (ed.), *L'homme et l'environnement*, p. 65–79. Lausanne, Payot, 104 p.

Lindvall Thomas, 1992. Radon, *in:* George B. Leslie and Frank W. Lunau (eds.), *Indoor air pollution: problems and priorities*, p. 99–116. Cambridge, Cambridge University Press, 329 p.

Lutz Wolfgang, 1992. *Population and environment: What do we need more urgently: better data, better models or better questions?* Oxford, Oxford University, 16 p.

Lutz Wolfgang (ed.), 1994. *Population–development–environment: understanding their interactions in Mauritius.* Berlin, Springer Verlag, 400 p.

Lutz Wolfgang, Prieto Leonel, and Sanderson Warren (eds.), 2000. *Population, development and environment on the Yucatan Peninsula: from ancient Maya to 2030.* Laxenburg, International Institute for Applied Systems Analysis.

Lutz Wolfgang, Prieto Leonel, and Sanderson Warren, 2002a. *Population and environment: methods of analysis, a supplement to Vol. 28: population and development review*, 251 p.

Lutz Wolfgang, Prskawetz Alexia, and Sanderson Warren, 2002b. Introduction, *Population and environment: methods of analysis, a supplement to Vol. 28: population and development review*, p. 1–21.

Lutz Wolfgang, Sanderson Warren C., and Wils Annababette 2002. Conclusions: toward comprehensive P-E studies, *population and environment: methods of analysis, a supplement to Vol. 28, population and development review*, p. 225–250.

Mackellar F. Landis, Lutz Wolfgang, Prinz Christopher, and Goujon Anne, 1995. Population, households, and CO_2 emissions, *Population and Development Review*, vol. 21(4), p. 849–865.

Markos Ezra, 1997. *Demographic Responses to Ecological Degradation and Food Insecurity.* Amsterdam, Netherlands Graduate School of Research in Demography (PDOD), 373 p.

McCarthy James J., Canziani Osvaldo F., Leary Neil A., Dokken David J., and White Kasey S. (eds.), 2001. *Climate change 2001: impacts, adaptation & vulnerability.* Cambridge, Cambridge University Press, 1000 p.

McNeill John, 2000. *Something new under the sun: an environmental history of the 20th century.* London, Penguin, 421 p.

Nakićenović Nebojsa and Swart Rob (eds.), 2000. *Emissions scenarios.* Cambridge, Cambridge University Press, 570 p.

Namboodiri Narayanam Krishnam, 1994. The human ecological approach to the study of population dynamics, *Population Index*, vol. 60(4), p. 577–539.

Naylor Rosamond, Falcon Walter, and Zavaleta Erika, 1997. Variability and growth in grain yields, 1950–1994: does the record point to greater instability? *Population and Development Review*, vol. 23(1), p. 41–58.

O'Neill Brian, MacKellar Landis, and Lutz Wolfgang (eds.), 2001. *Population and climate change.* Cambridge, Cambridge University Press, 266 p.

Ouharon Aïcha, 1997. Population et environnement: une controverse toujours d'actualité, *La Chronique du CEPED*, no. 24, p. 1–4.

Pachauri Rajendra K. and Qureshi Lubina F. (eds.), 1997. *Population, environment, and development.* New Delhi, Tata Energy Research Institute, 357 p.

Petit Véronique, 1997. Sociétés d'origine et logiques migratoires: les Dogon de Shanga (Mali), *Population*, vol. 52(3), p. 515–543.

Prades José A., 1996. Éthique, environnement et développement: vers un paradigme de la connaissance collaborative, *in:* Robert Tessier and Jean-Guy Vaillancourt (eds.), *La recherche sociale en environnement: nouveaux paradigmes*, p. 9–18. Montreal, Presses de l'Université de Montréal, 176 p.

Quesnel Louise (ed.), 1995. *Social sciences and the environment: les sciences sociales et l'environnement.* Ottawa, Presses de l'Université d'Ottawa, 223 p.

Sanderson Warren C., Fuller Ben, Hellmuth Molly E., and Strzepek Kenneth M., 2001. *Namibia's future: modeling population and sustainable development challenges in the era of HIS/AIDS. Executive summary.* Laxenburg, International Institute for Applied Systems Analysis, 29 p.

Sanderson Warren C., Hellmuth Molly E., and Strzepek Kenneth M., 2001. *Botswana's future: modeling population and sustainable development challenges in the era of HIS/AIDS. Executive summary.* Laxenburg, International Institute for Applied Systems Analysis, 32 p.

Smil Vaclav, 1990. Planetary warning: realities and responses, *Population and Development Review*, vol. 16(1), p. 1–30.

Smil Vaclav, 1991. Population growth and nitrogen: an exploration of a critical existential link, *Population and Development Review*, vol. 17(4), p. 569–601.

Smil Vaclav, 1994. How many people can the earth feed? *Population and Development Review*, vol. 20(2), p. 255–299.

Szerszynski Bronislaw, Lash Scott, and Wynne Brian, 1996. Introduction: ecology, realism and the social sciences, *in:* Scott Lash, Bronislaw Szerszynski, and Brian Wynne (eds.), *Risk, environment and modernity: towards a new ecology*, p. 1–26. London, Sage, 294 p.

Tabutin Dominique, 1995. Pressions démographiques et environnement: des faits et doctrines aux données scientifiques, *in:* Louise Quesnel (ed.), *Social sciences and the environment/les Sciences sociales et l'environnement*, p. 173–198. Ottawa, Presses de l'Université d'Ottawa, 223 p.

Tessier Robert and Vaillancourt Jean-Guy (eds.), 1996. *La recherche sociale en environnement: nouveaux paradigmes.* Montreal, Presses de l'Université de Montréal, 176 p.

Thiltgès Évelyne and Dominique Tabutin, 1992. *Population et environnement, une synthèse des faits, doctrines et politiques dans les pays du Sud.* Louvain-la-Neuve, Academia, 85 p.

Tiffen Mary, Mortimore Michael J., and Gichuki Francis, 1994. *More people, less erosion: environmental recovery in Kenya.* New York, J. Wiley, 311 p.

Turner II Billie L., Clark William C., Kates Robert W., Richards John F., Mathews Jessica T., and Meyer William B., 1990. *The Earth as transformed by human action: global and regional changes in the biosphere over the past 300 years.* Cambridge, Cambridge University Press, 713 p.

United Nations Environment Program [UNEP], 2002. *Global environmental outlook 3: past, present and future perspectives.* London, UNEP and Earthscan, 446 p.

United Nations Fund for Population Activities [UNFPA], 1991. *Population, resources, and the environment: the critical challenges.* New York, UNFPA, 154 p.

United Nations Fund for Population Activities [UNFPA], 2001. *The state of the world population 2001: footprints and milestones: population and environmental change.* New York, UNFPA, 76 p.

Vaillancourt Jean-Guy, 1996. Sociologie de l'environnement: de l'écologie humaine à l'écosociologie, *in:* Robert Tessier and Jean-Guy Vaillancourt (eds.), *La recherche sociale en environnement: nouveaux paradigmes. Vol. 1*, p. 19–47. Montreal, Presses de l'Université de Montréal, 176 p.

Veyret Yvette and Pech Pierre, 1993. *L'homme et l'environnement.* Paris, Presses Universitaires de France, 423 p.

von Weiszäcker Ernst-Ulrich, Lovins Amory B., and Lovins L. Hunter, 1997a. *Factor four: halving resource use: the new report to the Club of Rome.* London, Earthscan Publication Ltd, 322 p.

von Weiszäcker Ernst-Ulrich, Lovins Amory B., and Lovins L. Hunter, 1997b. *Facteur 4. Deux fois plus de bien-être en consommant deux fois moins de ressources.* Mens, Terre vivante, 320 p.

Wackernagel Mathis and Rees William, 1996. *Our Ecological Footprint: Reducing human impact on the Earth.* Gabriola Island, New Society Publishers, 160 p.

Water Energy Health Agriculture and Biodiversity Working Group [WEHAB], 2002a. *A framework for action on water and sanitation.* New York, United Nations, 40 p.

Water Energy Health Agriculture and Biodiversity Working Group [WEHAB], 2002b. *A framework for action on energy.* New York, United Nations, 36 p.

Water Energy Health Agriculture and Biodiversity Working Group [WEHAB], 2002c. *A framework for action on health and the environment.* New York, United Nations, 36 p.

Water Energy Health Agriculture and Biodiversity Working Group [WEHAB], 2002d. *A framework for action on agriculture.* New York, United Nations, 32 p.

Water Energy Health Agriculture and Biodiversity Working Group [WEHAB], 2002e. *A framework for action on biodiversity and ecosystem management.* New York, United Nations, 36 p.

World Commission on Environment and Development [WCED], 1987. *Our common future.* Oxford, Oxford University Press, 400 p. (The Brundtland Report).

World Health Organization [WHO], 2000. *Air quality guidelines for Europe.* Copenhagen, WHO Regional Office for Europe, 273 p.

Wils Annababette, 1996. *PDE-Cape Verde: a systems study of population, development and environment.* Laxenburg, International Institute for Applied Systems Analysis, 144 p.

Wils Annababette, da Costa Gaspar Manuel, Hellmuth Molly E., Ibraimo Maimuna, Prommer Isolde, and Sebastiao Emidio, 2001. *Mozambique's future: modeling population and sustainable development challenges in the era of HIS/AIDS. Executive summary.* Laxenburg, International Institute for Applied Systems Analysis, 32 p.

World Resources Institute [WRI], 1992. *World resources 1992: a guide to the global environment.* Oxford, Oxford University Press, 385 p.

World Summit on Sustainable Development [WSSD], 2002a. Plan of implementation of the World Summit on Sustainable Development, *in: Report of the World Summit on Sustainable Development*, p. 6–72. New York, United Nations.

World Summit on Sustainable Development [WSSD], 2002b. Political declaration, *in: Report of the World Summit on Sustainable Development*, p. 1–5. New York, United Nations.

World Summit on Sustainable Development [WSSD], 2002c. *Key outcomes of the summit.* New York, United Nations, 7 p.

Zamoun Slimane, Tabutin Dominique, Yaakoubd Abdel-Ilah, and Kouaouci Ali, 1995. *Population et environnement au Maghreb.* Paris and Louvain-la-Neuve, Academia-Erasme and L'Harmattan, 301 p.

APPENDIX: STATE OF THE ENVIRONMENT AT THE BEGINNING OF THE 21ST CENTURY

The following data are extracted from the *GEO-3* (UNEP, 2002) report, except where otherwise indicated.

Twenty percent of the richest of the world population account for 86% of the total private consumption expenditure; consume 58% of the world's energy (on the basis of 5 tons of oil equivalent per person per year) (WEHAB, 2002b), 45% of the consumption of meat and fish, and 84% of that of paper; and possesses 87% of the cars and 74% of the telephones.

Twenty percent of the poorest of the world population consume less than 5% of each of the goods and services. About 2000 million people have no access to electricity, mainly in the rural zone (WEHAB, 2002b).

Almost one-third of the world population live in a country that experiences moderate or high water shortage (i.e., the consumption of water is more than 10% of the renewable resources): 1100 million people have no access to drinking water, 2400 million have no access to sanitary installations, and 2200 million people die every year of diarrhea allied to the quality of the water.

Sewers constitute the most important source in volume of contamination of marine and coastal waters. Contributions of nitrogen and agricultural fertilizers from the sewers provoke outbreaks of eutrofication or explosive growth of toxic algae. The total volume of fishing is decreasing. The principal danger to the coastal environment is the direct physical destruction of the habitats. The latter is one of the main factors that bear down on coral, 58% of which are regarded as being threatened.

Nearly 2000 million hectares (mha), that is, about 15% of the continental surface, have been degraded by human activities (overgrazing, 35%; deforestation, 30%; agricultural activities, 27%; overexploitation of vegetation, 7%; industrial activities, 1%).

The net loss of afforested surface is 9.4 mha a year. Natural forests are losing about 16 mha per year, of which 15 mha are in tropical areas. About 70% of the deforestation has taken place in order to extend agriculture.

Biodiversity is decreasing at a speed that is several times higher than the historic rate. 24% of the mammal species and 12% of bird species are considered to be threatened on a global scale. The main factors in these changes are the degradration and loss of habitats (linked to agricultural development, exploitation of forests, construction of dams, mining activity and urbanization), climatic changes, pollution, nonsustainable exploitation of natural resources, and the introduction of exotic species.

Air pollution, both external and internal, is considered to be responsible for 5% of global morbidity. Air pollution aggravates (and perhaps causes) asthma and other respiratory allergic diseases. Almost 2 million people in developing countries die every year owing to exposure to the ultrafine particles of smoke from domestic biomass fires in dwellings. A third of the world's population depends on traditional fuels (wood, dung, and agricultural waste) to ensure their daily needs for heating and cooking (WEHAB, 2002b). The burning of fossil fuels is at the origin of emissions of sulphur components that cause acid rain and deposits. These emissions are rapidly declining in developed countries but are rapidly increasing in the Asia–Pacific region. Nitrogen oxide emissions lead to the formation of ozone in the lower atmosphere (where it is a pollutant that attacks both the lungs and vegetation).

Persistant organic pollutants, such as pesticides, which are chemically stable, are carried by the atmosphere over very long distances and accumulate particularly in the polar regions. They contaminate the food chain there and are to be found in animal fat, which constitutes a risk both for human health and the ecosystems.

The ozone layer in the stratosphere, which protects living organisms from an excess of ultraviolet radiation, has been reduced above the polar regions by chlorofluorocarbon gases (CFCs) used especially in refrigeration. Following the measures taken to progressively ban the production of CFC, the ozone layer in the stratosphere should recover its former state within some decades from now.

Because of the massive burning of fossil fuels, the concentration of carbon dioxide is 30% higher than its preindustrial value. The average global surface temperature has increased by 0.6°C during the 20th century. More than half the warming observed over the last 50 years is owing to the increase in the gases with greenhouse effect of human origin, of which carbon dioxide is the most important. Most of the rise in average sea levels observed in the 20th century (10 to 20 cm) is probably linked to the increase in temperature. The first consequences of these climatic changes for certain fragile ecosystems have been observed.

P A R T

II

THE POPULATION IN ITS BIOLOGICAL, ECONOMIC, SOCIAL AND CULTURAL SETTING

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

The determinants of change in demographic behavior were extensively discussed in Volumes I and II, and its consequences were more succinctly examined in Section II, Part I of this volume. Part II will now examine a set of relationships between population dynamics and various aspects of the biological, economic, social and cultural settings in which they take place; here it is more a matter of interactions than clear-cut causal relations. These contextual factors deserve demographers' consideration both in themselves and with a view to understanding their reciprocal influences. Here again, we have not attempted an exhaustive examination. That would be particularly futile since many fields that could be examined are virtually unexplored, and methods for studying others are still in their infancy. Our aim, once again, is simply to illustrate the range of possible approaches and to suggest some avenues for reflection.

As was said in Volume I, the basic principle of demographic analysis is to consider the population as a collection of individuals with no particular feature other than their membership of whatever category is selected for analysis. But we have also repeatedly pointed out that this is too reductionist a principle to enable us to grasp all aspects of the reality. It is sometimes even useful to take the opposite approach. Putting the individual at the center, we can obtain a more practical view of his or her development in time and space within the population(s) in which he or she is born, lives, moves, and dies. This is Paolo De Sandre's approach in Chapter 88, "From the Life Cycle to Life Paths and to Life Transitions."

A similar but more specific approach is to consider the way an individual's habitual environment changes as a result of the demographic transition. In Chapter 89, for example, Alain Monnier and Sophie Pennec describe how, at different stages of the transition, an individual (*Ego*) undergoes changes in "Experiencing Death."

The next two chapters examine households and families, the basic social units that structure all human populations and deserve the demographer's consideration for several reasons. They are the units of observation through which individuals can be approached to count a population or monitor change. They are also the units within which many economic and social decisions are formed and taken, and they can themselves sometimes be considered as elements of a statistical universe to which demographic analysis methods can be applied. They are also prime locations for new demographic behaviors to emerge or traditional behaviours to be conserved. Chapter 90, by Marc Pilon, sets out some general points and gives a practical illustration based on the experience of developing countries. In Chapter 91, Nico Keilman addresses the same question for developed countries.

But although the family (or household) is the basic unit of the social fabric, it is only one element. Four chapters examine different factors, all of which greatly transcend the individual and the family, and which are all more or less closely connected with population dynamics. In Chapter 92, François Héran addresses "Social Mobility, Then and Now"; in Chapter 93, Antonella Pinnelli reviews the relatively recent knowledge of relations between "Gender and Demography"; in Chapter 94, Giuseppe Gesano addresses the complex relationship between "Economic Activity and Demographic Behaviors"; in Chapter 95, Jean-Louis Rallu, Victor Piché and Patrick Simon discuss "Demography and Ethnicity: An Ambiguous Relationship."

To round off, referring back to the first chapter of this volume dealing with the genetic consequences of demographic change, in Chapter 96 Enzo Lucchetti and Alberto Soliani give a more general overview of the relationship between "Genetics and Demography."

CHAPTER

88

From the Life Cycle to Life Paths and to Life Transitions

PAOLO DE SANDRE

Dipartimento di Scienze Statistiche, Università di Padova, Padova, Italy

INTRODUCTION

Classical demographic analysis is characterized by a macrosocial approach based essentially on *officially exhaustive data* (censuses, vital statistics, registration of changes of residence). To investigate the mechanisms by which populations evolve, priority was given to the exploration of differences in behaviors, *beginning with specific events* (births, deaths, migration) and continuing with essential *prior states* (assumed to be closely correlated with or causally linked with the events). But causal analysis may also rely, in changing the perspective, on observation of populations (or subpopulations) that are *susceptible to experience* (or not) these events, in following the *principal biographic elements in the context of the temporal and spatial relationships that individuals maintain among themselves and with their life environment*. This approach is essentially based on exploitation of retrospective or prospective sample surveys, which, based on detailed questionnaires, are much richer sources of information than are the exhaustive sources.

The aim of this chapter is to lay out a sort of *inventory of points of contact* (at the level of *concepts*, of *information* to collect, of *methods of analysis*) between a microdemographic approach (that of the life cycle or life paths) and the macrodemographic approach. Hence, without claiming to be exhaustive, I will present a cross-sectional overview, with a number of elements in common with other chapters of this treatise.

We begin with the following empirical observation. In interaction with the radical evolution of population dynamics that has taken place in Western societies during the past 2 centuries, there have been important *structural changes in the stages of an individual's life*, very easily perceptible and still far from having produced all of their effects in the areas of life style, culture, or the economy. To have a broad knowledge of such radical transformations, it is sufficient to follow the evolution at the macrosocial level of a few key demographic variables. But to get to the heart of things, *longitudinal observation of individuals* and an in-depth *analysis of the systems* in which they live are required. Paradigms and conceptual categories need to be rethought and methodological approaches renewed, in order to combine, in an integrated and interdisciplinary manner, state characteristics, events, and individual behaviors, as well as the interactions among the individual paths and the contexts within which they unfold. From this last perspective, one should not lose sight that the contribution of demographic variables to the transformation of the society takes place in *an historical context of economic and social changes*: from the industrial revolution to postindustrialism (Coale and Watkins, 1986; Chesnais, 1986; Chasteland and Chesnais, 1997).

The chapter is divided into two parts, the first dealing with the empirical description of the life paths and the second presenting the methodological orientation appropriate for the study of these paths. After the first two sections devoted to changes in the nature of the different ages of life under the effect of the demographic transition and to certain consequences of this transformation of the *life cycle* that can lead to policy interventions, I will address several methodological aspects concerning the observation and analysis of life paths, including systems for collecting dynamic information and the opportunity to use multidimensional transition models that link up life trajectories and explanatory variables.

I. DEMOGRAPHIC REGIME AND STRUCTURAL EVOLUTION OF THE AGES OF LIFE

In thinking about the past 2 centuries or even, for the Mediterranean countries, the past 100 years (corresponding to from seven to three intergenerational intervals, respectively), one is struck by a certain number of major changes in the conditions of life linked to the evolution of the demographic regime and to that of the resulting population structures.

In the pretransitional regime with high mortality and high fertility (until the end of the 18th century for France and until the end of the 19th for Italy), the rules of population renewal were such that, from its birth until the mean age of childbearing, a cohort of women was reduced by half. The women who survived married, on average, at around age 20 years, and managed to give birth to 6 children before becoming infertile at around age 38 years, an age by which they had only about 20 years to live (80% became widows by age 50) (Livi Bacci, 1978).

In the current Western regime with low mortality and low fertility, nearly all women survive to the ages at which fertility is highest. On average, a woman only gives birth to less than two children. The mean age of first marriage of women has risen to 26 or 27 years (sometimes more), and although births are infrequent, they are also delayed. Conversely, at age 35 a woman's childbearing is 90% complete and her remaining life expectancy exceeds 45 years. Widowhood at adult ages has become quite rare, owing to improvement in the survival of the two sexes. However, voluntary disruptions of unions have become much more common (especially in the non-Mediterranean countries) (Véron, 1993, AIDELF, 2002).

The structural changes of the life paths underlying these transformations affect all of the traditional age groups.

1. Childhood

Childhood no longer represents a wager against death: to have a child born has become an irreversible choice, almost certainly durable over time. Although when mortality was very high more than one-third of newborns died before their fifth birthday, it is today only the case for one child in 100. Direct family and social investments in raising and educating children have become absolute and costly imperatives; at the same time, for the parents (and especially the mother), the opportunities they must forgo to confront these investment costs increase (in terms of employment and of different uses of time and money). Today, in certain countries of southern Europe (but also of central Europe and the ex-Soviet Union), the model of the only child is even gaining ground, alongside the two-child model, with profound changes in the ways of raising and socializing young children.

2. The Passage to Adulthood

Adolescence is extended to a *youth* who increasingly encroaches onto adulthood. With the passage from nearly direct access to the world of work at the time of preadolescence—typical of agricultural societies—to the *obligatory schooling* for all (most often to age 16, legally, but in reality up to the threshold of adulthood for most youth), which is characteristic of industrial societies, there has been a radical change in the social image of adolescents and of youth (INED, 1982, 1983). The increased educational investment seeks and often succeeds in guaranteeing upward socioeconomic mobility.

However, between the Scandinavian countries and the Mediterranean countries (two extreme situations), enormous *differences* exist in the forms of *passage to adulthood*. In the Scandinavian countries, the departure from the family of origin is *traditionally early*, with the formation of a family unit or of an own neolocal household (with an independent dwelling unit, whether living alone, as a married couple, or cohabiting). Entry into the world of work is likewise early, sometimes with study and work combined. All of this takes place in the context of an overall social insurance and public support system that contributes to maintaining this kind of behavior. In Mediterranean countries, by contrast, departure from the family is *traditionally late*; it still takes place almost exclusively for reasons of marriage, especially for women, and without any important public support (Corijn and Klijzing, 2001). The fertility and family surveys of the comparative European program Fertility and Family Surveys (UN-ECE, 1996–2001) have recently provided an overview of these differences: in Sweden in 1992–1993, only 8.4% of women age 23 were living with their parents, whereas in Italy in 1995–1996, 86.8% of women aged 20 to 24 still resided in their family of origin.

Nevertheless, in all Western countries, we have recently seen a general slowdown in the departure of youth from their families. Multiple factors seem to be at work in favor of a growing diffusion of the practice of staying in the family of origin for an extended period: (1) the reduction in the size of family units, which tends to maintain stronger centripetal cohesion between the parents (or the single parent) and the only child or only the two children (in contrast with large families); (2) the lengthening of periods spent as a student and the growing complexity of the connection between schooling and the evolving economic system and, furthermore, the difficulties of access in the labor market (more precarious prospects for getting established, less professional stability, stagnation of careers); (3) the tendency to delay marriage, whether because of new freedom to live affective and sexual experiences outside of traditional conjugal forms or the growing practice of nonmarital cohabitation; and (4) the emergence among the young of new needs that often cannot be satisfied by their own resources and are better covered by the economies of scale that can be realized within the family of origin. These are thus factors that accompany *a more complex and more conflicted passage to the autonomy and the responsibilities of adulthood*. At the same time that it is necessity to rethink what is meant by old age, because of demographic aging, and where a new schema of taking into account social, health, and economic aspects of the top part of the population pyramid is being laid out, it appears to be essential to not ignore the *extension of the ages* of youth nor the disquieting difficulties of the passage to *adult autonomy*, and to not penalize social investments in favor of youth (Cheal, 1983; Coale, 1987). The indeterminacy and precariousness of social norms concerning the transition to adulthood (Galland, 1984; Bozon and Villeneuve-Gokalp, 1994), which affect youth beyond simple short-term economic pressures, require a more precise knowledge and the definition of new equilibria in the *rules of intergenerational transmission of material and nonmaterial resources*, the old rules no longer being supported or legitimized by traditional demographic structures (Sgritta, 1995).

3. Adults

But *adulthood* is itself expanding and changing. At ages 60 to 65 years, can a person in good health and with 20 years of remaining life expectancy be considered as old? The high proportions of individuals in a cohort who attain this age in good economic and health conditions produce a new social stratum with its own expectations, an important capacity for consumption and investment, and non-negligible political power. But what has undoubtedly most completely changed the *lives of women*, in both timing and content, is the change in *ages of procreation*. Although the duration of their lives is getting longer, women have seen births reduced in number and concentrated within a shorter age interval, considerably reducing the portion of their lives devoted to reproduction and liberating them earlier from the direct obligations of child care. Assuming that these obligations last up to the end of adolescence of the last-born, about 60% of the remaining life after marriage (as opposed to 20% in the pre-transition regime) can thus be devoted to other activities. This is a *new condition*, which has made possible the massive presence of women in the labor market and has given women more continuity in their careers (conforming better to growing expectations in this area). More generally, this evolution has also given women a *different bargaining power between the sexes* in all domains, from access to *education* (new generations of women tend to have more schooling than men) and in decisions regarding *reproduction* (including decision on contraceptive means to use, on recourse to induced abortion, on access to assisted fertility), and the household. The emancipating pressure experienced by women is so strong that it becomes *difficult to reconcile maternal and professional roles* without first experiencing *a revision of the economic system regarding work* (flexibility of paid work time and of forms of support and of social and health services for mothers of small children), and without *also asking men* for greater flexibility on these same subjects and particularly a more equitable distribution of family tasks.

If the new reproductive paths offer new opportunities in life for women, the ways in which they are managed (control of all of the process of reproduction; segmentation of the stages of the reproductive process, which will be discussed again below) have also required profound cultural transformations that lead to the multiplication of models of sexual and conjugal behavior. Figure 88–1 reproduces a classical synthesis (Bongaarts, 1982) of the importance of the evolution of the principal intermediate variables in the change of potential fertility and in the lowering of effective fertility that marks the passage from traditional socio-economic contexts (level 6) to increasingly developed contexts (level 1).

4. Older Ages: Displacement of the Social Barycenter to Old Age

The aging of populations is well on its way to becoming the most visible and striking aspect overall of the demographic transition, whether it flows from

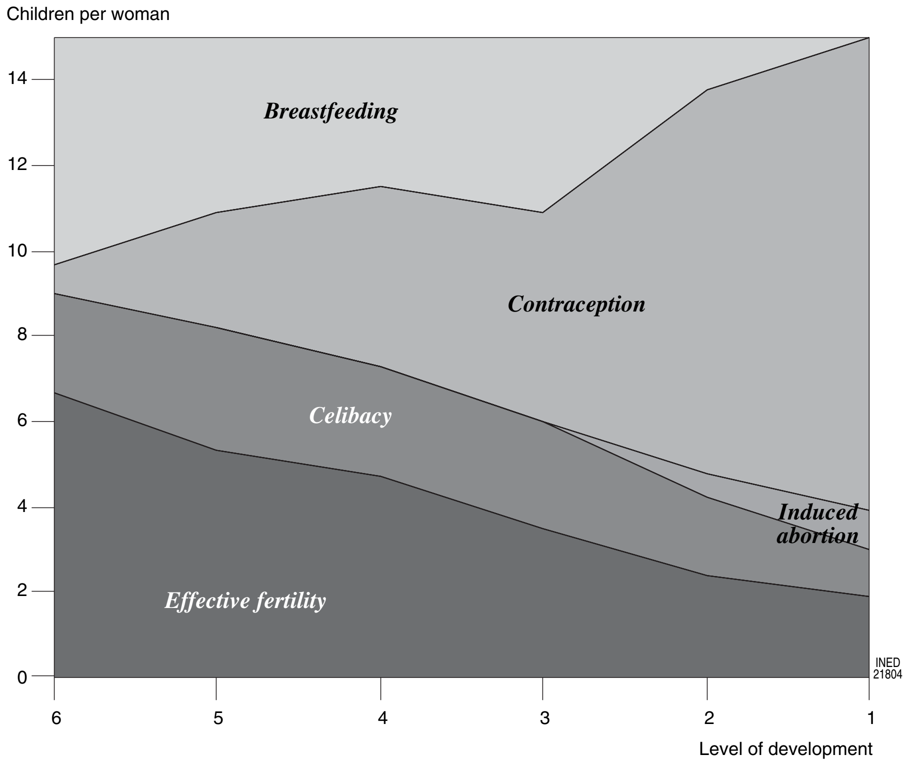


FIGURE 88–1 Effects of the intermediate variables on potential and effective fertility according to the level of development of a country. Scale: 6, less developed; 1, more developed. (From Bongaarts, 1982.)

the scarcity of births, which now tend to be below the replacement level and correspondingly reduce the base of the population pyramid, or the spectacular *increases in life expectancy being realized at older and older ages*. Before the transition, when mortality was high, only 20% of a generation reached age 65; with current survival conditions, it is now the case for nearly 90%. Life expectancy at birth has gone from 35 years, at the beginning of the transition, to 65 years, 50 years ago, and then to 75, 80, or higher (with differences by sex) at present. Just the effect of this prolongation of life, by itself, would have resulted in a demographic transition with more than a doubling of the population. Moreover, the number of people over age 60 has increased because of the prolongation of life and the growing concentration of deaths around the modal age at death (today close to 80). *The socioeconomic level* of the elderly has improved from one generation to the next, under the effect of the accumulation of human capital investments realized since the middle of the 20th century. At older ages health deteriorates, but the *severe loss of autonomy* only becomes important after age 70, becoming even more so after age 80. In France, for example, one-third of individuals over the latter age were estimated to be suffering from a medium incapacitation during the 1980s (Colvez and Robine, 1986). New thresholds need to be adopted for the description of the qualitative characteristics of the oldest age groups (e.g., ages 65 to 74, 75 to 79, and 80 and over for the very elderly), and a moving threshold should be used for old age (measured by the age at which remaining life expectancy is, e.g., 10 years), more consistent with the actual or expected conditions of good functional autonomy (life expectancy in good health; Fig. 88–2) (Caselli and Egidi, 1993). The conception of old age as a relatively brief and residual period of *passive enjoyment* of the goods accumulated during earlier years devoted to work and reproduction appears to be qualitatively obsolete and economically nonviable, in that it weighs entirely on the reduced numbers of people at the traditionally productive ages. The social and cultural adaptation of the idea of old age, in more *active* terms (Laslett, 1989), is getting started, albeit with great resistance and a lot of uncertainty.

5. New Links Regarding Intergenerational Exchanges

At the same time that it has contributed to transforming *individual life paths*, the change of demographic regime that has taken place in the course of the transition, in the economic context of the Industrial

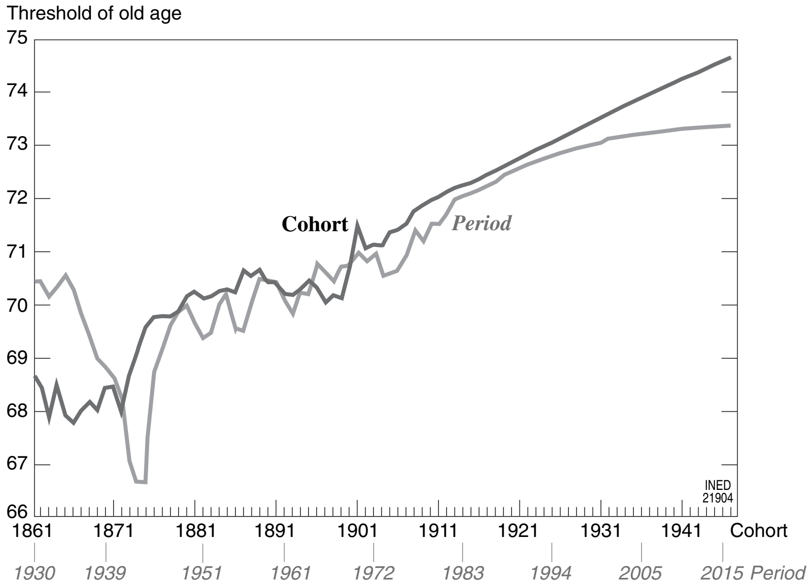


FIGURE 88–2 Threshold of old age (age with 10 years of remaining life expectancy). Comparison of Italian men's life tables by cohort and by period, the year of birth plus threshold of old age. (From Caselli and Egidi, 1993.)

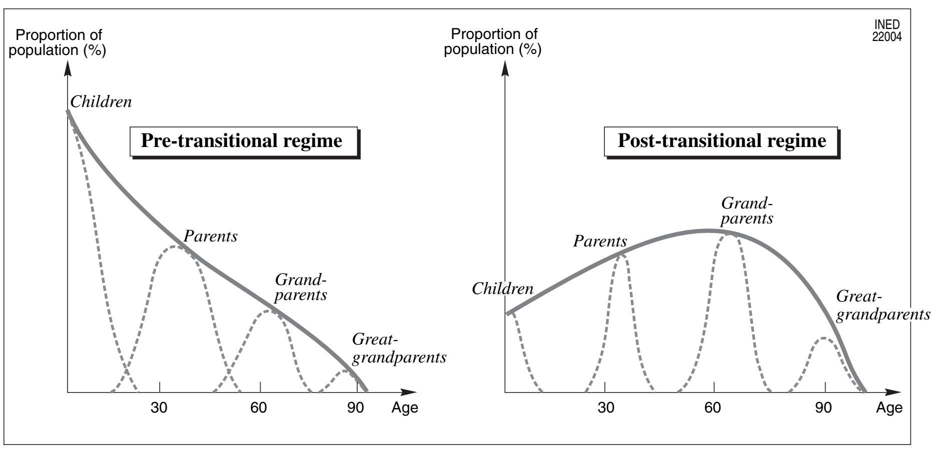


FIGURE 88–3 Chain of relatives.

Revolution, has also radically transformed *inter-generational relations* among children, parents, and grandparents.

The comparison (Fig. 88–3) between the age structure of a *young*, pretransitional population (sharp and regular reduction in numbers going from the youngest to the oldest ages) and that of a contemporary *aged* population (reduced number of the youngest, growing number of adults and elderly, sharp fall at the oldest ages) is eloquent. The *multimodal distribution* of the chain of *relatives* (children, parents, grandparent, great-grandparents), separated by about 28 years (the average interval between generations) and giving thought to the sequence of successive links, results in two major, inter-related observations.[1]

1. In traditional populations (with high fertility and high mortality), the *relatives* (children, parents, and grandparents; brothers and cousins; nephews and uncles), *are numerous and dispersed*, as regards a reference person, in terms of kinship groups and

[1] A good example of this was given by microsimulation, shaping the kinship network around a reference person (Ego) (Le Bras, 1982). For numerical data on the situation in Italy see Barbagli *et al.* (2003:201).

age classes, such that they are distributed with rather *regular frequency across age groups*, although rapidly reduced by mortality. In other words, kinship groups are numerous, especially related to young people. Conversely, in contemporary aged populations (low mortality, which is declining at the highest ages; low fertility, which is sometimes still declining), reproductive experiences are much less frequent and *concentrated* on a few ages, whereas mortality hardly eliminates individuals except at old age. The distribution of relatives is strongly concentrated around *three or four modal age values* and is characterized by successive waves of children, parents, grandparents, and great-grandparents (and relatives from parallel cohorts), spaced by the intervals between generations. The system of relationships within family networks is changing: horizontal linkages (brothers and cousins) are less frequent, vertical linkages (parents and children, grandparents, and nephews) are *more durable*, and in going from older to younger cohorts, there is a diminution in the number of descendants and an increase in the number of ancestors. At the same time, the ways in which children are socialized are changing, notably because of the changes in the structure of families and households (which are becoming simpler). The case of the only child, more and more common, is quite typical: in the absence of siblings, the child must broaden his relationships outside of the household, with cousins (if he has any) and especially with his adult relatives, and the child makes up for the absence of fraternal linkages by more friendly linkages. However, couples and young families find it more difficult to take in grandparents and great-grandparents.

2. In traditional populations, *young cohorts are proportionally much more numerous than those that gave birth to them*. The rules of intergenerational resource transfers, debit–credit relationships, and material and nonmaterial relationships all rest on this large demographic base. First, parents transfer resources to their young children, then these children become the intermediate generation, giving back to their now-elderly parents their filial debt accumulated in the previous stage (when the parents provided everything necessary for their children). *The viability of the filial debt* is assured, from the demographic point of view, by the reduced proportion of elderly compared with the number of adults and by the relatively short duration of life at higher ages and, from the economic point of view, by the rather reduced consumption requirements of the elderly. In current elderly populations, in contrast, *cohorts of parents have become more numerous than those of children*, at least so long as the reduction in annual numbers of births and improvement in survival at the highest ages continue. The changing of demographic relationships between cohorts is aggravated by the financial burden that weighs on the intermediate generations, at an age and condition to work, so as to pay the costs of assistance and retirement of the elderly (living to be older and with much more in the way of consumption needs than those previously, especially with regard to health). This happens especially in the countries where redistribution by the state is important (as is the case in nearly all European countries) (Esping-Andersen, 1999). In such a context, the traditional intergenerational system of the *filial debt becomes impossible* and is *fundamentally called into question*.

In any case, the exchange agreement of the pretransitional demographic regime (alternating between the filial debt accumulated by the young and then given back to the elderly by adults) was equally valuable for all generations in the cyclical evolution of their lives. The current demographic change no longer allows, with a perspective of a few generations, a guarantee to new generations that they will recover, in old age, the credits accumulated in the productive and reproductive phase of their lives, according to the traditional rule. Hence, there is concern about *iniquities regarding intergenerational transfers*. Reform of systems of pension funds, social security, and health—widening the subpopulation of taxpayers including raising the retirement age and drawing on other resources beyond wages and salaries—are among the paths to explore urgently in order to confront the crises heralded by the *conjunction between a regressive demographic dynamic and a new socioeconomic dynamic*.[2]

Throughout the evolution of the rules of intergenerational transfer of resources, it should be desirable to take advantage of a slowing up of the speed with which populations age. The issue is very complex, because it brings into play the cultural bases of collective life. This complexity can, however, be attenuated if it is possible to simultaneously meet the needs for investments in the young (in order to guarantee effective opportunities for education, employment, family autonomy) and support those who take care of children and family. In this way, some of the current obstacles to establishing new conditions for transferring material and nonmaterial resources between generations in the context of the redistributive policies of the welfare state, may be reduced (Demeny, 1987; Sgritta,

[2] We will not dwell here on the specific effects of the current postindustrial economic transformation, which, among other things, tends to reduce creation of employment and break the traditional national control barriers.

1995; Esping-Andersen, 1999; see also Chapter 81 of this Volume).

6. Multiplication of the Life Paths: Segmentation of the Conjugal and Reproductive Processes

Beyond issues related to the evolution of longevity, one can, from a demographic point of view, identify a plot of the changes that were just quickly described and that have been especially apparent during the recent decades, in considering three principal dimensions and their intersections (van de Kaa, 1987, 1988; Roussel, 1993, 1994):

1. The growing importance attributed to *the limiting of fertility* via a nearly complete mastery of the reproductive process (the radical nature of this control of fertility appears clearly in the recourse to induced abortion, which has become largely accessible and widespread in recent decades).
2. The priority given to the links of *the couple* compared with other family relationships such as the child–parent linkage, in a context of growth of women's economic independence. The importance of marital agreement is so great in and of itself that its weakening, even unilaterally, is legitimate grounds for dissolving the couple's union (the massive diffusion of *divorce* over the course of a few decades appears to be one of the most significant events of recent centuries, from a social point of view [Stone, 1994]); in addition, a couple's bonds can increasingly be established (or re-established, after a previous break-up) without *public contractual visibility* (consensual unions).
3. The simplification and the multiplication of *types of households, with or without a family nucleus*, have become highly dependent on individual biographies. From this point of view, the effects are strongly felt according to the way that, concretely, individuals realize and live the two conjugal and reproductive processes that were just alluded to (family adjustments after a divorce with minor children provide some varied examples) (Condran *et al*., 1994; Ni'Bhrolchain *et al*., 1994).

The *diversification of the life paths* in which conjugal and reproductive behaviors come together and evolve is an important phenomenon that also has methodological implications, notably with regard to observation (Roussel, 1980; De Sandre, 1981; Blayo, 1990). Figures 88–4 and 88–5 show possible paths for formation and dissolution of couples, fertility, and family building. In the pretransitional social and demographic regime, a *uniform idea* of marital and reproductive life prevailed, characterized by *biological constraints*: couples had a precise social mandate to have a large number of children to compensate for the high mortality and thereby guarantee both survival of the family and replacement of the population. The institution of marriage was strictly regulated so as to simultaneously ensure both the procreation function of the parents and the social identity of the child. Within the couple there was a hierarchy of asymmetric roles, with a very strong obligation for women to take care of domestic and reproductive tasks. Fertility control sought to space births rather than to prevent them. In recent decades, on the contrary, we've witnessed a *strong dissociation of ideas about marriage and reproduction*. Each dimension—concerning sex, coupling, reproduction—now follows its own path, both in terms of cultural representation and of behaviors, largely *autonomous* in relation to the others.

1. Sexuality, protected from the risk of unwanted conception, can be seen for itself—before, within, or outside of conjugal life.
2. The reproductive process, even when a conception is begun, can be fully controlled also via induced abortion; massive recourse to abortion is the sign, among others, of a cultural reversal *in the representation of the fetus* (someone who voluntarily has an abortion either resolves in this way a *conflict of subjective rights* in favor of the mother or else assumes that the fetus does not *have* any primary subjective *rights*, at least while abortion is allowed). Moreover, in addition to contraception and abortion, new forms of intervention on procreation have begun to appear, such as *assisted fertility* and genetic engineering.
3. The traditional conjugal model, legitimated by a public and legal pact, progressively moved away from its one-way and constraining pattern to give way to a free choice that couples make based on a private and reversible pact foreseeing a permanent search for consensus between the partners. Failure to achieve this consensus can lead to disruption of the union. The frequency of disruption of legal unions has greatly increased, as has already been noted, and may attain more than 50% of marriage cohorts, except in the Mediterranean countries, where this proportion remains clearly much lower.

The *interferences and interdependencies* between the particular types of management of each dimension of the conjugal and reproductive process—*combined in the life paths*—are numerous, complex, and not always easy to explore. Such interdependencies are engendered either *within individual biographies* or *at the intersection between biographies* of people who live together

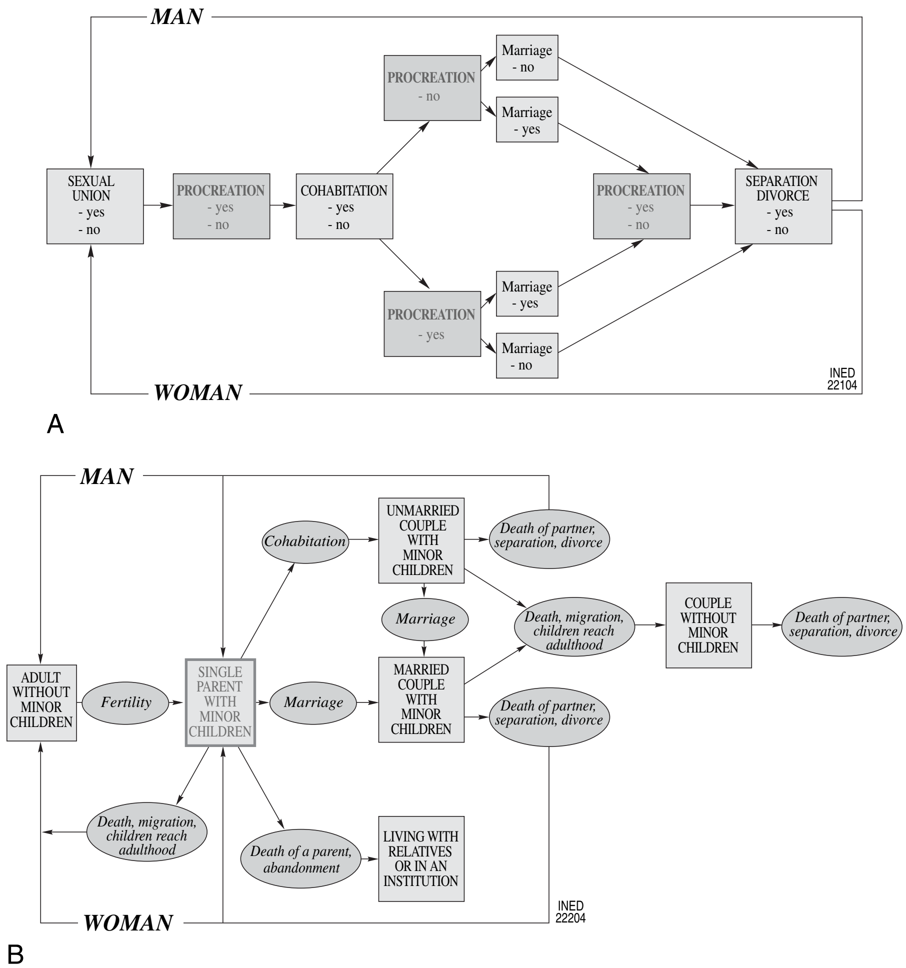


FIGURE 88–4 **(A)** Formation and dissolution of couples: diagram of principal possible paths (the effect of mortality, which can break up couples at any point, is ignored here). **(B)** Formation and dissolution of single-parent families.

during certain portions of their lives (e.g., couples and children). A good example of *interdependence internal to an individual biography* can be given by the adjustments of behaviors linked to a reduction in desired fertility, induced not only by an actual disruption of the couple but also by the simple sentiment of instability of the relationship (Bumpass, 1990). With regard to the *interdependence between biographies*, one example that can be cited concerns the family trajectories of children of separated parents (Hofferth, 1985) or the numerous economic woes of women with children after a divorce (Duncan, 1984; Bianchi and Spain, 1996). For the Western countries, recent data (including longitudinal data) document the deterioration of the well-being of minors involved in the union disruptions of their parents and justify the growing attention given today to the rights of children, even if there remains uncertainty about the long-term effects (Dumon, 1994). New consequences at the individual or collective levels may emerge from the *interdependencies between stages of life* in the difficult transition of the young toward adult status, especially in Mediterranean countries.

All things considered, the *segmentation of the conjugal and reproductive processes* implies a *profound discontinuity* between the institutional rules and the hierarchy that weighed uniformly and globally on couples, marriages, sexuality, conception, and birth (and extended throughout the life cycle) and *the emergence of new rules*. Their elaboration resides in the *sense that one gives* (at every moment) *to each of the dimensions of the process* and in the *protagonists' ability to negotiate* each of them (on the theme of the rights to formulate and safeguard as opposed to the duties to fulfill, and also on the capacity of individual control as opposed to being subject to social control, etc.). All of this evolves in a context of expanding *social tolerance* for different ideas that can lead to previously-stigmatized behaviors.

Such rules, which accompany other, more traditional ones and which don't only follow rational criteria of utilitarian calculation,[3] seem (1) to legitimate the *removal* not only of the material obstacles arising during an individual's life but also the interpersonal agreements (with the emblematical cases of recourse to divorce and to abortion); (2) to give priority to *reversible choices*, circumscribed in time and in their consequences; and also (3) *irreversible choices*, if in themselves they open new spaces of autonomy and opportunity. Whatever the causes of the differentiation of the paths and their rules, the fact is that *the possible paths are multiplying*, under the effect of different combinations of solutions concretely adopted, during their lives, in the domains of sexual experiences, the couple and reproduction.

Thorough investigation of the dynamics of marital life (or better, of life in a couple), reproductive life, family life, and linkages between generations should evidently account for the profound modifications of behaviors and attitudes, which in turn alter their empirical expressions in quantity, quality, and social meaning.

II. CONSEQUENCES OF CHANGES OF THE LIFE PATHS AND POLICY IMPLICATIONS

The structural changes of the components of demographic dynamics and how the population renews itself, the new disequilibria between young and adult generations, are so profound, new, and massively diffused that not only do they influence the persons directly involved, in changing life styles, but they weigh even more on *the evolution of social and economic rules and of the agreement between generations*, *on the basis of which this renewal is itself managed* (notably in terms of care and functions of old-age security and assistance) (Davis *et al*., 1987). In this context of change, the evolution of the rules happens because of *the initiative of individuals* who experience the changes and because of the *collectivity* (market, system of relationships, culture) and the *public sector*. New norms must be set forth regarding individual and social rights that require resources to be effective; the distribution of financial means must be modified, and services need to be reallocated to take account of the precarious economic conditions in which certain segments of the population—as a function of age, sex, family type, or number of children—are found. We must react to the perverse effects of aging of the age structure, for which the barycenter is made fragile by the decreasing share of the labor force in the population; the criteria for actions to equalize individual chances need to be revised. This entails important efforts that need to be innovative and that involve delays and lead to more or less coherent successive adjustments. To a great degree, this entails *a posteriori adjustments*, taking into account demographic modifications; sometimes interventions are explicitly decided on in the goal of *confronting demographic changes*. One must, therefore, evaluate the *economic impact* of certain demographic changes and notably their consequences for public finances (as well as the expected effects of interventions seeking to modify demographic trends by influencing fertility, health, migration).

[3] On the different types of logic of rational behavior, see Micheli (1995).

As a simple example, let us quickly emphasize the most current aspects of policy attention to the changes in this domain, a few interventions and international orientations, notably of the European Union, during the 1990s.[4]

1. The year 1993 was dedicated *to the elderly and to solidarity between generations* (CCE, 1994). In the context of maintaining the European social model (which gives more emphasis than do others to state intervention in favor of equalizing opportunities and addressing the struggle against social exclusion and poverty), it is noted that aging of the population requires improvement in the role and contribution of working retirees, that some form of insurance be provided for long-term care for dependent persons, and that old-age security systems be made more adequate. An *observatory of aging and of the elderly* was thus set up.

2. Since 1989, an observatory, renamed in 1999 the European Observatory on Family Matters, has also dealt, at the community level, with *national family policies*. Pertinent national policies assume a collective engagement (most often not systematic as the European Union would like) for increasing the value attached to families, which are essential social resources although partially compromised by difficult situations. In fact, households and families, in evolv-

[4] For an extension of the context to international policies, see Finkle and McIntosh (2002); Harbison and Robinson (2002); and, with regard specifically to fertility, Ofosu (1995). The subject of public policy seeking to adapt to changes in the life paths or to try and modify the demographic basis of these changes merits special and expanded attention, which is the principal focus of volume VII of the treatise. The stage of socioeconomic development here takes on crucial importance.

ing, have had to take on themselves numerous burdens, especially concerning the consequences of *divorces* (single-parent families, children with problems), *women's disadvantaged situations* (women with children in poverty, women without sufficient resources), and the *precarious positions of persons asking for support and care*. The measures proposed and adopted, in a very diversified manner and with variable intensity (Dumon, 1994), essentially deal with direct support for individuals so that their *family may satisfy its own needs* (e.g., support for the costs of minor children; facilitating women's work, especially for single mothers; forms of reconciling work and fertility; lightening of the burden of care, including health care). Initiatives are often public, but the growing difficulty of managing the welfare state requires solicitation of market services and voluntary associations. Special attention is paid to policy programs that can promote equal opportunity between women and men (CCE, 1994).[5] Interventions seeking to reduce the costs of children for the family and to reconcile work of mothers (and fathers) and care of children become more widespread as the number of births diminishes to well beneath the level of fertility necessary for replacement.

3. The *absolute priority* of the 1995–1997 community social action program was *employment*. In fact, this priority seeks not only to respond to the effects of the competition resulting from globalization of the economy and to the fact that technological innovations tend to reduce employment (reasons why one must *promote employment creation*), but also to facilitate the *access of the young to the labor market* thanks to new *training* programs and the immediate and future (permanent) improvement of the skills of those who work (EC, 1995). It would thus be possible that such a program could counter the current tendency of the young to prolong their stay in the family of origin and to delay their entry into a couple.[6]

[5] See also the specific contributions published on France, Hungary, and Sweden (Golini *et al*., 1991) and the papers presented at the Bologne Colloque on France (Chauviere, 1994), Germany (Ostner, 1994), England (Bradshaw, 1994), the United States (Kamerman, 1994), and Italy (Saraceno, 1994).

[6] To complete the information on the orientations of the European Union and the countries that constitute it, as well as to bring it up to date, it is useful to consult the European Union's Web sites: for society, demography, and family, see *www.europa.eu.int/comm/employment_social/eoss/index-eu.html*; for social security and social integration, *www.europa.eu.int/comm/employment_social/missoc_info_eu.htm*. One can also refer to the *White Book on Youth* (2000) and to the 2000–2006 program of action associated with it, as well as to *Social Policy in Europe* (2000), which gives particular attention to employment (which remains an absolute priority), to the problems of immigration, to equality between men and women, to social security, to the funding of retirees, to social integration, and to health.

III. OBSERVATION AND ANALYSIS OF THE LIFE PATHS (TOWARD AN INTEGRATED APPROACH)

This rapid proposed synthesis on the upsetting of the equilibria among age groups and the diversification of the life itineraries that characterize the replacement of Western generations (adorned by several notes on the interaction among socioeconomic components and on the asked-for policy adjustments) is based on solid and shared evidence. Emerging microdemographic knowledge integrates the macrodemographic background and context. It stems from conceptual research frameworks and modes of observation and analysis that add themselves to the macrodemographic approach (see Chapter 23 of Volume I), with convergent goals. How does one arrive at this knowledge? Can classical demographic methods be developed as a function of this type of observation? Some replies will be given in what follows, with emphasis on three aspects of microdemographic research: conceptual frameworks concerning (1) the individual life paths; (2) the spatial, temporal, social, and historical contexts in which they develop; and (3) appropriate methods of data collection and analysis.

The idea of distinguishing the different ages of life is rather old in the human sciences, whether it involves biomedicine (which extends from neonatology to auxology and to gerontology), psychology (which makes distinctions according to the stages of individual development), anthropology and sociology (which study the norms, roles, and forms of stratification by age), etc. The study of the evolution of generations or of intergenerational relationships has notably been the source of remarkable contributions in biodemography (from the description of mortality and fertility by age to measures of the intrinsic dynamics of the resulting generations) and in sociology (analysis of social mobility in comparing the social positions of parents and children or in following the evolution of their status so as to compare them in successive eras). A new step has, however, been taken recently with *the thorough analysis of biographies*. Since World War II and especially since the 1960s, scientific debates have developed between neighboring disciplines; useful field research has emerged, which has increasingly emphasized the importance of a longitudinal perspective, of analysis by generations or actual cohorts, at first *from aggregated data* and then in going more deeply (since the 1970s) into knowledge and analysis on the *life paths*, notably thanks to use of micro (individual) data with retrospective reconstructions, prospective locating of events, and development of specific methods of analysis (Elder, 1985; Courgeau and Lelièvre, 1989;

Magnusson *et al*., 1990; Mayer and Tuma, 1990; Trussell *et al*., 1992; Bogue *et al*., 1993). The observation of the life paths requires the possibility of *precisely situating events in their context* (environmental characteristics, groups to which people belong and reference groups, institutional and normative, social and economic, systems) to better approach the setting and the behavioral mechanisms (Anderton, 1993; Blalock, 1993; GRAB, 1999).

Some details of a conceptual nature can be useful, before recalling the most pertinent modalities of collection and analysis of biographic data of demographic and social interest.

1. The life paths form *elements of the social structure* produced by individual action, which is influenced by the processes of social and economic organization and by institutional and historical forces. It's not a question of considering them as isolated individual biographies, but as *socially modeled paths* (Mayer and Tuma, 1990): the complete representation of such paths requires the combination of observational and analytical approaches at both the individual and collective (micro and macro) levels.

2. The life paths are *multidimensional*, each one being a combination of complex stories (they are also called life domains, careers, or trajectories). Each of them is defined by a particular dimension of action, by a specific temporal structure, and by a particular sequence of events: history of schooling, professional life, life as part of a couple, reproductive histories (fertility and its control), health, migration and residency, military and civic life. The lesser attention given to certain of these trajectories for practical reasons should not make us forget the reality of the plurality of the set of trajectories, which are certainly interdependent. The *heterogeneity* of itineraries makes observation of the pertinent variables important, and it makes *unobserved heterogeneity* (measurement errors of explanatory variables, omission of key variables or important random period shocks) potentially misleading.

3. The recent diversification and increasing complexity of the life paths resulting from the multiple entanglements of trajectories at the level of individual experiences, leads to *abandoning the terms cycle and career*, which evoke regular and progressive sequences (or at least unidirectional ones) that are repeated from one subject to another or even sometimes for the same individual, whereas the actual dynamic can no longer be described in these terms. More precisely and in a more neutral manner, *trajectories* or paths will be used to describe the different stages of a life or a portion of life, or else *transitions from one state to another as influenced by an event or several competing events*, if only one of these stages is considered (Elder, 1985). From this point of view, it also seems reasonable, for example, to speak of the *family life cycle* (conjugal and reproductive), according to the classical expression of Paul Glick, only for the fraction of the population that is living the stages of family expansion and contraction via the sequence marriage–birth of children–departure of children from the family–widowhood. Table 88–1, using retrospective data from the Canadian censuses of 1971 and 1981, gives an example of the changes that took place in the course of several stages of the life cycle for family units with 15 to 35 years of marriage: stability of the unit, departure of children, couple without children. Overall, it is better to avoid using the word *cycle* for other types of sequences with disappearance, alteration, or conversion from one or several stages of the process (Fig. 88–5).

4. The *dates* at which the *state* conditions of the subject are observed and the *time intervals* during which the *events* that modify them take place can be identified in two ways: by the individual (biographic time) or with reference to the context (historical time). For isolated individuals *age* is a temporal variable of *major significance*, because it is a marker of the evolution of biological characteristics in a given context. But the *durations* spent in a state (e.g., as a spouse, mother of parity two, employed in the labor market) between the event that creates this state and the arrival of the event that will change it, also allow for, at a given age, following reality very closely and aggregating indi-

TABLE 88–1 Average Number of Years Lived by the Family Unit in Each Phase of the Life Cycle, after 15 Years of Marriage and before the 35th Anniversary (Women in a Couple), Canada 1971 and 1981

Family size	Duration of each phase of the cycle						
	Stability		Contraction		Without children (postparental)		1971
	1971	1981 (Δ%)	1971	1981 (Δ%)	1971	1981 (Δ%)	
1	11.1	6.3	—	—	8.9	–7.9	20.0
2	9.4	3.2	5.0	–4.0	5.6	–1.8	20.0
3	8.0	–3.7	8.7	–2.3	3.3	15.2	20.0
4	7.2	–12.5	10.6	7.5	2.2	4.6	20.0
5	6.2	–24.2	12.5	8.0	1.3	58.0	20.0
6+	7.9	0.0	8.3	2.4	3.8	–5.3	20.0

Δ difference, in percentages, in years lived as of the 1981 census, in relation to the years lived as of the 1971 census.

From Pêron *et al*., 1986: table 3.

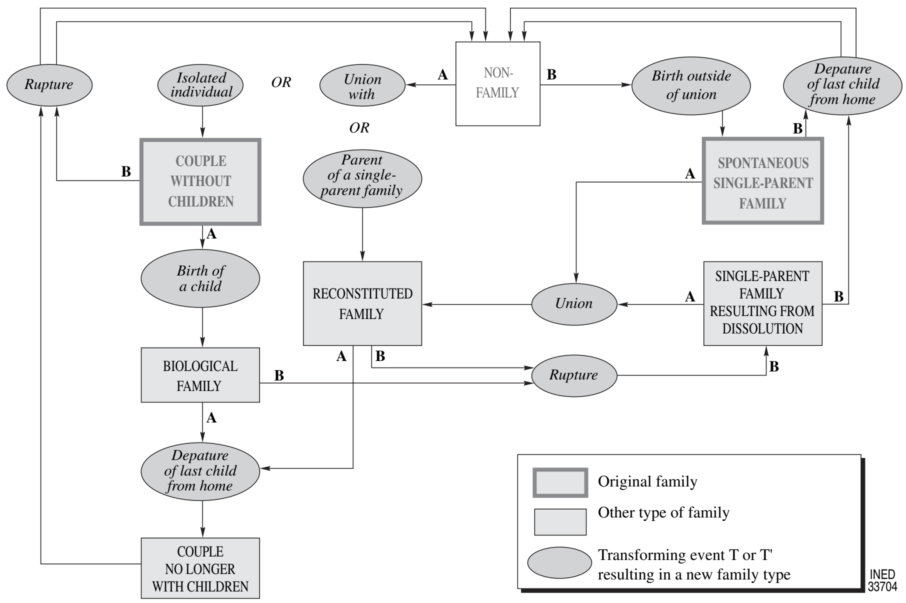


FIGURE 88–5 Family types and transforming events that constitute a family type. After a separation, the two partners in a family can form two single-parent families if each of the parents has at least one child, or one single-parent family and one nonfamily if only one of the parents is taking care of the children. To avoid overcrowding the diagram, the path of the parent going to the nonfamily situation is not shown. (From Blayo, 1990.)

vidual biographies that explain in part the heterogeneity of the population. *Historical*, *chronological time*, on the contrary, *both regarding individuals and the context*, has a *conventional meaning of proximate variable,* which incites us to study the underlying factors that are presumed to influence the life paths. Thus, attention *to the year during which a cohort of individuals is formed* (who all experience the same original event) makes sense not only for individuals but also for groups of individuals *exposed*, *at the same time*, *to the same environmental*, *economic*, *social*, *and institutional factors* that can give them a certain homogeneity at the outset and that may influence them for the rest of their lives. Moreover, individuals can be exposed to *specific* conditions that are important from a demographic point of view: as much as possible, it would be good to be explicit both about the general conditions (collective contextual factors) and about the specific conditions. In the same way, the *year in which an event takes place* can indicate propitious or unfavorable personal or contextual conditions, which should be measured by using pertinent indicators. To identify, with a goal of collecting data, the *factors underlying the stages of life* probably linked to the behaviors being studied is a complex exercise, open to debate and in continuous experimentation, which *results in the updating of personal characteristics* in the course of biographical time and *the construction of individual*, *environmental or social contextual variables*, which can be situated either all along individual biographies or as a reference to given periods. For example, at the *individual* level, asking interviewees the community or the type of family to which they belonged up to the end of adolescence seeks to capture contextual elements, important for explanatory purposes, during the phase when subjects were being educated. At the *collective* level, the study of events that affect individuals during a given year can usefully benefit from pertinent associated indicators, taken also from external sources, peculiar to the same period, dealing, for example, with the economic or labor market situation or with the political context.

5. In thus considering the *multidimensional* and *multitemporal* character of life trajectories, it is apparent that they require a *multilevel* frame of reference (Goldstein, 1995; see Chapter 24 of Volume I). It is necessary, therefore, in principle, to locate the data deemed to influence the variables that are the subjects of the study at multiple observational levels, not simply at the individual level, and also at the level of contextual units that may interact (family, social, normative, economic, territorial systems), which are themselves defined *in time and in space* in a coherent and pertinent manner.

1. Age-Period-Cohort and Beyond

The position acknowledged here implicitly takes part in the polemic on the supposed theoretical advantage of giving priority to cross-sectional (period) analysis compared with longitudinal analysis (by cohort) (Ni'Brolchain, 1992); the alternative here seems to be ambiguous. There is no doubt that the importance attributed by Norman Ryder (going back to the mid-1950s) to generations (cohorts) refers to the moment at which a group of subjects experienced an original event (and in the process refers to the interactive contextual conditions to which the members of the generation or cohort were exposed at that point) and focuses on the impact, generally specific and distinctive, of the historical vicissitudes on the life paths of different cohorts. The French demographic school was created and developed, under the direction of Louis Henry and Roland Pressat (INED, 1995), in harmony with this thesis. They emphasized that longitudinal description of sequential demographic behaviors (and other types of behaviors) constitutes the most *direct and natural* approach (i.e., closer to the unfolding of the life paths of individuals), whereas cross-sectional descriptions (across fictional cohorts) are purely conventional and lead to synthetic measures that are generally biased. But we do not see how this type of proposition, *from a general principle imposed on observation and analysis*, could be opposed to the empirical observation that, especially in periods of rapid behavioral changes, the variability of behaviors themselves is more attributable to a period effect than to a cohort effect. Nor is there opposition to the definition of current measures that draw a greater part from the *duration* in a state than from the *generation* (Rallu and Toulemon, 1993a,b). Such an observation should, if necessary, lead us to simultaneously and distinctly take account of the effects on behaviors both of the *generation or cohort* as well as the *period* in which the behaviors are expressed, rather than to neglect cohort measures, the importance of which can be felt in other temporal variations.

In this sense multivariate age-period-cohort (APC) analysis (Hobcraft *et al*., 1982; see also Chapter 18 of Volume I) of fertility and mortality (measured by year of observation and by age) seeks to distinguish, other than the (fixed) effect of the age structure of the phenomenon (strongly conditioned by biological factors), the (fixed) effect of the generation to which one belongs (variable across cohorts, but unaffected by period factors) and period effects. If the result was a nearly total absence of any period effects, this would mean that we are looking at cohorts for which the future life path is entirely predetermined from the outset. Such a drastic condition hardly seems reasonable. It is more realistic to think that a particular weight is exercised by the original context (there is more of a difference as the distance between successive generations is greater, because of the intervening historical changes), all the more so by *other significant moments* lived by the generations before they experience the events happening during the period of observation, which is equivalent to assuming that onto the specific history of generations is grafted *the influence of time throughout the life path and within the period during which the events being studied occur*. It is not surprising that APC analysis of the recent (1958–1982) variations in fertility by age in Italy show a period effect more distinct than the cohort effect at the time when first the *baby boom* and then the *baby bust* took place (the generations concerned being those born between 1918 and 1967), but in the same APC analysis *extended to other variables* (Martinengo, 1994), the result of the *differential effect of the education received in adolescence and the effect of territorial distribution* are still more distinct. This shows the importance of the influence of inequality of opportunity for access to education that they experienced in their youth, beyond that of the context in which they (in large part) lived the events studied. In the final analysis, it seems important not so much to reduce the time points taken into consideration but rather to *enlarge the scope* of investigation.

In this sense the question of the possible disappearance, owing to obsolescence, of simultaneous APC analyses (Mayer and Huinink, 1990) suggests some distinctions. The APC combination is a basic crossing of temporal references, for which the APC analytical techniques seek to discern the respective weights in exploiting macrodemographic data. This minimal combination may extend to other significant temporal variables that, if one has pertinent microdemographic data, can be exploited (with enhanced efficacity) via multidimensional analytical models of individual biographic data (see also Chapter 18 of Volume I).

2. Individual Trajectories (Life Course) and Family Trajectories (Family Cycle)

The demographic and social study of the evolution of the life paths thus requires observations as extended as possible in time and taking the context into account. Demographic events almost always bring into play the household and family contexts (Bongaarts *et al*., 1987). If *individuals* are the protagonists of behaviors, *families* (or households) appear themselves as the original context of demographic renewal, and it is reasonable to think that, via the systematic interpersonal relationships that characterize them, they influence the decision-making processes underlying individual conjugal and reproductive trajectories. But the development of data collection and analysis regarding individual and family units must follow different criteria. Is it possible, and how, to link the two approaches?

The tendency to reconstruct life paths cannot not *give priority to individuals*, *from the point of view of observation and of analysis*, although they will be gathered from the interactions with the higher-level pertinent social unit. In this sense the *family characteristics are used as an attribute* in the course of the personal life (e.g., in describing the types and characteristics of the household and the *family* of an individual: the family into which he is born, the family he left for the first time as a youth, the new family that he formed or in which he entered when he left his family of origin, the family in which he was living at the time of his divorce or after the divorce of his partner, the family in which he lives, or the family that he anticipates forming). It is also possible to seek to collect elements relative to the specific unit of the *couple* in which the individual may be a part. The *events* that change the configuration of families are thus considered as starting points of *transition* for individuals. One must acknowledge that, in this manner, a representation at multiple levels is supplied but focused on the individual himself, whereas the life paths of the family as a collective unit could be more appropriate to explain the specific individual paths. Nevertheless, the attempt to collect all of the important family changes assumes that complex distinctions have to be made among a multiplicity of nontraditional family paths, linked in particular to interferences between reproductive dimensions, on the one hand, and recent union types and union disruption of new couples, on the other hand,[7] but also to coresidential family passages (evolution of the household as such, as a higher-order unit compared with the component family nuclei).

Moreover, to simultaneously follow the entanglement of individual biographies that intersect with one another throughout the family paths (as unit of study), even if they can be reconfigured as operations of maximum adherence to the actual processes, brings with it specific data collection problems. And for the analysis, problems of *coherence between interdependent individual itineraries* (complicating the classical problems to reconcile, e.g., measures of male and female fertility) and of *identification in time of family units* (as sets of several subjects) must be resolved.[8] Identification of family units in time entails an arbitrary definition of the *rules of continuity*, of the type elaborated by the Survey of Income and Program Participation in the United States (McMillen and Herriot, 1985; De Sandre, 1986). The transition of individuals from one family type to another and, more infrequently, the transition of families from one type to another can usefully be studied with survival models, relying on different explanatory variables (White and Tsui, 1984). However, family typologies and the rules of continuity, even if only a little different, lead to substantially different results that are not comparable for the same situation and are difficult to reduce to the usual cross-sectional typologies.

One can quite rightly give priority to individuals, couples, or families (taken as units of observation and analysis) in research objectives, in either an alternative or integrated fashion. *Today it seems simpler to investigate the individual paths* by associating with them, as attributes (even to define a state or its variations), characteristics or events that take place at the collective level. The problems left hanging for the dynamic study of *units of two people or more* are constraining (measuring collective characteristics and interactions, identifying transitions and of competing risks, evaluating the selection of families that are outside of the field of observation, controlling heterogeneity, decomposing the effects linked to the stage of family life and to time, etc.), but that should not prevent us from seeking to disentangle them.

IV. DYNAMIC INFORMATION ON BIOGRAPHICAL STATES AND LIFE TRANSITIONS

The manner in which data collection is organized strongly conditions the possibilities for analysis of the life paths. This is why, quite fortunately, ad hoc surveys are organized. These surveys can provide information as appropriate as possible for the analytical methods. However, in certain cases, new analytical possibilities have been suggested, allowing for the use of information not specifically collected for the study of demographic and social biographies.

Exhaustive census or vital statistics data can be used to analyze partial biographies. For example, vital statistics data are often tabulated as a function of the

[7] Charlotte Höhn (1987), for example, considers it useful to take account of 12 to 40 types of paths.

[8] Nico Keilman and Nathan Keyfitz (1988), as well as Michael Murphy (1995), speak of dynamic definitions of families as elusive concepts.

cohort and the duration between consecutive events (marriage and birth) so that historical series by cohort can be constructed. However, the use of explanatory variables is then problematic, owing to the usual lack of annual information on the population according to exposure to the risk of experiencing the event being studied. The use of census microdata can allow for reconstruction of the births of women with minor children (who have not yet left the family of origin) in associating other characteristics of the woman or of the family pattern at the time of the survey. It is likewise possible to link individual (and family) microdata from successive censuses or census microdata and current birth data and data on subsequent deaths.

The most pertinent data for the study of biographical transitions and life trajectories come, however, from (1) retrospective or (2) prospective sample surveys of individuals or households (with or without a nucleus), which can be linked up to information from other sources. Prospective surveys include *panels* of representative samples of the population and of samples of birth cohorts. A well-known example of a panel of this type is the Panel Study on Income Dynamics, which covers 5000 to 6000 American families that have been followed and integrated since 1968 (Duncan, 1984; Duncan *et al*., 1994). An example of a sample by generation is the study that covers people born in 1958 in England, interviewed at 1-year intervals throughout their lives (Fox and Fogelman, 1990). It should be noted that even in these prospective surveys, interviews in each round include retrospective questions, but they are limited to the interval since the previous round. Although I refer the reader to other works for more detailed comments on the specifics of different types of data collection (Murphy, 1995; GRAB, 1999; see also, in Volume IV, the chapter by Dominique Tabutin (Chapter 121) on systems of observation), and with due emphasis on the care required regarding the sample frame when life histories in terms of transitions between states are being studied (Hoem, 1985), I wish to emphasize two points here.

In a number of Western countries during the 1980s, family *panels* were established in order to study the demographic and socioeconomic dynamics via follow-up observations at brief intervals (at least in Switzerland, France, Germany, England, Luxembourg, and The Netherlands). In some cases (Survey of Income and Program Participation in the United States, since 1983), the repeated data collection on the same individuals is limited to a small number of years so as to reduce one of the major drawbacks of panels (and even more of samples by generation)—the reduction of the sample (attrition), which over time biases the sample via a selection effect.

Retrospective cross-sectional sample surveys, which may be periodic but with independent samples, are, however, the most used. One of their principal limitations, the drawbacks of reconstructed data stemming from errors in memory (Auriat, 1996), does not seem to be a very pertinent objection for infrequent and very important events (such as marriage, divorce, or birth) for which there are means of verification, but they do present real problems for more frequent behaviors that are difficult to verify, especially for opinions and attitudes. The system of surveys that was the most profoundly innovative regarding retrospective data collection and that strongly stimulated applied analyses is the World Fertility Survey, carried out in more than 60 countries during the 1970s and 1980s (Cleland and Scott, 1987). Among the more recent surveys that constitute a new and important effort at retrospective description of different dimensions and trajectories is the National Survey of Families and Households of the University of Wisconsin (1988) (Sweet *et al*., 1989) and the Fertility and Family Surveys under the auspices of the Economic Commission for Europe of the United Nations (a survey done in the Economic Commission for Europe countries during the 1990s; see UNECE, 1996–2001). This latter set of surveys provides information, at the individual level, on retrospective characteristics pertaining to (1) the parents of the interviewee; (2) the parents of the interviewee's partner, (3) the places of birth and residence up to adolescence; (4) the first departure from the family of origin, in relation to the history of events with systematic dating; (5) schooling; (6) unions of couples and possible disruptions; (7) pregnancies; (8) contraception and induced abortion; (9) births and adoptions; (10) work; and (11) migration (UN-ECE, 1992). One of the major objectives of these surveys was, clearly, to reconstruct interrelated life trajectories, surveyed in much greater detail,[9] to evaluate their effect on marital and reproductive behaviors. Reference to variables other than at the individual level (couple, own family, family of origin, places lived as a youth) is only provided within the individual questionnaires, and not from collective variables measured in a comparable way in their context.

[9] The quality of the responses, above all regarding the exhaustiveness and the dating of events, seems to be satisfactory. With regard to the debatable international comparability of the surveys of the Fertility and Family Surveys project, see Festy and Prioux (2002). Concerning a new, longitudinal project, *Generations and Gender Programme*, see UNECE-UNPF (2000).

V. RELATIONSHIPS BETWEEN TRAJECTORIES AND DEPENDENT VARIABLES, MULTI-STATE TRANSITION TABLES AND HAZARD MODELS

There are two different analytical strategies that are used in the study of trajectories of events that follow one another in time (formation of a couple, contraceptive behavior, pregnancies, births, employment, etc.). One approach is to measure the *incidence of each nonrepeatable event* within a trajectory (unions, pregnancies, births, etc., *according to the order* of appearance) as a function of the *population at risk* (e.g., second-order births in relation to first-order births), in assuming, in this case, that the exposure to risk continues for someone who has experienced the initial event (first-order births, which defines the initial state of reference) even when he has experienced the event being studied. Appropriate analytical techniques are based on empirical probabilities and progressive rates, and they may partially alleviate the reduced attention to the temporal dynamic of the process in specifying it for periods before the state of reference (state dependence).

The other approach is to measure the *transition between one state and the following one* by virtue of *a nonrepeatable event* (or of several competing events), while taking account of the *duration between the initial state and the event* experienced. In this case it is assumed that the exposure to risk changes after the event being studied occurs. The appropriate analytical techniques are *survival functions by duration in the state* (life tables) and hazard or waiting-time models (Hobcraft and Murphy, 1986), with the useful specifications of the space of the states being studied and of the transitions between states.

The pertinent quantitative analytical models seem to have developed in a parallel fashion, only recently coming to a substantial integration: (1) one approach consists of analytical models of *relationships among variables* (linear models with latent variables based on the analysis of the covariance structure, linear structural relationships, and nonlinear extensions) applied to longitudinal data with a continuous metric; (2) another approach comprises the extensions of the classic demographic *survival functions* (multistate processes with multiple decrements, i.e., with several competing causes of exit or transition, and with multiple increment-decrements, i.e., with the possibility to increase the population at risk in addition to decrease it) applied especially to aggregated data; and (3) a third approach, based on the definition of the individual biography as a stochastic process with discrete states and continuous time, combines *survival models with regression analysis* (*hazard* models or multivariate survival analyses) and applies them to sample microdata. The last direction of research is destined to profoundly integrate the second. In particular, the exploitation of individual biographic data facilitates remarkable progress in linking up longitudinal and cross-sectional approaches (Palloni and De Sandre, Chapter 135): for example, by means of the application of multistate survival models to fertility by age-parity-duration and by other characteristics or in trying to separate out the effects of timing and intensity in synthetic period measures. Important methodological problems have been the subject of research into the riddle of what hypotheses may provide solutions (notably, measuring unobserved heterogeneity and evaluating the fitting of analytical models to biographies); the diffusion of technical applications, very modest at the outset (Trussell, 1992), is promising (Billari, 2000; see also Chapter 23 of Volume I). Innovative applications to complex real data, facilitated by the availability of biographic microdata and by the existence of manuals giving the methodology and providing directions for computer analysis (Courgeau and Lelièvre, 1989; Blossfeld and Rohwer, 1995; Lelièvre and Bringé, 1998), will enrich these techniques even more, techniques that have the merit of encouraging explicit dealing with a number of *sensitive problems posed both by the sample data* (exit of observations; selection of persons by the effect of the cross-sectional sampling frame; selection of the processes studied by the [right-]censoring of the life paths at the time of the survey; selection owing to sample attrition in panels; effect of omitted variables and omitted or erroneous responses) *and by the longitudinal approach* by transition durations between states, especially sensitive to biasing effects of an imperfect knowledge of the processes (Hobcraft and Murphy, 1986; Palloni, 1993; Blossfeld and Rohwer, 1995).

In a way, the rigor of the analytical procedures requires exercising a new vigilance even on the kinds of observations of behaviors, encouraging ultimately a *more prudent reading* of the underlying mechanisms of sociodemographic changes and *a more established and a more profound reading*.

CONCLUSION

The study of human behaviors, to progress in the analysis of causal mechanisms and long-term trends, should necessarily be based on multitemporal approaches (life paths and historical sequences), multidimensional approaches (taking into account multiple characteristics and factors and taking advantage of interdisciplinary synergies), and multilevel approaches (the results studied at the aggregate level depend on the integration of individuals into significant contexts of primary human relationships, of social, economic, legal, and geographic ties).

Demographic research has a rich and solid experience of studying vital processes based on aggregated data on actual cohorts (longitudinal analysis) or fictional cohorts (cross-sectional analysis). More recent is the direct utilization of individual biographic microdata. Reconciling the two research paths and making them converge represents a rather good investment, one that will take a long time. To link up demographic dynamics with their underlying factors, they must be studied by embracing the overlaps of the principal life trajectories of the individuals who produce them, reconstructing the significant contexts of human relationships (couples, families, communities to which individuals belong, and a reference system) at the level at which they develop. The field is open, especially today, to new contributions at all levels of research: conceptual revision of research projects, renovation and improvement of the ways of collecting information (and of using existing data), procedures and analytical models that are well-suited and pertinent, and updating of empirical reconstructions in terms of both time and space.

Acknowledgment

This chapter was translated from the French version by David Shapiro.

References

Anderton Douglas L., 1993. Contextual and multi-level effects, *in*: Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton (eds.), *Readings in population research methodology, vol. 6: advanced basic tools*, p. 31–41. Chicago, UNFPA, SDC, 414 p.

L'Association Internationale des Dêmographes de Langue Francaise [AIDELF] (ed.), 2002. *Vivre plus longtemps, avoir moins d'enfants, quelles implications?* Paris, PUF, 737 p.

Auriat Nadia, 1996. *Les défaillances de la mémoire humaine*. Paris, INED and Puf, 203 p.

Barbagli Marzio, Castiglioni Maria and Dalla Zuanna Gianpiero, 2003. *Fare famiglia in Italia. Un secolo di cambiamenti*. Bologna, il Mulino, 308 p.

Bianchi Suzanne and Spain Daphne, 1996. Women, work and family in America, *Population Bulletin*, vol. 51(3), p. 48.

Billari Francesco, 2000. *L'analisi delle biografie e la transizione allo stato adulto: aspetti metodologici e applicazioni ai dati della Seconda Indagine sulla Fecondità in Italia*. Padova, Cleup, 308 p.

Blalock Hubert M., 1993. Cross level analysis, *in*: Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton (eds.), *Readings in population research methodology, vol. 6: advanced basic tools*, p. 3–21. Chicago, UNFPA, SDC, 414 p.

Blayo Chantal, 1990. De l'application des principes d'analyse démographique à l'étude de l'évolution des familles, *Population*, vol. 45(1), p. 63–86

Blossfeld Hans-Peter and Rohwer Götz, 1995. *Techniques of event history modeling: new approaches to causal analysis*. Mahwah, Lawrence Erlbaum, 294 p.

Bogue Donald J., Arriaga Eduardo E. and Anderton Douglas L. (eds.), 1993. *Readings in population research methodology. vol. 6: advanced basic tools.* Chicago, UNFPA, SDC, 414 p.

Bongaarts John, 1982. The fertility inhibiting effects of the intermediate fertility variables, *Studies in Family Planning*, vol. 13(617), p. 179–189.

Bongaarts John, Burch Thomas and Wachter Kenneth (eds.), 1987. *Family demography: methods and their applications.* Oxford, Clarendon Press, 365 p.

Bozon Michel and Villeneuve-Gokalp Catherine, 1994. Les enjeux des relations entre générations à la fin de l'adolescence, *in*: L'enfant dans la famille, vingt ans de changement, *Population*, vol. 49(6), p. 1527–1556.

Bradshaw Jonathan, 1994. *Family policy in the UK.* Bologna, Comune di Bologna, 27 p.

Bumpass Larry, 1990. What's happening in the family? Interactions between demographic and institutional change, *Demography*, vol. 27(4), p. 483–490.

Caselli Graziella and Egidi Viviana, 1993. *New frontiers in survival: the length and quality of life*. Rome, DSD, 36 p.

Commission des Communautés européennes [CCE], 1994. *La politica sociale europea.* Luxembourg, CCE, DG5, 25 p.

Chasteland Jean-Claude and Chesnais Jean-Claude (eds.), 1997. *La population du monde: enjeux et problèmes.* Paris, PUF, INED, 638 p.

Chauvière Michel, 1994. *Jalons pour une socio-histoire de l'action publique dans le champ familial en France*. Bologna, Comune di Bologna, 17 p.

Cheal David, 1983. Intergenerational family transfers, *Journal of Marriage and the Family*, vol. 45, p. 805–813.

Chesnais Jean-Claude, 1986. *La transition démographique: étapes, formes, implications économiques.* Paris, INED and PUF, 582 p.

Cleland John and Scott Christopher (eds.), 1987. *The World Fertility Survey: an assessment.* Oxford, Oxford University Press, 1058 p.

Coale Ansley J., 1987. Demographic effects of below-replacement fertility and their social implications, *in*: Kingsley Davis, Mikhail S. Bernstam, and Rita Ricardo-Campbell (eds.), *Below replacement fertility in industrial societies: causes, consequences, policies*, p. 203–216. New York, The Population Council, 360 p.

Coale Ansley J. and Watkins Susan Cotts (eds.), 1986. *The decline of fertility in Europe.* Princeton, Princeton University Press, 484 p.

Colvez Alain and Robine Jean-Marie, 1986. L'espérance de vie sans incapacité en France en 1982, *Population*, vol. 41(6), p. 1025–1042.

Condran Gretchen and Furstenberg Frank, 1994. Évolution du bien-être des enfants et transformation de la famille américaine: l'enfant dans la famille, vingt ans de changement, *Population*, vol. 49(6), p. 1613–1638.

Corijn Martine and Klijzing Eric (eds.), 2001. *Transition to adulthood in Europe.* Dordrecht, Kluwer and Academic Publisher, 340 p.

Courgeau Daniel and Lelièvre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p.

Crimmins Eileen M., 1993. Demography: the past 30 years, the present, and the future, *Demography*, vol. 30(4), p. 579–591.

Davis Kinsley, Bernstam Mikhail S. and Ricardo-Campbell Rita (eds.), 1987. Below replacement fertility in industrial societies: causes, consequences, policies. *Population and Development Review*, vol. 12, 360 p.

Demeny Paul, 1987. Pronatalist policies in low-fertility countries: patterns, performance and prospects, *in*: Kingsley Davis, Mikhail S. Bernstam, and Rita Ricardo-Campbell (eds.), *Below replacement fertility in industrial societies: causes, consequences, policies*, p. 335–358. New York, The Population Council, 360 p.

De Sandre Paolo, 1981. Transformation des modèles de formation et de dissolution de la famille dans les pays industrialisés et conséquences démographiques, *in*: IUSSP (ed.) *Congrès International de la population, Manille, 1981*, vol. 4, p. 127–139. Liège, IUSSP, 377 p.

De Sandre paolo, 1986. Esperienze recenti ed esigenze conoscitive in tema di famiglie, in Istat, *Annali di Statistica*, vol. 9(6), p. 47–69.

Dumon Wilfred (ed.), 1994. *Changing family policies in the member States of the European Union.* Brussels, Commission of the European Communities DG-V, 238 p.

Duncan Greg J., 1984. *Years of poverty, years of plenty: the changing economic fortunes of American workers and families.* Ann Arbor, The University of Michigan, 184 p.

Duncan Greg, Yeung Jean W. and Rodgers Willard, 1994. Les familles monoparentales aux États-Unis: dynamique, niveau de vie et conséquences sur le développement de l'enfant: l'enfant dans la famille, vingt ans de changement. *Population*, vol. 49(6), p. 1419–1436.

Elder Glen H. (ed.), 1985. *Life course dynamics: trajectories and transitions, 1968–1980.* Ithaca, Cornell University Press, 345 p.

Esping-Andersen Gösta, 1999. *Social foundations of post-industrial economies.* Oxford, Oxford University Press, 334 p.

European Commission [EC], 1995. *Medium term social action programme, 1995–97: social Europe*, vol. 1. Brussels, EC, 61 p.

Festy Patrick and Prioux France, 2002. *An evaluation of the Fertility and Family Surveys project.* New York, UNECE-UNPF and INED, 55 p.

Finkle Jason L. and McIntosh Alison C., 2002. United Nations population conferences: shaping the policy agenda for the twenty-first century, *Studies in Family Planning*, vol. 33(1), p. 11–23.

Fox A. John and Fogelman Ken, 1990. New possibilities for longitudinal studies of intergenerational factors in child health and development, *in*: David Magnusson and Lars R. Bergman (eds.), *Data quality in longitudinal research*, p. 233–248. Cambridge, Cambridge University Press, 285 p.

Galland Olivier, 1984. Précarité et entrées dans la vie, *Revue française de sociologie*, vol. 25(1), p. 49–66

Goldstein Harvey, 1995. *Multilevel statistical models.* London, E. Arnold, 178 p.

Golini Antonio *et al*., 1991. *Famiglia e società in Europa. Crisi della natalità e politiche per la popolazione.* Turin, Fondazione G. Agnelli, 362 p.

GRAB (Groupe de réflexion sur l'approche biographique), 1999. *Biographies d'enquêtes: bilan de 14 collectes biographiques*. Paris, INED, PUF, 340 p.

Halli Shiva S. and Rao Vaninadha, 1992. *Advanced techniques of population analysis.* New York, Plenum Press, 226 p.

Harbison Sarah F. and Robinson Warren C., 2002. Policy implications of the next world demographic transition, *Studies in Family Planning*, vol. 33(1), p. 37–48.

Hobcraft John, Menken Jane and Preston Samuel, 1982. Age, period and cohort effects in demography: a review, *Population Index*, vol. 48(1), p. 4–43.

Hobcraft John and Murphy Michael, 1986. Demographic event history analysis: a selective review, *Population Index*, vol. 52(1), p. 3–27.

Hoem Jan, 1985. Weighting, misclassification and other issues in the analysis of survey samples of life histories, *in*: James Heckman and Burton Singer (eds.), *Longitudinal analysis of labour market data*, p. 249–293. Cambridge, Cambridge University Press, 425 p.

Hofferth Sandra L., 1985. Children's life course: family structure and living arrangements in cohort perspective, *in*: Glen H. Elder Jr. (ed.), *Life course dynamics: trajectories and transitions, 1968–1980*, p. 75–112. Ithaca, Cornell University Press, 345 p.

Höhn Charlotte, 1987. The family life cycle: needed extension of the concept, *in*: John Bongaarts, Thomas Burch, and Kenneth Watcher (eds.), *Family demography: methods and their applications*, p. 65–80. Oxford, Clarendon Press, p. 365.

Institut National d'Etudes Demographiques [INED], 1982. *Les âges de la vie. Actes du Colloque: tome 1.* Paris, INED and PUF, 212 p.

Institut National d'Etudes Demographiques [INED], 1983. *Les âges de la vie. Actes du Colloque: tome 2*. Paris, INED and PUF, 176 p.

Institut National d'Etudes Demographiques [INED], 1994. L'enfant dans la famille. Vingt ans de changements, *Population*, vol. 49(6), p. 1245–1636.

Institut National d'Etudes Demographiques [INED], 1995. Cinquante années de *Population*, *Population*, vol. 50(6), p. 1335–1726.

Kamerman Sheila B., 1994. *Family policy in the United States.* Bologne, Comune di Bologna, 20 p.

Keilman Nico and Keyfitz Nathan, 1988. Recurrent issues in dynamic household modelling, *in*: Nico Keilman, Antonie Kuijsten, and Ad Vossen (eds.), *Modelling household formation and dissolution.* Oxford, Clarendon Press, 298 p.

Laslett Peter, 1989. *A fresh map of life: the emergence of the third age.* London, Weidenfeld & Nicolson, 213 p.

Le Bras Hervé, 1982. Évolution des liens de famille au cours de l'existence, *in*: INED (ed.), *Les âges de la vie: actes du VIIe col-loque national de démographie, tome 1*, p. 27–45. Paris, PUF, 212 p.

Lelievre Eva and Bringé Arnaud, 1998. *Manuel pratique pour l'analyse statistique des biographies. presentation des modèles de durée et utilisation des logiciels SAS, TDA, STATA.* Paris, INED, 189 p.

Livi-Bacci Massimo, 1978. Changement démographique et cycle de vie des femmes, *in*: Éveline Sullerot (ed.), *Le fait féminin*, p. 443–457. Paris, Arthème Fayard, 499 p.

Magnusson David and Bergman Lars R. (eds.), 1990. *Data quality in longitudinal research.* Cambridge, Cambridge University Press, 285 p.

Martinengo Anna, 1994. *Analisi congiunturale della fecondità in Italia: criteri e applicazioni.* Padoue, Facoltà di Scienze Statistiche, Università di Padova.

Mayer Karl U. and Huinink Johannes, 1990. Age, period and cohort in the study of the life course: a comparison of classical APC analysis with event history analysis or farewell to Lexis? *in*: David Magnusson and Lars R. Bergman (eds.), 1990. *Data quality in longitudinal research*, p. 211–232. Cambridge University Press, 285 p.

Mayer Karl U. and Tuma Nancy B. (eds.), 1990. *Event history analysis in life course research.* Madison, University of Wisconsin Press, 297 p.

Mcmillen David B. and Herriot Roger, 1985. Toward a longitudinal definition of households. *Journal of Economic and Social Measurement*, vol. 13(3–4), p. 349–360.

Micheli Giuseppe A. (ed.), 1995. *La società del figlio assente. Voci a confronto sulla seconda transizione demografica in Italia.* Milan, Angeli, 251 p.

Murphy Michael, 1995. The collection and comparability of demographic and social data in Europe: the family life cycle, *in*: Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et com-*

parabilité des données démographiques et sociales en Europe, p. 149–182. Louvain-la-Neuve, Academia and L'Harmattan, 672 p.

Ni'Bhrolchain Maire, 1992. Period paramount? A critique of the cohort approach to fertility, *Population and Development Review*, vol. 18(4), p. 599–629.

Ni'Bhrolchain Maire, Chappell Roma and Diamond Ian, 1994. Scolarité et autres caractéristiques socio-démographiques des enfants de mariages rompus: l'enfant dans la famille, vingt ans de changement. *Population*, vol. 49(6), p. 1585–1612.

Ofosu Dikson Y., 1995. Politiques et transitions de fécondité dans le Tiers Monde: des échecs, des réussites, de l'incertain, *in*: Dominique Tabutin, Thierry Eggerickx, and Catherine Gourbin (eds.), *Transitions démographiques et sociétés*, p. 603–638. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 692 p.

Ostner Ilona, 1994. *Family policy and family in post-unification Germany.* Bologna, Comuni di Bologna, 17 p.

Palloni Alberto, 1993. Hazard models and their applications to the study of demographic phenomena, *in*: Alberto Palloni, Massimo Attanasio, Maria Castiglioni, Alberto Cazzola, Alessandra De Rose, Fausta Ongaro, Goetz Rohwer, and Nicola Torelli (eds.), *Applicazioni di modelli di sopravvivenza nella ricerca demografico-sociale*, p. 7–12. Padova, Cleup, 102 p.

Peron Yves, Lapierre-Adamcyk Evelyne and Morisette Denis, 1986. Départ des enfants et contraction des familles d'après les recensements canadiens de 1971 et de 1981, *European Journal of population/Revue européenne de démographie*, vol. 2(2), p. 155–175.

Rainwater Lee and Smeeding Timothy M., 1994. Le bien-être économique des enfants européens: une perspective comparative, *in*: L'enfant dans la famille, vingt ans de changement, *Population*, vol. 49(6), p. 1437–1450.

Rallu Jean-Louis and Toulemon Laurent, 1993a. Les mesures de la fécondité transversale, II: construction des différents indices, *Population*, vol. 48(1), p. 7–26.

Rallu Jean-Louis and Toulemon Laurent, 1993b. Les mesures de la fécondité transversale, I: application à la France de 1946 à 1989, *Population*, vol. 48(2), p. 369–404.

Roussel Louis, 1980. Mariages et divorces: contribution à une analyse systématique des modèles matrimoniaux, *Population*, vol. 35(6), p. 1025–1040.

Roussel Louis, 1993. Fécondité et famille, *in*: *Conférence européenne sur la population*, p. 35–110. Strasbourg, Conseil de l'Europe, 447 p.

Roussel Louis, 1994. *La famille en Europe occidentale dépuis 1950.* Bologna, Comune di Bologna, 17 p.

Saraceno Chiara, 1994. *Un familismo ambivalente: le politiche della famiglia in Italia dal dopoguerra ad oggi.* Bologna, Comuni di Bologna, 29 p.

Sgritta Giovanni Battista, 1995. New forms of social organization and interpersonal relationships in ageing societies, *in*: EAPS-IUSSP (eds.), *Evolution or revolution in European population, European Population Conference, Milano, 1995*, vol. 1, p. 155–199. Milan, Fra Angeli, 266 p.

Stone Leroy, 1994. *Marital separation and divorce in England from the seventeenth to the twentieth century.* Bologna, Comune di Bologna, 19 p.

Sweet James, Bumpass Larry and Call Vaughn, 1989. *The design and content of the National Survey of Families and Households.* Madison, CED, University of Wisconsin, 62 p.

Trussell James, 1992. Introduction, *in*: James Trussell, Richard Hankinson, and Judith Tilton (eds.), *Demographic application of event history analysis*, p. 1,7. Oxford, Clarendon Press, 276 p.

Trussell James, Hankinson Richard and Tilton Judith (eds.), 1992. *Demographic application of event history analysis.* Oxford, Clarendon Press, 276 p.

United Nations, Economic Commission for Europe [UN-ECE], 1992. *FFS/Fertility and Family surveys in Countries of the ECE Region: questionnaire and codebook.* New York, UN-ECE, 124 p.

United Nations, Economic Commission for Europe [UN-ECE], 1996–2001. *Standard Country Report*, series with a report for each country participating in the project, New York, UN-ECE (Economic Studies no. 10).

United Nations, Economic Commission for Europe [UN-ECE]/ United Nations Population Fund [UNPF], 2000. *Generations and Gender programme. Exploiting future research and data collection options.* New York, UNECE/UNPF, 166 p.

US Census Bureau, 1993. *An aging world II.* Washington, Census Bureau, 85 p.

Van de kaa Dirk, 1987. Europe's second demographic transition, *Population Bulletin*, vol. 42(1), p. 1–59.

Van de Kaa Dirk, 1988. *The second demographic transition revisited: theories and expectations.* Florence, European University Institute, 55 p.

Véron Jacques, 1993. *Arithmétique de l'homme.* Paris, Le Seuil, 237 p.

White Michael J. and Tsui Ami O., 1984. A panel study of family-level structural change, *Journal of Marriage and the Family*, vol. 48(2), p. 435–446.

C H A P T E R

89

Experiencing Death

A Demographic Approach

ALAIN MONNIER AND SOPHIE PENNEC

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

Demography began with the study of mortality, which remains one of the main cores of the discipline. However, research on mortality focuses less on death than on the study of deaths, that is, on sudden events that have already happened; these events are seen as a collective fatality and are not considered from the point of view of the individual. Studies on medical, cause-specific mortality come a little closer to the reality of death, but nevertheless, they are just as limited as general mortality studies because they look at individuals only in terms of social identity, nationality, socio-occupational background, region, etc.

In reality, death is not an instantaneous or an abstract phenomenon but a concrete process, an experience corresponding to the last stages of life, *a process that concerns people*—the dying person as well as those who know him/her, that is, family, friends, colleagues, and, increasingly, the medical staff,—*and that has an impact on those who remain behind.* This point of view has been widely investigated in other disciplines, especially anthropology and history—in particular the work of Paul-Louis Vincent, Philippe Ariès, Michel Vovelle, and Edgar Morin—through studies on death rituals, or, in the field of psychology (in the line of research conducted by Elisabeth Kubler-Ross), through studies on the support given to the dying, to their family, and to the bereaved.

In demography, however, such an approach to death entails a conceptual leap, because it means considering death as a founding event, a starting point, and not as a final event as it is usually considered by demographers. Death, or its impending arrival, defines the dying person and has a specific impact on the persons close to him or her, who experience death in the sense that they become widows or widowers or orphans, or lose a grandparent, a brother, or sister, etc. Very few demographic studies have been conducted from that point of view. We can mention those concerning the death of relatives (Fourastié, 1959; Le Bras, 1973), on the consequences of widowhood on the mortality of the surviving spouse (Thierry, 1999), or on the subject of elderly persons (Delbès and Gaymu, 1999).

The way death, and all that accompanies it, is experienced by the dying person or by the survivors relies on social trends and, in particular, on demographic trends. In the first part of this chapter, we will examine certain basic characteristics of death from a macrodemographic perspective: in what ways ages at death, causes of death (and consequently the changes in the image of death tied to these trends), and places of death have changed. In the second part, thanks to a microsimulation model, we will describe how the experience of death has changed within families; these changes are due not only to mortality decline but also to changes in fertility trends and in family structures. As will be shown, the experience of death in families occurs not only at later ages but also more often.

Another reason why it is important for demographers to study death is that in the coming decades, the number of deceased and thus of dying persons (and very elderly dying persons), which in the past 50 years has not significantly changed in industrialized countries, will increase considerably. In France, for example, the estimate is that in less than 25 years from now, the number of deaths per year of elderly persons aged over 60 years will be about 550,000, that is, a little over the current number of deceased at all ages; in 2040, this figure will correspond to the number of deaths after 80 years. Overall, in 2050, there should be nearly 700,000 deaths of persons aged over 60 years, among which are 600,000 persons over 80 years. The number of deaths of very old persons, about 250,000 in France today, will more than double during the next 50 years, although the total population will not increase much.[1]

These trends will probably mean that there will be an increasing *presence* of death in society and also, probably, in the lives of individuals. This will entail new public health requirements and will have an incidence on ethical debates concerning the end of life, palliative care, the legalization of euthanasia, or medically assisted suicide.

I. THE DYING: AGE AND SEX, CAUSE AND PLACE

We do not have much statistical information on what is conventionally called the end of life, the last stage of life; in fact, it is difficult to determine when this last stage actually begins.[2] However, one can draw a general picture of how the individual characteristics of dying persons have changed, as well as the circumstances of their death: indeed, these have been considerably modified in recent years, owing to advances made in fighting mortality. Today, dying persons are mainly old persons, even very old persons, and in the latter case, they are often women. The hierarchy of causes of death has undergone a reorganization, differentiated according to age, with a sharp decline in deaths owing to infectious diseases, a rise in accidents and AIDS among young people, and a rise in cancers among mature men and women. Last, people now often die in a medical institution. The description of these changes will be based on macrodemographic data concerning deaths, which will be examined from the perspective of the information they also provide on the dying.

1. The Dying Are Increasingly Old

Today, the dying are most often old persons. This has not always been the case, and it may even be said that in the history of humanity it is a very recent phenomenon, observed only in the 20th century in the countries farthest ahead in the struggle against mortality.

If we look at the distribution of deaths (and thus of the dying) according to age in France (Fig. 89–1; Table 89–1), we see that in the early 19th century, owing to the high infant and child mortality rates, deaths after the age of 60 represented barely one out of three deaths. In 1825, there were more deaths before the age of 15 (37.6%, 20% during the first year) than after 60 (35%). The situation remained the same up to around 1880, with a proportion of deaths in each of the large age groups fluctuating between 30% and 40%. Thus, there were about as many deaths of children as of adults and older persons. With the decline of infant and child mortality at the turn of the century, this distribution changed quite rapidly, and just before World War II, the proportion of deaths of children under 15 fell below 10%, whereas that of persons over 60 reached 60%. As for the deaths of persons aged 15 to 59, the proportion remained remarkably stable at 30% for over a century, except during the periods corresponding to the 1870 war and World War I. Thus,

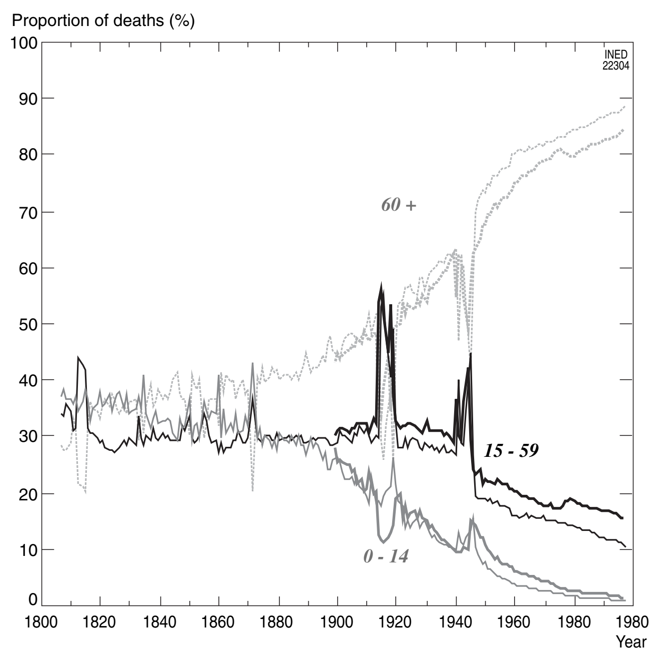


FIGURE 89–1 Age-specific distribution of deaths in France, 19th and 20th centuries. Fine line indicates distribution calculated on the basis of deaths recorded in life tables; bold line, distribution calculated on the basis of observed deaths. (From Vallin and Meslé, 2001.)

[1] These calculations were made on the basis of the mortality projections of Jacques Vallin and France Meslé (2001).

[2] Health economists avoid the problem by focusing on the last year or the last month of life and its costs.

toward 1940, the ages at death and thus of dying persons were still quite heterogeneous.

After World War II, the age-specific distribution of deaths once again rapidly changed: thanks to medical advances and the improvement of living conditions, infant and child deaths became very rare and deaths between ages 15 and 60 years diminished by half; today, the latter represent only 10% of deaths. The consequence of these trends is that the proportion of deaths of old or very old persons has considerably increased: over 80% of deaths occur after 60 years and even 50% after 80 years.

TABLE 89–1 Age-Specific Distribution of Deaths in France

Year	Age			
	0–14 years	15–59 years	60 years and over	
			Total	80 years and over
1825	37.6	27.4	35.0	9.1
1850	29.2	29.8	41.0	10.3
1875	30.1	30.1	39.8	8.6
1900	25.1	30.6	44.3	10.5
1910	20.4	32.1	47.6	11.3
1920	18.5	32.4	49.1	11.9
1930	14.5	32.2	53.3	13.1
1940	9.2	36.4	54.4	15.7
1950	10.4	22.2	67.4	20.8
1960	5.6	19.8	74.6	27.9
1970	3.9	17.4	78.7	30.6
1980	2.2	18.3	79.4	36.3
1990	1.6	16.3	82.1	45.7
1995	1.1	15.6	83.3	49.0
2000	1.0	14.8	84.2	47.1

Distribution in years 1825, 1850, and 1875 calculated on the basis of life tables.

From Vallin and Meslé, 2001; National Institute for Statistics and Economic Studies (INSEE) for 2000.

In addition, the age of dying persons observed through ages at death (bold lines of Fig. 89–1) is lower than that observed through life table (fine lines of Fig. 89–1), because throughout the 20th century the age structure of the French population was younger than that of the stationary population corresponding to the life table.

2. More Women among Very Old Dying Persons

Of course, the overall picture must be qualified according to sex, because the excess mortality of males has increased. Whereas in the beginning of the 20th century, the proportion of deaths over 80 was close to 10% among both men and women, the gap widened over the years, in particular between the early 1960s and the mid-1980s. Today, nearly two out of three female deaths (63.8% in 1995) concerns women aged over 80 (Table 89–2).

We can see that the proportion of women among very old dying persons has increased. Indeed, during the first half of the 20th century, there were about as many deaths of men aged 60 to 79 years as of women of the same age, because the proportion of women among older dying persons had increased from 56% to 61%. Over the second half of the 20th century, we observed a significant decline of the proportion of women among those dying between 60 and 79 years, (as well as under 60), whereas the proportion of women increased among dying persons aged over 80, up to two-thirds in the 1970s and representing about 63% in 1995.

To sum up this overall trend, we can underscore the fact that for those under 80, about two-thirds of the deceased are men, whereas for those over 80, it is the opposite, nearly two-thirds of the deceased are women.

TABLE 89–2 Age- and Sex-Specific Distribution of Deaths

Year	0–14 years		15–59 years		60–79 years		80 years and over	
	Males	Females	Males	Females	Males	Females	Males	Females
Per 100 persons of each sex								
1900	25.9	24.2	32.1	28.9	33.0	34.8	8.9	12.1
1950	11.7	9.1	26.0	17.1	46.2	46.9	16.0	25.8
2000	1.1	0.8	20.2	9.1	45.4	28.5	33.3	61.5
			No. of male deaths per 100 deaths: 64.6				No. of female deaths per 100 deaths: 63.7	

From Vallin and Meslé, 2001.

3. Medical Causes of Death and Changes in the Face of Death

Not only have the dying grown older, but the face of death has also changed over the past 50 years. Indeed, after World War II, the distribution of medical causes of death underwent deep transformations, in particular owing to medical advances. Indeed, the medical cause of death determines not only the circumstances under which the dying person spends the last moments of his or her life but also those experienced by the dying person's family and friends, both before and after the person's death. Dying suddenly in an accident, dying after several weeks because of an acute disease, or dying after a long and painful illness are obviously different experiences, both for the person who dies and for the survivors. To assess the changes that have occurred in the frequency of deaths (depending on medical cause) and to understand the main trends, we will study the changes that have occurred between 1950 and 1995 in the most striking causes of death among the large age groups corresponding to times of life: early childhood, childhood, adolescence, adulthood, old age, and very old age.[3]

If we calculate how the deaths that have occurred in France in each age group are distributed according to medical cause, we can get a picture of the relative importance of the main causes of death in 1950 and 1995.[4]

The contrast between the two periods is striking. In 1950, infant and child mortality was still dominated by infectious diseases, in particular respiratory diseases, which represented 48% of deaths during the first year of life and between the ages of 1 and 14. Deaths owing to congenital or perinatal causes were also quite frequent, whereas deaths owing to accidents seem to have been rather infrequent among children. In 1995 on the contrary, mortality during the first year of life is dominated by congenital or perinatal causes (70% of deaths), causes that concern the very first days of life, whereas accidental deaths dominate deaths from age 1 to 14. The sudden infant death syndrome had not yet been identified in 1950, and for that reason, it does not appear at that time (see below, for more on this issue). However, a new type of mortality owing to malignant tumors has emerged, whereas mortality owing to (usually infectious) diseases of the respiratory tract have disappeared.

In 1950, among adolescents and young adults, 40% of deaths were caused by an infectious disease, either general or of the respiratory tract (tuberculosis), and only 20% were caused by accidents. In 1995 however, accidental deaths represented 70% of the total number of deaths.

Depending on the medical cause of death, in 1950 adult ages and old age were still dominated by degenerative diseases, that is, diseases of the circulatory system, which caused one out of four deaths between 30 and 59 and one out of two or even more beyond that age. At those ages, deaths caused by malignant tumors represented, respectively, only one out of five and one out of ten deaths. Once again, the contrast is striking: in 1995, up to the age of 80 years 40% of deaths were caused by a tumor, and deaths owing to diseases of the circulatory system represented only 15% of deaths between 30 and 59 years and 30% between 60 and 79 years.

In the past 50 years, the overall picture of causes of death has been marked by the disappearance of young children's infectious diseases (and the near disappearance of children's deaths), the increase in the number of accidental deaths among young adults, and the newly leading place of cancers, up to the beginning of very old age. These general lines must be completed first by establishing a distinction between causes of death among males and among females and by specifying the role of certain specific causes as concerns 1995.

There are two main distinctions between female and male death causes, both in 1950 and in 1995: accidental deaths are less frequent among females than among males (however, in 1995 this cause predominates for both sexes), and mortality owing to malignant tumors between 30 and 59 years is a little higher among women, but the tumors have different locations for men and women. Among men of that age group, mortality owing to malignant tumors is caused mainly by smoking, whereas for women the tumors are mainly tumors of the genital system (breast, ovaries, uterus), which cannot be tied to any specific behavior.

Previous groupings according to large age groups and causes tend to conceal certain specificities tied to certain age subgroups or to one or the other sex. To provide an accurate rendering of the overall landscape of death in France at the end of the 20th century, several causes must be pinpointed more specifically. These causes are different from the others because they are particularly striking for public opinion, because

[3] The relative weight of each cause of death is calculated in comparison to the number of deaths, without standardization according to age, because the image of death at a given moment is precisely tied to the distribution according to the age of the dying.

[4] The groups of causes of death globally correspond to the section of the ninth revision of the *International Classification of Diseases*, and statistical continuity since 1950 was ensured to the best of our ability. The data are taken from a database on causes of death accessible on the Institut National d'Etudes Démographiques Web site (http://www.deces.ined.fr/).

their frequency makes them more visible, or both. We selected a few that seem to have a specific impact on the general perception of death today: the sudden infant death syndrome, suicide, AIDS, lung, and breast cancer.

The sudden infant death syndrome is specific in that although its etiology is unknown, it is quite prominent among parents' perceptions of the death risks threatening their newborn baby. This attributed importance corresponds to statistical reality since during the first year of life, one out of five deaths can be ascribed to this syndrome.

Suicide also represents a large share of adult deaths (20% between 15 and 29 years and 10% between 30 and 59 years). As to AIDS, it is not too frequent among young adults (5% of deaths), but it gains in prominence between ages 30 and 39 (one out of five deaths among men and one out of eight deaths among women).

As concerns all tumors, the prominent position of which for ages 30 to 59 years has been mentioned above, some are particularly visible owing to their statistical weight: cancer of the respiratory system (malignant neoplasm of trachea, bronchus, and lung) among men and breast cancer among women. The former represents 10% of deaths among men aged 30 to 79 years, and the latter represents 15% of deaths among women aged 30 to 59 years.

The impact of these transformations (the increasing weight of chronic diseases) on perceptions of death cannot be studied in detail in this chapter and would require specific observations obtained through interviews and surveys. However, it seems that two main images of death tend to predominate: on one hand, sudden death, something quite rare that strikes (accident, acute heart attack, sudden infant death syndrome) the individual or his/her relatives without warning and, on the other hand, death after a long disease, which goes along with serious physical and sometimes psychic alterations of the person. The fact that the latter representation of death is so widespread can be linked to the development of various movements in favor of the legalization of euthanasia and medically aided suicide.

4. Where People Die

In France, the changes in death mentioned above went along with a complete change in place of death—usually a hospital or other institution (clinic, nursing home, etc.), whereas only 30 years ago over half of all deaths still occurred in the dying person's home. The dying process has thus moved from the private to the public sphere.

The study of French statistics shows that people die in three categories of places: home, various institutions (mainly the hospital) and public places, with a miscellaneous category: other places and unreported places. One may leave aside the last two categories, which represent one out of twenty deaths, are closely linked to certain specific causes (accidents), and are special in the sense that the place of death has not been chosen. One must mainly consider the two alternatives: dying at home or dying in a medical institution, with a rather indefinite entity of nursing homes, which are both homes and institutions.

The place where people die has drastically changed over the past 50 years (Table 89–3): in 1950, people died mainly at home, a trend that continued until the mid-1970s, after which there was a change to institutions, in particular hospitals. Today, 50% of deaths take place in hospitals (compared with only 15% in 1950).

One can wonder about the reasons for this trend, which reflects a profound change in attitudes, in the sense that it marks the disappearance of the negative image of the hospital as a dying place for the poor, an image still strong in the 1950s. Various factors have been suggested: urbanization, the disappearance of extended families, isolation at the end of life, housing conditions, etc. An analysis of recent French data

TABLE 89–3 Distribution (%) of Deaths According to Place in France (1950–1995)

Place of death:	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Home	71.6	68.9	65.8	62.6	57.1	53.2	32.8	30.4	28.6	25.9
All institutions	20.7	25.4	29.8	34.2	39.1	43.8	63.7	65.8	66.1	68.4
Hospitals and clinics	17.6	21.3	28.7	32.9	37.3	38.8	58.4	60.1	58.1	59.3
Nursing homes	0.6	0.7	1.2	1.5	1.9	5.0	5.3	5.8	8.0	9.1
Public places, roadways	0.9	1.0	1.2	1.5	1.7	1.5	1.9	2.0	2.3	2.2
Other places and unreported	6.8	4.8	3.3	1.6	2.1	1.4	1.5	1.7	3.0	3.5
All places	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Data from National Institute for Statistics and Economic Studies (INSEE).

shows that few differences depend on the characteristics of the deceased person or on the circumstances of death. Concerning deaths owing to illness, the proportion of persons dying in a hospital or institution are about the same regardless of whether the person is young or old, married or unmarried, and living in a rural town or in a large city.

However, to better understand the meaning of this change in behavior, it is necessary to know more about the period referred to as the end of life, from home (or nursing home) to hospital. Dying in a hospital does not mean that the entire end of life period was spent in a hospital, and conversely, dying at home does not mean that some time was not spent in a hospital. The opposition between dying at home and in an institution is not as clear-cut as it may seem, especially as new forms of hospital care dispensed in the person's home are now increasingly being developed.

In addition, very few countries publish data on the place of death, and one may surmise that situations in Europe still vary. In Greece, for example, at the end of the 1990s, nearly half of all deaths still took place at home.

On the basis of the observations presented above, a few general remarks can be made concerning the conditions under which dying persons experience their last moments. These observations should be completed by more precise measurements to better apprehend the reality of this ultimate experience. But the dying person's death is also an experience affecting the person's entourage, first and foremost his or her family.

II. DEATH AS EXPERIENCED BY THE DYING PERSON'S FAMILY

Although the dying process concerns the dying person and his/her close family and friends, in a sense, death is only an event for those who survive. As opposed to the usual conception in demography, whereby death is seen as a final event (the ultimate point of each life line of a Lexis diagram), one can also consider death as a starting point, an element belonging to the biography of the surviving relatives. This event, which can be termed "experiencing the death of another person" can take many forms and concerns death of friends, relatives, work relations, or (and this is what we will focus on) close family members. Each discipline that studies death has developed its own specific approach: anthropologists study rituals, psychologists study mourning processes, economists measure the cost of death, etc. The demographic approach to the experience of another person's death consists in measuring the frequency, over the years, with which a person, *Ego*, witnesses the death of his/her grandparents, parents, spouse, siblings, or even his own children.

The experience of a family member's death changed radically during the 20th century, owing to the mortality decline and changes in family structures. The traditional instruments of demography—calculations based on aggregated vital statistics and population observations—make it possible to assess some of these changes, in particular those concerning orphanhood. But to have an overall view of the experience of death within the family, it is necessary to use other approaches that take into account the numerous dimensions of the phenomenon.

1. A Simple Case: Orphanhood

To measure orphanhood, we have at our disposal the findings of Alfred Lotka (1939), who showed that the proportion of orphans in a cohort depended only on the mortality of the parents and could be easily calculated by using the appropriate life tables. Indeed, the probability of becoming an orphan at a given age is equal to the probability of the father or mother's death during the period lasting from birth to that age.

Lotka's equation was established by considering that during a given year, the distribution of births according to the age of the mother or father is correctly summed up by the mean age of mothers and fathers at birth (mean age calculated on the basis of the size of the birth cohort and not on rates). Thus the probability Ω_x for a child reaching the age x of losing his father or mother is equal to the complement to unity of the probability of survival of his mother or father between the mean age at childbirth and her or his age reached x years later.

$$\Omega_x = 1 - S(\bar{a} + x)/S(\bar{a})$$

with:

S = survivors of the life table;
$\bar{a}$ = mean age at motherhood/fatherhood;
x = age reached by the child.

Thanks to the existence of cohort life tables (Vallin and Meslé, 2001), we can add some precision to these calculations (Alfred Lotka only had period life tables to work with), and modern computer tools make it possible to perform complete calculations without simplifying by using the mean age. If we consider that the proportion of orphans is equal to the complement to unity of an average of survival probabilities tied to each age of the mother (father) between the birth and age x of the child, an average weighted by the distribution of births according to the age of the father (mother):

$$\Omega_x = 1 - \sum_{a-\min}^{Max} \frac{Na}{N} \cdot \frac{Sa+x}{Sa}$$

S = survivors of the life table;
N_a = Number of births according to the age of the mother (father);
N = total number of births;
a = age at birth of the child;
x = age reached by the child.

The descriptions of orphanhood thus obtained are taken from the point of view of a surviving *Ego*, thus the mortality of *Ego* is not taken into account. However, not all children become orphans, because some children die before their parents. Indeed, it is possible, if one combines the life tables of the children with those of their parents, to calculate the proportion of children who die before their parents; but the method is insufficient when one looks at other family relations, because other phenomena must be accounted for. Studying orphanhood is quite simple due to the extent that one necessarily has two parents and no more than two, if we limit ourselves to biological relationships. But if we look at other family relations, things are more complex and do not only depend on mortality: the proportion of persons who become widows also depends on nuptiality; the proportion of persons who lose a child also depends on fertility, just as the proportion of persons who lose a sibling, etc.

2. A Microsimulation Model

To take into consideration the numerous phenomena that together define the experience of another person's death as seen from the point of view of an individual *Ego* who is himself liable to die, one must refer to a microsimulation model thanks to which, for each type of family relation, one can measure the frequency of the three variants in the experience of another person's death, the other person being any member of *Ego*'s family:

1. The other person may have died before *Ego*'s birth.
2. The other person may die during *Ego*'s life, and it is thus possible to know *Ego*'s age at the time of his death.
3. *Ego* can die before the other person.

These possibilities can be observed when the other person is a grandparent, a sibling, an uncle, an aunt, father-in-law or mother-in-law, a nephew, or niece. However, the first possibility is not applicable to *Ego*'s mother (she can die during childbirth but not before *Ego*'s birth) or to *Ego*'s spouse or to a child, because they obviously cannot have died before *Ego*'s birth (before the union or the birth, respectively).

It must be noted that these data cannot be obtained through traditional surveys, in particular retrospective surveys. Indeed, by definition, the respondents of these surveys are necessarily survivors. In the first place, we cannot exclude that there is a dependence between *Ego*'s mortality and that of his relatives, and this can introduce a selection bias for the studied phenomenon. But above all, an important aspect of the phenomenon is lost to the extent that one cannot study the frequency of the cases in which *Ego* dies before his close relatives.

A microsimulation model makes it possible to reconstruct the family of an individual of reference and the deaths occurring within this family, including the death of the reference individual him/herself.

This simulation method was developed by Guy Orcutt in 1957 in order to simulate the American socioeconomic system. It has since been applied in various fields. Demographers also used this method in the 1960s and 1970s, first to study fertility trends and the impact of contraception (Perrin and Sheps, 1963; Jacquard and Leridon, 1974) and to integrate the different members of such networks as family, relatives, or the employees of a company. Thus we have, among many others, the Socsim model of Eugene Hammel and Kenneth Wachter, the more historical Camsim model of Roger Schofield, Peter Laslett, and Jim Oppen; and the Kinsim model of Wendy Post and Evert Van Imhoff (Le Bras, 1973; Wachter *et al.*, 1978; Smith, 1987; Hammel *et al.*, 1989; Reher and Schofield, 1993; Smith and Oeppen, 1993; Van Imhoff and Post, 1997). Microsimulation is based on the idea that studying the sum of individual behaviors provides a more accurate representation of reality than does studying changes in aggregates. In addition, the study of these networks requires the ability to recreate them, and that is not possible with more aggregating techniques. The method uses stochastic processes to determine for each person the events that he will experience during the simulated period. The probability of events is determined according to known or reconstructed trends when dealing with the past, or according to estimated trends when dealing with projections for the future.

The results presented here are limited to what could be called the first circle of the family: *Ego*'s grandparents, his parents, siblings, spouse, and children, but other relatives may also be included (e.g., grandchildren, in-laws, uncles and aunts). Of course, the simulations are sex specific. The model functions on the basis of individual microsimulations based on mortality and fertility data.

a. Mortality

The French data used here are age-specific probabilities of dying for each birth cohort since 1806, either calculated on the basis of observations or projected (Vallin and Meslé, 2001). For the cohorts born before 1806 (grandparents of *Ego* if the latter was born in 1850 or 1860), we supposed that their mortality was the same as that of the 1806 cohort. The mortality of cohorts born in the 20th century is based on observations and projections: thus, the mortality of the 1950 cohort was observed from the year of birth until 1997 and projected beyond; for the children of parents belonging to that cohort, the estimated part is even higher.

Table 89–4 shows life expectancies at birth of the cohorts of *Ego* only, but his/her parents, grandparents, spouse, children, etc. can belong to any other cohort: the simulation thus involves 300 series of probabilities (from the end of the 18th century, for the grandparents of an *Ego* born in 1850 to the end of the 21st century for the children of *Ego* born in 1950, for both sexes).

b. Fertility

We assessed, for each cohort born between 1850 and 1950, the age-specific and parity-specific fertility rates. The rates for the 1850 cohort were applied in order to reconstruct the cohorts born before 1850 (*Ego*'s parents, grandparents, siblings if any). The completed fertility of each cohort and the mean age at childbearing are shown in Table 89–4.

The reconstruction thus takes into account the real mortality of the different cohorts under consideration, as well as a partly estimated fertility (as concerns parity-specific fertility) whose general trend, in terms of completed fertility, corresponds to reality.

TABLE 89–4 Fertility and Mortality Parameters of the Model for the Cohorts of *Ego*

Cohort	Average life span		Completed fertility	Mean age at Childbirth
	Males	Females		
1850	39.8	43.2	2.78	29.8
1860	40.8	44.3	2.67	28.8
1870	38.8	43.4	2.54	28.9
1880	40.8	48.4	2.32	28.6
1890	38.7	51.5	2.09	28.3
1900	48.3	56.0	2.10	28.5
1910	53.1	61.5	2.27	28.6
1920	57.3	67.2	2.48	28.4
1930	64.0	73.6	2.62	27.4
1940	67.3	76.6	2.41	26.4
1950	72.9	82.0	2.11	26.5

Data from Festy, 1979; Vallin and Meslé, 2001.

3. Experiencing Death in the Family from 1850 to the Present

Thanks to the implementation of this model, it was possible to obtain data concerning the trends in the experience of death in the family for cohorts born between 1850 and 1950.[5]

The curves illustrating the experience of death in the family, that is, all the curves pertaining to each degree of relationship, provide a general picture for a given cohort. Graph A of Fig. 89–2 thus gives an overall view of the frequency of deaths in the family of a woman born in 1900.

In addition to period variations, owing to which this graph is a little difficult to read (in particular, World War I and World War II caused mortality peaks affecting fathers and brothers, respectively, spouses or children), three waves of deaths can be distinctly observed, corresponding to the deaths of the grandparents, parents, and contemporaries of *Ego*, whether the spouse or the siblings. We also observe the high impact of the death of children, either siblings or children of *Ego*.

A comparison with the same graph elaborated for the 1950 cohort (Fig. 89–2B) sheds some light on the changes that occurred throughout the 20th century, especially during the second half, even if some of the results shown here are still hypothetical (the 1950 cohort is not yet extinct). Still more clearly we can see the three waves, but whereas they were declining from the generation of the grandparents to that of contemporaries in the 1900 cohort, here their volume is about constant, if not increasing. In addition, advances made in the field of infant and child mortality have almost eliminated mortality at young ages, which was still high among the siblings and children of women born in 1900. This fact makes the graph more easily understandable. In fact, the graph illustrates the successive order of the deaths of close relatives, which follow each other in much the same order as the generations. It also provides a general overview of the overall aging of the age at death as concerns the death of other people, viewed as an event in the life of the survivors. Last, it seems to show that the number of deaths in the family of a woman born in 1950 is higher than that observed among the relatives of a woman born in 1900.

[5] Because of a different implementation of the microsimulation method, the results presented here differ from those we presented in various articles and papers.

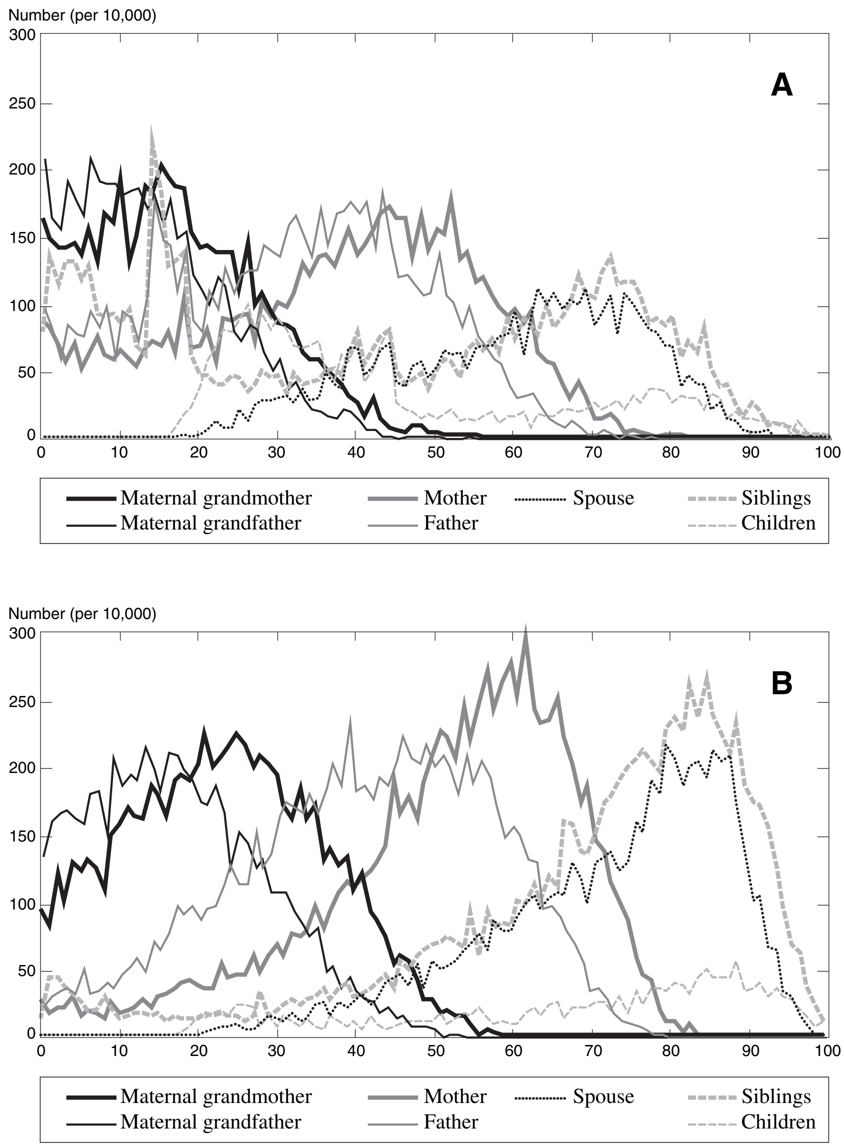


FIGURE 89–2 Numbers, according to age, of deaths in the family of a woman born in 1900 and in that of a woman born in 1950. **(A)** *Ego* is a woman born in 1900. **(B)** *Ego* is a woman born in 1950.

4. Death in the Family Is Experienced Later in Life

In this general context, in what way does the experience of death change when a person has reached old age? We will attempt to answer this question by looking at the deaths of *Ego*'s parents, spouse, siblings, and children (by the age of 60, for all cohorts, *Ego*'s grandparents have already died).

As concerns parents, there are two possible cases: either *Ego* dies after his/her parents or his/her parents die after him. The latter possibility was quite frequent during the 19th century, when infant and child mortality were very high. Thus in the 1850 cohort, we may assess that in more than 40% of the cases, the person died before his/her mother (42% of the females and 45% of the males). This proportion considerably declined, especially after 1900, and in the 1950 cohort it was only 11% (Table 89–5A). As concerns the death of the father (Table 89–5B), the trend was comparable. We can note that men die a little more often than do women before their parents and that this difference has increased over time with the rise in the excess mortality of males.

Alongside this trend, the mean age at which a person loses his/her father or mother has increased: in the 1850 cohort, a person was orphaned around the age of 30, whereas in the 1950 cohort the mean age of *Ego*

TABLE 89–5A The Death of the Mother

Cohort of *Ego*	*Ego* is a woman				*Ego* is a man			
	Mother died before *Ego* (%)	*Ego* dies before mother (%)	Mean age of *Ego* at death of mother	*Ego* is over 60 at death of mother (%)	Mother died before *Ego* (%)	*Ego* dies before mother (%)	Mean age of *Ego* at death of mother	*Ego* is over 60 at death of mother (%)
1850	58	42	31.7	3.1	55	45	31.5	2.8
1860	59	41	32.4	3.3	56	44	32.1	2.9
1870	57	43	33.0	4.4	53	47	32.2	3.5
1880	62	38	34.5	5.3	54	46	32.9	4.1
1890	64	36	35.9	6.5	49	51	33.1	5.2
1900	69	31	37.4	9.3	62	38	36.3	8.0
1910	72	28	39.8	13.5	66	34	38.5	11.3
1920	77	23	42.8	18.1	69	31	41.4	15.6
1930	83	17	45.6	23.0	75	25	44.4	20.2
1940	85	15	48.4	30.2	77	23	47.4	27.7
1950	89	11	52.5	38.1	81	19	51.3	35.2

TABLE 89–5B The Death of the Father

Cohort of *Ego*	*Ego* is a woman				*Ego* is a man			
	Father dies before *Ego* (%)	*Ego* dies before father (%)	Mean age of *Ego* at death of father	Father dies before *Ego* (%)	Father dies before *Ego* (%)	*Ego* dies before father (%)	Mean age of *Ego* at death of father	Father dies before *Ego* (%)
1850	60	40	29.6	1.3	57	43	29.3	1.2
1860	61	39	29.9	1.3	58	42	29.7	1.2
1870	59	41	29.2	1.5	56	44	28.9	1.2
1880	64	36	30.4	1.7	58	42	29.1	1.2
1890	67	33	31.0	2.0	53	47	28.5	1.6
1900	71	29	31.3	2.7	66	34	30.6	2.2
1910	76	24	28.9	3.1	72	28	28.2	2.6
1920	80	20	34.7	5.1	74	26	33.9	4.5
1930	86	14	35.9	6.7	81	20	35.2	5.9
1940	87	13	37.5	9.8	83	17	37.1	8.8
1950	92	8	41.7	13.9	87	13	41.1	12.7

at his/her mother's death was around 50 and at his/her father's death around 40. Last, the proportion of persons who are orphaned after 60 has also increased, in particular as concerns the death of the mother: for the 1850 cohort, only 2% to 3% of individuals lost their mother after 60, whereas this proportion represented, respectively, 35% and 38% for men and women born in 1950. As concerns the death of the father, the figures pertaining to the 1950 cohort are lower: 13% to 14% of men and women born in 1950 will lose their father after the age of 60.

The death of the spouse also occurs at later ages for both sexes, but the gap between the sexes has widened. We know that widowhood affects mainly women: in the cohort born in 1950, 75% of the women are or will become widows, against only 27% of the men (see Table 89–6). But for men and women, the mean age at widowhood has increased by about 20 years at least in the space of 100 years: the women born in 1850 became widows at 53 on average, and those born in 1950 at 73; for men the figures are, respectively, 54 and 77. Widowhood is now a phenomenon occurring mainly after 60 and even increasingly tends to happen after 80: for the 1950 cohort, 42% of the women and 54% of the men will be widowed after the age of 80, according to mortality hypotheses. As for the death of parents, these changes took place mainly after 1900.

The death of a sibling is an event that is similar to the death of a spouse in the sense that they are close to *Ego* in age. However (and this is a significant difference),

TABLE 89–6 The Death of a Spouse

Cohort of *Ego*	*Ego* is a woman				*Ego* is a man			
	Women who become widows (%)			Mean age at widowhood	Men who become widowers(%)			Mean age at widowhood
	Regardless of age	At and over 60	At and over 80		Regardless of age	At and over 60	At and over 80	
1850	60	40	2	53.3	42	42.7	4.7	54.4
1860	61	42	2	54.2	40	43.8	4.9	55.0
1870	63	44	3	55.1	39	45.3	5.8	55.7
1880	66	42	4	53.9	33	46.2	9.0	56.3
1890	70	42	6	52.3	31	50.9	11.7	58.2
1900	69	55	9	59.5	31	57.9	15.5	60.9
1910	72	61	13	61.4	30	62.3	22.5	63.4
1920	72	66	19	64.1	28	72.8	32.8	68.7
1930	74	71	26	67.6	27	77.9	39.7	72.0
1940	75	76	33	70.1	28	80.8	46.9	74.3
1950	75	80	42	72.5	27	83.4	54.4	76.8

TABLE 89–7 Deaths among Siblings

Cohort of *Ego*	*Ego* is a woman				*Ego* is a man			
	Women who lose a sibling (%)			Mean age at death	Men who lose a sibling (%)			Mean age at death
	Regardless of age	At and over 60	At and over 80		Regardless of age	At and over 60	At and over 80	
1850	46	20	1	33.9	42	17	1	31.7
1860	47	23	2	35.5	41	18	1	32.8
1870	44	25	2	37.1	39	20	1	34.2
1880	48	26	4	38.9	40	18	2	33.9
1890	49	29	5	39.7	37	19	2	32.6
1900	50	37	8	44.0	41	28	4	38.1
1910	51	46	13	49.5	41	35	6	43.0
1920	53	54	19	55.1	40	44	11	48.6
1930	55	64	26	63.0	41	54	17	56.5
1940	55	71	32	66.6	41	62	23	61.3
1950	57	77	40	72.0	42	71	30	67.4

siblings can die while *Ego* is still a child, which is not the case for a spouse. This explains why the mean age of *Ego* at the time of the death of sibling is always younger than that at the time of a spouse's death.

In addition, because the birth order of *Ego* has not been controlled (*Ego* can be of any parity), *Ego* is, on average, in the middle and thus witnesses the death of about one half of his siblings (between 46% and 57% for females, about 40% for males[6]). The dominant feature of this changing phenomenon is its aging: an individual born in 1850 experienced the death of one or several siblings before the age of 60 in 80% of the cases, whereas for the 1950 cohort, the death of a sibling happens after 60 in 77% of cases if *Ego* is female and 71% of cases if *Ego* is male (see Table 89–7). What was then the exception has now become the norm. This aging process is now clearly continuing: the death of a sibling after the age of 80 occurs in 40% of cases if *Ego* is female and in 30% of cases if *Ego* is male.

To experience the death of one or several children was quite frequent in the 19th century: 43% of the women and 40% of the men born in 1850 had lost at least one child (see Table 89–8). Today this has

[6] When *Ego* is male, the sex ratio of the other members of the group of siblings decreases (on average, *Ego* has more sisters than brothers); conversely, if *Ego* is female, the ratio increases. This, together with the excess mortality of males, is the reason why females experience the death of over half of their siblings and males less than half.

TABLE 89–8 The Death of Children

Cohort of *Ego*	*Ego* is a woman				*Ego* is a man			
	Women who lost a child			Mean Age	Men who lost a child			Mean age
	Regardless of age	At and over 60	At and over 80		Regardless of age	At and over 60	At and over 80	
1850	43	18	2	40.2	40	44	1	40.6
1860	44	12	2	40.7	40	9	1	40.9
1870	40	15	3	40.3	34	12	1	39.8
1880	33	20	5	41.9	27	14	2	39.8
1890	30	19	7	42.6	24	11	3	39.7
1900	23	29	12	45.7	18	18	5	41.5
1910	19	37	17	49.5	15	24	8	43.4
1920	17	44	21	53.0	12	32	12	47.5
1930	14	56	29	60.0	9	43	17	53.5
1940	12	62	36	64.5	8	50	23	58.5
1950	10	69	41	68.5	6	59	29	63.0

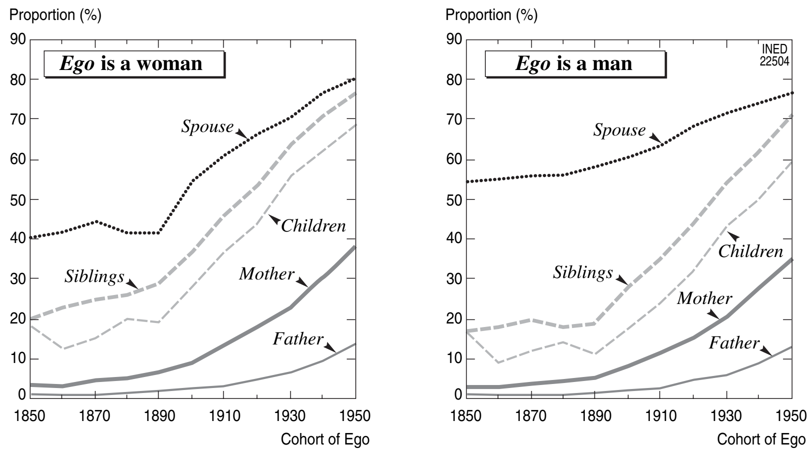


FIGURE 89–3 Proportion of deaths among close relatives when *Ego* is over 60.

become quite rare, because only 10% of the women and 6% of the men born in 1950 will have experienced it. But at the same time, not only has this phenomenon become more rare, but it has changed in nature. Indeed, in the 19th century, when parents experienced the death of a child, it was a small child, and these parents were thus young adults (the curve of deaths of children according to the age of the mother at the time of the child's death is thus very similar to the fertility rate curve). Today, the death of a child concerns older parents, even very old parents: for the 1950 cohort, 69% of the women and 59% of the men will be over 60 if ever one of their children dies (41% and 29%, respectively, will be over 80).

In all the cases examined above, the deaths of members of the family occur later in the life of individuals. The increase in the mean age of an individual, male or female, born in 1950, at the death of his/her mother, spouse, sibling or child is 20 to 35 years; the age at the death of the father has increased less, but it is still 12 years older for both men and women.

The changes in the proportion of deaths in the family occurring when *Ego* is over 60 shows that the age at which one experiences the death of another person has increased (Fig. 89–3). Beginning with the 1900 cohort, the curves rise quite clearly upward as concerns the death of the spouse, a sibling, and children, and a little less sharply upward as concerns father and mother.

5. Death Is a More Frequent Experience

These findings pertain to the trend in the schedule of the experience of death, at the time when *Ego* experiences the death of a close relative. But as we have seen, this phenomenon has other implications: *Ego* can die before any of his/her relatives (in which case he is no longer concerned by this schedule), and some of his/her relatives (a grandparent, brother, or sister) may have died before his/her birth (in which case they are not taken into consideration). The total frequency, the intensity of the experience of death, varies according to these factors: if *Ego* dies less often before his/her parents (father or mother), he will more often experience his/her parents' death; the same is true if the death of his/her grandparents occurs less frequently before his/her birth. In addition, over time, the number of parents (fathers and mothers) and of spouses (in this case the first spouse) remains, the same but the number of siblings or children can change. Here again we observe a factor that can change the frequency of the experience of death: mortality being equal, the greater the number of children, the greater the risk of losing a child.

To take into account these aspects of the phenomenon, one can calculate the mean number of deaths (converted for each individual at birth) for each degree of kinship; this calculation was made by establishing a distinction between deaths occurring when *Ego* is over 60 and those occurring before that age (Table 89–9). From females born in 1850 to those born in 1950, the total mean number of deaths within the family (as defined here) has significantly increased, especially during the 20th century, from 5.20 for the 1850 cohort to 5.43 for the 1900 cohort and 7.27 for the 1950 cohort. Thus, the experience of death in the family happens at a later age but has become more frequent.

OVERVIEW

In France and probably in other developed countries, from the point of view of the dying person and that of the persons close to him/her, the experience of death has considerably changed during a period that can be determined quite precisely: the second half of the 20th century. In 1959, Jean Fourastié wrote that "death was at the center of life in the same way as the graveyard was in the center of the village." One might say today that death is at the center of old age, because

TABLE 89–9 Mean Number of Deaths (for One Person)

Relationship with *Ego*	Cohort								
	1850			1900			1950		
	Death occurring		Potential deaths	Deaths occurring		Potential deaths	Deaths occurring		Potential deaths
	Before 60	After 60		Before 60	After 60		Before 60	After 60	
Ego is a woman									
Deaths of:									
Grandparents	1.71	0.00	2.41	1.93	0.00	2.49	2.63	0.00	2.80
Parents	1.14	0.04	2.00	1.27	0.13	2.00	1.12	0.69	2.00
Spouse	0.19	0.12	0.51	0.17	0.21	0.55	0.12	0.50	0.80
Siblings	1.26	0.31	3.30	0.96	0.53	3.50	0.51	1.60	3.70
Children	0.35	0.08	1.80	0.16	0.07	1.70	0.03	0.07	1.90
Total	4.65	0.55	10.02	4.49	0.94	10.24	4.41	2.86	11.20
Regardless of age	5.20		5.43		7.27				
Total except grandparents	2.94	0.55	7.61	2.56	0.94	7.75	1.78	2.86	8.40
Regardless of age	3.49		3.50		4.64				
Ego is a man									
Deaths of:									
Grandparents	1.65	0.00	2.41	1.87	0.00	2.49	2.57	0	2.79
Parents	1.09	0.03	2.00	1.18	0.10	2.00	1.09	0.59	2.00
Spouse	0.11	0.08	0.48	0.07	0.10	0.52	0.03	0.18	0.80
Siblings	1.19	0.20	3.31	0.90	0.32	2.99	0.49	1.08	3.73
Children	0.16	0.01	1.72	0.14	0.03	1.64	0.03	0.04	1.70
Total	4.20	0.32	9.92	4.16	0.55	9.64	4.21	1.89	11.02
Regardless of age	4.52		4.71		6.10				
Total except grandparents	2.55	0.32	7.51	2.29	0.55	7.15	1.64	1.89	8.23
Regardless of age	2.87		2.84		3.53				

those who die are mainly old but also very old persons, and their close relatives—their spouses or their children—are no longer young or are even close to old age. That death is no longer in the center of the village because nowadays, as a general rule, people die in hospitals or institutions.

Several aspects of these trends deserve special attention because the transformations observed have implications both for the individual and for society that should be examined in greater detail.

In the first place, a distinction was made according to the age of the dying persons; those of today are older than those of yesterday. In addition, dying women are much older than are dying men. But does this trend, which is positive as concerns the duration of life, have an impact in terms of the conditions surrounding a person's death? In other words, even though it is better to die at 85 than at 75, does the fact of being older make a difference for the dying person, for his family if he has one, and for the medical staff? As concerns the latter, whose thoughts on the subject are better known, we must remember that as Renée Sebag-Lanoë (2002) has pointed out, "age is one of the worst criteria for making a decision" as regards the choice of appropriate palliative care.

Given the increase in the age at death, brought about by the decrease of deaths at young ages, the succession of deaths within the family increasingly follows the order of generations. This ordering might seem to be a factor of greater *acceptance* in the general picture of death that we have briefly drawn, if we admit (but this remains to be proved) that death is more acceptable when it affects each in his own time. But conversely, because early deaths—deaths of children, adolescents, young people, or mature adults—happen more seldom, have they not become even more difficult to accept? What significance can be ascribed to the fact that we now live to experience a greater number of deaths in the family than in the past? Does this situation make it easier to accept death or, on the contrary, does it make it more difficult?

Next, we found that today, death occurs most often in institutions, because only a minority of persons now die at home. This trend seems to run counter to other trends that seem to give precedence to the private sphere over the public sphere. Thus, the decreasing frequency of marriages, the increasing frequency of consensual unions, and the rise in births out of wedlock are often interpreted as a sign that individuals prefer to withdraw their marital and family life from the influence of institutions. Are things different with death, in which the need for institutional support is very high, or can the fact that death is increasingly medicalized, that is, medical staff is needed, be seen as the sign of a more general phenomenon, that of the extension of the medical influence? This extension has been observed in the field of procreation, with the massive spread of contraception, the legalization of abortion, and the development of medically assisted procreation. Last, will the expected increase in the number of dying persons have an influence on medical practices pertaining to the end of life? It is to be expected that such issues as palliative care, staff resources, and the role of families in the dying person's last moments and euthanasia will become increasingly acute when a country such as France will be counting 250,000 more deaths each year.

Acknowledgment

This chapter was translated from the French version by Zoé Andreyev.

References

Delbès Christiane and Gaymu Joëlle, 1999. Les retraités en France, d'hier à demain, *Cahiers de sociologie et démographie médicales*, vol. 56(2), p. 113–131.

Festy Patrick, 1979. *La fécondité des pays occidentaux de 1870 à 1970*. Paris, INED, PUF, 399 p.

Fourastié Jean, 1959. De la vie traditionnelle à la vie "tertiaire": recherches sur le calendrier démographique de l'homme moyen, *Population*, vol. 14(3), p. 417–432.

Hammel Eugene A, Mason Carl, and Wachter Kenneth W, 1990. *SOCSIM II: a sociodemographic microsimulation program rev. 1.0.: operating manual*, Berkeley, University of California/Graduate Group in Demography.

Jacquard Albert and Leridon Henri, 1973. Simulating human reproduction: How complicated should a model be?, *in:* Bennett Dyke and Jean W. MacCluer (eds.), *Computer simulation in human population studies*, p. 241–250. New York, Academic Press, 518 p.

Le Bras Hervé, 1973. Parents, grands-parents, bisaïeux, *Population*, vol. 28(1), p. 9–38.

Lotka Alfred J., 1939. *Théorie analytique des associations biologiques. Deuxième partie: analyse démographique avec application particulière à l'espèce humaine.* Paris, Herman & Cie, 139 p.

Orcutt Guy H, 1957. A new type of socio-economic system, *Review of Economics and Statistics*, vol. 58(2), p. 116–123.

Reher David and Schofield Roger (eds.), 1993. *Old and new methods in historical demography.* Oxford, Oxford University Press, 426 p.

Sebag-Lanoë Renée, 2002. Vingt ans d'expérience de soins palliatifs et de réflexion éthique en gériatrie à l'hôpital Paul Brousse, *Revue française des Affaires sociales*, 56ème année, no. 3, p. 163–171.

Sheps Mindel C., 1969. Simulation methods and the use of models in fertility analysis, *in:* IUSSP/UIESP (ed.), *International Population Conference, London 1969: proceedings*, vol. 1, p. 53–64. Liège, IUSSP/UIESP, 844 p.

Smith James E., 1987. The computer simulation of kin sets and kin counts, *in:* John Bongaarts, Thomas Burch, and Kenneth Watcher (eds.), *Family demography: methods and their applications*, p. 249–166. Oxford, Clarendon Press, 365 p.

Smith James E. and Oeppen Jim, 1993. Estimating numbers of kin in historical England using demographic microsimulation, *in:* David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 280–317. Oxford, Oxford University Press, 426 p.

Thierry Xavier, 1999. Risque de mortalité et de surmortalité au cours des dix premières années de veuvage, *Population*, vol. 54(2), p. 177–204.

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX^e^ et XX^e^ siècles et projections pour le XXI^e^ siècle.* Paris, INED, 102 p. + CD-rom.

Wachter Kenneth W., Hammel Eugene A., and Laslett Peter, 1978. *Statistical studies of the historical social structure*. New York, Academic Press, 229 p.

C H A P T E R

90

Household and Family Demography

Application to Developing Countries

MARC PILON

Institut de recherché pour le development (IRD), Burkina Faso

INTRODUCTION

It has become increasingly important to demographers to place population behavior within its family context (Tabutin and Bartiaux, 1986). The capacity to statistically document family structure and dynamics, shed light on factors that influence them, explain their variations (within societies and between different societies), and follow their development over time has created a new knowledge base that has contributed to the emergence of household and family demography. This new field of research is important enough to justify significant space in this volume of the treatise.

Two chapters have been devoted to this subject in order to better emphasize the diversity of the situations, both sociological and statistical, of different cultures. The first chapter discusses developing countries and the second centers on the European experience. It is therefore appropriate for this first chapter to begin with a general introduction to the concepts, the theories, and the development of household and family demography, before applying this knowledge to developing countries.

I. HOUSEHOLD AND FAMILY DEMOGRAPHY

First of all it is necessary to agree on the terms used. This is all the more important for concepts that have very different meanings according to the circumstances and cultures in which they are used. This done, it is easier to evaluate the various current theories of family development and then introduce the relatively recent scientific dimension of household and family demography.

1. Concepts: Household and Family, What Are They?

Common sense could lead us to believe that the only difference between family and household is in their nature: the family being a natural biological unit and the household an economic and residential unit. However, common sense is not always enough and believing so in this specific case frequently leads to considerable errors in interpretation. The family is not a natural entity, and the household might only be a demographic tool for census officials.

a. The Family Is Not a Natural Entity

As stressed in the *Dictionnaire de l'ethnologie et de l'anthropologie* (Bonte and Izard, 1991:273), "everyone believes that they know what a family is: it seems to stem from the natural order. . . . But it is interesting to note that however vital, essential, and apparently universal the family institution may be, just as is the case for marriage, no strict definition exists for it."

Acceptance of the term "family" changes according to the protagonists and the institutional contexts, as well as according to society and era. It refers to representations within each society at a given period in its history; frequently not explicitly defined, the term family is an interpretation of the situation that is considered as the norm, the socially acceptable model (the extended family or kinship in Africa[1]), or is presented as such in cases in which the state has imposed itself and interferes in people's private lives. Indeed, the polysemic character of the family means that it only truly develops a meaning once it has been qualified.

Regardless of the definition, whether we consider a couple with children or a kinship group, it is necessary to remember that "the family is not, contrary to the belief of certain demographers and sociologists, a base unit or a cell of society. A family can not exist and be reproduced through generations independently of other families" (Godelier, 1973:8). The family is not an independent, autonomous unit.

We must agree with Pierre Bourdieu (1993:36) that although it is doubtless necessary to "stop considering the family as an immediate fact of social reality but rather as a tool for constructing this reality," we should also "ask ourselves who created these construction tools." In Western societies the state has played a crucial role in this area, so that the family appears as "the product of a long period of legal and political construction." Some even ask whether "the current statistical concept of the family still corresponds to an economic and sociological reality" (Desplanques, 1994:97); comparative studies on Europe are always faced with numerous obstacles (CEPR, 1991; Lefranc, 1995; Wall, 1996; see also the following chapter by Nico Keilman). A similar process, reinforced by the influence of international organizations, is underway in developing countries. Therefore, the family is not an immediate natural entity but a progressive "constructed category" (Bourdieu, 1993:36).

It can be defined or analyzed in terms of structure (family formation), through its functions (e.g., production, reproduction, socialization), and its relational and transactional dimensions (intra and extrafamily).

b. The Household: A Data Collection Tool That Has Become an Object for Analysis

The concept of household was created in Western societies by statisticians and demographers in search of a statistical unit for operational observation that makes it possible to identify individuals without omission or double counting during censuses and surveys; gathering information on kinship being above all a method of identifying individuals. The objective of these population collection operations was not (and is not) to study the family.

However, as Alain Collomp (1992:13) states "of the different meanings that can be given to the word 'family' (from a conjugal couple and their children to a large kinship group including uncles and cousins), it is that of household to which the history of the forms of the family refers." It is at the level of households that statistics on family are produced.

The pertinence of the household has been widely discussed (Netting *et al.*, 1984; AMIRA, 1987; Sala Diakanda, 1988; Lacombe and Lamy, 1989; McDonald, 1992; Burch, 1993). On the one hand, it is possible to cohabit without being related (particularly in urban areas); on the other hand, it is possible to be related and to live apart: for example, the phenomenon of noncohabitation of spouses and dependent children is relatively common in African cities, particularly in cases of polygyny (Locoh, 1990). Moreover, the residential unit does not necessarily coincide with production and consumption units, which themselves can be distinct and whose rules of composition vary according to societies; several levels of production and consumption can coexist within the same society (Gruenais, 1981). For the rural Bwa in Mali, the proportion of units headed by a female falls from 27% to 3% according to whether the household or the domestic group is considered (Hertrich, 1996). The reality thus observed is very different, particularly concerning the concept of the household head.

Conveying above all residential arrangements,[2] the household is one of the highly significant ways for grouping individuals: that individuals within a society, whether they are related or not, gather in one location to live daily for a given time does not occur at random; it definitely signifies something that is necessarily translated into *one* social reality and *one* real-life experience, a reality that can be both domestic and economic. Indeed, "the problem here is not so much the type of unit but the shifts in meaning to which it can be subject during interpretation" (Pilon and Hertrich, 1995:2).

Regardless of the society and the chosen definition, it is clear that household and family cannot be interchangeable concepts, nor can they be totally autonomous and independent of the rest of the kinship

[1] According to R. Mbuh (1994:64), "to bring this notion [of family] to traditional African society refers to a system of an extended family, favoring lineage." According to Christophe Sawadogo (1992:63), "in the traditional African context a definition of family should surpass the scope of household and home and should be placed at the level of lineage and clan."

[2] But which also convey different realities according to the reference population used during data collection: the approaches differ depending on the country as some consider the *de jure* population, whereas others consider the de facto population.

group and/or of society. Even if they coincide historically or occasionally, they are fundamentally different, but not exclusive, realities. It is therefore necessary each time to specify what is meant by the terms household and family.

2. Theories on Family Evolution

The family began to be studied in a scientific context during the second half of the 19th century (see the works of Le Play, 1871; Morgan, 1887; Durkheim, 1888; Engels, 1948) and was developed during the 1920s under the influence of American sociologists of the Chicago *interactionist* school (Michel, 1986). The post–World War II period saw the emergence and generalization of a global theory of social change, the modernization theory, of which the demographic transition theory (see Chapters 68, 69) and the nuclearization of the family theory continue to nourish scientific debate.

The argument of a convergence of family systems toward the nuclear model originated in Talcott Parsons' work on the American family, and from a structuralist-functionalist approach (Parsons and Bales, 1955), it was expanded by William Goode (1963), in particular in *World Revolution and Family Patterns.* According to this theory the modernization process, through industrialization and urbanization, ineluctably results in the transition from the traditional extended family to the modern nuclear family. This development is both a change in the structure and function of the family and a change in the male and female roles within it. The nuclear family, which is disconnected from the rest of the kinship group, is presented as the family model best adapted to the economic conditions of contemporary (American) society.

However, as stated by Patrice Vimard (1995:145), "Since 1970, family studies that have been developed according to relativistic and historical perspectives make it possible to strongly qualify or invalidate the nuclearization theory by showing that on the one hand, the extended family was the dominant, but not exclusive, model of the old family (Laslett and Wall, 1972), and on the other hand, that the nuclear family is not the ultimate and universal model of the modern family (Rao, 1973)." It is not necessarily where we expect to find it.

For industrialized Western countries, other theoretical avenues have been and are being explored in the light of significant and rapid changes in the family sphere over recent decades: the increase in the number of people living alone, of single-parent families (Lefaucheur, 1992), and of reconstituted families (Thery, 1987) are examples. The role of kinship and family solidarity has been rediscovered and reaffirmed: nuclear families do not seem so independent of the rest of the kinship group (Segalen, 1984, 1996). Some authors, such as Edward Shorter (1975, 1977), have announced the advent of the postmodern family, whereas others (Roussel, 1986) are leaning toward a theory of diversification and of multiplicity in family models.

For developing countries, family situations and their evolutions are also characterized by complexity and diversification that are a far cry from a linear path (McDonald, 1992; Vimard, 1993). In Nigeria (Oke, 1986), the Arab world (Behnam, 1985), and India (Lardinois, 1986), extended families and traditional family values can accompany the modernization process.

All things considered, we can only agree with Peter McDonald (1992:25) that the "Convergence theory not only failed to predict the path of changes in family systems in the non-Western context, it also failed to predict the movement away from the idealized version of the conjugal family that occurred in the West itself in the past 20 years." In the absence of a satisfactory new theoretical scope, much remains to be done in family studies before any certainties that do not reiterate the appearance of new myths and stereotypes can be put forward (de Oliveira, 1992).

3. Household and Family Demography

The study of the family is resolutely at the crossroads of several disciplines. In the wake of sociology and history, other social sciences have progressively joined this field: psychoanalysis, psychology, law, political sciences, ethnology, economy, and demography (de Singly, 1992) have all shed light on the issue.

Research in historical demography began in the 1950s (Henry, 1953–1955; Rosental, 2003) and has greatly contributed to improving the understanding of the dynamics and evolution of families in European countries, making it possible to expose a series of false impressions of the past and to call into question the myth of the extended family as a prop for high fertility, thereby rejecting the nuclearization theory (Le Bras, 1992; Segalen, 1996).

Demographers' more recent interest in the household and the family is partly associated with the need to contextualize demographic phenomena that are generally understood and analyzed at an individual level; this has meant crossing from a basically descriptive approach to a more explanatory approach in the hope of finding an observation unit that would be the entrance to an explanation of demographic behavior.

The relationships between family structure and fertility are assuredly those that have been most studied (Ryder, 1984; Bartiaux and Tabutin, 1986).

Demographic research on household and family as a separate research theme emerged in the 1960s and had grown in volume by the end of the 1970s. Meetings and international scientific programs dedicated to this theme multiplied: Committee for International Cooperation in National Research in Demography's program in 1979–1982 (CICRED, 1984) and the creation in 1982 of an International Union for the Scientific Study in Population (IUSSP) scientific committee on Family Demography[3] that organized two international seminars, the first in New York in 1983 on the methods and applications of family demography (Bongaarts *et al.*, 1987) and the other in Berlin in 1984 on the latter phases of the family life cycle (Höhn *et al.*, 1989). In 1986, the IUSSP created a new scientific committee on the same theme,[4] which in its turn organized three international seminars, the first in 1987 in Honolulu on changing family structures and life course in developing countries (Berquo and Xenos, 1992); the second in Vaucresson, also in 1987, on new forms of family life in developed countries (Prioux, 1990); and the third in Tokyo in 1988 on theories on family evolution.

An entire field of research called *household and family demography* that produced concepts and methods was thus progressively created. One of the leaders, Thomas Burch (1993:1), defined its outlines, "Union formation and dissolution; family and household structure and change; and kinship" (more details can be found in Burch, 1979). All the demographic phenomena obviously affect the size, structure, and dynamics of households and the family: fertility, nuptiality, mortality, and migration each have their own effects.

Although research is essentially centered around the household and the (conjugal) family, demographic sources make it possible to analyze different types of units of individual membership at a given time. We can list them as follows: the residential unit, the residential family group, the family nucleus (combination of husband/wife[s]/children that can be single parent, nuclear, or polygynous) residential or not (as it can be distributed between several residential units), the domestic group (that can be spatially fragmented), and the matricentric (mother–children) unit.

As a complement to other disciplines, demography seems better equipped to evaluate structures and their dynamics. For example, it can give statistical information on the proportion of individuals living in or belonging to a unit (residential or domestic) where polygyny is practiced (and this according to sex, age, and other individual characteristics), on the evolution of the residential forms of domestic units, on the significance and characteristics of situations of noncoresidence of spouses, on matched characteristics of spouses (age, ethnic group, religion, education, employment etc.), and on households with elderly people, etc. Such quantitative aspects if correctly interpreted, can be a mine of information. Demography can also grasp changes in the family and population structure of residential and domestic units, as a consequence of changes in demographic parameters (mortality, fertility, nuptiality, and migration), structural changes that can influence collective functions, roles, and individual status.

With reference to domestic groups the sociologist Martine Segalen (1996:34) rightly states that their "structure . . . is interesting as it reveals a certain form of organization that regulates the transmission of cultural values and practices which articulate family and work, family and power, family and knowledge." She then illustrates her words with a quotation by William Goode from his book *The Family*:

> "The various forms of the household have a number of implications for family interaction. They help to determine, for example, the chances of more or less intimate social relations among members of the kinship group. Thereby, these structural patterns shape in part the process of strain and adjustment among relatives. Various role relations may have to be spelled out in detail, if the household includes certain relatives. For example, if the household usually includes a man and his mother-in-law, there may be rules requiring much reserve or noninteraction between the two.
>
> Socialization patterns are also affected by who is included in the household. A mother-in-law, or a young boy may go to his mother's brother's house to grow up. A young child in a polygynous household sees a wider range of adult models intimately than he could observe in a nuclear family. Those who share the same household are likely to share the same budget, and thus economic exchanges are partly determined by the forms of the household." (Goode, 1964:44–45).

The 1980s produced "A plethora of formal methodological, as well as substantive-descriptive, studies of the family" (Höhn, 1992:4). Much progress was made in several fields: in conceptualization, notably with the recognition of the difference between household and family (Burch, 1993)[5]; in measurement, with the development of comparative indicators (Burch,

[3] This committee's president was John Bongaarts, and its members were Thomas Burch, Kenneth Wachter, Gustav Feichtinger, Hervé Le Bras, and Elza Berquo.

[4] Elza Berquo was the president of this committee, and the members were Charlotte Höhn, Louis Roussel, Jan Trost, and Peter Xenos.

[5] "A key step has been the full realization of the difference between household and family, a distinction somewhat blurred in early demographic work, due to a Western tendency to identify the two, and resulting census conventions . . ." (Burch, 1993:3).

1980; Burch *et al.*, 1987); in classification (Laslett and Wall, 1972); in modeling the interrelations between basic demographic factors and household and family dynamics at both the macro- and microlevels (Goodman *et al.*, 1974; Le Bras, 1992; Bongaarts *et al.*, 1987; Mason, 1987); and in attempts at generalization using empirical studies.

Despite this, all the problems are far from being overcome. Comparative studies of the household and the family in Europe, and more so at a global level, still come up against numerous obstacles: the great heterogeneity of definitions, classifications that are not always made explicit and that often make comparison very difficult, insufficient published statistics, etc. (Locoh, 1988a; CEPR, 1991; de Oliveira, 1992; Lefranc, 1995; Wall, 1996). Moreover, household and family demography essentially refers to Western societies and thus to the corresponding acceptance of the (nuclear) family. Many of these methodological and analytical tools have often been shown to be unsuitable for other societal contexts, particularly in Africa.

Research on household structures and their development remains rudimentary in developing countries, and the lack of knowledge is particularly evident in sub-Saharan Africa. This could be owing to several reasons: the weight of studies on kinship systems, the emphasis placed on fertility studies, and an often immutable perception of African societies as "eternally rural and traditional" (Pilon, 1996a; Cordell and Piché, 1997).

a. Kinship Ties: Essential But Neglected Information

Regardless of the definition retained for the household or the family, information on kinship ties is obviously essential in order to analyze family structures and household formation. Its degree of precision to a great extent determines the possibilities for elaborating classification, which then have varying degrees of analytical pertinence.

Kinship ties, combined with other variables—such as sex, age, marital status, and type of activity (educational or economic)—are required to identify family status and roles.

To know, as precisely as possible, who lives or resides within a household can thus make it possible to study certain phenomena, such as the circulation of children and the migration of adults that give an idea of the level of family solidarity. Who are the fostered children and the migrants within the household? To whom are they related: the household head, his (or one of his) spouse(s), or another member? What are their characteristics (age, marital status, level of education, activity)? These questions are particularly relevant to urban households, who often play host, and it is important for these households to be able to perform this role, if only partially.

Even though numerous societies experience high levels of matrimonial instability, and in Africa polygyny is widely practiced, it is important within households where the head is married to be able to distinguish the couple's children from those belonging to each spouse from a previous union, on the one hand, and those belonging to different spouses, on the other hand.

For censuses and large surveys, the fact that information on kinship ties is first and foremost collected in order to facilitate the identification of individuals during data collection and to help control the coherence of other variables rather than for analysis, largely explains both the loss of information and the underexploitation of data.

Very frequently, the types of kinship relations used for codification (in the absence of precodification) is far less detailed than is the information gathered during data collection; this inflates the figures of the "other relatives" category. The examination of 32 African censuses shows that for countries that have modified their systems between the last two censuses, the trend has been simplification rather than refinement.

Regardless of the precision in kinship ties, statistical tabulation on this variable remains rare and far below the possibilities. The simple cross-tabulation of kinship ties with age, sex, and marital status can be a method of refining the codification of kinship ties that are too basic and a way of helping to identify and characterize the family status of individuals within the household. A United Nations document (1980) on principles and recommendations for censuses proposes the production of several such tables.

b. Typologies

In a comparative, temporal, and spatial theorization perspective, sociologists and historians have attempted to define household and family types. But there is no unique universal typology. Depending on the phenomena or the logic that we wish to expose, the criteria used are very diverse and can be demographic, economic, social, cultural, etc.: for example, rules and practices regarding authority, inheritance, residence, kinship system, nature of interpersonal relations, family roles (in particular between husband and wife); duties undertaken by the household or family, production method, and external relations.

Although the study of household structure is not an end in itself and is not adequate to take into account all the dimensions of the family (a household or family

TABLE 90–1 The Typology of Households According to Peter Laslett (1972)

1. Solitary	a. Widows or widowers
	b. Single people
2. Household without a family	a. Brothers and sisters co-residing
	b. Other relatives co-residing
	c. Unrelated co-residents
3. Simple household	a. Married couple without children
	b. Married couple with children
	c. Widows with children
	d. Widowers with children
4. Extended household	a. Extended upward
	b. Extended downward
	c. Extended laterally
	d. Combination of the above
5. Multiple household	a. Multiple upward
	b. Multiple downward
	c. Multiple in both directions
	d. Frereches (coresident siblings)
	e. other
6. Household of indeterminate structure including some kinship ties	

cannot be reduced to its structure only), the household typology introduced by Peter Laslett (1972) at the beginning of the 1970s is a scientific contribution of great value (Table 90–1). Historical research that has resulted from it has shed light on the pre-existence of nuclear structures at the time of the industrial revolution, which has contributed to the nuclearization theory being called into question.

Demographers essentially use this type of typology in order to analyze the morphology of households and families at a given time (who lives with whom?); the term "structure" is also used. The expressions family structure, household or family types, and family system are often used indifferently but incorrectly.

Developed for a historical study of European societies, Peter Laslett's classification has some shortcomings and ignores a certain number of existing family formations (grouped in his no. 6 category). It does not take into consideration the practice of polygyny (an essential phenomenon in African family systems) and also leaves the status of "other relatives" and the unrelated unclear.

As Patrice Vimard (1995:9) states, "there are no typologies that are a priori good or bad"; it all depends on our objectives and the specificities of the populations studied. Multiple criteria can be taken into consideration: the number of generations represented and their organization, the number of conjugal families (and their monogamous or polygynous character), the presence of other related or nonrelated members, the marital status and sex of the household head, etc. As previously mentioned, the level of precision of the kinship ties determines the level of refinement for the typology. The number of categories used can also vary significantly (from two to several dozen), depending on whether a more descriptive or overall character is desired.

Typologies developed at the end of an analysis frequently aim to expose the level of nuclearization, which is conveyed by the opposition between nuclear or restricted households, on the one hand, and extended or enlarged households (or families), on the other hand, adjectives that are frequently not defined.

The diversity of the typologies used as well as the heterogeneity and the vagueness of the terms that define them frequently make comparisons between populations very difficult. Without undermining the value of this diversity (each classification is pertinent to the population under question, depending on the issue under discussion) and while waiting for a minimal typology acceptable to all, efforts to clarify the terminology should be undertaken in order to facilitate comparative studies, which are essential for increasing knowledge.

c. For a Dynamic Approach

Analyses of the distribution of household types at a given date are often justifiably reproached for causing incorrect interpretations, conclusions that are too hasty in terms of family models or in terms of changes by comparing the distribution at different dates (Carter, 1984; McDonald, 1992): "the typological classification annihilates the temporal dimension" (Segalen, 1996:37). Demographic dynamics, in particular, result in the fact that, even in a society marked by the predominance of the extended family system, there is always a greater proportion of simple families owing to life-cycle events (birth, marriage, mobility, death) and their timing, as well as constraints in the age and sex structure of kinship groups. Furthermore, the same family configuration can result from different processes and can also be accompanied by different functions, be it between different populations or different periods, and inversely. Therefore, "reintroducing the long-term makes it possible to understand whether the configurations . . . are structural or conjunctural. Are the households always multiple or always conjugal? Do they experience a typical succession of phases, and if so, what is its model? How can we explain the divergence between the norm and the average insofar as a determined model has been acknowledged as preponderant in a given society?" (Segalen, 1996:37).

The simple matching of household types with the household head's sex and age, too infrequently performed, makes it possible to undertake an initial evaluation of the dynamic character of their structures. But the longitudinal approach (and including other types of data) is obviously the best. Specific concepts and methodologies have been developed in this direction, by sociologists and historians: the concepts of "family life cycle," which refers to a succession of stages, and "life course," according to which the family experiences a process made up of transitions and in which individual, family, and social periods overlap (Hareven, 1977, 1991) (see also Chapter 88 of this volume, which discusses the life cycle).

II. HOUSEHOLDS AND FAMILIES IN DEVELOPING COUNTRIES: STATE OF THE ART

An attempt at producing a synthesis of knowledge on the size, composition, and evolution of households in the developing countries is confronted by the paucity and inadequacy of the existing quantitative data (under exploitation of census and survey data, heterogeneity of definitions, terminology, and classifications). Data on the past (even the recent past, but the more distant past in particular) is particularly lacking and limits the possibilities of a historical approach (de Oliveira, 1992; Cordell and Piché, 1997).

By using the most recent statistics available at time of writing,[6] I will attempt to describe the overall situation, incorporating household size, family composition, female household heads, and "fostered children."

1. Household Size

The three indicators used here to characterize the sizes of households, the average size, the proportion of one-member households, and the proportion of households with at least nine members[7] vary significantly according to the large developing regions of the world and between the countries.

The average household size is between four and six people for the large majority of countries (Table 90–2; Fig. 90–1). The countries of Latin America and east Asia have the smallest households, with relatively few variations between countries. In contrast, diversity is particularly significant in Africa (where the highest values are most frequently found) and in western and central Asia.

The average size of African households varies between 3.6 in Ghana and 9 in Senegal. Households seem to be generally smaller in east and south Africa compared with north and west Africa (with the exception of Ghana).

The differences between rural and urban zones also vary. They are quite small in Latin America; the greatest differences are in Africa and central Asia (Turkmenistan, Kyrghyzstan and Kazakhstan). It is moreover interesting that among the countries considered here, a significant proportion (40%) of them have larger urban than rural households.

The proportion of one-member households (Fig. 90–2) vary even more significantly than does the average size: it is only 1.1% in rural Bangladesh, but 30.2% in urban Ghana. This type of household is found most frequently in urban areas: this is the case for three-quarters of the countries and for nearly all the African countries. All the Asian countries (with the exception of the three central Asian countries) and the north African countries have lower proportions. sub-Saharan Africa has significant variations, both between the countries and between urban and rural areas. However the differences according to area of residence are greatest in Turkmenistan, Kyrghyzstan, and Kazakhstan, with much higher proportions in the cities.

The proportion of large households (made up of at least nine members) also varies significantly: from 1% in rural Kazakhstan to 47.5% in rural Senegal (Fig. 93–3). In the large regions, the lowest proportions are found in Latin America, Asia, and east and south Africa. The countries where the proportion of large households is more than 20% are mainly in west Africa: Yemen, Namibia, Guinea, Côte d'Ivoire, Burkina Faso, and Senegal.

Although rudimentary, the variations in these three indicators—whether between large regions, among countries within the same region, or between rural and urban areas—suggest a great diversity in residential family structures.

According to a study by the United Nations (1990), which makes a comparison between the 1970s and the 1980s, the average size of households has declined in every region with the exception of north Africa and west and south Asia. However, other studies (Locoh, 1988a) and the results of the most recent Demographic and Health Surveys reveal the opposite in sub-Saharan Africa, with an increase in household sizes, with the exception of Kenya, Zimbabwe, and Botswana, where fertility decline is now well underway.

[6] The Demographic and Health Surveys supply the database that offers the widest geographic cover (50 countries). See the Macro International Web site at *www.measuredhs.com*.

[7] Indicators from tables published in national survey reports.

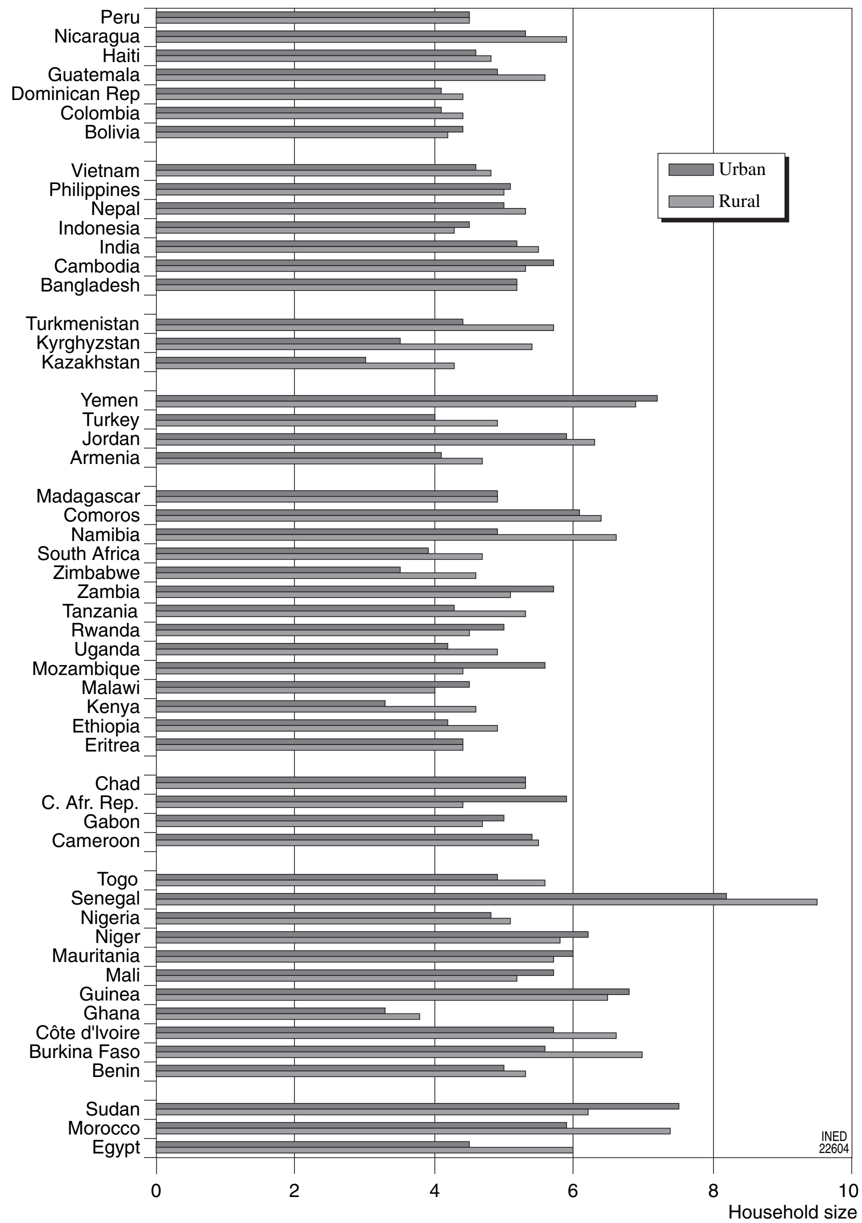


FIGURE 90–1 Average household size according to the Demographic and Health Surveys of 1990–2000.

2. The Family Composition of Households

The paucity of appropriate data is here again more apparent. Few publications of census and survey results provide tables on household structure and particularly tables that match it with the household head's sex and age. The United Nations' assessment (1993) published in its *Demographic Yearbook* is eloquent (Tables 90–3, 90–4).

Beyond this initial difficulty, the comparative approach also faces the problem of diversity in definitions and terminology. The possibilities for statistical synthesis therefore are very limited. In addition, books and articles offering an overall view of household and family, at a regional or world level, provide little quantitative syntheses on household composition, and they generally rely on the results of occasional studies or on limited data (Netting *et al.*, 1984; de Vos, 1987; Locoh, 1988b; Ekouevi *et al.*, 1991; Berquo and Xenos, 1992; Vimard, 1993; Cho and Yada, 1994; Pilon, 1996a; Adepoju, 1997; Pilon *et al.*, 1997). Generalization of the observed results involves obvious risks of once more producing stereotypes or misleading ideas. Nevertheless, knowledge that already makes it possible to draw

TABLE 90–2 Average Household Size and Proportion of One-Member Households and Households with at Least Nine Members According to the Demographic and Health Surveys of 1990–2000

Country	Average Size		Proportion (%) of households				Survey Year
			One-member		At least 9 members		
	Rural	Urban	Rural	Urban	Rural	Urban	
North Africa							
Egypt	6	4.5	3.7	6.2	15.6	3.7	2000
Morocco	7.4	5.9	3.8	5.6	30	13.3	1995
Sudan	6.2	7.5	2.8	1.9	20.6	34.7	1989–1090
West Africa							
Benin	5.3	5	11.5	12.4	14.6	12.5	2001
Burkina Faso	7	5.6	2.9	11.9	29.3	17.4	1998
Cøte-d'Ivoire	6.6	5.7	10.8	12.9	27.5	19	1998–1999
Ghana	3.8	3.3	23.9	30.2	4.9	2.7	1998
Guinea	6.5	6.8	2.8	9.2	23	27.7	1999
Mali	5.2	5.7	6.6	9.5	14.1	18.3	2001
Mauritania	5.7	6	7.3	4.1	15.7	19.7	2000/2001
Niger	5.8	6.2	4.1	8.4	17.1	22.7	1998
Nigeria	5.1	4.8	11.2	11.4	12.8	9.7	1999
Senegal	9.5	8.2	1.7	8.1	47.5	39.1	1997
Togo	5.6	4.9	11.4	14.5	17	12.7	1998
Central Africa							
Cameroon	5.5	5.4	11.8	14.4	17.6	17.3	1998
Central African Republic	4.4	5.9	15.3	10.2	8.4	21.4	1994–1995
Chad	5.3	5.3	8.6	12.9	13.8	16	1996–1997
Gabon	4.7	5	19.3	19.6	14.6	16.8	2000
East Africa							
Comoros	6.4	6.1	1.6	4.5	22.1	21.8	1996
Eritrea	4.4	4.4	9.6	12.1	4.6	7.6	1995
Ethiopia	4.9	4.2	3.6	12.8	6.8	6.4	2000
Kenya	4.6	3.3	12.5	26.3	7	2.9	1998
Madagascar	4.9	4.9	6.6	6.9	9.3	9.9	1997
Malawi	4	4.5	8	8.1	5.2	6.7	2000
Mozambique	4.4	5.6	10.2	7.1	7.2	15.6	1997
Rwanda	4.5	5	5.9	6.6	4.1	10.2	2000
Tanzania	5.3	4.3	7.1	15.4	12.2	6.4	1999
Uganda	4.9	4.2	10.1	16.4	10.4	7.8	2000–2001
Zambia	5.1	5.7	6.7	4.9	11.5	16.9	1996
Zimbabwe	4.6	3.5	12.2	21	7.6	3.4	1999
Southern Africa							
Namibia	6.6	4.9	7.2	11.1	25.4	12.8	1992
South Africa	4.7	3.9	10.7	12.9	8.8	4.2	1999
Western Asia							
Armenia	4.7	4.1	7	9.3	3.4	2.9	2000
Jordan	6.3	5.9	7.1	4.1	24.7	18.4	1997
Turkey	4.9	4	5	5.3	10.8	3.6	1998
Yemen	6.9	7.2	2.9	2.7	29.3	31.3	1997
Central Asia							
Kazakhstan	4.3	3	5.9	18.8	2.9	1	1999
Kyrghyzstan	5.4	3.5	2.5	16.1	9.2	1.2	1997
Turkmenistan	5.7	4.4	3.9	9.5	12.1	4.8	2000
South Central Asia							
Bangladesh	5.2	5.2	1.6	1.1	7.4	7.7	1999–2000
India	5.5	5.2	3.1	3.2	11	8.8	1998–1999
Indonesia	4.3	4.5	4.3	6.1	2.5	4.2	1997
Nepal	5.3	5	3.8	5.1	9.7	8.1	2001

TABLE 90–2 (*continued*) **Average Household Size and Proportion of One-Member Households and Households with at Least Nine Members According to the Demographic and Health Surveys of 1990–2000**

Country	Average Size		Proportion (%) of households				Survey Year
			One-member		At least 9 members		
	Rural	Urban	Rural	Urban	Rural	Urban	
Southeast Asia							
Cambodia	5.3	5.7	1.9	1.8	7.6	11	2000
Philippines	5	5.1	4.3	3.8	7.7	8.3	1998
Vietnam	4.8	4.6	4	3.5	3.9	3.6	1997
Central America and the Caribbean							
Dominican Republic	4.4	4.1	6.2	8.5	3.8	2.3	1999
Guatemala	5.6	4.9	3.5	4.8	14.4	8.1	1998–1999
Haiti	4.8	4.6	8.8	7.1	8.5	7	2000
Nicaragua	5.9	5.3	4.2	3.8	17	11	1997–1998
South America							
Bolivia	4.2	4.4	12.8	8.2	4.9	4	1998
Colombia	4.4	4.1	8.5	7.4	5	3.3	2000
Peru	4.5	4.5	10.5	6.9	6.2	4.9	2000

TABLE 90–3 Number of Censuses for which Data on Households According to Size and Number of Family Nuclei Are Available in the Database of the Demographic Yearbook, by Large Geographic Regions (Censuses for the 1946–1990 period)

Continent	Country or territory	Total	No. of censuses		
			No. of households	Households according to size	No. of family nuclei
World	183	513	70	373	70
Africa	36	55	11	40	4
North America	36	113	17	82	14
South America	13	36	8	27	1
Asia	36	101	10	77	14
Europe	39	140	9	108	23
Oceania	23	68	15	39	14

From UN, 1993:64.

TABLE 90–4 Number of Countries for which Data on Household Characteristics According to the Household Head's Sex and Age Are Available, by Large Geographic Region

Region	Households according to the household head's age, sex, and household size	Households according to the household head's age, sex, and marital situation	Households according to the household head's sex, household size, and relation with the household[a]	Households according to household head's age, sex, and the household composition[b]
Total	38	30	20	39
Africa	3	3	2	–
North America	17	4	6	14
South America	4	2	2	6
Asia	4	5	3	6
Europe	9	15	5	10
Oceania	1	1	2	3

[a] Some countries specified the relation but not the size of the household.
[b] Composition means nuclear, enlarged, or complex households.
From UN, 1993:65.

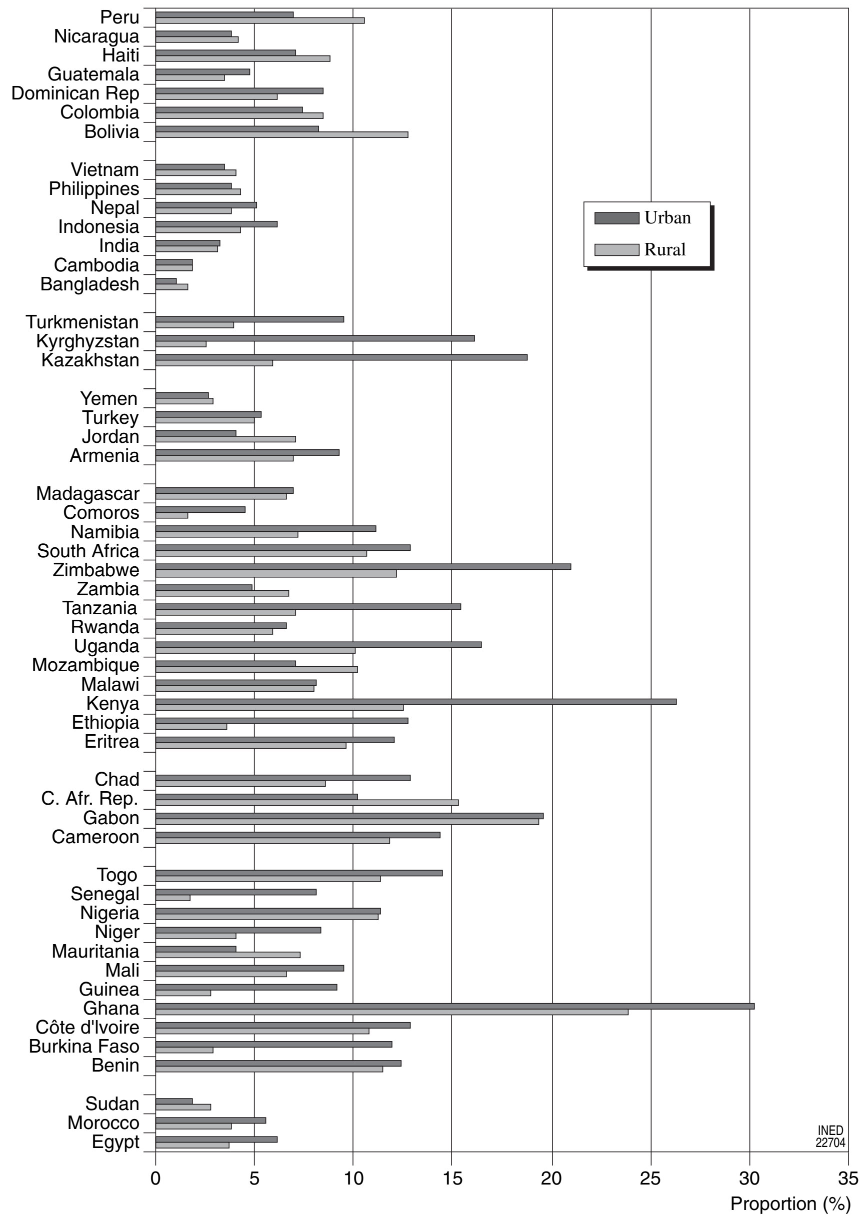


FIGURE 90–2 Proportion of one-member households according to the Demographic and Health Surveys of 1990–2000.

some conclusions, particularly concerning the convergence theory of family systems toward the nuclear model, is slowly building up.

In rural areas and particularly in private income economy zones, households and individuals are demanding more autonomy, be it financial or regarding marriage (Dozon, 1986; Berquo and Xenos, 1992; Gautier and Pilon, 1997). This process—which results in increasing tension in domestic group/rest of the kinship group, older/younger relations, and male/female relations—does not, however, result in the spread of the conjugal family model.

The role of host frequently performed by urban households, which maintains their interdependence with the rest of the kinship group (fostering of boys for schooling and girls for domestic help, which will be discussed below; accommodating adult migrants; etc.), often results in a structure that exceeds the nuclear structure. Generally smaller than those in rural areas, these urban households are easily composed of more extended and diversified relations. Furthermore, studies carried out in Abidjan show that "while the nuclear family is predominant at the lowest social levels (and here it represents more of a solution forced

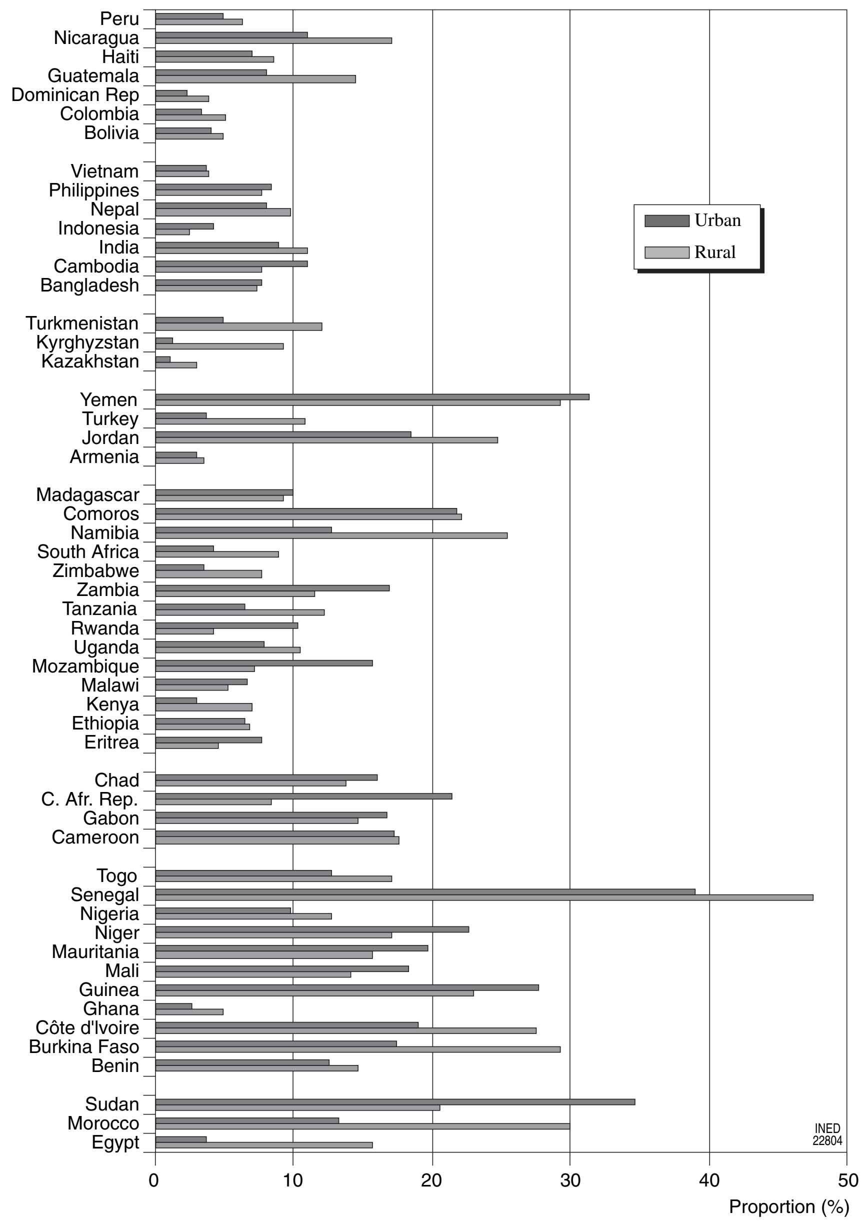


FIGURE 90–3 Proportion of households with at least nine members according to the Demographic and Health Surveys of 1990–2000.

by poverty than an actual choice), it is the groups that fit best into modern socioeconomic structures that perpetuate the practices of family interdependence and that recreate enlarged families" (Vimard, 1993:105–106).

However, the economic crisis significantly calls this logic into question. The stretching of host duties to the extreme in urban households affects children as much as it does adult migrants who arrive into an increasingly difficult and uncertain job market. Thus, the surveys conducted in the cities of Benin in 1990 and 1992 showed that the proportion of households that have accommodated people arriving from rural areas had fallen in 2 years from 24.1% in 1990 to 6.5% in 1992. Indeed, the situations are very diverse. According to the society and the context, the economic crisis has resulted in selective demographic offloads within the kinship group (Abidjan) (Vimard, 1997) and in the reinforcement of family relations (Dakar) (Antoine *et al.*, 1995).

The emergence of new types of unions, notably accompanied by non-coresidence of spouses but also by an increasing weakening of unions, the accentuation of temporary migration (male and also female) contributes, along with other factors, to the increase in the number of single-parent households in most countries (Ono-Osaki, 1991) and to the emergence of female household heads (see below).

Two other phenomena should also be mentioned because of their specific effects on household structure, namely, AIDS and the movement of refugees and displaced persons.

AIDS significantly affects households in multiple ways. The sexual aspect of the disease particularly contributes to the weakening of unions, especially for women; single-parent situations are increasing. Interfamily mobility is accentuated, whether of sufferers themselves or children being taken into the care of their grandparents or others. Circumstances can also force AIDS orphans on to the streets: whether it be as a result of exclusion or abandonment, the family's inability to cope financially, or the absence of other solutions. Faced with family destructuralization, strategies for reconstituting families are reactivating interdependence practices, but the increasing financial burdens are making this more selective and difficult (Hunter, 1990; Barnett and Blaikie, 1992; Delcroix and Guillaume, 1997).

The family consequences of the phenomenon of refugees and displaced persons (which has particularly increased in Africa) are also very significant: deaths, broken households, abandonment of the weakest, women raped and abandoned in camps, difficulty or impossibility of reconstituting conjugal families, and returning to the area of origin, etc. (Khasiani, 1989; de Renty, 1993). The increase in this phenomenon (UNHCR, 1994), and its near structural character in some regions, must lead us to further question its long-term effects.

Therefore, beyond the above-mentioned problems of the heterogeneity of definitions, plurality and complexity are without question the key words for the current state and development of households and families in the developing countries. A study by Economic Commission for Latin America and the Caribbean (ECLAC, 1994:78) on the social panorama of Latin America thus describes "the growing heterogeneity of household structure. Family structure is heterogeneous and varies according to the country, the urban or rural location of the family, and the level of poverty."

Nuclear households are not a new situation (Kabir, 1980; Charbit, 1987). Historical studies often call into question perceptions of the traditional family, particularly in terms of its structure. This was the case for China, for example, where it appears that the extended forms were not predominant in the past (Cho and Yada, 1994). The association between modernization and the nuclear family model is far from being confirmed everywhere. In the numerous countries experiencing a decline in fertility, the reduction in household size that it entails does not result in a process of nuclearization. Indeed, new domestic and residential arrangements and family reconstitution can be observed, associated with a redefinition of social relations and family roles (between sexes and between generations).

3. Female Household Heads

Although it has recently increased, this phenomenon is not new. It has existed in several societies since ancient times, notably in those in which matrilinearity is proven, in which women have the right to inherit and to possess land (Ono-Osaki, 1991; El Khouri, 1996). Diverse factors are behind it: husband's death, matrilinearity, marital instability, absence of migrant husband or the wife's own migration, practice of non-cohabitation of spouses,[8] households without an adult male (Charbit and Petit, 1996), and the woman's own decision (in terms of strategy of autonomy), and in a more artificial way, census and survey methods[9] contribute to females being considered as household heads, although sex roles usually confine them to a status of dependence on males. Indeed "according to the United Nations' recommendations, the male should systematically be declared household head within a couple" (Tichit, 1994:19) unless the female lives alone or without the presence of an adult male. These situations result in extremely diverse family and economic realities. Becoming a female household head can be the result of choice or circumstances, it can also reflect a strategy of autonomy as well as socioeconomic precarity (Locoh, 1988b; Ono-Osaki, 1991; Bisiliat, 1996).

Whatever the case, in both urban and rural zones an increasing number of women are taking on the role of household head. The development in male labor migration, growing marital instability, increasing practice of non-coresidence of spouses in cities, and a general female emancipation process make up the factors that explain the development of this phenom-

[8] In a study of the Bwa of Mali, who practice non-cohabiting polygyny, Vêronique Hertrich (1996) shows that the difference in the increase of "female households" could be considerable according to whether domestic groups or residential units are considered: the status of head is given to females in only 3% of cases for the latter, but in 27% of cases for the former.

[9] Because of the residential dimension of the household and the criteria of duration of absence, which leads to the statistical exclusion from a household of any person absent for more than a certain period (often 6 months).

enon over recent decades. However, the "tendency is possibly much older, dating to the socioeconomic changes initiated by colonization" (El Khouri, 1996:21).

The proportion of female household heads varies significantly depending on the region, the country, and the area of residence, from 5% in rural Burkina Faso to 51% in urban Botswana (Table 90–5; Fig. 90–4). In the majority of countries, the proportion of female household heads exceeds 20%, a fact that confirms the significance of the phenomenon. The proportions are highest in Latin America, east Africa, south Africa, and in the three central Asian countries.

The distinction according to residential area shows that the phenomenon is more marked in cities in nearly all the regions; this is particularly the case in South America and central Asia (as well as in Vietnam). Some countries in south Africa (Botswana, Zimbabwe, Malawi, Zambia, Kenya) are exceptions, with higher proportions in rural areas, a situation that is largely owing to the scale of male labor migration.

Although published statistics rarely provide information on the structure of female-headed households, three main types of composition can be identified: women alone, single-parent households (with or

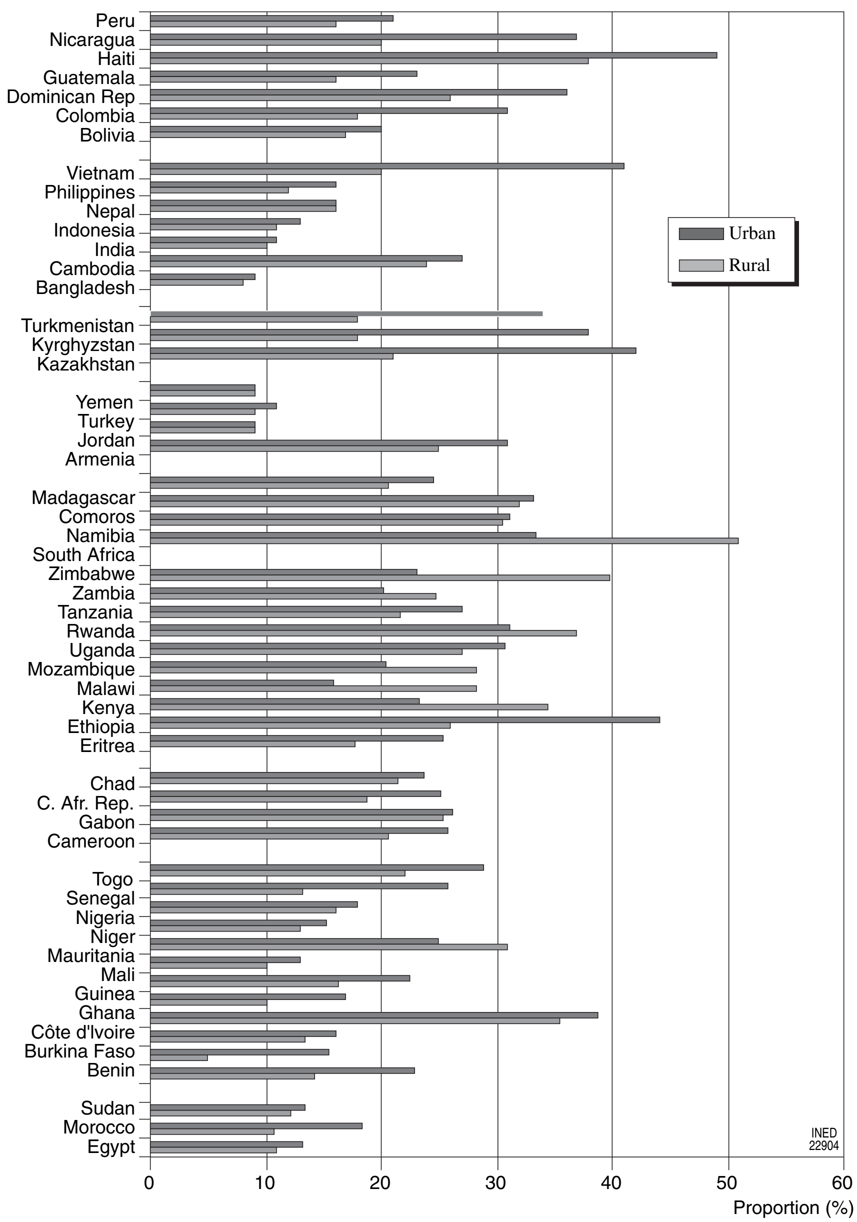


FIGURE 90–4 Proportion of households with a female head according to the Demographic and Health Surveys of 1990–2000.

TABLE 90–5 Proportion of Households with a Female Head According to the Demographic and Health Surveys of 1990–2000

Country	Rural	Urban	Total	Year
North Africa				
Egypt	10.9	13.1	12	2000
Morocco	10.8	18.4	14.5	1995
Sudan	12.1	13.3	12.6	1989–1990
West Africa				
Benin	14.2	22.9	17.5	1996
Burkina Faso	5	15.5	6.8	1998
Cøte d'lvoire	13.3	16.1	14.4	1998–1999
Ghana	35.4	38.8	36.6	1998
Guinea	10	17	12	1999
Liberia	16.3	22.4		1986
Mali	10	13	11	2001
Mauritania	31	25	29	2000/2001
Niger	12.9	15.3	13.2	1998
Nigeria	16	18	16	1990
Senegal	13.1	25.7	18.4	1997
Togo	22.1	28.9	24.3	1998
Central Africa				
Cameroon	20.7	25.8	22.4	1998
Central African Republic	18.8	25.1	21	1994–1995
Chad	21.5	23.7	22	1996–1997
Gabon	25.4	26.2	26	2000
East Africa				
Burundi	17.8	25.3	18.1	1987
Comoros	31.9	33.1	32.2	1996
Eritrea	25.9	44.2	30.8	1995
Kenya	34.4	23.2	31.7	1998
Madagascar	20.6	24.5	21.6	1997
Malawi	28.3	15.9	2.6	2000
Mozambique	28.2	20.5		1997
Rwanda	36.9	31.1	36.1	2000
Tanzania	21.7	27	23.1	1999
Uganda	27	30.8	27.5	2000–2001
Zambia	24.8	20.2	23.1	1996
Zimbabwe	39.8	23.1	33.5	1999
Southern Africa				
Botswana	51	33.3	45.9	1988
Namibia	30.6	31.2	30.8	1992
Western Asia				
Armenia	25	31	28	2000
Egypt	10	13	11	2000
Jordan	9	9	9	1997
Turkey	9	11	10	1998
Yemen	9	9	9	1997
Central Asia				
Kazakhstan	21	42	33	1999
Kyrghyzstan	18	38	26	1997
Turkmenistan	18	34	26	2000
South central Asia				
Bangladesh	8	9	8	1999–2000
India	10	11	10	1998–1999
Nepal	16	16	16	2001
Southeast Asia				
Cambodia	24	27	25	2000
Indonesia	11	13	12	1997
Philippines	12	16	14	1998
Vietnam	20	41	24	1997
Central America and the Caribbean				
Dominican Republic	26	36	32	1999
Guatemala	16	23	19	1998–1999
Haiti	38	49	42	2000
Nicaragua	20	37	30	1997–1998
South America				
Bolivia	17	20	19	1998
Colombia	18	31	27	2000
Peru	16	21	19	2000

without other persons, related or not), and women without children living with other persons (related or not). The specificity of female-headed households results in households being smaller than those headed by males.

The demographic profile of female household heads also varies significantly. Although their age structure generally shows a high proportion of elderly women (but it is generally lower than in industrialized countries), quite frequently there are also young women in some regions—in sub-Saharan Africa, Latin America, and the Caribbean (Arias and Palloni, 1999). Widowhood and spinsterhood are obviously associated with these two age categories.

According to a United Nations synthesis (1992:18) of the situation of females in the world, "In Africa, in the developed regions, and in the Asia Pacific region, approximately half of female household heads are widows. In Latin America and the Caribbean, only 28% are widows and 36% are single. In all regions, there are more married and single women than divorcees. The proportion of divorced female household heads is particularly low in Asia and the Pacific: 6% against 13% in Latin America and the Caribbean, 14% in Africa, and 16% in the developed regions."

4. Fostered Children

The common extended structure of households in developing countries is owing to the cohabitation of several family nuclei and/or the presence of others, related or not. Children are often among the former, notably owing to fosterage. The circulation of children is an old phenomenon found in numerous societies all over the world (Silk, 1987; Lallemand, 1993:41–48). Children who do not live with both parents fall into the following categories: mother present/father absent, mother present/father deceased, mother absent/father present, mother absent/father elsewhere, mother deceased/father present, and mother deceased/father deceased. A first distinction to establish is whether there is child mobility or not: does he/she come from another household, or have the parents left the household (or who are deceased)?

There are multiple reasons for all these scenarios, and the corresponding family situations are extremely diverse: illness, death, divorce, separation or parents' non-cohabitation, family support, socialization, education, relations through marriage, etc. (Lallemand, 1993). For the societies concerned, the circulation of children is a characteristic element of family systems and conforms to the logic of family interdependence and to the system of rights and responsibilities. In addition to a reinforcement of social ties, this practice, via a wider distribution of the economic costs of raising children, also appears to contribute to maintaining high fertility behavior. It is this association with fertility that has led demographers to become more interested in the phenomenon of fostered children. Regardless of the reasons for this, fosterage is one of the components of the structure and dynamics of households.

Adoption, loan, gift, exchange, transfer, circulation, mobility, and fostered child are some of the terms used[10] to indicate this plural reality; this is a little confusing because the same terms are used to describe different situations. The heterogeneity of the definitions makes comparative studies particularly difficult, as shown by the following examples taken from the literature:

Children aged 0 to 14 years, other than those of the household head, not living with their mother (Page, 1989)
Children aged 0 to 10 years, other than those of the household head, not living with their parents (Isiugo-Abanihe, 1985)
Children aged 0 to 19 years, other than those of the household head (even if one of the parents is present) (Vimard and Guillaume, 1991)
Children aged 7 to 14 years, other than those of the household head, not living with their parents but not orphaned (Ainsworth, 1990)
Children aged 1 to 12 years, not living with their mother (Bledsoe and Brandon, 1989)
Children aged 0 to 14 years, not living with their parents (Blanc and Lloyd, 1994).

In the absence of precise information on the situations observed, the expression "children's living arrangements" (Lloyd and Desai, 1992) seems the most appropriate, or at least the most neutral, term.

Specific studies and analyses carried out on demographic survey data have multiplied, shedding new light on the matter (Vandermeersch, 2000). Beyond the measurement of the scale of the phenomenon, authors have attempted to discover its principal determinants and to see to what extent this practice affects fertility behavior.

The results of the Demographic and Health Surveys currently provide the most comprehensive statistical information, with the possibility of temporal comparative analyses for a growing number of countries. The results published generally provide a household level indicator (the percentage of households in which children live without their parents), available according to place of residence, and a table giving the distribution of children according to survival and their parents' residence according to their age, sex, and area of residence. The percentage of households with children living without their parents (whether they are alive or deceased) gives an approximation of the scale of the phenomenon of child fosterage at the household level (Table 90–6; Fig. 90–5).

Existing studies concur that fosterage is particularly widespread in Africa (Antoine and Guillaume, 1986; Page, 1989; Isiugo-Ibanihe, 1991); which confirms the contents of Table 6 and exposes a large variety of situations within the African continent. The phenomenon seems to have a lesser importance in Latin American countries, even more so in Asian countries. In the majority of African countries, the percentage of households with at least one child aged under 15 years living without his/her parents is above the threshold of 20%.

CONCLUSION

This overview of household size and structure in developing countries does not really make it possible to come to any firm conclusions. On the one hand, the results available are still only partial, and on the other hand, the comparative analysis is limited owing to the data source; any interpretation must therefore be a

[10] Terms include fosterage, fostering, adoption, child relocation and transfer, child circulation, child migration, child rearing delegation (Isiugo-Abanihe, 1985, 1991), and exchange of children (Pennington and Harpending, 1993:171–199). Fostering and fosterage are the most commonly used terms.

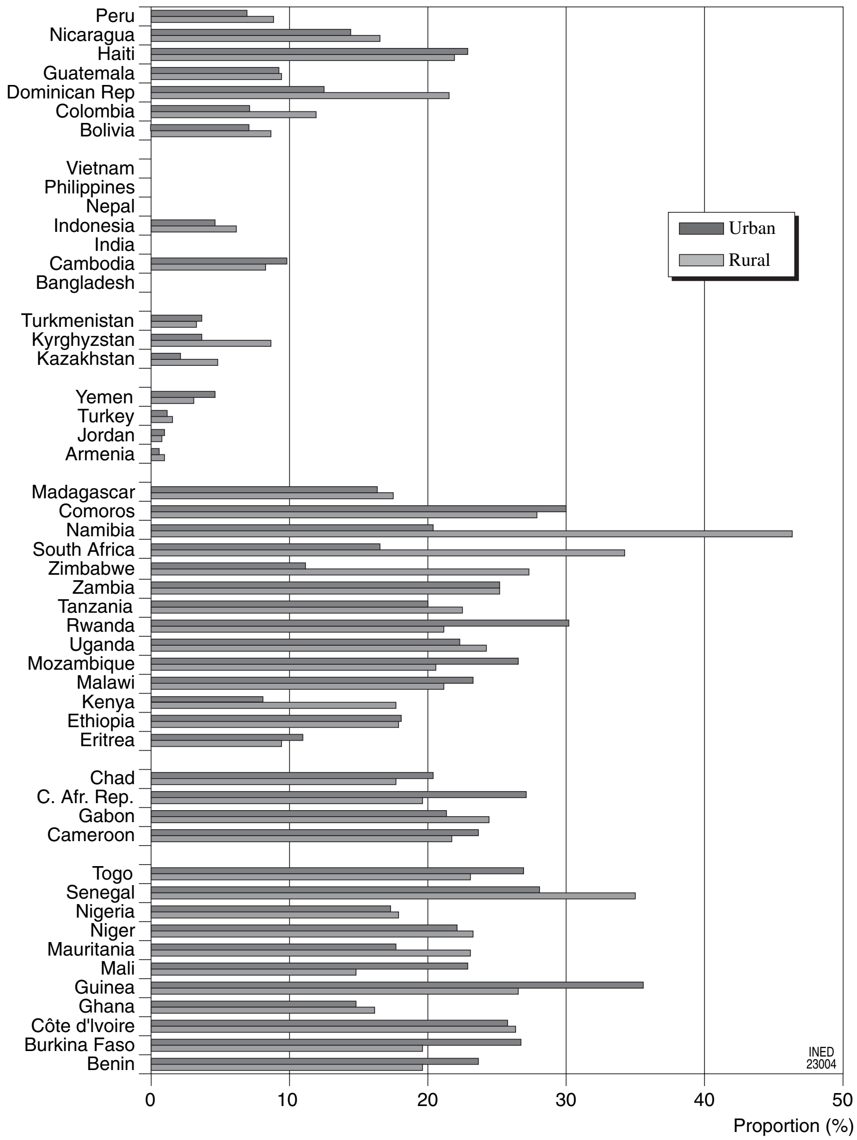


FIGURE 90–5 Proportion of households with children aged under 15 years without their parents, according to the Demographic and Health Surveys of 1990–2000.

TABLE 90–6 Proportion of Households with Children Aged under 15 without their Parents, According to the Demographic and Health Surveys of 1990–2000

Country	Rural	Urban	Total	Year
North Africa				
Egypt	1.2	1.1	1.1	2000
West Africa				
Benin	19.6	23.7	21.1	2000–2001
Burkina Faso	19.7	26.8	21.1	1992–1993
Cøte d'Ivoire	26.4	25.7	26.1	1994
Ghana	16.1	14.9	15.7	1998
Guinea	26.5	35.5	29.1	1999
Mali	14.9	22.8	16.9	2001
Mauritania	23.1	17.6	20.8	2000–2001
Niger	23.2	22.2	23	1998
Nigeria	17.9	17.4	17.7	1999
Senegal	35	28.1	32.1	1992–1993
Togo	23.1	26.9	24.4	1998
Central Africa				
Cameroon	21.8	23.7	22.4	1998
Central African Republic	19.7	27.2	22.4	1994–1995
Chad	17.6	20.3	18.3	1996–1997
Gabon	24.4	21.3	22.1	2000
East Africa				
Comoros	27.9	30	28.5	1996
Eritrea	9.5	10.9	9.8	1995
Ethiopia	17.9	18.1	17.9	2000
Kenya	17.6	8.1	15.3	1998
Madagascar	17.5	16.4	17.2	1997
Malawi	21.2	23.2	21.5	2000
Mozambique	20.5	26.5	21.6	1997
Rwanda	21.2	30.2	22.4	2000
Tanzania	22.5	20	21.8	1999
Uganda	24.2	22.4	23.9	2000–2001
Zambia	25.1	25.1	25.1	1996
Zimbabwe	27.4	11.2	21.2	1999
Southern Africa				
Namibia	46.4	20.3	37	1992
South Africa	34.3	16.6	23.9	1999
Western Asia				
Armenia	0.9	0.6	0.7	2000
Jordan	0.8	0.9	0.9	1997
Turkey	1.6	1.2	1.3	1998
Yemen	3.1	4.7	3.5	1997
Central Asia				
Kazakhstan	4.8	2.2	3.3	1999
Kyrghyzstan	8.6	3.6	6.6	1997
Turkmenistan	3.2	3.6	3.4	2000
South central Asia				
Bangladesh	–	–	–	1999–2000
India	–	–	–	1998–1999
Nepal	–	–	–	2001
Southeast Asia				
Cambodia	8.3	9.9	8.6	2000
Indonesia	6.1	4.7	5.7	1997
Philippines	–	–	–	1998
Vietnam	–	–	–	1997
Central America and the Caribbean				
Dominican Republic	21.6	12.5	15.7	1999
Guatemala	9.5	9.3	9.4	1998–1999
Haiti	21.9	22.9	22.3	2000
Nicaragua	16.5	14.4	15.2	1997–1998
South America				
Bolivia	8.6	7.1	7.7	1998
Colombia	11.9	7.1	8.4	2000
Peru	8.8	7	7.6	2000

cautious one. It is no doubt necessary to increase the refinement of analytical and data collection concepts which remains to be undertaken in the field of household and family demography in developing countries.

Acknowledgment

This chapter was translated from the French version by accenta Ltd.

References

Adepoju Aderanti (ed.), 1997. *Family, population and development in Africa*. London, Zed Books Ltd, 230 p.

Ainsworth Martha, 1990. *Economic aspects of child fostering in Côte d'Ivoire*. Washington, World Bank, 44 p.

Amira, 1987. *Les unités d'observation*. Paris, AMIRA, 283 p.

Antoine Philippe, Bocquier Philippe, Fall Abdou Salam, Guisse Youssouf M., and Nanitelamio Jeanne, 1995. *Les familles dakaroises faceàla crise*. Dakar and Paris, IFAN, CEPED, and ORSTOM, 209 p.

Antoine Philippe and Guillaume Agnès, 1986. Une expression de la solidarité familialeà Abidjan: enfants du couple et enfants confiés, *in:* AIDELF (ed.), *Les familles d'aujourd'hui, démographie et évolution récente des comportements familiaux*, p. 17–29. Paris, PUF, 600 p.

Arias Elizabeth and Palloni Alberto, 1999. Prevalence and patterns of female headed households in Latin America: 1970–1990, *Journal of comparative family studies*, vol. 30(2), p. 257–279.

Barnett Tony and Blaikie Piers, 1992. *AIDS in Africa: its present and future impact*. London, Belhaven Press, 194 p.

Bartiaux Françoise and Tabutin Dominique, 1986. Structures familiales et fécondité dans les pays en voie de développement. Problèmes de mesure et éléments d'explication, *in:* AIDELF (ed.), *Les familles d'aujourd'hui, démographie et évolution récente des comportements familiaux*, p. 245–262. Paris, PUF, 600 p.

Behnam Djamchid, 1985. The Tunis Conference: the Muslim family and the modern world, *Current Anthropology*, vol. 26(5), p. 555–556.

Berquo Elsa and Xenos Peter, 1992 (eds.). *Family systems and cultural change*. Oxford, Clarendon Press, 222 p.

Bisilliat Jeanne (ed.), 1996. *Femmes du Sud, chefs de famille*. Paris, Karthala, 410 p.

Blanc Ann K. and Lloyd Cynthia, 1994. Women's work, child-bearing and child-rearing over the life cycle in Ghana, *in:* Aderati Adepoju and Christine Oppong (eds.), *Gender, work and population in sub-Saharan Africa*, p. 112–131. London, James Currey and Heinemann, 246 p.

Bledsoe Caroline H. and Brandon Anastasia, 1989. Le placement des enfants et son influence sur la mortalité, *in:* Gilles Pison, Étienne van de Walle, and Mpembele Sala Diakanda (eds.), *Mortalité et Société en Afrique du sud du Sahara*, p. 271–293. Paris, INED and PUF, 446 p.

Bledsoe Caroline H. and Brandon Anastasia, 1992. Child fosterage and child mortality in sub-Saharan Africa, *in:* Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa*, p. 279–302. Oxford, Clarendon Press, 450 p.

Bongaarts John, Burch Thomas K., and Wachter Kenneth W. (eds.), 1987. *Family demography: methods and their application*. Oxford, Oxford University Press, 365 p.

Bonte Pierre and Izard Michel, 1991. *Dictionnaire de l'ethnologie et de l'anthropologie*. Paris, PUF, Paris, 755 p.

Bourdieu Pierre, 1993. À propos de la famille comme catégorie réalisée, *Actes de la Recherche en Sciences Sociales*, no. 100, p. 32–36.

Burch Thomas K., 1979. Household and family demography: a bibliographic essay, *Population Index*, vol. 45(2), p. 173–195.

Burch Thomas K., 1980. The index of overall headship: a simple measure of household complexity standardized for age and sex, *Demography*, vol. 17(1), p. 93–107.

Burch Thomas K., 1993. *Theories of household formation: progress and challenges*. London, University of Western Ontario, Population Studies Centre, 27 p.

Burch Thomas K., Halli Shiva, Madan Ashok, Thomas Kausar, and Wai Lokky, 1987. Measures of household composition and headship based on aggregate routine census data, *in:* John Bongaarts, Thomas Burch, and Kenneth W. Wachter (eds.), *Family demography: methods and their applications*, p. 19–39. Oxford, Clarendon Press, 365 p.

Carter Anthony T., 1984. Households histories, *in:* Robert M. Netting, Richard R., and Wilk Arnould Eric J. (eds.), *Households: comparative and historical studies of the domestic group*, p. 44–83. Berkeley, University of California Press, 480 p.

Centre for Economic Policy Research [CEPR], 1991. *Conference proceedings: Beyond national statistics: household and family patterns in comparative perspective*. Paris, INSEE, 164 p.

Charbit Yves, 1987. *Famille et nuptialité dans la Caraïbe*. Paris, INED and PUF, 412 p.

Charbit Yves and Petit Véronique, 1996. Migrations et rôles familiaux: mesure démographique et contextes culturels, *Revue européenne des migrations internationales*, vol. 12(2), p. 25–42.

Cho Lee-Jay and Yada Moto (eds.), 1994. *Tradition and change in the Asian family*. Honolulu, University of Hawaii Press, 470 p.

Collomp Alain, 1992. Les formes de la famille: approche historique, *in:* François de Singly (ed.), *La famille: l'état des savoirs*, p. 13–21. Paris, La Découverte, 449 p.

Committee for International Cooperation in National Research in Demography [CICRED], 1984. *Demography of the family/Démographie de la famille*, *Projet no.2*, *Rapport final*. Paris, CICRED, 202 p.

Cordell Dennis and Piché Victor, 1997. Histoire de la famille en Afrique: au delà de la fabrication des modèles, *in:* Marc Pilon, Thérese Locoh, Kokou Vignikin, and Patrice Vimard (eds.), *Ménages et familles en Afrique: approches des dynamiques contemporaines*, p. 55–74. Paris, CEPED, ENSEA, INS, ORSTOM and URD, 402 p.

Delcroix Sylvie and Guillaume Agnès, 1997. Le devenir des familles affectées par le sida: une étude de cas en Côte d'Ivoire, *in:* Marc Pilon, Thérese Locoh, Kokou Vignikin, and Patrice Vimard (eds.), *Ménages et familles en Afrique: approches des dynamiques contemporaines*, p. 345–369. Paris, CEPED, ENSEA, INS, ORSTOM and URD, 402 p.

De Oliveira Maria-Coleta F. A., 1992. Family change and family process: implications for research in developing countries, *in:* Elsa Berquo and Peter Xenos (eds.), *Family systems and cultural change*, p. 201–214. Oxford, Clarendon Press, 222 p.

Desplanques Guy, 1994. La famille: histoire d'un concept, *Revue française des affaires sociales*, vol. 48(4), p. 95–98.

De Vos Susan, 1987. Latin American households in comparative perspective, *Population Studies*, vol. 41(3), p. 501–517.

Dozon Jean-Pierre, 1986. En Afrique, la famille à la croisée des chemins, *in:* André Burgière, Christiane Klapisch-Zuber, Martine Segalen, and Françoise Zonabend (eds.), *Histoire de la famille, 3: le choc des modernités*, p. 393–442. Paris, Armand Colin, 736 p.

Durkheim Émile, 1888. Introduction à la sociologie de la famille, *Annales de la faculté des lettres de Bordeaux*, 10.

Economic Commission for Latin America and the Caribbean [ECLAC], 1994. *Social panorama of Latin America*. Santiago, ECLAC, 206 p.

Ekouevi Koffi, Ayad Mohamed, Barrere Bernard, and Cantor David C., 1991. Household Structure from a comparative perspective, *in:* Macro International (ed.), *Demographic and Health Surveys World Conference, August 5–7 1991*, p. 1547–1577. Washington DC, Macro International, vol. 3, p. 1547–2185.

El Khouri Arlene Doumit, 1996. Les femmes chefs de famille: état de la recherche et réflexions méthodologiques, *in:* Jeanne Bissiliat (ed.), *Femmes du Sud, chefs de famille*, p. 13–47. Paris, Karthala, 410 p.

Engels Friedrich, 1948. *L'origine de la famille, de la propriété privée et de l'État*, Paris, A. Costes.

Gautier Arlette and Pilon Marc (eds.), 1997. *Familles du Sud*. Paris, ORSTOM and l'Aube, 175 p.

Godelier Maurice, 1973. Modes de production, rapports de parenté et structures démographiques, *La Pensée*, no.172, p. 7–31.

Goode William J., 1963. *World revolution and family patterns*. London, The Free Press of Glencoe, 432 p.

Goode William J., 1964. *The family*. Englewood Cliffs, Prentice-Hall, 120 p.

Goodman Leo A., Keyfitz Nathan, and Pullum Thomas, 1974. Family formation and the frequency of various kinship relationships, *Theoretical Population Biology*, vol. 5(1), p. 1–27.

Gruenais Marc-Éric, 1981. *Famille et démographie de la famille en Afrique*. Paris, ORSTOM, collectif de travail sur la famille, 52 p.

Hareven Tamara K., 1977. The family life cycle in historical perspective, *in:* Cuisinier Jean and Ségalen Martine (eds.), *Le cycle de la vie familiale dans les sociétés européennes*, Mouton, Paris-The Hague.

Hareven Tamara K., 1991. The history of the family and the complexity of social change, *The American Historical Review*, vol. 96(1), p. 95–124.

Henry Louis, 1953. Démographie du passé. Registres paroissiaux et généalogies, *Bulletin de l'association Marc Bloch de Toulouse*, p. 31–36.

Hertrich Véronique, 1996. *Permanence et changements de l'Afrique rurale. Dynamiques familiales chez les Bwa du Mali*. Paris, CEPED, 548 p.

Höhn Charlotte, 1992. The IUSSP programme in family demography, *in:* Elsa Berquo and Peter Xenos (eds.), *Family systems and cultural change*, p. 3–7. Oxford, Clarendon Press, 222 p.

Höhn Charlotte, Mackensen Rainer, and Grebenik Eugene (eds.), 1989. *Later Phases of the Family Life*. Oxford, Oxford University Press, 249 p.

Hunter Susan S., 1990. Orphans as a window on the aids epidemic in sub-Saharan Africa: initial results and implications of a study in Uganda, *Social Science and Medicine*, vol. 31(6), p. 681–690.

Isiugo-Abanihe Uche C., 1985. Child fosterage in West-Africa, *Population and Development Review*, vol. 11(1), p. 53–73.

Isiugo-Abanihe Uche C., 1991. Parenthood in sub-Saharan Africa: child fostering and its relationship with fertility, *in:* Thérèse Locoh and Véronique Hertrich (eds.), *The onset of fertility transition in sub-Saharan Africa*, p. 163–174. Liège, Ordina Éditions and IUSSP, 308 p.

Kabir Mohammad, 1980. *Demographic characteristics of household population*. London, International Statistical Institute, 64 p.

Kaufmann Giorgia, Lesthaeghe Ron, and Meekers Dominique, 1988. Les caractéristiques et tendances du mariage, *in:* Dominique Tabutin (ed.), *Population et sociétés en Afrique au Sud du Sahara*, p. 217–248. Paris, L'Harmattan, 551 p.

Khasiani Shaniyisa A., 1989. Incidence de la présence des réfugiés dans les pays d'accueil en Afrique: le cas des femmes et des cadres, *in:* Reginald Appleyard (ed.), *L'incidence des migrations internationales sur les pays en développement*, p. 269–284. Paris, Centre de développement de l'Organisation de coopération et de développement économique [OCDE], 459 p.

Lacombe Bernard and Lamy Marie-Josée, 1989. Le ménage et la famille restreinte, illusion méthodologique de la statistique et de la démographie d'enquête, *Cahiers des sciences humaines*, vol. 25(3), p. 407–414.

Lallemand Suzanne, 1993. *La circulation des enfants en société traditionnelle. Prêt, don échange*. Paris, L'Harmattan, 224 p.

Lardinois Roland, 1986. En Inde, la famille, l'État, la femme, *in:* André Burguiere, Christiane Klapish-Zuber, Martine Segalen, and Françoise Zonabend (eds.), *Histoire de la famille, volume 2: le choc des modernités*, p. 267–299. Paris, Armand Colin, 736 p.

Laslett Peter and Wall Richard, 1972. *Household and family in past time.* London, Cambridge University Press, 623 p.

Le Bras Hervé, 1992. La démographie historique, *in:* François De Singly (ed.), *La famille: état des savoirs*, p. 347–355. Paris, La Découverte, 449 p.

Lefaucheur Nadine, 1992. Les familles dites monoparentales, *in:* François De Singly (ed.), *La famille: état des savoirs*, p. 67–74. Paris, La Découverte, 449 p.

Lefranc Christophe, 1995. *Où en est la statistique des ménages en Europe?* Paris, CEPED, 13 p.

Le Play Frédéric, 1871. *L'organisation de la famille suivant le vrai modèle signalé par l'histoire de toutes les races et de tous les temps*. Paris, Tequi, 318 p.

Lloyd Cynthia and Desai Sonalde, 1992. Children's living arrangements in developing countries, *Research and Policy Review*, vol. 11(3), p. 193–216.

Locoh Thérèse, 1988a. L'analyse comparative de la taille et de la structure des ménages, *in: African Population Conference, Dakar, 7–12 November 1988, vol. 2*, p. 5.2.17–5.2.42. Liège, IUSSP, pag. mult.

Locoh Thérèse, 1988b. L'évolution de la famille en Afrique, *in:* Étienne van de Walle, Mpembele D. Sala Diakanda, and Patrick O. Ohadike, *L'état de la démographie africaine*, p. 45–66. Liège, IUSSP, 156 p.

Locoh Thérèse, 1990. *Changement social et situations matrimoniales: les nouvelles formes d'union à Lomé*. Paris, INED, 23 p.

Locoh Thérèse, 1997. Structures familiales: l'apport de la démographie, *in:* Marc Pilon, Thérèse Locoh, Kokou Vignikin, and Patrice Vimard (eds.), *Ménages et familles en Afrique: approches des dynamiques contemporaines*, p. 11–29. Paris, CEPED, ENSEA, INS, ORSTOM and URD, 402 p.

Mason Andrew, 1987. *HOMES: a household model for economic and social studies*. Honolulu, East-West Population Institute, East-West Center, 114 p.

Mbuh R., 1994. La famille africaine dans le processus de développement, *UNESCO-Africa*, no.8, p. 64–74.

McDonald Peter, 1992. Convergence or compromise in historical family change? *in:* Elza Berquo and Peter Xenos (eds.), *Family systems and cultural change*, p. 15–30. Oxford, Clarendon Press, 222 p.

Michel André, 1986. *Sociologie de la famille et du mariage*. Paris, PUF, 263 p.

Morgan Lewis, 1971. *La société archaïque*, Paris, Anthropos (1st edition 1887).

Netting Robert M., Wilk Richard R., and Arnould Eric J. (eds.), 1984. *Households: comparative and historical studies of the domestic group*. Berkeley, University of California Press, 480 p.

Oke E. Adewale, 1986. Kinship interaction in Nigeria in relation to societal modernization: a pragmatic approach, *Journal of comparative family studies*, vol. 17(2), p. 185–196.

Ono-Osaki Keiko, 1991. Female headed households in developing countries: by choice or by circumstances? *in:* Macro International (ed.), *Demographic and Health Surveys World Conference, Columbia, August 5–7, 1991: proceedings*, p. 1603–1621. Washington, Macro International, vol. 3, p. 1547–2185.

Page Hilary J. 1989. Child-rearing versus child-bearing: co-residence of mother and child in sub-Saharan Africa, *in:* Ron Lesthaeghe (ed.), *Reproduction and social organization in Africa*, p. 401–441. Berkeley, University of California Press, 556 p.

Parsons Talcott and Bales Robert F., 1955. *The family: socialization and interaction process*. New York, Free Press of Glencoe, 422 p.

Pennington René and Harpending Henry, 1993. The structure of an African pastoralist community. *Demography, history and ecology of the Ngamiland Herero*. Oxford, Clarendon Press, 268 p.

Pilon Marc, 1996a. *Pour une remue-ménage de la démographie de la famille en Afrique*. Paris, CEPED, 25 p.

Pilon Marc, 1996b. Les femmes chefs de ménage en Afrique: contribution à un état des connaissances, *in:* Jeanne Bissiliat (ed.), *Femmes du Sud, chefs de famille*, p. 235–256. Paris, Karthala, 410 p.

Pilon Marc and Hertrich Véronique, 1995. Aller au delà du ménage: pour de nouvelles approches. Paris, CEPED, 18 p.

Pilon Marc, Locoh Thérèse, Vignikin Kokou, and Vimard Patrice (eds.), 1997. *Ménage et famille en Afrique: approches pluridisciplinaires des dynamiques contemporaines.* Paris, CEPED, ENSEA, INS, ORSTOM and URD, 424 p.

Prioux France, 1990. *La famille dans les pays développés: permanences et changements: actes du Séminaire sur les nouvelles formes de vie familiale, Vaucresson, octobre 1987*. Paris, INED, 316 p.

Rao L. J., 1973. Industrialization and the family: a world view, *International Journal of Sociology of the Family*, vol. 3(2), p. 179–189.

Renty (de) Patricia, 1993. *Pour une démographie des réfugiés de la corne de l'Afrique*. Paris, Institut de démographie de l'Université de Paris I, 123 p.

Rosental Paul-André, 2003. La nouveauté d'un genre ancien: Louis Henry et la fondation de la démographie historique, *Population*, vol. 58(1), p. 103–136.

Roussel Louis, 1986. Évolution récente de la structure des ménages dans quelques pays industriels, *Population*, vol. 41(6), p. 913–934.

Ryder Norman B., 1984. Structure familiale et fécondité, *Bulletin démographique des Nations unies*, no. 15-1983, p. 17–39.

Sala Diakanda Mpembele, 1988. *Problèmes conceptuels et pratiques liés aux informations disponibles sur la structure des ménages en Afrique ainsi qu'à son analyse: analyse*, African Population Conference, Dakar 1988, vol. 2, p. 5.2.1–5.2.15. Liège, IUSSP, pag. mult.

Sawadogo Ram Christophe, 1992. Famille africaine d'hier et de demain: quelle compréhension pour quelles interrogations? *in:* UAPS (ed.), *Structure et dynamique de la formation des familles en Afrique*, p. 61–76. Dakar, UAPS, 76 p.

Segalen Martine, 1984. Nuclear is not independent: organization of the household in the Pays Bigouden Sud in the nineteenth and twentieth centuries, *in:* Robert McC. Netting, Richard R., and Wilk Arnould Eric J. (eds.), *Households: comparative and historical studies of the domestic group*, p. 163–186. Berkeley, University of California Press, 480 p.

Segalen Martine, 1996. *Sociologie de la famille*. Paris, Armand Colin, 296 p.

Shorter Edward, 1977. *Naissance de la famille moderne, XVIII^{e}–XX^{e} siècles*. Paris, Seuil, 379 p.

Silk Joan B., 1987. Adoption and fosterage in human societies: adaptations or enigmas? *Cultural Anthropology*, vol. 2(1), p. 39–49.

Singly (de) François, 1992. *La famille: l'état des savoirs*. La Découverte, Paris, 448 p.

Tabutin Dominique and Bartiaux Françoise, 1986. Structures familiales et structures sociales dans le tiers-monde, *in:* AIDELF (ed.), *Les familles d'aujourd'hui*, p. 231–243. Paris, PUF, 600 p.

Théry Irène, 1987. Remariage et familles composées: des évidences aux incertitudes, *L'Année sociologique*, vol. 37, p. 119–152.

Tichit Christine, 1994. *La montée des femmes chefs de ménage en Afrique au Sud du Sahara: examen à partir des données publiées sur les ménages*. Paris, École des hautes études en sciences sociales, 73 p.

United Nations High Commissioner for Refugees [UNHCR], 1994. Focus: année internationale de la famille, *Réfugiés I*, vol. 95, 30 p.

United Nations [UN], 1980. *Principes et recommandations concernant les recensements de la population et de l'habitation*. New York, UN, 130 p.

United Nations [UN], 1990. *Demographic yearbook*. New York, UN, Statistical Division, 1020 p.

United Nations [UN], 1992. *Les femmes dans le monde. 1970–1990: des chiffres et des idées*. New York, UN, Population Division, UNICEF, FNUAP, UNIFEM, 122 p.

United Nations [UN], 1993. Demographic Yearbook/*Annuaire démographique*. New York, UN, Statistical Division, 1032 p.

Vandermeersch Céline, 2000. *Les enfants confiés au Sénégal*. Paris, Institut d'études politiques de Paris, 502 p.

Vimard Patrice, 1993. Modernité et pluralités familiales en Afrique de l'Ouest, *Revue Tiers Monde*, vol. 34(133), p. 89–115.

Vimard Patrice, 1995. *De quelques typologies des ménages*. Paris, CEPED, 17 p.

Vimard Patrice, 1997. Modernisation, crise et transformation familiale en Afrique subsaharienne, *in:* Arlette Gautier and Marc Pilon (eds.), *Familles du Sud*, p.143–159. Paris, Revue Autrepart, ORSTOM, and l'Aube, 175 p.

Vimard Patrice and Guillaume Agnès, 1991. Mobilités familiale et spatiale des enfants en Côte-d'Ivoire, *in:* André Quesnel and Patrice Vimard (eds.), *Migration, changements sociaux et développement, III^{es} journées démographiques de l'ORSTOM*, p. 243–260. Paris, ORSTOM, 338 p.

Wall Richard, 1996. Comparer ménages et familles au niveau européen: problèmes et perspectives, *Population*, vol. 51(1), p. 93–116.

CHAPTER

91

Households and Families

NICO KEILMAN

Department of Economics, University of Oslo, Oslo, Norway

INTRODUCTION

Statistics Norway reported in its family statistics that there were 501,000 families consisting of a married couple with one or more children on January 1, 1999 (Statistics Norway, 2000). However, in another table the number of married couples with children for that date is much lower: 359,000 (Statistics Norway, 2000). Both figures are correct, but the definition of a family consisting of a married couple and one or more children differs between the two. For the first figure, children can be of any age, and for the second one, children must be younger than 18 years of age. Introduction of an upper age limit for children reduces the number of married-couple families by about 30%.

The November 1990 Population and Housing Census for Norway reported 102,000 consensual unions (Statistics Norway, 1999). But sample surveys carried out in 1988 and 1994 indicate that the numbers in those years were at least 130,000 and 200,000, respectively (Statistics Norway, 1995a).[1] The three numbers suggest a strong fall between 1988 and 1990 and a steep rise thereafter. In reality, this has not been the case: consensual unions have gradually become more and more popular in Norway since the end of the 1970s, and there are no indications for major fluctuations in the numbers. The main reason why the census number is much lower than what could be expected on the basis of the information for 1988 and 1994 is that the census number relates to the household situation according to the Central Population Register, that is, the *de jure* place of residence. However, the surveys are based on the de facto place of residence. The register does not always reflect the de facto place of residence accurately, and for partners in a consensual union in particular there is a strong underregistration in the register. This explains why the census figure is much lower than that obtained on the basis of a survey.

The two examples given above illustrate some of the difficulties one will encounter when family and household trends are investigated. Different definitions and different measurement methods make a comparison over time or across countries problematic. In this treatise, both the analysis and the synthesis of population developments are described. Therefore, the aim of this chapter is not only to give a broad overview of trends in family and household developments in Europe but also to discuss concepts and definitions concerning the household, the family, and their members: what constitutes a household, a family; what is a consensual union, a child, a one-parent family, a reconstituted family?

Next I review various issues connected to measuring household and family developments: de facto or *de jure* place of residence, measure of household and family structure at one point in time, measure of household and family dynamics over a certain period, the

[1] The estimate for 1988 has been computed on the basis of age-specific shares cohabiting women for ages 22, 27, 32, 37, and 42 from the survey (see Statistics Norway, 1995a). These shares were assumed to be the average shares for age groups 20 to 24, 25 to 29, 30 to 34, 35 to 39, and 40 to 44, and next these were multiplied with the registered numbers of women in these age groups for the country as a whole on January 1, 1988. This results in 129,000 cohabiting women aged 20 to 44. Cohabiting women outside that age group are not included in this estimate.

individual or the group as unit of measurement, the problem of longitudinal households, relationships between the events that several members of the same household may experience (e.g., a lone mother becomes a one-person household when her last child leaves the parental home), and the representativeness of the data. The strengths and weaknesses of various data sources often used to map household and family developments are discussed later in the chapter. Finally, the main trends in family and household developments in Europe after World War II are summarized. Notoriously lacking is a review of relevant behavioral theories. A decade ago, Burch (1995) noted that family and household demography have made considerable progress, but that theory has greatly lagged behind. This may reflect the complexity of the subject matter.[2]

I. DEFINITIONS AND CONCEPTS

The notions of household and family are closely connected, yet there are strong differences. A formal definition of both will be given below, but one can say that an important aspect of the household definition is coresidence irrespective of consanguinity or conjugality, whereas consanguinity and conjugality are major ingredients for the family definition, irrespective of place of residence. However, as we will see below, the family in the more restrictive sense is limited to persons related by blood or marriage, who live in the same household. In that respect every family is also a household, but some households are not family households, for instance, one-person households.

The 47th edition of the *Demographic Yearbook* of the United Nations (UN, 1997) contains summary tables for the definitions used by countries that participated in the 1990 round of censuses. The same type of information for the 1980 census is found in the 39th edition (UN, 1989). Much of the international experience that is given below is drawn from a survey among member and associated countries of the UN Economic Commission of Europe (ECE), which was carried out in the spring of 2004 (see Keilman and Kinsella, 2004). Forty-six countries gave details on their experience with the 2000 round of censuses.

1. The Household

For the 2000 round of censuses, the United Nations recommended using the following definition of a private household: a household is "(1) a one-person household, i.e., a person who lives alone in a separate housing unit or who occupies, as a lodger, a separate room (or rooms) of a housing unit but does not join with any of the other occupants of the housing unit to form part of a multi-person household as defined below; or (2) a multi-person household, i.e., a group of two or more persons who combine to occupy the whole or part of a housing unit and to provide themselves with food and possibly other essentials for living" (UN, 1998). This definition centers on two important notions: common housekeeping and living in the same house. Both conditions must be fulfilled when two or more persons can be said to belong to the same household, according to this so-called *housekeeping-unit concept* of the private household. A somewhat less restrictive definition centers on the *household-dwelling concept*, which says that all persons who occupy the same housing unit live in the same household (UN, 1998). Although in practice some countries use the household dwelling unit definition, the housekeeping unit definition may prove to be more useful in socioeconomic analyses of household structure. Two-thirds of the ECE countries applied the housekeeping unit definition in their census carried out around 2000 (Keilman and Kinsella, 2004). Many of the countries that used the household dwelling concept based their census on a population register, for instance, Denmark, Finland, France, Norway, and Switzerland. Similar shares (two-thirds of the countries using the housekeeping definition, one-third using the dwelling definition) were also observed in European countries for the 1980 round and for the 1990 round of censuses (Keilman 2004; Keilman 1995:table 5.1).

Household data may be obtained not only from a census, but also from other sources. Sample surveys are commonly used, for instance, a Labor Force Survey. Kormendi (1995:table 1) has shown that 8 of the 12 member states of the former European Community (EC) employed the housekeeping unit concept in their Labor Force Survey in the 1990s. Belgium, Denmark, Spain, and France based their household data on the household dwelling concept.

Whether one uses the housekeeping unit definition or the dwelling unit definition has little effect on the *total* number of households: differences of 2% to 3% have been reported for Austria, Germany, and Norway (see Ås, 1990; Galler and Ott, 1993; Gisser, 1995).[3] But

[2] Although, as Burch notes, a good theory makes things look simple(r).

[3] Although the numbers for Austria, Germany, and Norway cited here are small, this does not hold more generally, as is demonstrated by the case of Czechoslovakia: the difference in the number of households implied by the dwelling unit definition, as compared to the housekeeping definition was 8% in the Czech Republic, both in 1970 and 1980; in the Slovak Republic the difference was 9% in 1970 and 11% in 1980 (Kalibova 1991).

one-person households may be affected more strongly: lodgers are counted as one-person households according to the housekeeping unit concept, whereas they are member of the household that they share the housing unit with when the dwelling unit concept is used.

Housekeeping unit and dwelling unit are not the only two alternatives; other possibilities fall somewhere in between these two. For instance, Todd and Griffiths (1986) carried out a study into the effects of a household definition change on numbers of households in England. In many major household surveys of the Office of Population Censuses and Surveys up to 1981, the housekeeping unit concept was used to define a household. But in 1981 it was considered more appropriate to bring the instructions for interviewers in line with census practice, in which persons who shared a living room were also included in the household, even if they did not share catering arrangements. The effect of this change in definition was that the overall estimate of the number of households was reduced by 108,000, or 0.6%. Most of the excluded households were in the private rented sector in outer London and the southeast of the country, comprising one-person households and households with a resident landlord.

In addition to the private household defined above, the UN also has a recommendation concerning institutional households:

> "An institutional household comprises persons whose need for shelter and subsistence are being provided by an institution. An institution is understood as a legal body for the purpose of long-term inhabitation and provision of institutionalized care given to a group of persons. The institution's accommodation is by nature of its structure intended as a long-term accommodation for an institutional household. The great majority of institutions consists of one of the following: dormitories of educational institutions, hospitals, convalescence establishments, establishments for the disabled, psychiatric institutions, old people's homes, nursing homes, welfare institutions, hotels, motels, tourist homes, military installations, correctional and penal institutions, religious institutions, boarding houses, worker dormitories" (UN, 1998).

2. Head of Household

Until the 1990 round of censuses, the UN recommended using the notion "head of the household," that is, that person in the household or the family who is acknowledged as such by the other members. It implicitly involved the person who bears the chief responsibility for the economic sustenance of the household or the family. In a number of European countries, this concept is outdated. The UN formulated it quite cautiously: "In countries where spouses are considered equal in household authority and responsibility or shared economic support of the household, the concept of head of household was no longer considered valid even for family households" (UN, 1989). Thus, the concept of household head no longer reflects social reality in many European countries, and moreover, it is thought to be offensive to a large fraction of the population if they were asked to report one household member as head in a statistical enquiry (Murphy, 1991a). Therefore, starting with the 2000 round, the UN suggested that countries might prefer that the members of the household designate one among them as a *reference person* with no implication of headship:

> "There is some evidence though to suggest that the following criteria for selection of the reference person will yield the most fruitful range of explicit kin relationships:
>
> Either the husband or the wife of a married couple living in the household (preferably from the middle generation in a multi-generational household);
> Either partner of a consensual union couple living in the household where there is no married couple present;
> The parent, where one parent lives with his or her sons or daughters of any age;
> Where none of the above conditions apply, any adult member of the household may be selected" (UN, 1998).

Next, the structure of the household may be explored by tracing the relationship between the remaining members and the reference person, or the household head in countries where the traditional definition is considered appropriate.[4] Following UN practice, the term *householder* is used to denote household head or reference person.

Whether a householder is called head or reference person is of less importance than the fact that there is heterogeneity among countries in determining the householder. Among 46 ECE countries, all but two (Denmark and Norway) reported that they identified a reference person in the household (Keilman and Kinsella, 2004). However, for purposes of household structure mapping, the two countries just mentioned define the oldest person in the household as a point of reference. In 22 of the 46 countries, census respondents chose the reference person freely among the adults living in the household, using various criteria.

Does this large variability in the international practice concerning household head and reference person render international comparisons useless? Not for all cases, probably. First of all, in practice, the age distribution of householders seems to differ very little when alternative definitions are used (Murphy, 1991b). Second, many households consist of a married couple

[4] Brynin and Smith (1995) discuss various methods for mapping such relationships on the basis of the British Household Panel Survey.

(with or without children), a one-parent family, or a single person. In the case of a married couple, the husband will most frequently be denoted as (one of the) householder(s). In case a one-parent family is recorded, the householder will often be a woman. International comparisons are therefore still possible for a limited group of countries, provided that one distinguish between two-parent families, one-parent families, and other households, and that male and female householders be considered separately. When no such distinction is made, the comparison becomes hazardous, in particular for women. Indeed, as an UN review shows, female headship rates for Austria, Norway, and Sweden are almost consistently higher than are those in Italy, Portugal and Spain, and some of the differences may be interpreted in terms of economic criteria that are used to identify the householder (UN, 1989).

The ultimate solution is of course to standardize the definition of the reference person (Wall 1996). But this is virtually impossible to achieve in practice, and therefore Wall (1996) has suggested to base international household comparisons on a detailed set of standardized tables. The basic idea is to cross-classify age, sex, and marital status with a person's household position: whether he or she lives in a household as a couple, as a lone parent, as a child, as one person, etc. Such detailed tables have been compiled for Ireland (the Census of 1979) and Canada (the census of 1986), and can be constructed for a number of other countries as well (e.g., England and Wales, Great Britain, The Netherlands, Italy, Norway).

3. The Family

The UN recommend to define, for census purposes, a family as "two or more persons within a private or institutional household who are related as husband and wife, as cohabiting partners, or as parent and child. Thus a family comprises a couple without children, or a couple with one or more children, or a lone parent with one or more children" (UN, 1998). Among 46 ECE countries, 39 used a definition that complies with the recommendation. In Canada, England and Wales, and Northern Ireland, grandparent(s) living with one or more grandchildren but without the grandchild(ren)'s parents are also regarded as a family. Canada and Ireland restricted the family to those living in private households. Norway complies with the definition when compiling international tables, whereas persons living alone are also counted as families (one-person families) in national tables. The Swiss census did not include family information. The United States restricts families to two or more persons related by birth, by marriage, or by adoption to the householder. This differs from the recommended definition in two respects: cohabiting partners who are not married to each other are not counted as families, and a household consisting of a household reference person (householder) and two or more persons who form a family (according to the UN definition), but who are not related (birth, marriage, adoption) to the reference person, is not considered a family household.

Better comparability with previous censuses or with other statistical surveys was mentioned often as the main reason for a different definition.

4. Consensual Union

Starting with the 2000 round of censuses, the UN-ECE has formulated a recommended definition of a consensual union: "Two persons are understood as partners in a consensual union when they have usual residence in the same household, are not married to each other, and report to have a marriage-like relationship to each other." (UN, 1998). Note that partners can be of the same sex. Because of problems connected to developing criteria for marriage-like relationship, it must be left to the respondents to report themselves as partners in consensual union or not (for a review of issues connected to defining cohabitation, see Trost, 1988). All 46 ECE countries for which we have information practiced this so-called phenomenological definition (owing to Trost) in the 2000 round of censuses (see Keilman and Kinsella, 2004). Using a phenomenological definition will introduce a bias for various reasons (Trost, 1988): social acceptance, or tax avoidance, or differences in perception between the partners (one considers the relationship as marriage-like, the other one looks on it as more casual). As a result, two nonmarried adults living in the same household may be recorded as a consensual union at one occasion and as two nonrelated adults at the other. In case there are dependent children as well in the household, the alternative registration will often be a one-parent family, and a nonrelated adult living in the same household.

5. Child

The UN (1998) recommends to define a child "as any person with no partner and no child who has usual residence in the household of at least one of the parents." Children also includes stepchildren and adopted children, but not foster children. A child that alternates between two households (e.g., after the parents' divorce) is counted at only one of these households, for instance, on the basis of the *de jure* place of usual residence (see Section "De Facto or *De Jure* Place of Residence?") or the number of nights spent at either of the households. Only five countries among the 46 for which we have information deviated from the recommended definition (Keilman and Kinsella, 2004). The census in the Czech Republic required that children be economically dependent (economically not active) and not older than 25 years of age. There is no restriction on partners or own children. Denmark required that children be less than 25 years old. In Switzerland, sons-in-law and daughters-in-law living in the same household were also considered as children. Moreover, there was no restriction on marital status in the Swiss definition. Thus, a child could be married and living with his or her spouse in the household of the parent(s). The United States did not have any restrictions regarding the child's own children or partner.

The case of three or more generations living in the same household may create problems. A woman, who lives in the same household as her children *and* her parents, but without a partner, should be regarded as being member of the same family as her children, but not of the same family as her parents: "A three-generation household consists of two or more separate family nuclei or one family nucleus and (an)other family member(s). A woman who is living in a household with her own child(ren) should be regarded as being in the same family nucleus as the child(ren) even if she is never-married and even if she is living in the same household as her parents; the same applies in the case of a man who is living in a household with his own child(ren). Thus, the youngest two generations constitute one family nucleus." (UN, 1998).

6. One-Parent Family

The UN has not formulated a separate definition of one-parent family, but one possibility is one adult living in the same household with one or more of his or her children (by blood or adoption). Problems in connection with the notion of "child" have been discussed above. But other problems may arise when there is a second adult member in the household. When this person is a relative of the lone parent (mother, father), the situation is clear. But another possibility is that this person is the lone parent's cohabitee, and then there is the situation of a household consisting of a cohabiting couple plus the child of one of the partners. How can we distinguish between this household type and "one-parent family plus other adult"? A possible marriage-like relationship between the two adults is not enough for the household to be called a consensual union with child(ren). A further condition is that *both* adults take parental responsibility for the children. In Trost's words: the number of "parent–child units" should be twice the number of children, in addition to one "conjugal unit" (Trost 1990). If only one of the adults forms a parent–child unit with the children and a marriage-like relationship with the other adult, it will be unclear whether or not this is a one-parent family.

The situation described here is not at all hypothetical: Höpflinger (1991) indicates that in the 1980s in Germany, 10% of women who declared themselves as "lone mother" were in fact cohabiting; for "lone fathers" the figure might even have been as high as 28%.

7. Reconstituted Family

The UN regards a reconstituted family as "a family consisting of a married or cohabiting couple with one or more children, where at least one child is a non-common child, i.e., either the natural or adopted child of only one member of the couple. If the child (natural or adopted) of one partner is adopted by the other partner, the resulting family is still a reconstituted family" (UN, 1998). When a lone parent starts cohabiting with a partner who takes parental responsibility for the former lone parent's child(ren), we have one example of a reconstituted family or stepfamily. The other example of reconstituted family is when the two adults are legally married. Although such reconstituted families historically referred to family formation through remarriage after widowhood, we nowadays also regard separation or divorce, followed by the start of a consensual union, as events leading to a reconstituted family (Deven, 1995). Issues connected to reconstituted families have been discussed by Leridon (1993).

Only 19 of the 46 ECE countries for which information exists reported that they are able to identify reconstituted families by using census data. One possible explanation for the poor performance across countries on this item is the fact that many countries mapped household structures by means of the relationship of each household member to the household reference person, but not by means of the relationship to other household members. This implies that not all reconstituted families can be identified. For example, in case a father is the reference person and the mother is his second wife, this family can not be identified as reconstituted because the child is referenced to only one adult, viz. the father. It would be tallied as being a married-couple family with a biological child of the reference person but of unknown connection to the wife. However, if the father were listed as the "husband" of the reference person (the reference person being the second wife), then this child would be listed as the stepchild of the householder. This family could be identified as a reconstituted family in which the householder was not the biological mother of the child. Thus identification of reconstituted families requires more detailed mapping of household structures than by means of the relationship to just one person in the household. A full matrix of relationships between all household members is one possibility. Such a matrix was used by 13 of the 44 countries that reported that the relationship of each household member to the household's reference person was asked: Albania, Cyprus, Estonia, Greece, Hungary, Kazakhstan, Lithuania, Portugal, the Russian Federation, Serbia and Montenegro, England and Wales, Scotland, and Northern Ireland. An intermediate solution is to map, for every child in the household, the relationship with all adults (biological or nonbiological parent).

II. MEASUREMENT OF FAMILY AND HOUSEHOLD DEVELOPMENTS

Given a particular definition of the various household and family variables, the analyst can start measuring them. Before doing so, however, several questions have to be answered.

1. De Facto or *De Jure* Place of Residence?

Do the data to be measured refer to the *de facto* or the *de jure* situation? A number of countries have a population register, which is used directly or indirectly for collecting information on households and families. The situation according to the register (*de jure*) may differ from the actual situation (de facto). An important aspect in this respect is the place of usual residence. The UN (1998) recommends defining as place of residence: "the geographic place where the enumerated person usually resides; this may be the same as, or different from, the place where he/she actually is at the time of the census; or it may be his/her legal residence. A person's usual residence should be that at which he/she spends most of his/her daily night-rest."

Of 46 ECE countries, 43 reported that they complied closely or completely with this definition. Scotland applied a different definition, because a person's usual residence was preferred to be at their family residence if they worked away from home during the week. In Austria, the usual residence of a person is her/his main residence as registered in the population register. Week commuters are registered at the family's residence, similar to the case of Scotland.

Given this recommendation it must be concluded that the UN recommend collecting de facto information regarding membership in families and households. In practice, a number of countries have probably used a *de jure* definition of place of residence and hence of family and household membership, most often based on a population register. Although this may be good practice in case there is little difference between *de jure* and de facto place of residence, it creates difficulties when the register shows considerable differences with the actual situation. Below I give some examples for Norway, Switzerland, and Germany connected to the 1990 round of censuses. For these countries, we have detailed statistical data on the differences between *de jure* and de facto places of residence.

Registration rules for the Norwegian population register until 1998 stated that a never-married person who resides outside the home of the parents because of education should be registered as living at the parents' address (Statistics Norway 1985; Skiri 1995). Data collection for the November 1990 Population and Housing Census of Norway was closely linked to the register. One of the census questions was "With whom do you share this dwelling?," and respondents had to answer *in conformity with the register*, irrespective of the persons actually living in the respondent's dwelling. One consequence was that the census severely underestimated the de facto number of consensual unions and the number of one-person households: an estimated 22% for consensual unions (Keilman and Brunborg, 1995), and 19% for one-person households (Schjalm, 1996). At the same time, the census reported far more young adults living with their parents than in reality. The overall result was that the 1990 census underestimated the de facto number of private households by 4% to 6% (Statistics Norway, 1994). Registration rules were changed in 1998, when students were given the possibility to register as their *de jure* place of residence the place where they actually lived. But the 2001 census found that still 83,000 students had not done so. The consequence was that the census underestimated the share of persons living in a one-person household by two percentage points.

The December 1990 Census of Switzerland distinguished, for the first time, between legal and economic residence (M. Buscher, personal communication). The economic residence is the place where the individual concerned usually (the largest part of the week) lives.

The legal residence is where taxes are paid, political rights are exercised, etc. The legal place of residence is different from the economic one for 3% of the population.

Finally, note that difficulties may be encountered when a child alternates between two households, for instance, after the parents' divorce. In general, an individual may theoretically be a member of two households, but with different degrees of membership (Berge, 1988). In practice, however, when classification is attempted, only one household will be assigned to each individual. But in German household statistics, persons with two residences are counted twice (Linke, 1988). The population in private households in a certain municipality includes all persons counted there, even if some of them live most of the time in a different municipality. Schwarz (1983) estimates that the number of one-person households in 1961 had to be adjusted downward by no less than 30% in order to reflect reality.

2. Individual or Group Level Data?

Do the data apply to *individuals* or to the *group* level? One should clearly distinguish between an analysis for persons and an analysis for households (or families or married couples). For instance, whereas 38% of private households were a one-person household in Norway in 2001, only 17% of the population lived in such a household (see *http://www.ssb.no/emner/02/01/fobhushold/*).[5] Both types of data have their merits: for instance, some purchases are directly linked to the individual level (food, clothing), others to the group level (housing, capital goods). Furthermore, the household can make decisions as a group (e.g., the decision to marry taken by a cohabiting couple) or by individual household members (e.g., home leaving of young adults).

Data that reflect the household and family situation at one point in time (stock data) are routinely given both at the individual and the household level. For instance, the UN recommend to tabulate, among others, the *population* by household status (persons in various types of households) and, at the same time, *households* by various types. Note that knowledge of the type of household an individual lives in is not sufficient for determining this person's individual household status. For example, a person living in a household of type "married couple with children" may have household status "child" or "spouse." Conversely, age, sex, and marital status may be added to an individual's household status, and this often helps in determining that person's household type given household status, but not always—it depends on the actual classifications chosen for both household type and household status.

Contrary to stock data, the individual level may be preferred when information is collected concerning household and family *dynamics* over a certain period (*flow* data). The reason is that the group level perspective may create some problems; compare the discussion below on longitudinal households.

In between the individual and the complete household is what Ermisch and Overton (1985; see also Ermisch, 1988) have called the minimal household unit (MHU). It is the smallest group within the household that constitutes a demographically definable entity. This means that an individual, over his lifetime, moves from one type of MHU to another by means of simple demographic events only. The four MHUs defined by Ermisch and Overton are (1) childless, nonmarried adult; (2) one-parent family; (3) childless married couple; and (4) married couple with dependent children. Demographic events that cause individuals to move between MHUs are marriage, loss of spouse, birth of child, loss of last child, and divorce with loss of children. The status "married" and the event "marriage" are to be interpreted as including de facto marriage. Hence MHU types 2 through 4 are equivalent to a narrow definition of the family, whereas type 1 corresponds with a person living alone. Households can be viewed as consisting of one or more MHUs.

3. Adult or Child as the Unit of Analysis?

When family or household structures are investigated, the focus is often on the relationship between children and adults. If the data apply to individuals, the choice is between the perspective of the child and that of the adult. Most often, the adult is chosen as the unit of analysis, because most data are supplied in that form (OECD, 1979). But in a few cases, the focus has been on children. The results may differ, sometimes considerably, according to which viewpoint is chosen. The first reason is that most children live in families and households with only one or two adults, whereas the range in number of children seen from the perspective of the adult is much larger: from one to six or even more. The second reason is particularly relevant when *relative* numbers are compared, for instance, the share of lone parents among all parents to the share of children with a lone parent compared with all children. In this case the findings will differ because the reference groups differ (all parents versus all children).

[5] Note that the ratio of these two numbers equals mean household size; for Norway 38/17 is 2.3 persons per private household on average.

As an example consider the Children Statistics compiled by Statistics Norway (1996:table 1), from which we can compute that 17.3% of all *children* (under 18) lived with a lone parent on January 1, 1995. On the other hand, only 13.7% of all *parents* (with children under 18) were lone parents at that time (Statistics Norway 1995b:table 3).

The few demographic family and household studies in which children are the unit of analysis have most often analyzed family break-up and the extent to which children experience divorce by the parents. Examples are findings reported for Denmark, England, France, The Netherlands, New Zealand, Norway, and the United States.

4. Structure or Dynamics?

Is the interest in the household and family *structure* at one point in time or in *dynamics* over a certain period? Cross-sectional data can only give a snapshot of the situation. Changes in that structure may be mapped at two levels: that of the population as a whole, and that of the individual or the household. Repeated cross-sectional surveys can provide us with repeated measurements for population indicators, for instance, the development in the number of households of various types, in the share of children living in a one-parent family, and so on. Only the net change between two points in time is measured this way—hence we may also speak of comparative statics. Yet this will generate useful data. For instance, changes in the composition of households by size and type may help us in understanding developments in housing preferences, or consumption patterns. But when the interest is in a more causal analysis, data on net changes obtained from repeated cross sections are only of help during a first stage in the analysis—changes at a continuous time scale (gross flows) measured for individual persons or households should be analyzed next. This would facilitate exploring the consequences of certain policy actions for subsequent household and family behavior, or to study the impact of events in one arena of life (e.g., education) for those in a different arena (e.g., childbearing). The point is that the *timing* of certain events is recorded for *the same* individual or household during a particular time period.

The amount of household change that is found during a particular period increases when more detail is distinguished in the breakdown by household type or household status. When households are only divided into family households and nonfamily households, a woman who becomes a lone mother will not change household type. This will only be the case when an additional distinction is made between couple households and lone-parent households. Compare the distinction between interregional migration and intraregional migration in studies of regional mobility.

5. Longitudinal Households

An individual person is born, dies, and experiences during his life several events such as home leaving, start of a partnership, marriage, childbearing, etc. Events of this type reflect household and family dynamics at the *individual* level. We can also define events at the *group* level: household formation, household dissolution, and household change. When the focus is on the development of households at the group level, one may be confronted with the problem of *longitudinal households*: when a household splits, it is often difficult to determine which part is the original household and which part constitutes a new one. In other words, when can two households that exist at two points in time be considered as the same household (McMillan and Herriott 1985)? For example, when a married couple with four children separates and the man leaves the household, the result is a lone mother and a one-person household. Are these two truly new households? Or is the one-parent family a continuation of the old household (although of a different type)? To answer such questions, continuity rules have to be formulated. One possibility is to call a household a continuous one as long as 50% or more of its members are the same at two points in time. In that case the lone mother and her children mentioned above would not change household. Another one is the rule that two households that have the same head are the same household. This may lead to the decision (depending on the definition of head of household) that the man who now lives alone in fact lives in the same household. The continuity rule chosen for a particular situation is often arbitrary, and an important objection is that different rules lead to different results (Citro, 1985; McMillan and Herriott, 1985). Therefore, when household dynamics are investigated, the unit of analysis is likely to be the individual rather than the household (Duncan and Hill, 1985; Murphy, 1995). In such an approach, individuals move between household states. As an example, consider the following sequence: from “child with two parents” to “cohabiting” to “married” to “married, one child” to “separated” to “deceased.”

6. Interrelated Events for Members of the Same Household

To focus on individuals rather than on households when analyzing household and family dynamics circumvents the problem of longitudinal households, but it leads to another problem: many household events for members of the same household are interrelated. An event for one household member often induces an event for other members. To continue the example given in the previous section, when a person moves from “married, one child” to “separated,” the spouse becomes a lone parent. If we want to understand household dynamics at the individual level, these and other links between events experienced by members in the same household must be taken into account. How complex these interrelationships are depends on the level of detail in the household state classification for individuals. For instance, with only four marital states (never married, currently married, divorced, widow[er]) we just have to trace marriage and divorce for both partners, and death of one spouse together with transition to widowhood. But with a reasonable set of proper household states (e.g., child, living alone, cohabiting, married, in institution, and other) there are many interrelationships of various types.

7. Representativeness

When the data come from a sample survey, one should consider carefully which larger population they represent. At survey date, the sample should be representative of the population both in terms of individuals *and* of households. In other words, the distribution of sample members by household status should be reasonably close to the corresponding distribution for all persons in the population, and at the same time, the distribution of households by some characteristic should be close to the corresponding distribution for households in the population. This may be problematic when the sample frame is in terms of individual persons, of which a sample of persons is drawn who next report on the household situation of all other members in their household. This leads to an overrepresentation of large households, because a randomly drawn person is more likely to live in a large household than in a small household. Thus, a representative sample of individuals will not give a representative sample of households. The same problem may occur when generations are linked. A representative sample of children will not give a representative sample of the previous generation: mothers with large families will be overrepresented, and childless women will not be included at all (Murphy, 1995). Weights by household size or family size (sampling proportional to size) must be used to achieve representativeness in such situations.

The problem becomes even more complex when the data do not apply to one point in time, but when they cover a certain period, for instance, in the case of a panel survey. Because of selectivity in the exit and entrance processes for the relevant population, the sample may become gradually less and less representative of the whole population. Selective panel nonresponse only adds to the problem. Time-dependent (longitudinal) weights may be necessary to achieve representativeness for each wave, both at the individual and at the household level (Duncan, 1994). This may also be useful for retrospective data—there is no nonresponse, but exit and entry processes in the past may have been different for the population compared with the sample.

In any case, one has to decide whether the sample should be representative of the larger population:

1. Only in terms of individuals, or also in terms of families or households
2. Only at one point in time or over a longer period

III. DATA SOURCES

1. Census

A large part of the data that we have on families and households stems from censuses. Indeed, as was indicated in the previous sections, the UN formulates recommendations for measurement and tabulation of such kind of data. Important aspects of census data are (Eggerickx and Bégeot 1993):

1. The characteristics of the whole population are mapped, not just a sample.[6]
2. Measurement in the form of a snapshot takes place at regular intervals, typically every decade.

These aspects point at the same time at a number of advantages and disadvantages. Because census data cover the whole population, they can be broken down by region, socioeconomic group, etc. Regular measurement facilitates an analysis of changes in household and family structure over a relatively long period. However, these changes can only be investigated in the form of net flows, as argued above. And although regional and socioeconomic detail can be given for the household and family variables, the huge undertaking each census is implies that it is rather costly and, hence, that only a few variables can be measured. Most often will we be able to give the household and family structure at census time only and not changes at the

[6] This is not the case for the few countries in which the census is a sample survey. This was the case in Norway in 1990 (sample of between 8% and 20%), and in 1991 in Germany (the 1% microcensus).

level of individuals or households. Another problem is that because of its administrative character, some household variables may be difficult to measure accurately, for instance, whether or not a person cohabits. Consequently, numbers of consensual unions based on a census have been underestimated by 30% to 50% in Norway, France, and England and Wales in the 1980s and 1990s (Keilman, 1995).

2. Cross-Sectional Sample Survey

A cross-sectional sample survey can be held just once, or it can be repeated a number of times. A one-time cross-sectional sample survey is relatively cheap, and therefore, it usually contains a large number of questions. The survey is most often organized for a specific purpose, and hence, it will be relatively easy, compared with the case of a census, to build up good rapport between respondent and interviewer. This increases the quality of the data. Yet well-known problems connected to the sample character remain: the sample can be biased, and the nonresponse can be selective. For instance, the proportion of women aged 20 to 24 years in Norway who live in a consensual union varies between 24% and 34% in five sample surveys around 1990 (Gulbrandsen and Gulbrandsen, 1993).[7] Another drawback is that the information obtained from a pure cross-sectional sample is limited: no information is obtained concerning changes. If the interest is in dynamics, and not only in the structure at one point in time, retrospective elements have to be included or the survey has to be repeated. If the same sample is used for each round, we speak of a panel. In case a fresh sample is drawn in each round, we have a repeated cross-sectional survey. An example is the annual household statistics for The Netherlands, which started in 1988 and is based on annual averages of data from the monthly Labor Force Surveys (Visser, 1995).

There are numerous cross-sectional sample surveys repeated on a regular basis that routinely collect past and present household information. The main difference with a multiround panel survey is, of course, that links between the household situations between two subsequent interview dates can only be given for the aggregate, not for individuals. Klijzing (1988) noted that few trend studies attempt to link up the differences in household composition, as observed between subsequent rounds. Part of the problem is that, from round to round, sample designs are frequently adapted to changing research objectives.

A repeated cross-sectional survey for which the household data are certainly not underutilized is the General Household Survey (GHS) in Great Britain. Established in 1971, the GHS has a standard sample size of around 10,000 private households each year (Harrop and Plewis, 1995). Originally, the GHS primarily produced stock data on household composition, but since 1979 questions on family formation were included as well. These questions covered aspects such as length of current cohabitation and type of premarital cohabitation (Brown and Kiernan, 1981; Haskey and Coleman, 1986). Hence, since 1979 the GHS may be typified as a multiround retrospective survey.

3. Retrospective Survey

In a retrospective survey, the sample is approached only once. However, the questions cover not only the present situation but also a certain period in the past. Retrospective surveys have gained popularity among demographers in the past few decades, not the least because they provide life-history (biographical) data. Murphy (1995) provides a useful overview. A prime example is the set of so-called Family and Fertility Surveys carried out in several European countries and co-ordinated by the UN-ECE (ECE 1992; Klijzing 1995). During the period 1988–1997, 20 countries in Europe, plus the United States, Canada, and New Zealand carried out the field work for such a Family and Fertility Survey, and all of them have included life-history data on partnership formation and dissolution (although with somewhat variable level of detail).

Compared with its competitor the panel survey, a retrospective survey is relatively cheap, and it produces immediate data on dynamics. A problem, however, is the quality of the data. In particular, retrospective data may be subject to errors of omission and misplacement—in particular for events that the respondent experienced as unpleasant (divorce, abortion), for which the timing is not clear-cut (start of a consensual union, leaving the parental home), or for which the occurrence took place in the remote past (Courgeau, 1991; Poulain *et al.*, 1991). In addition, to collect reliable retrospective information on individual values is next to impossible, except for those values that are so fundamental to the individual (characterizing perhaps his or her personality) that they are constant or almost constant over time. A final problem is

[7] Although the surveys cover a long period (1987–1992), this fact can not explain the large differences in the proportions. The high share of 34% dates from 1988 and the low one (24%) from 1992. Moreover, the earliest survey (1987) reports 30%. An apparent ***decrease*** in cohabitation is not plausible; it must be caused by sample errors of various kinds.

that of sample selection bias (Tuma and Hannan, 1984). Only persons alive and living in the region of interest at the time of the interview can answer the retrospective questions, and thus, the sample may become selective because of survival and emigration. Even if the sample is representative of the population *at interview time*, the *retrospective data* may be biased when the process of interest depends on mortality and emigration. When the sample is restricted to a special subgroup of the population (e.g., only married couples), the problem becomes even larger because there are additional exit processes (divorce and transition to widowhood in this example).

4. Panel Survey

If one would want to improve on the quality of retrospective data, one could rely on a panel instead. Panel data record the situation of sample members at two or more points in time. Sometimes, retrospective questions are added, which cover the period between two panel waves. This reduces memory effects clearly and facilitates the collection of data on current norms and values. Drawbacks, however, are that a panel is relatively costly, that the panel has to be followed-up over time, that selective panel drop-out may introduce a bias in the estimates, and that repeated interview may distort the answers or even the actual behavior of the respondents (Kasprzyk *et al*., 1989).

Usually a panel comes in the form of a sample survey, but the sample size varies a great deal. For instance, in the panel investigation on changing fertility, parenthood and family formation in Nordrhein-Westfalen (Federal Republic of Germany [FRG]), the sample dropped from 2620 women aged 18 to 30 years in the first round (late 1981/early 1982) to 1472 women 2 years later (Kaufmann *et al*., 1982). In contrast, we have the British Panel Household Survey, with annual waves between 1991 and 1994 (Buck *et al*., 1994). This source contains household and family data from nearly 13,000 individuals in about 5000 households in the first round. Other examples of panel surveys that may be used to trace family and household developments are the annual Panel Study of Income Dynamics in the United States, with a sample size of 5000 to 6000 families since 1968 (Duncan and Morgan, 1976), and the German Socioeconomic Panel, with annual waves for 6000 households since 1984 and an additional 2000 households from the former German Democratic Republic since 1990 (Ott, 1995).

During the years 1994–2001, the European Union (EU) carried out the so-called EC Household Panel, a multipurpose survey that covers demographic characteristics as well as a number of other topics (income, labor force, health, education, housing, migration) in the EU. Annual waves have been carried out based on a sample of approximately 61,000 households (about 127,000 individuals) in 12 EU countries. All members in the sample household aged 16 years or over were interviewed. The achieved sample size ranged from some 3500 households in Denmark to nearly 7500 households in Spain (except for Luxembourg, where the sample size was 1010 households). Austria, Finland, and Sweden were also included in wave 3 in 1996. An illustrative analysis based on the results of two waves of the pilot survey shows how data of this kind may be used to map dynamics in household composition and in individual relationships within households, among others (Verma, 1994).

A specific type of panel data is that which is obtained by linking the records of successive censuses. Data of this type are available for Sweden for the period 1975–1980 (Dellgran *et al*., 1984), for France for 1968–1995 (the so-called *Échantillon démographique permanent*), and for England and Wales for 1971–1991. The latter data source is known as the 1% Longitudinal Study. It is a record linkage study covering about 500,000 people in England and Wales (1% of the population). It includes a sample of the census records from 1971, 1981, and 1991. But register information for sample members regarding births (women only), deaths, and cancer has been linked to the census data (Murphy *et al*., 1988; Craig, 1990; Dale, 1993). The French *Échantillon démographique permanent* covers a little over 1/200 of the population. It is based on the census forms from the censuses of 1968, 1975, 1982, and 1990 and the civil registration forms for major demographic events (births, marriages, deaths, adoptions) for the years 1968–1975 (Ekert-Jaffé *et al*., 2002).

5. Register

Another possibility is to analyze population register data. There is no selectivity connected to this approach, but there are other problems. For instance, the *de jure* situation that the register reflects is sometimes only a crude approximation of the de facto situation that one is interested in. (This has implications not only for statistics on household membership, but also for migration statistics.) Furthermore, the number of variables is usually very limited. As most of the existing population registers record (changes in) formal marital status but not (those in) household status, only limited information on families and households can be obtained from such a source. Informal living arrangements such as consensual unions are difficult to measure with a register.[8] But exceptions are the Scandinavian countries (Denmark, Finland, Iceland, Norway, and Sweden) and The Netherlands. Concerning families, statistics in the form of stock data are published with regular intervals for the Scandinavian countries, for instance, annually for Finland and for Norway. Statistics Netherlands carried out register counts in 1987 and 1992 with ample information on families (Prins and Levering, 1994). Concerning households, Denmark has a household register (Noordhoek and Petersen, 1984; Petersen, 1985). Finally, Statistics Sweden combined data from the censuses of 1980 and 1985 with information from the registers of marriages, that of internal migration, and that of deaths. This resulted in stock and flow data for couple formation and dissolution (Statistics Sweden, 1992; Prinz *et al*., 1995).

IV. MAIN TRENDS IN FAMILY AND HOUSEHOLD DEVELOPMENTS IN WESTERN EUROPE AFTER WORLD WAR II

The main developments in numbers and types of households in Western Europe and other industrialized countries after World War II have been documented extensively (Hall, 1986; Burch and Matthews, 1987; Hoffmann-Nowotny, 1987; Keilman, 1987; Schwarz, 1988; Linke *et al*., 1990; Gonnot and Vukovich, 1989; Höpflinger, 1991; Haskey, 1993; Kuijsten, 1995, 1996). Therefore, I can restrict myself to a short summary. The extent to which the various countries experienced these trends is not always the same—in some countries the *levels* are lower than in others. This may, at least partly, be explained by the fact that definitions, concepts, and measurement of household and family variables differ among countries and within countries over time. However, the *direction* of the trends seems to be similar (except for a few policy-induced temporary effects, such as marriage peaks in Austria in 1983 and 1987 and in Sweden in 1989). However, at the end of this section,, I shall discuss the apparent uniformity of the trends.

[8] Previous address of both spouses is recorded when a marriage takes place in Austria, France, or Spain. This is one possibility for compiling statistics on consensual unions (De Santis and Santini, 1995), but the application is restricted to cohabiting couples who eventually marry. A second possibility is the practice followed by Statistics Finland and Statistics Denmark. When two persons of opposite sex are registered at the same address, do not have a common biological parent, are not married to each other, and have an age difference of less than 16 years (and, for Denmark, no other persons live at that address), the two are considered as a cohabiting couple.

1. Falling Average Household Size

There has been a continuous *fall in average household size*. Two demographic causes stand out: (1) numbers and proportions of persons living alone have greatly increased, and (2) family sizes were reduced, owing to the fall in fertility rates, especially those of higher birth orders (in addition to a modest increase in voluntary childlessness in some countries). Changes in the age structure of the populations concerned have also contributed to this development. An aged population has relatively many couples whose children have left the parental home and also relatively many elderly persons who live alone, but such changes contribute much less to the fall in the average size of households than the first two factors.

Information on average household size covering Europe as a whole is only available since 1950. Estimates by MacKellar *et al.* (1995) indicate that average household size in Western Europe dropped from 3.5 in 1950, via 3.1 in 1970, to 2.6 in 1990. The figures for eastern Europe were higher by 0.2 to 0.3 persons per household for those years. The average for developed countries in the year 2000 was 2.5, with a range between 2 and 4 (UN, 1995; UN Centre for Human Settlements, 2001). Country-specific data for the old EC indicate that average household size was between 2.5 (Denmark and the FRG) and 3.7 (Ireland) in 1980 and dropped to between 2.2 (Denmark) and 3.3 (Ireland, Spain) in 1991 (EC, 1993). Households in Europe are smallest in Sweden (2.1 persons on average in 1990), closely followed by Denmark (2.2) and Switzerland (2.3) (see Eurostat, 1995). Whereas the fall in birth rates and excess mortality among men are among the main demographic explanations for the falling household size in Western Europe around 1970, the recent drop is caused to a large extent by more frequent divorce and separation and by increasing longevity. Countries in eastern Europe had relatively high and stable fertility until the mid-1980s, when strong declines occurred (Coleman, 1996). This explains the somewhat larger households in that region, compared with the western part.

MacKellar *et al.* expect a modest further decrease in average household size for eastern Europe, to a level of 2.4 to 2.8 persons per household on average in the year 2030. For Western Europe a slight increase might occur to a level of 2.7 in the year 2030 when fertility would increase again, but in three other scenarios the authors expect a further drop by 0.1 to 0.3 persons per household by 2030. This agrees with Bumpass' (1990) view that sociological factors such as a shift in tastes toward privacy, and economic factors such as higher income, all lead to more atomized living arrangements in industrialized countries. Indeed, the following factors are related to shrinking households:

First, all other factors remaining the same, falling birth rates reduce population size but do not affect the number of households; hence, household size is reduced.

Second, social, economic, and cultural theories of demographic behavior point to a variety of reasons why individuals prefer to live in small households. These include less adherence to strict norms; less religiosity and increased individual freedom on ethical issues; female education, which has led to women having greater economic independence and also facilitates divorce; more assertiveness in favor of symmetrical sex roles; the contribution of women to the labor market; increased economic aspirations; and greater residential autonomy among young adults (Van de Kaa, 1987; Lesthaeghe, 1995; Verdon, 1998).

Finally, population ageing reduces household size. This is a direct consequence of two facts: increased longevity leads to longer periods of time when children do not live with their parents and the greater mortality of men, together with the usual age difference between spouses, results in many widows who live alone.

Smaller households, then, are the result of processes that cannot be reversed (such as modern contraception and liberalization from norms) or that we value for a number of reasons (such as women's emancipation).[9]

2. Increasing Importance of Consensual Unions

The *traditional family* has lost its dominant position. Marriage has been postponed to higher ages, and ever lower proportions of successive birth cohorts have ever married (or may be expected to do so) at age 40, say. Griffin *et al.* (1995) note that the increase in mean age at marriage has been stronger for women than for men in many European countries.

Consensual unions have become accepted widely—not only as a prelude to marriage, but in some countries also (and increasingly so) as an alternative to married life. De Santis and Santini (1995) have analyzed trends in observed age-specific shares of women living in consensual union for nine industrialized countries between 1975 and 1990. In seven of these, they noted an increase over time in the average share (the average taken over 35 ages between 15 and 50 years).

[9] Some processes are irreversible *and* appreciated at the same time.

The rising popularity of consensual unions went together with increased levels of extramarital childbearing in some countries. In other countries childbearing remained largely restricted to married couples. Lesthaeghe (2000) plotted national percentages of nonmarital births against proportions of cohabiting women aged 20 to 24 years and found the following groups of countries:[10]

1. *Low extramarital fertility together with a low incidence of cohabitation* in Mediterranean countries, Poland, and Japan
2. *Low prevalence of cohabitation but high nonmarital fertility* in Eastern Europe, Portugal, Ireland the UK and the US;
3. *High prevalence of cohabitation but low nonmarital fertility* in Western European populations such as Belgium, The Netherlands, the FRG, and Switzerland;
4. *High prevalence of cohabitation combined with parenthood* in Scandinavian countries, and, to a certain extent also in France, Austria, Estonia, Slovenia, the former German Democratic Republic, and Canada.

The picture is not static, as Lesthaeghe notes. For instance, Portugal has recently moved from the first to the second group, and Belgium, the Netherlands, and the FRG are on their way to the fourth group.

At the same time De Santis and Santini (1995) have noted an increasing tendency to live without a partner (neither cohabitee nor spouse) in Sweden (1975–1985), Canada (1984–1990), The Netherlands (1982–1988), Austria (1981–1989), Italy (1983–1988), and the FRG (1982–1986). For Norway a similar rise has been observed in the age group 25 to 44 years (1980–1995) (see Barstad, 1996). Also, Klijzing and Macura (1997) note increasing shares of women who, by the age of 25, never had entered a first partnership. The increase is observed for women born between 1950 and 1970 in Italy, Spain, Belgium (Flanders), the FRG, New Zealand, France, and The Netherlands.[11] Hence the fall

[10] Information stems from the Fertility and Family Survey project, with data collected in the various countries between 1988 and 1999. Note that Lesthaeghe's fertility indicator applies to women of all ages, whereas cohabitation is restricted to young women.

[11] The indicator used by Klijzing and Macura is different from that employed by De Santis and Santini. Entrance into partnership *after* age 25 will not be reflected by the former two authors' indicator, as opposed to the latter two authors' indicator. This explains why Sweden, Canada, and Austria appear in De Santis and Santini's list, whereas Klijzing and Macura note *no* increase for these countries. Nor is it observed in Eastern European countries such as Poland, Lithuania, Latvia, the German Democratic Republic, Estonia, central Europe (Switzerland, in addition to Austria), or northern Europe (Norway and Finland, in addition to Sweden).

in marriage rates generally observed for European countries has not fully been compensated by increasing shares of women living in a consensual union (see Haskey, 1993). Part of the rise in the proportions living alone is an indirect consequence of the increased popularity of consensual unions, as these unions have relatively high separation rates. This has been documented for Sweden (Hoem and Hoem, 1992; Trussell *et al.*, 1992a), The Netherlands (Klijzing, 1992; Manting, 1994), and Norway (Keilman and Brunborg, 1995; Texmon, 1999). But also the role of education and of psychological factors has been stressed (Van Hoorn, 1994).

Several authors have noted that cohabiting couples do not form a homogeneous group. Rindfuss and VandenHeuvel (1991) have stressed that the behavior of some cohabiting couples is close to that of married couples, whereas that of others is much more similar to the behavior of persons who live alone. Noack (1996) discusses whether a cohabiting couple should be regarded as "just lovers," as "engaged to be married," or as a de facto married couple.

3. Growing Numbers of Lone Parents

The nature of *one-parent families* has changed dramatically. Decreasing mortality and increasing divorce has led to fewer lone parents who are a widow(er), and many more divorced lone parents, in particular lone mothers. Although the share of births outside marriage went up sharply since the 1970s, this is not a major cause for the increase in one-parent families. The reason is that most of these children are born in a consensual union.

Data for the former FRG on the distribution by marital status of lone mothers indicate that the share of divorced lone mothers went up from 24% in 1961 to 46% in 1985. The share of widows dropped from 46% to 18% over the same period. For Switzerland the figures show a similar development between 1970 and 1980; for The Netherlands the changes during the years 1960–1983 are even stronger: from 14% to 60% for married lone mothers and from 71% to 21% for widows.

As a consequence of the growth in divorce rates, larger shares of children spend more and more time of their childhood in a one-parent family, usually with their mother. Studies have been carried out for Denmark, England and Wales, France, The Netherlands, Norway, and the United States, among others (Hofferth, 1985; Brown, 1986; Kuijsten and Voets, 1986; De Jong, 1989; Voets and Kuijsten, 1989; Haskey, 1990; Jensen *et al.*, 1991; Villeneuve-Gokalp, 1993; Duncan *et al.*, 1994; Festy, 1994). Longitudinal data collected for adolescents indicate that roughly between 15% and 30% of them will have lived in a one-parent family, depending on the age limit for these young adults (often between 18 and 21) and the type of family they were born in (two married parents, two cohabiting parents, or a lone mother). The figures are particularly high for blacks in the United States.

4. Higher Levels of Childlessness

A number of countries in northern, western and southern Europe experience growing levels of childlessness for subsequent birth cohorts of women.[12] Among the 11 countries listed by Höpflinger (1991), there are five in which women born in the 1950s have higher shares with parity zero than do women born around 1940. Prioux (1990a) notes that seven countries in western and northern Europe experienced low childlessness for women born between 1935 and 1947 (between 8% and 15% of these women are childless). Later cohorts have higher levels. Beets (1995) showed that for 17 European countries, the mean percentage childless at age 30 increased from 16% for women born in 1945 to 22% for the 1955 cohort and 27% for the 1960 cohort. Childlessness at age 30 *decreased* between cohorts born in 1935 and 1945 for England and Wales, France, Italy, The Netherlands, Norway, and Portugal. De Beer (1995) found that the share of childless women is particularly high in southern Europe in the early 1990s: over 30% as measured on a period basis, compared with 20% in Nordic countries. De Jong (1995) has analyzed births by parity in the 17 countries of the European Economic Area (EEA)[13] and notes that first birth rates were high around 1970. Then a decrease took place that lasted until the mid-1980s, with southern Europe following the pace set by northern and western Europe. In Ireland, Spain, and Italy the fall continues at least until 1990. Because of the period perspective adopted by De Beer and De Jong, these figures exaggerate the fall in cohort shares, particularly as the age at birth of the first child is rising in many European countries.[14] This implies that some of the women born

[12] Leridon (1999) presents evidence that suggests that proportions childless were more or less constant at a level close to 10% for women in Hungary, the Czech Republic, Poland, the former German Democratic Republic, and Yugoslavia who were born between 1940 and 1960.

[13] The 15 member states of the European Union plus Iceland and Norway.

[14] When birth generations 1950 and 1960 are compared, the mean age at motherhood (irrespective of parity) went up from 26.7 to 27.8 years for the average of the 17 EEA countries. Greece and Portugal are the only countries in which no rise took place. (see De Jong, 1995:table 4).

at the end of the 1960s and during the early 1970s may still have a first child after age 30. Indeed, first birth rates went slightly up again in the 1980s in Sweden, Norway, Denmark, Portugal, West Germany, The Netherlands, and France. Yet the figures indicate that a rise in cohort childlessness has to be expected, as current period levels of first birth rates are lower than those 20 years ago. This is also confirmed by recent data compiled by Prioux (2002). She suggests that shares of childless women born in the early 1960s in countries in northern, western, and southern Europe may increase to 15% to 20%.

How can these developments be explained? Prioux (1990a) discusses strictly demographic factors: the strong fall in first marriage rates, the relatively low marital first birth rates early in marriage, and, as a factor counteracting the latter two trends, the rise in extramarital fertility. Beets (1995) stresses the fact that postponement of the first birth may lead to higher infecundity: 30-year-old women who wish to become pregnant will succeed after approximately 3 months, but for women aged 35 it will take about 7 months. Hobcraft and Kiernan (1995) present a thoughtful discussion of the existing literature about the transition to parenthood from various perspectives: they elaborate both pronatalist forces and constraints on becoming a parent, involving biology, time, money, ideas, and security. They argue that fertility was high in the 1950s and 1960s for several reasons: costs of rearing children (education, health, welfare) were increasingly covered by the state rather than directly by the parents themselves; economic growth and permanent employment added to the anticipated security for parenthood; reconstruction of housing after the war improved quality and access; and the accelerated establishment of independent living, which usually involved marriage, contributed to earlier entry into parenthood. After 1970, the scene changed fundamentally (although Hobcraft and Kiernan also argue that some of the key changes can be dated back to the war time period or even further back). Increased female labor force participation reduced the time available for childrearing. Reliable means of fertility control broke the intimate connection between sex and reproduction. Patterns of partnership changed, with marriage becoming a more fragile institution and cohabitation becoming more widespread. In addition, employment insecurity delayed the achievement of the basic requirements for becoming a parent and increased the levels of uncertainty about future security. The authors are rather pessimistic for the future, although they assume that generous provisions by Scandinavian welfare states have led to higher entry rates into parenthood than in southern Europe, where commitment to marriage for women is relatively great, attachment to the labor force is lower, and state support for parenthood is less.

The analysis by Hobcraft and Kiernan confirms that prolonged education for women and their increased labor force participation per se are insufficient explanations. Indeed, Hoem (1992, 1995) argues for the case of Sweden that mothers with high levels of education also had the highest fertility in the 1970s and the first part of the 1980s. And in subsequent years, the labor force participation of women has increased (further), simultaneously with a substantial rise in fertility for *all* birth orders. Hence she questions the roles of improved educational level and higher labor force participation as the main explanatory factors in the intensity and timing of childbirth. Instead, she argues that new attitudes as to what is suitable at different ages in life should be analyzed further as potential explanations: childbearing is regarded as an important part of full life, but it is a part that many young people see as belonging to a later phase than earlier cohorts did. This argument points into the direction of *activity status* as an important determinant for childbearing, rather than educational level. This is confirmed by Kravdal (1994), who finds that the effect of educational level on first birth rates of Norwegian women is small, whereas being in education has a strong fertility-inhibiting impact. In other words, what matters for childbearing is not educational attainment, but educational participation. Being employed reduces the chances to become a mother somewhat.

5. Fluctuations in the Age at Leaving Home

During the 1960s and 1970s, young adults left the parental home at progressively younger ages. This has been documented for Sweden, Germany, The Netherlands, Great Britain, France, and Norway (as well as the United States and Australia) (for a summary of the literature, see Texmon, 1996). Among the reasons that explain this trend are changes in attitudes and values, increasing individualism, a decreasing tendency for living in families, improved economic conditions, and the increased availability of (student) dwellings. This trend stopped up in the 1980s, and it has turned into an increase in home leaving ages for some countries (Sweden, The Netherlands, the FRG, and England), in particular for men. Moors and Van Nimwegen (1990) present figures for the EU as a whole in 1982 and 1987 that show a trend toward later home leaving at ages between 21 and 24 for men, and at all ages between 15 and 24 for women. Economic difficulties in the housing market and the labor market are mentioned as factors behind this reversed trend. But part of the effect may also be explained by the fact that surveys are not always comparable (*first* or *final* leave, problems with measuring living with parents, etc) (see the particularly thoughtful analysis by Texmon, 1996). It is unclear whether the tendency for young adults to leave the parental home at higher age as observed in the 1980s has continued in recent years. Figures presented by Visser (1995) for The Netherlands indicate that at least for men this seems *not* to be the case: in 1988/1989 63% of men aged 20 to 24 lived with their parents, and in 1993/1994 only 60%. (For women the decrease was only two percentage points.) On the other hand, Italian men and women *did* leave their parents at a higher age between 1983–1985 and 1995–1996 (Manting and Alders, 1998).

6. Two Qualifiers

It should be stressed that many of the trends noted here have a long history, which started well before World War II. Drops in fertility and family size, and an increase in divorce were first in evidence in the 19th century. The movement of women in the labor force and the growth of sexual permissiveness became noteworthy in the first half of the 20th century (Popenoe, 1991). More surprising, however, was the fact that in the 1930s birth rates began to rise, and that immediately after World War II many Western countries experienced a "family renaissance." Although family life was also valued before 1950, there was an increasing emphasis in the decade thereafter on marriage and parenthood as central to fulfilling life. Popenoe argues that one explanation for this phenomenon is that the unparalleled affluence and upward mobility after the war gave the working classes the opportunity to achieve the bourgeois family ideal. With less education and access only to inferior jobs, which they held out of economic necessity, the women of the working class had long looked with envy at the nonworking wife middle-class wife. Other explanations point to political circumstances, such as the Allied victory, which led to a renewed confidence in the stability of the democratic state, or the Cold War, which undermined confidence in the future and led people to retreat to the safety of home and family (Cherlin, 1992). However, it is problematic that all these arguments point to circumstances after the war. They fail to explain why fertility and marriage rates began to rise in the 1930s already, immediately after the Great Depression, and why the increase continued during the war in a number of western countries. More generally, when one tries to explain developments in household and reproductive behavior, attention should also be given to what happened to families and households between 1930 and 1950 and not only to the events since the mid-1960s. But others have argued that the 1960s *did* imply a break. Lesthaeghe and Van de Kaa have been among those who most strongly have stressed the role of tastes and aspirations and of norms and attitudes in the explanation of family and fertility changes in industrialized countries after World War II (Lesthaeghe, 1983; Van de Kaa, 1987; Lesthaeghe and Surkyn, 1988). During the 1960s and thereafter, a strong shift in norms toward progressiveness and individualism, and a focus on self-fulfilment implied increased individual freedom from religious and political institutions, from peers, and from parents. At the same time, the economic position of young men has deteriorated (Oppenheimer, 1994). These developments, which have been in operation for the past 3 decades, have had led to lower levels of fertility and marriage.

A second qualifier is that the family and household trends described above are rather similar for various countries as long as an analysis at only a superficial level is employed, for instance, in the form of aggregate demographic rates for family and household dynamics. But more detailed investigations reveal a great deal of pluriformity. Hobcraft and Kiernan (1995) stress the difference between Scandinavia and southern Europe when they analyze the constraints around parenthood. Lelièvre (1994, 1995) gives a detailed comparison of couple formation and the birth of the first child in Great Britain and France. She concludes that the recent rise of births outside marriage and the extent of extramarital cohabitation are very different between the two countries, despite apparently similar trends. Mellens (1999) notes that female labor force participation is high both in eastern and in northern Europe, but for different reasons. In former communist countries, many women belong or belonged to the labor force out of economic necessity—in the north, the explanation is to be found in nontraditional sex roles. Kuijsten (1996) criticizes expectations of a continuing convergence of demographic and family trends in Europe toward uniformity, as expressed by various authors. The material he presents for ten European countries on combined family status/labor market status for women shows a picture of substantial variation between the countries. These differences cannot be explained as intercountry differentials in the speed toward a uniform model of family and household patterns. This leads Kuijsten to the belief that the Swedish model—with high shares of cohabiting couples, lone mothers and one-person households, and high labor force participation among women—does not necessarily predict the future situation in other countries. For instance, the family life in Germany (in particular in the former FRG) is highly polarized. There is a strong and growing nonfamily sector (consensual unions are not interpreted as families in Kuijsten's analysis), but at the same time there are fewer employed married mothers aged 25 to 29 than traditional housewives in that group. In general, Kuijsten endorses the view that industrial countries have entered an era of new biographical models with a pluralization of life styles, both over the life course of the individual and in the cross-sectional picture at one point in time.

References

Ås Dagfinn, 1990. Det er husholdninger som bor i boligene: om å telle og beskrive husholdninger. *in*: Ole Gulbrandsen and Bjørg MOEN (eds.) *Husholdninger: data og definisjoner. Prosjektrapport 63.* Oslo, Norwegian Institute for Building Research, 97 p.

Barstad Anders, 1996. Flere aleneboende, men færre venneløse. *Smafunnsspeilet*, no. 1/1996, p. 24–29.

Beets Gijs, 1995. Does the increasing age at first birth lead to increases in involuntary childlessness? The Hague, NIDI.

Berge Erling, 1988. *The household: a conceptual discussion.* Oslo, Institutt for Sosialforskning, 35 p.

Brown Audrey, 1986. Family circumstances of young children. *Population Trends*, no. 43, p. 18–23.

Brown Audrey and Kiernan Kathleen, 1981. Cohabitation in Great Britain: evidence from the General Household Survey. *Population Trends*, no. 25, p. 4–10.

Brynin Malcolm and Smith Rachel, 1995. Mapping the household. *Journal of Economic and Social Measurement*, vol. 21(2), p. 127–144.

Buck Nick, Gershuny Jonathan, Rose David and Scott Jacqueline, 1994. *Changing Households: the BHPS 1990 to 1992.* Colchester, ESRC Research Centre on Micro-Social Change, 319 p.

Bumpass Larry, 1990. What's happening to the family? Interactions between demographic and institutional change. *Demography*, vol. 27(4), p. 483–498.

Burch Thomas, 1995. Theories of household formation: progress and challenges, *in*. Evert Van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo Van Wissen (eds.), *Household demography and household modeling*, p. 85–108. New York, Plenum Press 369 p.

Burch Thomas and Matthews Beverly, 1987. Household formation in developed countries. *Population and Development Review*, vol. 13(3), p. 495–511.

Cherlin Andrew, 1992. *Marriage, divorce, remarriage: revised and enlarged edition.* Cambridge, Harvard University Press, 178 p.

Citro Constance, 1985. Alternative definitions of longitudinal households in the Income Survey Development Program: implications for annual statistics, *in*: Delma Frankel (ed.), *Survey of Income and Program Participation 1985*, p. 21–26. Washington, US Bureau of the Census, 91 p.

Coleman David, 1996. New patterns and trends in European fertility: international and sub-national comparisons, *in*: D. Coleman (ed.), *Europe's population in the 1990s*, p. 1–61. Oxford, Oxford University Press, 346 p.

Courgeau Daniel, 1991. Analyse des données biographiques erronées. *Population* vol. 46(1), p. 89–104. (Also published in English as Impact of response errors on event history analysis, *Population, An English Selection*, 1992, vol. 4, p. 97–110.)

Craig John, 1990. Some reflections on the history of the OPCS Longitudinal Study. *Population Trends*, 61, p. 43–45.

Dale Angela, 1993. The Office of Population Censuses and Surveys Longitudinal Study. Special Issue of *Environment and Planning A*, 25(10).

De Beer Joop, 1995. *New long-term population scenarios for the European Economic Area: general outline*. Paper Eurostat/Statec Seminar New Long-Term Scenarios for the European Economic Area, Luxembourg, November 1995.

De Jong Andries, 1989. Een op de zes kinderen maakt voor de 21-ste verjaardag echtscheiding ouders mee. *Maandstatistiek van de Bevolking*, vol. 37(3), p. 14.

De Jong Andries, 1995. *Long-term fertility scenarios for the countries of the European Economic Area.* Paper Eurostat/Statec Seminar New Long-Term Scenarios for the European Economic Area, Luxembourg, November 1995.

Dellgran Peter, Holmberg Ingvar and Olovsson Paul, 1984. *Hushållsbildning och Bostadsefterfrågan*. Demographic Research Unit Reports 16, Gothenburg University, 260 p.

De Santis Gustavo and Santini Antonio, 1995. Formation et dissolution des couples: mariage et cohabitation, *in*: Josianne Duchêne and Guillaume Wunsch (eds.) *Collecte et Comparabilité des Données Démographiques et Sociales en Europe*, p. 107–132. Louvain-la-Neuve: Academia L'Harmattan, 671 p.

Deven Freddy, 1995. *What about "reconstituted" families?* Paper workshop Household and family statistics at the dawn of the 21st century organized by Eurostat and the New University of Lisbon, Lisbon, February 1995.

Duncan Greg, 1994. *Weighting and quality control measures in the EC Panel Project*. Luxembourg, Eurostat, 17 p.

Duncan Greg and Hill Martha, 1985. Conceptions of longitudinal households: fertile or futile? *Journal of Economic and Social Measurement*, vol. 13(3/4), p. 361–375.

Duncan Greg and Morgan James, 1976. *Five Thousand American Families.* Ann Arbor, Institute for Social Research, 436 p.

Duncan Greg, Yeung Jean and Rodgers Willard, 1994. Les familles monoparentales aux États-Unis: dynamique, niveau de vie et conséquences sur le développement de l'enfant. *Population*, vol. 49(6), p. 1419–1435.

Economic Commission for Europe [ECE], 1992. *Fertility and family surveys in countries of the ECE region: questionnaire and codebook.* Geneva: United Nations Economic Commission for Europe, Population Activities Unit, 124 p.

Economic Commission for Europe [ECE] and Eurostat, 1998. *Recommendations for the 2000 censuses of population and housing in the ECE region*. New York, United Nations Statistical Commission, ECE, 82 p.

Eggerickx Thierry and Bégeot François, 1993. Les recensements en Europe dans les années 1990: de la diversité des pratiques nationales à la comparabilité internationale des résultats. *Population*, vol. 48(6), p. 1705–1732.

Ekert-Jaffé Olivia, Joshi Heather, Lynch Kevin, Mougin Rémi and Rendall Michael, 2002. Fécondité, calendrier des naissances et milieu social en France et en Grande-Bretagne: politiques sociales et polarisation socioprofessionnelle. *Population*, vol. 57(3), p. 485–518.

Ermisch John, 1988. An economic perspective on household modelling, *in*: Nico Keilman, Anton Kuijsten, and Ad Vossen (eds.) *Modelling household formation and dissolution*, p. 23–40. Oxford, Clarendon Press, 298 p.

Ermisch John and Overton Elizabeth, 1985. Minimal household units: a new approach to the analysis of household formation. *Population Studies*, vol. 39, p. 33–54.

European Commission [EC], 1993. *Statistics on housing in the European Community 1993.* Brussels, EC, 138 p.

Eurostat, 1995. Ménages et familles dans l'Espace Économique Européen. *Statistique en bref: population et conditions sociales* Luxembourg, Eurostat, 11 p.

Festy Patrick, 1994. L'enfant dans la famille. Vingt ans de changement dans l'environnement familial des enfants. *Population*, vol. 49(6), p. 1245–1296

Galler Heinz-Peter and Ott Notburga, 1993. *Empirische Haushaltsforschung: Erhebungskonzepte und Analyseansätze angesichts neuer Lebensformen. Reihe Stiftung de Private Haushalt Band 16.* Frankfurt, Campus Verlag 214 p.

Gisser Richard, 1995. *Concepts and typologies of household and family in the 1981 and 1991 population censuses in Austria.* Paper ISEGI-Eurostat workshop Les statistiques sur les ménages et familles à l'aube du 21e siècle, Lisbon, 16–17 February.

Gonnot Jean-Pierre and Vukovich Gabriella, 1989). *Recent trends in living arrangements in fourteen industrialized countries.* Working Paper nr. WP-89–34. Laxenburg: IIASA, 37p.

Griffin T., Mikkelsen Lene and Kalampasut Parisudhi, 1995. A statistical profile of youth in Europe and North-America. *Statistical Journal of the United Nations ECE*, vol. 12, p. 329–368.

Gulbrandsen Lars and Gulbrandsen Ole, 1993. Registerdata og surveydata, *in*: *Statistik utan Gränser Statistics without Frontiers.* Proceedings 19th Conference of Nordic Statisticians, p. 310–323. Reykjavik, Conference of Nordic Statisticians, 517 p.

Hall Richard, 1986. Household trends within Western Europe, 1970–1980, *in*: Allan Findlay and Paul White (eds.) *West European Population Change*, p. 19–34. London, Croom Helm, 253 p.

Harrop Anne and Plewis Ian, 1995. Two decades of family change: secondary analysis of continuous government surveys. *Journal of the Royal Statistical Society A*, vol. 158(1), p. 91–106.

Haskey John, 1990. Children in families broken by divorce. *Population Trends*, vol. 61, p. 34–42.

Haskey John, 1993. Formation and dissolution of unions in the different countries in Europe, *in*: Alain Blum and Jean-Louis Rallu (eds.) *European population II: demographic dynamics*, p. 211–229. Paris, John Libbey Eurotext, 526 p.

Haskey John and Coleman David, 1986. Cohabitation before marriage: a comparison of information from marriage registration and the General Household Survey. *Population Trends*, vol. 43, p. 15–17.

Hobcraft John and Kiernan Kathleen, 1995. Becoming a parent in Europe, *in*: *Evolution or Revolution in Europe. European Population Conference Milano September 1995. Vol. 1: Plenary Sessions*, p. 27–65. Milano, Franco Angeli, 298 p.

Hoem Britta, 1992. *Recent changes in family formation in Sweden. Stockholm Research Report in Demography no. 71.* Stockholm, Stockholm University, 42 p.

Hoem Britta, 1995. Sweden, *in*: Hans-Peter Blossfeld (ed.), *Family formation in modern societies and the new role of women; a cross-national comparative study of nine countries*, p. 35–55. Boulder, Westview Press, 266 p.

Hoem Britta and Hoem Jan, 1992. The disruption of marital and non-marital unions in contemporary Sweden, *in*: James Trussell, Richard Hankinson, and Judith Tilton (eds.) *Demographic applications of event history analysis*, p. 61–93. Oxford, Clarendon Press, 276 p.

Hofferth Sandra, 1985. Updating children's life courses. *Journal of Marriage and the Family*, vol. 47(1), p. 93–115.

Hoffmann-Nowotny Hans-Joachim, 1987. The future of the family, *in*: *Proceedings European Population Conference 1987*, p. 113–198. Helsinki, Central Statistical Office of Finland, 434 p.

Höpflinger François, 1991. The future of household and family structures in Europe, *in*: *Proceedings of the Seminar on present demographic trends and lifestyles in Europe*, p. 291–338. Strasbourg: Council of Europe, 437 p.

Jensen An-Magritt, Moen Bjørg and Clausen Stein-Erik, 1991. *Enebarn, delebarn, stebarn? Barns familier i lys av nye samlivsmønstre blant foreldre i 1970- og 80-åra.* NIBR-Rapport 1991:14. Oslo, Norsk institutt for by- og regionforskning, 95 p.

Kalibova Kveta, 1991. *Typology of households and families: Czechoslovak experience.* Paper 1991 Quetelet Seminar The collection and comparability of demographic and social data in Europe, Gembloux, Belgium, September 1991, 12 p.

Kasprzyk Daniel, Duncan Greg, Kalton Graham and Singh M. P. (eds.), 1989. *Panel surveys.* New York, Wiley, 592 p.

Kaufmann Franz-Xavir, Herlth Alois, Quitmann Joachim, Simm Regina and Strohmeier Klaus-Peter, 1982. Familienent-wicklung: Generatives Verhalten im familialen Kontext. *Zeitschrift für Bevölkerungswissenschaft* vol. 8(4) p. 523–545.

Keilman Nico, 1987. Recent trends in family and household composition in Europe. *European Journal of Population*, vol. 3, p. 297–325.

Keilman Nico, 1995. Household concepts and household definitions in Western Europe: different levels but similar trends in household developments, *in*: Evert Van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo Van Wissen (eds.) *Household demography and household modelling*, p. 111–135. New York, Plenum Press, 369 p.

Keilman Nico, 2004. Démographie des ménages et de la famille: Application aux pays développés, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.) *Démographie: Analyse et Synthèse Vol. VI Population et Société*, p. 345–388. Paris, Éditions de l'Institut d'Études Démographiques 582 p.

Keilman Nico and Brunborg Helge, 1995. Household projections for Norway 1990–2020. Oslo, Statistics Norway, 82 p.

Keilman Nico and Kinsella Kevin, 2004. Families and households in the 2000 round of censuses in ECE member countries. Paper Joint UNECE-Eurostat Work Session on Population Censuses, Geneva, November 23–25, 2004, 42 p.

Keilman Nico, Kuijsten Anton and Vossen Ad (eds.), 1988. *Modelling household formation and dissolution* Oxford, Clarendon Press, 298 p.

Klijzing Erik, 1988. Household data from surveys containing information for individuals, *in*: Nico Keilman, Anton Kuijsten, and Ad Vossen (eds.) *Modelling household formation and dissolution*, p. 43–55. Oxford, Clarendon Press, 298 p.

Klijzing Erik, 1992. "Weeding" in the Netherlands: first-union disruption among men and women born between 1928 and 1965. *European Sociological Review*, vol. 8, p. 53–72.

Klijzing Erik, 1995. *International comparability of fertility and family survey (FFS) data.* Paper Annual Meeting of the Population Association of America, San Francisco, April 1995, 10 p.

Klijzing Erik and Macura Miroslav, 1997. Cohabitation and extra-marital childbearing: Early FFS evidence, *in*: IUSSP *Proceedings of the International Population Conference Beijing 1997: Volume 2*, p. 885–901. Liège, IUSSP.

Kormendi Esther, 1995. *Household definitions in the EU Labour Force Survey.* Paper ISEGI-Eurostat workshop Les statistiques sur les ménages et familles à l'aube du 21e siècle, Lisbon, February, 16–17, 1995, 26 p.

Kravdal Øystein, 1994. The importance of economic activity, economic potential and economic resources for the timing of first births in Norway. *Population Studies*, vol. 48(2) p. 249–267.

Kuijsten Anton, 1995. Recent trends in household and family structure in Europe: an overview, *in*: Evert Van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo Van Wissen (eds.) *Household*

demography and household modelling, p. 53–84. New York, Plenum Press, 369 p.
Kuijsten Anton, 1996. Changing family patterns in Europe: a case of divergence? *European Journal of Population*, vol. 12(2), p. 115–143.
Kuijsten Anton and Voets Saskia, 1986. Relatievorming: het perspectief van de kinderen, *in*: D. J. van de Kaa and R. Lesthaeghe (eds.), *Bevolking: Groei en Krimp*, p. 37–58. Deventer, Van Loghum Slaterus, 115 p.
Lelièvre Éva, 1994. Formation des couples et fécondité hors mariage en Grande-Bretagne: divergences et similitudes avec la situation française. *Population*, vol. 49(1), p. 61–90.
Lelièvre Éva, 1995. Couple formation and fertility outside marriage in Great Britain: differences and similarities with the French situation, *Population, An English Selection*, vol. 7, p. 67–94.
Leridon Henri, 1993. Recomposer les familles dans les sources statistiques, *in*: Marie-Thérèse Meulders-Klein and Irène THÉRY (eds.) *Les Recompositions Familiales Aujourd'hui*, p. 51–66. Paris, Nathan, 350 p.
Leridon Henri, 1999. Les nouveaux modes de planification de la famille en Europe, *in*: Dirk Van De Kaa, Henri Leridon, Giuseppo Gesano, and Marek Okolski (eds.), *European Populations: Unity in Diversity*, p. 51–76. Dordrecht, Kluwer, 194 p.
Lesthaeghe Ron, 1983. A century of demographic and cultural change in Western Europe: an exploration of underlying dimensions. *Population and Development Review*, vol. 9(3), p. 411–435.
Lesthaeghe Ron, 1995. The second demographic transition, *in*: Karen Mason and An-Magritt Jensen (eds.), *Gender and family change in industrial countries*, p. 17–62. Oxford, Clarendon Press, 329 p.
Lesthaeghe Ron, 2000. Europe's demographic issues: Fertility, household formation and replacement migration. Interuniversity papers in demography no. IPD-WP 2000–6. Brussels: Vrije Universiteit Brussel/ Gent: Universiteit Gent.
Lesthaeghe Ron and Surkyn Johan, 1988. Cultural dynamics and economic theories of fertility change. *Population and Development Review*, vol. 9(3), p. 411–435.
Linke Wilfried, 1988. The headship rate approach in modelling households: the case of the Federal Republic of Germany, *in*: Nico Keilman, Anton Kuijsten and Ad Vossen (eds.) *Modelling Household Formation and Dissolution*, p. 108–122. Oxford, Clarendon Press, 298 p.
Linke Wilfried, De Saboulin Michel, Baldursson Gudni and Kuijsten Anton, 1990. *Household structures in Europe: Report of the Select Committee of Experts on household structures.* Population studies no. 22. Strasbourg, Council of Europe, 128 p.
Mackellar F. Landis, Lutz Wolfgang, Prinz Christopher and Goujon Anne, 1995. Population, households, and CO_2 emissions. *Population and Development Review*, vol. 21(4), p. 849–865.
Manting Dorien, 1994. *Dynamics in marriage and cohabitation: an inter-temporal life course analysis of first union formation and dissolution*. Amsterdam, Thesis Publishers, 224 p.
Manting Dorien and Alders Maarten, 1998. Household scenarios for the European Union. *Maandstatistiek van de bevolking*, vol. 46(10), p. 11–27.
McMillan David B. and Herriot Roger, 1985. Toward a longitudinal definition of households. *Journal of Economic and Social Measurement*, vol. 13(3/4), p. 349–360.
Mellens Martin, 1999. Uniformity and diversity defined, *in*: Joop De Beer and Leo Van Wissen (eds.) *Europe: one continent, different worlds. population scenarios for the 21st century*, p. 33–41. Dordrecht, Kluwer Academic Publishers, 189 p.
Moors Hein and Van Nimwegen Nico, 1990. Social and demographic effects of changing household structures on children and young people, *in*: *Council of Europe, Proceedings of the Seminar on Present Demographic Trends and Lifestyles in Europe*, p. 113–158. Strasbourg, Council of Europe, 437 p.
Murphy Michael, 1991a. Household modelling and forecasting: dynamic approaches with use of linked Census data. *Environment and Planning A* vol. 23, p. 885–902.
Murphy Michael, 1991b. Modelling households: a synthesis, *in*: Michael Murphy and John Hobcraft (eds.) *Population research in Britain*, p. 157–176. Supplement to *Population Studies*, 45.
Murphy Michael, 1995. The collection and comparability of demographic and social data in Europe: the family life cycle, *in*: Josianne Duchêne and Guillaume Wunsch (eds.) *Collecte et Comparabilité des Données Démographiques et Sociales en Europe*, p. 149–182. Louvain-la-Neuve, Academia L'Harmattan, 672 p.
Murphy Michael, Sullivan Oriel and Brown Audrey, 1988. Sources of data for modelling household change with special reference to the OPCS 1% Longitudinal Study, *in*: Nico Keilman, Anton Kuijsten and Ad Vossen (eds.) *Modelling household formation and dissolution*, p. 56–66. Oxford, Clarendon Press, 298 p.
Noack Turid, 1996. Familieutvikling i demografisk perspektiv, *in*: Berit Brandth and Kari Moxnes (eds.) *Familie for Tiden: Stabilitet og Forandring*, p. 11–29. Oslo, Tano Aschehoug, 286 p.
Noordhoek Johannes and Petersen Ole, 1984. Household and family concepts in Danish population registers and surveys. *Statistical Journal of the United Nations*, vol. 2(2).
Oppenheimer Valerie Kincade, 1994. Women's rising employment and the future of the family in industrial societies. *Population and Development Review*, vol. 20(2), p. 293–342.
Organization for Economic Cooperation and Development [OECD], 1979. *Child and family: demographic developments in OECD dountries*. Paris, Centre for Educational Research and Innovation, OECD, 218 p.
Ott Notburga, 1995. The use of panel data in the analysis of household structures, *in*: Evert Van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo Van Wissen (eds.), *Household demography and household modelling*, p. 163–183. New York, Plenum Press, 369 p.
Petersen Ole, 1985. Register-based family and household statistics, *in*: *The seventh Nordic Demographic Symposium*, p. 74–98. Helsinki, The Scandinavian Demographic Society (Scandinavian Population Studies 7) 199 p.
Popenoe David, 1991. *Disturbing the nest: family change and decline in modern societies.* New York, Aldine de Gruyter, 407 p.
Poulain Michel, Riandey Benoît and Firdion Jean-Marie, 1991. Enquête biographique et registre belge de population: une confrontation des données. *Population*, vol. 46(1), p. 65–88.
Prins Kees J. M. and Levering Jacques, 1994. Gezinnen naar gezinsfase, 1 januari 1992. *Maandstatistiek van de Bevolking* vol. 42(5), p. 30–39.
Prinz Christopher, Nilsson Åke and Sellerfors Hakan, 1995. Alternative options for living arrangement models: a sensitivity analysis, *in*: Evert Van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo Van Wissen (eds.), *Household demography and household modelling*, p. 227–250. New York, Plenum Press, 369 p.
Prioux France, 1990a. La fréquence de l'infécondité dans les cohortes, *in*: *La famille dans les pays développés: permanences et changements*, p. 161–175. Paris, INED, 316 p.
Prioux France (ed.), 1990b. *La famille dans les pays développés: permanences et changements.* Actes du Séminaire sur les nouvelles formes de vie familiale, Vaucresson, octobre 1987. Paris, INED, 316 p.
Prioux France, 2002. Avoir moins d'enfants: Une évolution inéluctable?, *in*: AIDELF (ed.), *Vivre plus longtemps, avoir moins d'enfants, quelles implications?*, p. 29–48. Paris, PUF, 737 p.

Rindfuss Ronald and Vandenheuvel Audrey, 1991. Cohabitation: a precursor to marriage or an alternative to being single? *Population and Development Review*, vol. 16(4).

Schjalm Arnfinn, 1996. *Kvalitetsundersøkelsen for Folke- og boligtelling 1990.* Rapporter no. 96/10. Oslo, Statistics Norway, 36 p.

Schwarz Karl, 1983. Les ménages en République Fédérale d'Allemagne, 1961–1972–1981. *Population*, vol. 38(3), p. 565–586.

Schwarz Karl, 1988. Household trends in Europe after World War II, *in*: Nico Keilman, Anton Kuijsten, and Ad Vossen (eds.), *Modelling household formation and dissolution*, p. 67–83. Oxford, Clarendon Press, 298 p.

Skiri Halvard, 1995. *Role and status of civil registration (population registration) and vital statistics systems in Norway.* Notater no. 95/41. Oslo, Statistics Norway, 25 p.

Statistics Finland, 1996. *Population Census 1995 Handbook.* Helsinki, Statistics Finland, 67 p.

Statistics Norway, 1985. *Forslag til standarder for kjennemerker knyttet til familier og husholdninger.* Interne Notater no. 85/31. Oslo, Statistics Norway, 36 p.

Statistics Norway, 1994. *Folke- og Boligtelling 1990 Evaluering. Instilling fra et utvalg som har vurdert tellingen.* Notater no. 94/3. Oslo, Statistics Norway, 46 p.

Statistics Norway, 1995a. De færreste samboere planlegger snarlig giftermål. *Ukens statistikk*, vol. 6, p 3.

Statistics Norway, 1995b. *-Familiestatistikk, 1. januar 1995. Ukens statistikk*, vol. 39, p. 15–20.

Statistics Norway, 1996. *Barnestatistikk, 1. januar 1995. Ukens statistikk*, vol. 10, p. 9–13.

Statistics Norway, 1999. *Folke: og boligtelling 1990.* Norges offisielle statistikk nr C 524. Oslo, Statistics Norway, 205 p.

Statistics Norway, 2000. *Familier og barn, 1. januar 1999. Aktuelle befolkningstall* vol. 3, 17 p.

Statistics Sweden, 1992. *Familjeförändringer kring 1980 och 1985.* Statistiska meddelanden Be15SM9201. Stockholm: Statistiska Centralbyrån, 142 p.

Texmon Inger, 1996. Leaving the parental home among young adults. *Economic Survey*, vol. 6(3), p. 32–42.

Texmon Inger, 1999. Samliv i Norge mot slutten av 1900-tallet. En beskrivelse av endringer og mangfold, *in*: *Samboerne og Samfunnet. Norges offentlige utredninger 1999*, p. 251–285. Oslo, Statens forvaltningstjeneste/Statens trykning, 304 p.

Todd J. and Griffiths D., 1986. *Changing the definition of a household.* London, Her Majesty's Stationery Office, 43 p.

Trost Jan, 1988. Cohabitation and marriage: transitional pattern, different lifestyle, or just another legal form, *in*: Hein Moors and Jeannette Schoorl (eds.), *Lifestyles, contraception and parenthood*, p. 3–14. The Hague and Brussels, Netherlands Interdisciplinary Demographic Institute (NIDI) and the Population and Family Study Centre (CBGS), 407 p.

Trost Jan, 1990. Stabilité et transformation de la famille, *in*: France Prioux (ed.), *La famille dans les pays développés: permanences et changements. Actes du Séminaire sur les nouvelles formes de vie familiale, Vaucresson, octobre 1987*, p. 25–39. Paris, INED, 316 p.

Trussell James, Hankinson Richard and Tilton Judith (eds.), 1992b. *Demographic Applications of Event History Analysis.* Oxford, Clarendon Press, 276 p.

Trussell James, Rodriguez German and Vaughan Barbara, 1992a. Union dissolution in Sweden, *in*: Trussell *et al.* (eds.), Trussell James, Hankinson Richard, and Tilton Judith (eds.), 1992b. *Demographic applications of event history analysis*, p. 38–60. Oxford, Clarendon Press, 276 p.

Tuma Nancy and Hannan Michael, 1984. *Social dynamics: models and methods.* Orlando, Academic Press, 578 p.

United Nations [UN], 1989. *Demographic yearbook/annuaire démographique: thirty-ninth issue/trente-neuvième édition.* New York, UN, 1399 p.

United Nations [UN], 1997. *Demographic yearbook/annuaire démographique: forty seventh issue/quarante-septième édition.* New York, UN, 1087 p.

United Nations [UN], 1998. *Recommendations for the 2000 censuses of population and housing in the ECE region.* New York, UN, 85 p.

United Nations Centre for Human Settlements, 2001. *Cities in a globalizing world.* London, Earthscan Publ. Ltd, 344 p.

Van De Kaa, Dick, 1987. Europe's second demographic transition. *Population Bulletin* vol. 42(1), p. 1–57.

Van Hoorn Wim, 1994. *Living alone: choice, coincidence or fate.* Paper ECE/Eurostat Joint Work Session on Demographic Projections. Mondorf-les-Bains, Luxembourg, June 1994, 26 p.

Verdon, Michel, 1998. *Rethinking Households.* London, Routledge, 220 p.

Verma V., 1994. *European Community Household Panel (ECHP): longitudinal analysis of the pilot survey.* Paper Eurostat. Luxembourg: Eurostat, 78 p.

Villeneuve-Gokalp Catherine, 1993. De la famille d'origine à la famille recomposée, *in*: Marie-Thérèse Meulders-Klein and Irène HÉRY (eds.), *Les recompositions familiales aujourd'hui*, p. 67–79. Paris, Nathan, 350 p.

Visser Harry, 1995. Sterke groei van niet-gehuwd samenwonenden: enkele belangrijke uitkomsten van de Jaarlijkse Huishoudensstatistiek, 1988–1994. *Maandstatistiek van de Bevolking* vol. 43(4) 13–18, p. 23–27.

Voets Saskia and Kuijsten Anton, 1989. Echtscheiding en leefsituatieveranderingen van kinderen: een overzicht van benaderingen. *Bevolking en Gezin* vol. 2, p. 73–101.

Wall Richard, 1996. Comparer ménages et familles au niveau Européen: problèmes et perspectives. *Population*, vol. 51(1), p. 93–116.

C H A P T E R

92

Social Mobility

Then and Now

FRANÇOIS HÉRAN

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

The sociology of networks has turned the study of social mobility into the most formalized branch of quantitative sociology and, as a result, the most international. Specialists in the field have a reputation of forming a closed circle. As quantitative sociology is, itself, a minority branch of sociology as a whole—in both Europe and America—the analysts of social-mobility now form a minority within a minority among sociologists. Hence the temptation to attach value judgments to this singular status: what some observers will see as a subsect of the discipline, others will describe as the elite of the elite.

Fortunately, this stark choice seems ever less valid. On the contrary, recent advances in the study of social mobility have promoted a broad opening of the specialty by lowering the entrance barriers that had been steadily erected in the 1970s and 1980s. The technical revolution introduced by the application of log-linear models simplified the approach to social mobility. It rapidly made obsolete most of the research efforts hitherto expended in an all-out race to sophistication. The change also made it possible to reintroduce social-mobility phenomena into the wider analysis of *life chances*—Max Weber's predilect *Lebenschancen*—which include educational attainment, the school-to-work transition, working career, and life expectancy, while allowing time comparisons at lower cost. It is no exaggeration to state that, after the log-linear turning point, the tool had ceased to hide the object and now enables it to be problematized with unmatched clarity.

As a result, social-mobility studies can once again attract interest from a social science such as demography, without reconnecting in any way with the dubious approaches (to say the least) that drove the first explorations of social heredity between the 1880s and 1930s—investigations that conjured up the specter of prolific proletarians overwhelming the Malthusian elites.

In the interest of full disclosure, it should be said at the outset that I am hardly a specialist of social mobility: my incursions into the field are remote. My viewpoint here is that of a practitioner of sociodemographic surveys who has had occasion to observe at a distance the growth of social-mobility studies while examining the history of social statistics. My attitude is one of a sympathizer, so to speak, but of a critical sympathizer who, after a period of discouragement, has regained hope.

I. GROWTH OF SOCIAL-MOBILITY STUDIES: QUALITY OF HISTORICAL RESEARCH

Someone seeking an overview of the history of studies on social mobility in the Western world can rely on a high-quality literature in French without equivalent in English.[1] Four authors currently dominate the field.

Charles-Henry Cuin, an academic close to Raymond Boudon, has examined the seminal book by Pitirim Sorokin, *Social Mobility* (1927), as well as the reasons why French sociologists from Émile Durkheim to Pierre Bourdieu overlooked or rejected the study of social mobility, until Raymond Boudon himself became directly involved in it (Cuin, 1987, 1988, 1993, 1995).

Laurent Thévenot, an *administrateur* (senior statistician) at the French National Institute of Statistics and Economic Studies (INSEE) who later joined the École des Hautes Études en Sciences Sociales, follows a totally different approach. He has focused on identifying the principles of political philosophy that underlay the first French surveys on social success and social mobility. These surveys were conducted at International Network of Economic Developers (INED) by Marcel Bresard and Alain Girard and at INSEE by Jacques Desabie (Thévenot, 1987, 1990).

By using a historical, critical, and pedagogic approach, Dominique Merllié has reconstructed the development of social-mobility surveys, starting from the first Italian, French, British, and Scandinavian experiments (Merllié, 1990, 1994, 1995; Merllié and Prévot, 1991). He takes his study up to INSEE's 1985 survey on education and occupational skills (Formation-Qualification Professionnelle: FQP). He was unable to include *The Constant Flux*, the wide-ranging international synthesis published in 1993 by Robert Erikson and John Goldthorpe. In all, he devotes only two pages to the introduction of odd ratios and log-linear models, which profoundly transformed the comparative study of social mobility (see next section and Appendix 92–1).

Gilles de La Gorce tried to fill this gap. His doctoral thesis, defended in Paris in 1993 at the École des Hautes Études en Sciences Sociales and still unpublished, gives an intellectual and technical history of social-mobility models in Western countries. The author's scholarship is less sure-footed than is that of Dominique Merllié, but perhaps it has a keener sense of the prevailing currents and significant breakpoints. In the current context, it is no doubt the best introduction to the subject. The article published before the thesis (La Gorce, 1991) reports only part of the findings.

[1] The historical summaries by Gosta Carlsson (1963), Anthony Heath (1981), and John Goldthorpe (1976) are less detailed.

II. RECENT DATA: A CONFIRMATION OF THE LOG-LINEAR MODEL

Since these historical and critical studies, the flow of empirical publications on social mobility has not ceased. To continue with the French example, the decisive role of log-linear modeling (see Appendix 92–1) in social-mobility analysis was fully vindicated by a remarkable series of theses and publications that echo the work being done in other Western countries such as Switzerland (Lévy *et al*., 1997) and Italy (Schizzerotto *et al*., 1995).

Louis-André Vallet, one of the few French academics to have been involved in official statistics before joining the French National Center for Scientific Research (CNRS), was the first large-scale user of the log-linear model in France, which he applied to the study of female social mobility. He has helped generously to disseminate the method in the research community (Vallet, 1991, 1992, 1995). More recently, he has switched to log-multiplicative models.

John Goldthorpe, who used surveys dating back to the 1970s in *The Constant Flux*, widened his analysis to include French and United Kingdom data of the 1980s, and amply corroborated his earlier findings (Goldthorpe, 1995).

Two INSEE statisticians with a keen interest in sociology, Dominique Goux and Éric Maurin, used the same techniques to analyze INSEE's latest FQP survey (1993) and, retrospectively, the complete series of previous FQP surveys since 1970 (Goux and Maurin, 1995, 1997a,b,c).

In the same spirit, Louis Chauvel, a researcher at a Paris-based economic institute, Office Français de Conjoncture Économique (OFCE), and a graduate of the INSEE statistics school (ENSAE), used the same methods to analyze the changes in social mobility in France between the 1964 FQP survey and the 1995 Labor Force survey. He separates the cohorts, however, into 5-year groups (Chauvel, 1997).

III. FRANCE–ENGLAND: THE DUAL ORIGIN OF THE SOCIAL-MOBILITY TABLE

The most widespread approach to social mobility relies on a simple technique: the construction of a *social-mobility table* that cross-tabulates sons' occupations with fathers' occupations (the occupations of daughters and mothers were seldom used until recently).

The first such table was published in England in the journal *Biometrika* in 1904. Prepared by Emily Perrin, an assistant to Karl Pearson, it draws on 750 entries of the *Dictionary of National Biography*. The cross-tabulation did not yet involve social categories but occupational sectors or disciplines (medicine, literature, theology, art, politics, military, etc.). The goal was to compute an association index derived from chi-square, the *contingency coefficient*, and measure the extent of its possible divergence from the coefficients previously determined for physical measurements, which ranged between 0.45 and 0.50. Emily Perrin (1904) found a value of 0.63, which led her to conclude without any further verification that biological inheritance accounted for nearly three-quarters of total occupational heredity. In its design, Emily Perrin's mobility table was not new. It was inspired by the contingency tables for physical measurements, which already tabulated sons' characteristics against their fathers'.

It would be wrong, however, to establish an intrinsic link between the emergence of the social-mobility table and the hereditarist or even eugenicist ideology. The counter-proof of this was supplied in the same year in Third Republic France by Paul Lapie (1869–1927). This heterodox disciple of Émile Durkheim, who became administrator of the elementary-school system, published a social-mobility table in 1904, without being aware of the work of Karl Pearson's team (Cherkaoui, 1979:250–251; table partly reproduced in Merllié, 1994:127). Lapie classified into nine categories the occupations of 772 alumni of an elementary school in Champagne, cross-tabulated them with their fathers' occupations, and calculated the access rates by origin. In 1911, he extended the analysis to two other schools, taking the opportunity to remove orphans from his classification.

The contrast with Karl Pearson's approach is striking. Paul Lapie spontaneously offered an interpretation with no trace of biologism. Interestingly, he introduced a third variable by duplicating the mobility table so as to distinguish between children possessing the "elementary-school certificate" (*certificat d'Études primaires*, or CEP) and those who did not. Comparing the two series, he concluded that "school sometimes succeeds in breaking the links of the net in which economic causes trap our fates."

But the overall inspiration of Lapie work's is consistently ambiguous. On the one hand, he ardently defended the meritocracy principle in his book entitled *La justice par l'État* (*Justice through the State*), arguing that children of poor families should be given scholarships to facilitate access to jobs compatible with their merits. On the other hand, he wanted to reassure the critics of the Third Republic's education policy, who opposed general compulsory schooling, by demonstrating that school did not foster excessive ambition among children of "the people": it did not "declassify" them up the social scale. Reversing the value judgments conveyed by the liberal approach to the issue, Lapie does not speak of *social mobility* (the term was first used by Pitirim Sorokin in his 1927 essay), but of "social instability." "Occupational heredity" was measured by "stability coefficients" or "instability coefficients," which are, in fact, the rates of access to current positions determined by social origin (Lapie, 1904).

IV. BENINI'S TWO INDICES

A third source for the mobility table was Italian (see Capecchi, 1967:297; Jones, 1985; Merllié, 1994:138, 142). It consists of the marriage tables combining the characteristics of the two spouses.

By 1897, the economist and demographer Rodolfo Benini (1862–1956) was teaching his students at the University of Pavia how to construct an index, called the "sympathy coefficient," which measures the ratio of the number of individuals in each cell to the number that would result from the assumption that spouses' origins were independent of each other. Benini used this initial index in an article of 1898 on "sympathetic combinations in demography."[2] But, in 1911, he corrected it in his *Principii di demografia* by deciding to bound it to the smaller of the table's two margins. He called this second indicator "index of attraction or repulsion." Benini was thus the first author who sought to neutralize structural effects in the measurement of social reproduction.

Benini's double contribution went largely unnoticed because it did not directly concern social mobility but was confined to marital mobility. Frank Jones (1985) and Dominique Merllié (1994), who rehabilitated this pioneer, show that most of the indices later used to study father—son mobility would be identical to the ones developed by Benini.

This is true of Benini's first index: the ratio of the number of observed individuals to the theoretical number under the independent-origins hypothesis. It was reinvented in 1948 by the American statistician Herbert Goldhamer under the name of *social distance mobility ratio*—more simply, *mobility ratio* (see also Goldhamer, 1968) His compatriot Natalie Rogoff, author of a social-mobility survey in Indianapolis, mentioned this origin in an article in *Population* in 1950. During her stay at INED in Paris, she communicated the index to Marcel Bresard, author of the

[2] *Le combinazioni simpatiche in demografia*, 1898.

first French social-mobility study, who used it immediately (Bresard, 1950). She then popularized it in the United States through several publications (Rogoff, 1953).

Benini's first index, also known as the Goldhamer-Rogoff index, was cited next in the publication of the first Anglo-Welsh survey on the topic, conducted by David Glass (1954). The British demographer states that he discovered it independently of Natalie Rogoff. Curiously, he mentions Benini (apparently a second-hand reference via the articles by Natalie Rogoff and Livio Livi in *Population*) without realizing that this was the same index, which he named *index of association or dissociation*. In his critical review of mobility indices, Raymond Boudon was to discuss "Glass's index" without being aware of its pedigree (Boudon, 1973).

Benini's second index had the same clandestine posterity. Glass's disciples criticized it without understanding it: they saw it as an imperfect version of the first index. Later, several authors, in an attempt to neutralize the effect of margins on the measurement of mobility, proposed adjusted indices that allowed for the differences in population size between rows and columns.

In a 1955 article in *Population Studies*, Wladyslaw Billewicz cataloged the shortcomings of the initial index: if N is the total sample size and max the larger of the two populations in the category of origin and category of arrival, the index will never exceed N/max. The article was followed by commentary by Elizabeth Durbin, who proposed an adjusted index called I for "inertia." In 1964, the Japanese Saburo Yasuda, in an often cited article of the *American Sociological Review*, repeated the operation by suggesting a "coefficient of openness." He acknowledged in passing his debt toward Elisabeth Durbin (his index being merely the complement to unity of Durbin's inertia index). But the literature would now refer to the *Yasuda index*, neglecting the contributions of Durbin and—before her—Benini.

V. SOROKIN: A MULTIPLE APPROACH TO SOCIAL MOBILITY

Rodolfo Benini's technical advances were all the more remarkable for having been unknown, 25 years later, to the man who would make the greatest contribution to defining the subject of *social mobility*: the Russian-born American sociologist Pitirim Sorokin. In his *Sociology of the Revolution* (1925), he described the "abnormal" forms of mobility introduced by the revolution, which had expelled him from Russia in 1922. He broadened the scope of his investigation to all forms of mobility in his 1927 treatise, *Social Mobility*.

Sorokin, however, was severely short of technical resources. When he set out to comment on a series of mobility tables culled from several sources, including his own monographs produced in Minneapolis, as well as data already published by the Italian Federico Chessa (1912), he was clearly unaware of Benini's indices of association (Sorokin, 1927:429–434). His indicator of "transmission of occupational status" was simply the percentage of immobile individuals of a given origin (Sorokin, 1927:419). He did not think to complement the table of destinations with a table of origins, although he provided a "literary" commentary on the latter. His conclusions oscillate between the strength of the link and the significance of dispersion: Western societies can be described neither as a caste system nor as a disorderly flow.

Of his five propositions on social mobility, the last is still quoted today: "As far as the corresponding historical and other materials permit seeing [. . .], there seems to be no definite perpetual trend toward either an increase or a decrease of the intensiveness and the generality of social mobility. This is proposed as valid for the history of a country, for that of a large social body, and, finally, for the history of mankind" (Sorokin, 1927:152). But Sorokin's horizon was not a few decades, as it is for John Goldthorpe or for Dominique Goux and Éric Maurin today: he was taking a multicentury view.

Another central theme of *Social Mobility* is that social mobility operates through "selection agencies" such as the family, school, the army, trade unions, occupations, and so on. This notion was hastily rejected because of its functionalism, but it deserves to be explored and tested. Sorokin developed a prophetic idea: the function of schooling is to sort and eliminate. Those who imagine that the expansion of education will inevitably increase social mobility are mistaken: the greater its spread, the more it will help to perpetuate inequality.

VI. SOCIAL MOBILITY AND DIFFERENTIAL FERTILITY

The first systematic investigations of social mobility, in the late 19th century, were haunted by the issue of differential fertility. Because they are more fertile than the upper classes, the lower classes may eventually submerge them. How can one protect the elite from the invasion of the masses? How can one shelter quality from quantity? Francis Galton (1869), Arsène

Dumont (1890), Alain Girard (1961), and many others have played variations on this theme.

Pitirim Sorokin fits into this approach. For him, the lesser fertility of the elites creates an imbalance that jump-starts the upward-mobility engine by triggering a gradual rise of the lower classes. He mentions (Sorokin, 1927:138) Arsène Dumont's famous "social capillarity." The abundant European data that he accumulated on the topic give no hint of a "U-curve" or "J-curve" that would indicate efforts by the middle-class to reduce costs in order to ensure their ascent (Charles Westoff described such a curve in 1953). Despite the detailed nature of the occupational classifications he refers to, Sorokin generally reduces his commentary to a dichotomy between upper classes and lower classes—the elite versus the masses—as Vilfredo Pareto had already done (1916). However, he intuits that mobility spreads from neighbor to neighbor, so that the inward currents generated in the upper strata of society initially benefit the strata immediately below it, that is, the middle classes (Sorokin, 1927:450). This "staircase mobility" concept was formalized by Daniel Bertaux (1977) and consistently validated by later studies.

While loudly proclaiming the need for a quantitative approach free of value judgments, Sorokin often freely indulges in ideological considerations: among the authors he cites most often are Otto Ammon and Georges Vacher de Lapouge, who later became references for racist theories of social reproduction (see Béjin, 1982, 1989). Another source of inspiration was Pareto. For lack of proof, Sorokin rejects the theory of the intrinsic superiority of the Nordic race, but he regards as "likely to be true" the negative social selection process "so brilliantly depicted" by Georges Vacher de Lapouge. *Social Mobility* concludes with two obsessive worries: by steadily reducing its fertility, the upper class will "waste" the high-quality genetic stock of which it is the depositary; meanwhile, the steady reduction of infant mortality among the lower classes in large cities will enable the weak elements of the population to survive at the expense of the strong elements, preventing the latter from periodically compensating the deficits of the upper class (Sorokin, 1927:499, 506).

By accumulating quantitative data from a wide range of sources demonstrating that the upper classes are healthier, taller, and longer-lived than are the lower classes, Sorokin concludes that "Social stratification, with some exceptions, is correlated and considerably coincides with *biological* stratification of the same population from the standpoint of the physical superiority" (Sorokin, 1927:268, author's emphasis). This correlation, he asserts, is "likely to be permanent." But whereas today's sociologists would see this as proof of an influence of living conditions on physical condition, he inverts the causality: "Other conditions being equal, physical superiority has been the condition which has favored the social promotion of individuals and has facilitated their social climbing, while physical inferiority has facilitated the 'social sinking' of individuals and their location in the social lower strata" (Sorokin, 1927:275). If this correlation should weaken, a circulatory mechanism consistent with Pareto's law will restore it naturally: an elite that loses its physical superiority is on the road to "degeneration" and will promptly be expelled by parvenus from the lower classes, provided that the latter are physically superior (Sorokin, 1927:275). Sorokin notes, however, that the physical factor no longer plays in our societies the role it still plays in societies without writing. His thinking on this point remains unsettled; he finds solutions based on voluntary action appealing but unrealistic. "Eugenics exists still only on paper, and there are not great hopes that it will be practiced efficiently in the future" (Sorokin, 1927:499).

Sorokin follows a similar line of argument in regard to moral and mental superiority. Drawing on Nietzsche, Michels, and Pareto for inspiration, he proclaims that, in ordinary times, members of the upper class are merely brutish, ambitious, cynical, and hard-hearted. In periods of decadence, they sink into limpness and humanitarian sentimentalism, at the risk of being overthrown by stronger-minded minorities of plebeian origin (Sorokin, 1927:312). In a revealing footnote, he accordingly prophesied the capture of power by fascists, monarchists, communists, and labor unions. That is what happened in Russia and Italy. Soon would come the turn of countries such as Germany, France, Spain, and even England (Sorokin, 1927:316).

The book ends with a dream of perfect mobility: "In a mobile society, in order to know a man, we cannot rely any more on the information he is a member of a given group [. . .]. Under the conditions of intensive circulation, to say that 'Mr. Smith is a business man, or mechanic, or clerk,' means almost nothing because Mr. Smith yesterday might be a minister and to-morrow may become a millionaire or a senator. We need to know his whole curriculum" (Sorokin, 1927:542). Yet extreme mobility carries the danger of a rootless cosmopolitanism that, by way of compensation, would seek to lock other people into fixed, abstract "social boxes," with no regard for human individuals. That, says Sorokin, is the driving force behind communist collectivism. The book concludes on a bitter note by describing the foreign intellectuals who, when the Red Terror is mentioned, do not want to hear anything about the "15 millions of lives, at least, sacrificed to this God of 'humanitarian communism'" (Sorokin, 1927: 546).

VII. 1948: A BREACH OPENED BY INED

Sorokin's book relied on an accumulation of monographs, which still relied mainly on the extraction of data from biographical dictionaries of eminent persons. Its publication did not trigger a significant series of surveys on the topic. Natalie Rogoff's 1950 panorama of mobility surveys in *Population* is based on two incomplete samplings whose size she carefully refrains from disclosing; she amalgamates them in her presentation, giving the average numbers. At the time, the United States had been overtaken by France.

In November 1948, INED launched France's first-ever sample survey of social mobility. The project was managed by Marcel Bresard, who released the initial results in 1950. The survey would be cited for many years by American sociologists in their international comparisons (see Lipset and Bendix, 1960; Blau and Duncan, 1967).[3]

Discussing Bresard's survey, Charles-Henry Cuin emphasized the fact that its purpose was not "to measure and analyze mobility flows in France, but rather [to assess] the influence of family size on the odds of its members' social ascent, and more precisely to study the relationship between social origin and educational attainment, which was only the first step in a mobility process" (Cuin, 1995:38). Cuin's assertions seemed to downplay the impact of this initial work, his main goal being to show that French sociology had missed every opportunity to study social mobility before the arrival of Raymond Boudon.

Reading Bresard's pioneering work today, we are surprised to note that, far from restricting the scope of social mobility, he addresses it in very open-minded manner, incorporating aspects that would still be worth exploring today.[4] The influence of family size on the odds of social ascent is only one of them. Bresard also analyzes the components of the occupational scale, career-choice factors (family business, personal calling, search for security, financial emergency, etc.), the inheritability of differential fertility from one's grandparents, and marital mobility.[5]

These topics are admittedly lined up in a row rather than interconnected. When Bresard built three-dimensional tables, it was to measure the influence of the number of siblings on the link between the son's occupation and the father's occupation, then the link between the father's occupation and the father-in-law's occupation (Bresard, 1950:546, 551). There was no technical obstacle to his using the same approach for the length of schooling, but he did not do so. The latter variable is effectively included in a three-entry table, but only at the intersection between social origin and family size. Incidentally, this table is most enlightening, because it shows that for an identical origin, a large family reduces the mean length of schooling, except among management-level workers and professional people.

With hindsight, we can assume that Bresard had the means to deepen his analysis of interactions between social mobility, fertility, and educational attainment; we can also regret that he was interested less in the link between education and employment than in the unfavorable status of large families. But if we bear in mind the historical context—the sorry state of French empirical sociology in the late 1940s—the most striking aspect of Bresard's undertaking is its innovativeness: his three-dimensional intersections were a major novelty in the sociological literature of his day. Bresard also contributed another key innovation, well known to historians of social classification: he was the first French researcher to build and use a classification of socio-occupational categories, before INSEE prepared its famous *nomenclature des catégories socioprofessionnelles*, introduced in the 1954 census.

VIII. 1953: INSEE TAKES UP THE BATON AND KEEPS IT

The 1948 survey was the first and last that INED would be able to conduct on social mobility. In 1953, INSEE took up the baton for good by introducing a one-time question on the father's occupation in that year's Labor Force survey (it did not reappear in the questionnaire until 1982). The results were published rapidly by Jacques Desabie in 1954. Convinced of their usefulness, INSEE administrators decided to repeat the same questions in the first survey on education and

[3] The survey, which covers 3076 economically active males aged 18 to 50, was not properly representative—a deficiency underscored many years later by Daniel Bertaux (1969:451, 1971:88): manual workers are substantially underrepresented owing to inadequate enforcement of the quota method, which resulted in greater overall mobility for the sample.

[4] Sibling social mobility is an issue that few authors have addressed in recent years. Two who have done so are Nicole Tabard (1984) and Bernard Zarca (1995a,b).

[5] In so doing, he introduced the term *homogamie* (homogamy) into French statistics (Bresard, 1950:542), well before Alain Girard's *Le Choix du conjoint*, "choosing one's spouse" (1964).

occupational skills (FQP), which followed the 1962 census.

Charles-Henry Cuin gave a curious interpretation of the interval between the 1953 Labor Force survey and the 1964 FQP survey. He saw it as proof that the social-mobility issue had remained "dormant" at INSEE for about 10 years, because the State Planning Commission (Commissariat au Plan) had assigned other tasks to the Institute, such as the assessment of skills levels (Cuin, 1995:39). But this reading neglects the fact that the timing of FQP surveys was closely linked to that of population censuses. Unlike standard surveys, based on address samples, the FQP survey took a sample of individuals stratified by social group. This could not be done until detailed census results were available. However, one could not wait too long, as this would entail a greater number of changes of address, and the cost of tracking down respondents at their new locations would become excessive. Hence a routine gap of 2 years between a specific census and the corresponding FQP survey. The year 1964 was therefore the first possible date for launching such a survey, which, as all census operations, obviously required several years' preparation.

Moreover, the FQP launch coincided with the financial bonanza from the State Planning Commission, which suddenly allowed INSEE to broaden its social-survey program in a spirit of great intellectual freedom. Without the conjunction of these new resources and the INSEE social statisticians' abiding sociological interest in changes in living conditions, the institute would never have been able to undertake the impressive series of FQP surveys, which continues to this day and provides a matchless source for social-mobility developments.[6] In reality, INSEE's interest in the topic has been remarkably constant since the 1950s, even though its pace has been set by the fluctuations in the census calendar.

IX. BOURDIEU'S CRITIQUE AND BOUDON'S WORK

In 1965, Michel Praderie, senior statistician (*administrateur*) at INSEE, presented the results of the FQP survey at the Arras colloquium, jointly organized by a sociologist, Pierre Bourdieu, and an institute statistician, Alain Darbel. In keeping with a principle that INSEE would consistently follow in the future, Praderie dedicated two separate contributions to the survey findings on *social* mobility (intergenerational) and the findings on *occupational* mobility (changes in occupational category during the respondent's working career, over a 5-year period).

The proceedings were published in 1966 under the title of *Le Partage des bénéfices* (*Sharing the Profits*). The volume was signed by the collective pseudonym Darras, an allusion to the conference venue. The organizers' goals included "examining how the economic and social privileges that hinder the equalization of chances are transmitted from one generation to another" (Darras, 1966:18). As a companion piece to Michel Praderie's article on "Social Inheritance and the Odds of Ascension," Pierre Bourdieu contributed a paper on "The Transmission of Cultural Heritage."

This text is remarkable in many ways. First, it is Bourdieu's clearest précis of his own thought, which was already highly structured. It is thus strongly recommended to all readers seeking a rapid introduction to this difficult author. Second, it is the text in which Bourdieu most openly acknowledges his debt toward INED research on school tracking, published in a series of articles by Alain Girard and Paul Clerc in *Population* in 1963 and 1964.[7] Third—and we shall focus on this aspect here—it took an explicit position on the classic approach to social mobility.

Pierre Bourdieu relies on INED data to emphasize that the school system, far from being a factor of social mobility, is "one of the most efficient factors of social conservation," because it legitimates social inequalities by giving its seal of approval to cultural heritage, treated as a natural gift. Even at this early stage, he stressed that families' cultural capital includes not only "verbal proficiency" and "the nonformal culture acquired in extra-curricular experiences," but also "information from parents on curriculum choices and careers"—a resource whose importance would rise steadily as the system became more diverse and complex (Bourdieu, 1966:394–395, 404). Consequently, Bourdieu argues, the study of social mobility is worthwhile only if it tests the egalitarian ideals of a "democratic society." This, however, requires "measuring the chances of accessing the institutionalized instruments of social ascent and cultural salvation that it offers to persons of different social classes" (Bourdieu, 1966:240). Chances of access imply unequal access costs. Bourdieu criticizes the analyses based

[6] There have been five FQP surveys to date, in 1964, 1970, 1977, 1985, and 1993, following the 1962, 1968, 1975, 1982, and 1990 population censuses. The survey consecutive to the 1999 census is due in 2004.

[7] In particular, Bourdieu used the three-variable tables such as scholastic achievement as a function of the father's educational attainment for an identical income level, and extension of child's education as a function of father's social category for an identical scholastic achievement. At the time, few French sociologists were capable of fully exploiting such tables.

exclusively on father-son social-mobility tables for bracketing these inequalities. He regards the "perfect mobility" postulate—which asserts total independence between fathers and sons—as, basically, an abstraction that ignores reality; however, he admits that this "myth" can serve a critical purpose by objectivating the gap between reality and democratic ideals.

To grasp the logic of this approach, it is useful to refer to the typology of political philosophies that, according to Luc Boltanski and Laurent Thévenot (1987), inform the different models of possible *polities*. Bourdieu's sociology of education consists in denouncing the contamination of the *civic* model of the school polity by a *domestic* model. Although the education system thinks it is assessing students on the basis of scholastic criteria, the education system actually classifies them on the basis of their family strengths, starting with the cultural proficiency forged in early childhood (Thévenot, 1990). This diagnosis of Bourdieu's position seems more interesting to us than that of Charles-Henry Cuin, who merely tags it with a hastily contrived sociopolitical label: a "structuralist metamorphosis of Marxism." We can better understand Bourdieu's special interest in Paul Clerc and Alain Girard's three-variable tables, which introduce the notion of a differential return on the diploma or educational attainment by social origin; the simple two-variable mobility table, as typically presented in the available literature, seems feeble indeed by comparison. On this point, Bourdieu is not far removed from Sorokin, who already saw education as the main *selection agency* capable of preserving the range of social-mobility odds.

By using a different approach, Raymond Boudon also set out to enhance the analysis of social mobility by introducing educational-achievement variables (Boudon, 1973). But the task was technically complex. He sought to catalog the second- and third-order interactions between the main variables of social and educational mobility, including differential fertility. He concluded his work with some 40 partial conclusions that cannot give a clear overview of the issue for readers in a hurry. (Significantly, Charles-Henry Cuin gave up on summarizing Boudon's theory, because he claimed that it could be stated not in natural language but only in model form.)

Paradoxically, we can now provide simple answers to the two French sociologists' questions by means of the log-linear interaction model applied to three-dimensional tables (Goux and Maurin, 1997). Bourdieu's model postulates that the relationship between social origin and social destination (which we shall denote as O and D, respectively) is concentrated by the school, in the form of a diploma or educational attainment (written E) that certifies social origin and determines destination. This is a simple-interactions model: (OED) = (OE)(ED). Dominique Goux and Éric Maurin show that it regresses the FQP survey data far better than a purely meritocratic model of the (OED) = (O)(ED) type, which postulates a relationship between education and destination that is independent of social origin. However, the model remains inadequate.

In other passages (not mentioned by Goux and Maurin), Bourdieu suggests on the contrary that, for the same diploma, social origin continues to differentiate social destinations (there exists a "differential return on diploma"), which gives an undersaturated model of the (OED) = (OE)(ED)(OD) type: social origin continues to weigh directly on destination, even if schooling acts as an intermediary factor. This model proves to be far more consistent with the facts observed in France in recent decades. It was used by Claude Thélot (1982), whose book *Tel père, tel fils?* (*Like Father, Like Son?*) was a landmark in social-mobility studies in France. The model was also the final choice of Dominique Goux and Éric Maurin. The two authors note that it contradicts one of Boudon's hypothesis, namely, that the social return on education differs from one period to another, depending on whether we are dealing with high-rank or low-rank graduates.[8] In reality, the combined impact of family heritage and educational attainment on social destination is corroborated for diplomas and degrees of all levels.

X. SOCIOLOGY/DEMOGRAPHY: IDEOLOGICAL OR INSTITUTIONAL DIVIDE?

In his history of social-mobility studies, Charles-Henry Cuin (1993, 1995) emphasizes the ideological impediments to the expansion of research by French sociologists. Similar to Merllié (1994) and Thévenot (1987, 1990) before him, he reminds readers of how the young Alain Touraine, still imbued with a Marxist-inspired sympathy for the working classes (*ouvriérisme*), had thwarted France's participation in 1954 in the planned international survey on social mobility launched by David Glass at the International Association of Sociology. For Touraine, the survey conveyed an individualistic concept of mobility. What counted, in his view, was "collective" social mobility, that is, the improvement of the fate of the working

[8] Indeed, under this hypothesis, family heritage would not be a factor for selecting highly educated candidates unless they were in oversupply in the labor market. If supply is weak, cultural heritage would be a selection factor among low-qualified candidates.

class obtained through labor-union struggles (Touraine, 1954). In a similar spirit, but 20 years later, the political scientist Nicos Poulantzas (1974), a disciple of Louis Althusser, denounced the "bourgeois" approach to social mobility.

These ideological taboos are indeed real, but do they suffice to explain the weakness of social-mobility studies among French sociologists? Factors of a more structural kind—concerning the organization of research in France—also played a role. When social mobility is described as having long remained "a preserve of economists and demographers" in France (Cuin, 1995:38), the term should be taken to mean a preserve of sociologists and survey statisticians working in the main official statistical institutes, INED and INSEE—a field that sociologists affiliated with universities or the CNRS did not enter.[9] But this separation is hardly specific to social-mobility analysis. It applies, more broadly, to all social statistics, including family expenditures, consumption surveys, fertility surveys, and the sociology of the family.

Unlike in the United States, United Kingdom, and Germany, there is a radical divide in France between academia and the official statistical system. The latter not only holds a monopoly on the collection resources needed to produce data with large-scale representativeness. It also analyzes the data, largely relying on its own staff, who are proficient in statistical methods and demographic analysis. And the system publishes the results itself, in institutional media. This is a far cry from the situation in the United States, where, as far back as the 1950s, official agencies such as the Census Bureau simply collected social-mobility data, subcontracting the data tracking and analysis to specialized academics (Robert Centers, Peter Blau, Otis Duncan, David Featherman and Robert Hauser, and others).

The French situation has admittedly begun to improve, but only recently. INSEE and CNRS did not begin exchanging a limited number of specialists until the early 1980s. The framework agreement signed by the two organizations on survey-file transmission, including FQP surveys, was implemented in the mid-1980s. And it is only in the mid-1990s that INED and INSEE began including doctoral students in their research teams.

As a result, French academics working on social mobility have never tackled the subject directly, either by taking part in INSEE questionnaire development or by conducting their own analyses of the data files.[10] Some have skillfully examined the history and epistemology of the specialty, at the risk of irritating practitioners, who readily view them as amiable arbiters of elegance. An even smaller number—for example, Raymond Boudon in his 1973 book—engaged in the secondary processing of the data. But harsh reality forced them to confine their analysis to the existing social-mobility tables, and they soon veered toward the construction of abstract models that were impossible to validate in the near term. Some of Raymond Boudon's central assumptions, for instance, were not submitted to experimental testing by INSEE researchers until 25 years later (Goux and Maurin, 1997).

The author of *L'Inégalité des chances* was himself aware of the limits of the exercise; he made an impassioned plea for the production of empirical materials covering the long term: "A satisfactory theory of mobility cannot be designed until we have mobility surveys repeated at fairly regular intervals over time. Only thus will we be able, for example, to introduce in a precise manner a fundamental variable into the mobility process, namely, the change in social structure. For our part, we have confined our work in this area to analyzing the consequences of the cautious and spare hypothesis that the social structure changes less quickly than the educational structure" (Boudon, 1973:219).

With hindsight, it is striking to see the extent of the divergence between the research strategies of Raymond Boudon and John Goldthorpe in the field of social mobility. Although the French academic remained dependent for data on the official statistical system, his British colleague tackled the data head-on. Boudon's work has been continued in France by young researchers who, similar to Louis-André Vallet, Michel Forsé, Louis Chauvel, Dominique Goux, and Éric Maurin, master the new techniques of mobility analysis. Boudon himself eventually switched to cognitive sociology: he decided not to update his textbook on mobility indices, published in English and never translated into French. John Goldthorpe, by contrast, has

[9] The distribution of authors by discipline is not as clearcut as Charles-Henry Cuin makes it out to be. Why classify Nicos Poulantzas among sociologists and Alain Girard among demographers? Poulantzas was a political scientist who never conducted an empirical survey—a good example of the "materialists without material" pilloried by Bourdieu. Girard played an important role at International Network of Economic Developers, but was not a demographer himself. He studied demographic behavior as a sociologist, with a predilection for sample surveys that sought to identify perceptions and opinions.

[10] Except for two sociologists at universities in western France, François de Singly (1977, 1983, 1984, 1986) and Louis-André Vallet (1991), who cooperated productively with Claude Thélot when he was working with the National Institute for Statistics and Economic Studies regional office in Nantes. Both de Singly and Vallet specialize in the analysis of female mobility.

sought to maintain contact with social statisticians. His team established links with a wide international network in order to collect—and, sometimes, inspire—the main mobility surveys in Western countries in the 1970s. He was helped in this task by the organic connections between social statistics and academic research in the United Kingdom, Germany, and the United States, whereas Bourdon remained a prisoner of his isolated status.

XI. THE "CONSTANT FLOW": GOLDTHORPE

Following the work of Leo Goodman (1979), Robert Hauser (1978), Michael Hout (1982), and a few others, John Goldthorpe imposed the use of *odds ratios* in the sociological profession as an indicator of overrepresentation of groups in a social-mobility table. Odds ratios offer two advantages. First—unlike any of the mobility indicators used until then—they remain independent of variations in the size of marginal groups. Second, they can be incorporated into a very simple modeling of mobility tables, that is, log-linear models, which use a logistic metric in which effects are additive rather than multiplicative as previously. In such a model, the number of individuals in each cell is obtained by adding together several effects (or multiplying them, if one renounces logarithmic transposition): a scale effect (parameter taking into account the total number of individuals in the table), a row effect (destination or origin, depending on the table's arrangement), a column effect, and an effect of the row-column interaction. The same logic can apply to tables with *n* dimensions. These properties make log-linear models the ideal choice for international and time comparisons. For his project, John Goldthorpe has steadily gathered individual files from the main industrialized countries (most notably thanks to funding from the Volkswagen Foundation in Hanover): United Kingdom, France, Netherlands, Sweden, United States, and Canada, soon followed by Japan and several countries of central/eastern Europe.

As a result, in his analyses and those of his critics, we have seen a recent flowering of some odd expressions (to say the least) consisting of dual metaphors, such as "inertia of social fluidity" and "core social fluidity." These alliances of words are incomprehensible if we refer to first-order truths. They denote second-order truths that require reasoning on a "some things being equal" basis.

John Goldthorpe calls the core assumption the "FJH hypothesis," in homage to the three U.S. researchers who formulated it back in 1975: David Featherman, Frank Jones, and Robert Hauser. It consists in stating that if the Western industrialized countries display an unquestionable overall mobility, the relative mobility (or social fluidity) is fairly stable from one society to another, just as it is stable in time. It is not the quantity of mobility that remains invariable, but the *competitive structure of flows* or, if one prefers, *differential mobility*.

Let us take, for example, two neighboring countries at very different stages in the process of reduction and concentration of the rural population—one with, say, a 20% share of farmers in its total population, the other with only 2%. Despite the gap, the access of farmers' sons to urban social strata will follow the same hierarchy in both countries. At the other end, the number of managerial jobs or higher-level intellectual professions may increase in different proportions and at different paces in the two countries, but the demand created at the top of the social ladder will be filled in different ways depending on the social origin. Applicants from each group will access these new positions in accordance with very similar odds hierarchies in both countries. Social groups do move, but without significantly modifying the system of connections or distances that unites them. A final example is the risk of downward mobility represented by exposure to unemployment, which now justifies introducing the unemployed into mobility tables: this phenomenon chiefly affects the most vulnerable groups, at the bottom of the ladder.

The thesis of a *constant flow* of social mobility goes back a long way. Sorokin had already intuited it (admittedly in the context of a multi-century time frame). It crops up more or less explicitly in the observation made by sociologists of all persuasions in the 1960s and 1970s: contrary to all expectations—most notably those of U.S. sociologists of social mobility, such as Richard Bendix and Seymour Lipset—the massive spread of education in the years of expansion was not matched by an equivalent reduction of economic and cultural inequalities; nor did it generate an overall rise in the odds of upward mobility. Twenty-five years later, Dominique Goux and Éric Maurin said the very same thing: "the level is rising, but the classifications are not changing. "And Michel Forsé gave the following title to an article published by OFCE: "The Decrease in the Inequality of Education Odds Is Not Enough to Reduce the Inequality of Social Odds" (Forsé, 1997). The more recent studies by Louis-André Vallet (1999), based on log-multiplicative models, show the following pattern for France over the long run (4 decades): the relative link between social origin and access odds has loosened very slightly, but not on a scale proportional to the changes in social structure during the intervening period.

Raymond Boudon's essay on social mobility took this observation as its starting point. This was the central significance of the "Anderson paradox" that he chose to showcase: as the school structure, driven by the pressure of education demand, evolves faster than the social structure, the length of schooling increases without changing the mobility structures. Consequently, a child whose educational attainment exceeds that of his or her father has no guarantee of rising higher than the father's position on the social ladder (Boudon, 1973:207). Formulated in a period when long-term data were scarce, this analysis has stood the test of time.

At the risk of reconciling authors commonly viewed as irreconcilable, we can say the same about Pierre Bourdieu's verdict on the issue, despite his different approach. In *Le Partage des bénéfices* (1966) as in *La Distinction* (1979) and some intermediate analyses (Bourdieu, 1973), Bourdieu consistently stressed that the preservation of intergroup competition structures entailed the constant renewal of content in the ranking struggles (e.g., in France, parents seeking to make their children study German in order to obtain the same result formerly achieved by opting for Greek, i.e., access to the more prestigious high schools). The notion of 'conversions" or "field homologies" that preserve relative statuses is a means for Bourdieu to identify the invariant elements concealed under the apparent fluidity. Can we say that, in so doing, he is distancing himself from social mobility? In fact, he only extends it—as Sorokin had already done (and as has since been forgotten)—to all the phenomena of unequal and distinctive diffusion of goods, information, habits, and tastes; in turn, these are incorporated into a general sociology of "access odds," of which social mobility in the strict sense is merely a specific expression. The overall process does indeed resemble a "constant flow." The discussion on this point would take us too far. Let us draw a provisional conclusion by suggesting that John Goldthorpe's general result contains nothing bluntly incompatible with Bordieu's sociology of social structure. More precisely, Goldthorpe's success has demonstrated that Bourdieu's presumptions against mobility analysis were no longer justified.

XII. A POSSIBLE CONVERGENCE: DIFFERENTIAL MOBILITY AND DIFFERENTIAL DIFFUSION

Why has *relative* social mobility remained so constant in our societies? We need to compare it with other differential diffusion phenomena that leave the overall pattern of inequality unchanged. For example, the uneven diffusion of medical innovations—new therapies, new equipment, and new knowledge, via the selective diffusion of recourse to specialized doctors—may partly explain the persistence or *relative* gaps in health or life expectancy. The current diffusion of personal computers in homes is recreating, before our eyes, new social inequalities between children of school age. Every innovation tends to initially favor the best-equipped social categories.

The identification of *constant flow* should not accredit the existence of an eternal sociological law. Constant flow is a remarkable regularity (which is already a lot) that will hold true as long as it has not been disproved. As Louis Chauvel rightly noted, we should not rule out the possibility that Goldthorpe's constant flow may one day undergo the same fate as the alpha coefficient of Pareto's wealth curve, which stood the test of fact for nearly a century before succumbing to it very recently. For the moment, constant flow applies to all cohorts born between 1900 and 1960. By chaining the data from FQP and Labor Force surveys produced by INSEE from 1970 to 1995, Chauvel established that social fluidity remained broadly unchanged in France for all male cohorts entering the labor market from the 1930s to the present.

In other words, there seems to be little change in the structure of mobility in time and space in the main Western countries—as demonstrated recently for Italy and Switzerland. The demonstration also disproves the widespread notion that North American society is more mobile than are European societies.

The criticisms directed at John Goldthorpe have spared the bulk of his work and focused on what he has called, not without humor, "core fluidity." What does this consist of? Goldthorpe takes the logistic comparisons of selected cells in the table and subjects them to a series of effects. For example, mobility out of or toward the category of farmers and farm workers is slowed by a "sector effect" that can be estimated (all other things being equal, switching to a farming occupation is particularly difficult, as is exiting from one). Other paths are influenced by effects that counteract mobility from a short distance or a long distance, and so on. The structure of attractions or repulsions between classes modulates the general phenomenon of relative fluidity via eight effects that, according to Goldthorpe, form the "core fluidity" of industrialized countries. France and England are the best embodiments of this core; Sweden diverges from it because of a slight excess fluidity that may be owing to its long-term implementation of a social-democratic social policy (but this needs to be demonstrated).

We are now witnessing an interesting reversal of approach, effectively noted by Dominique Merllié and Jean Prévot (1991) and Gilles de La Gorce (1993). A social-mobility study would typically begin by setting aside structural mobility in order to concentrate on net mobility, equated with true mobility. Because of the abundantly repeated observation of constant flow, the latest research starts by verifying the existence of a constant relative mobility before moving on to the study of absolute mobility tied to changes in social structure. Let us return to our earlier example of two countries in which farmers make up 2% and 20% of the population, respectively. However hard they try to distribute the social-mobility flows while preserving the structures of interclass competition and ranking—which are similar in both countries—the fact of the matter is that life is not the same in a country where farmers represent 2% rather than 20%. The portion of the population that changes economic sectors and ways of life is not the same. The log-linear analysis of social mobility thus seems to have completed a long side-trip in order to discover that, ultimately, social-mobility analysis is but a chapter of a broader venture: the study of the changes in social hierarchies in general.

Acknowledgment

This chapter was translated from the french version by Jaonathan Mandelbaum.

References

Agresti Alan, 1984. *Analysis of ordinal categorical data*. New York, John Wiley, 287 p.

Agresti Alan, 1990. *Categorical data analysis*. New York, John Wiley, 558 p.

Ammon Otto, 1895. *Die Gesellschaftordnung und ihre natürlichen Grundlagen. Entwurf einer Sozial-Anthropologie zum Gebrauch für alle Gebildeten, die sich mit sozialen Fragen befassen*. Jena, G. Fischer, 408 p. (2nd ed. 1900; French trans. by H. Muffang, *L'Ordre social et ses bases naturelles: esquisse d'une anthroposociologie*, Paris, Albert Fontemoing, 516 p.)

Anderson C. Arnold, 1961. A skeptical note on the relation of vertical mobility to education, *American Journal of Sociology*, vol. 66(5), p. 560–570.

Aron Raymond, 1964. *La Lutte de classes: nouvelles leçons sur les sociétés industrielles*. Paris, Gallimard.

Béjin André, 1982. Le sang, le sens et le travail: Georges Vacher de Lapouge, darwiniste social fondateur de l'anthroposociologie, *Cahiers internationaux de sociologie*, vol. 73, p. 323–343.

Béjin André, 1989. Arsène Dumont et la capillarité sociale, *Population*, vol. 44(6), p. 1009–1027.

Bendix Richard and Lipset Seymour M. (eds.), 1953. *Class, status and power: a reader in social stratification*. London, Routledge & Kegan Paul, 727 p. (2nd ed., 1990, *Class, status and power: social stratification in comparative perspective*).

Benini Rodolfo, 1898. Le combinazioni simpatiche in demografia, *Rivista italiana di sociologia*, vol. 2(2), p. 152–171.

Benini Rodolfo, 1901. *Principii di demografia*. Florence, G. Barbèra, 296 p.

Benini Rodolfo, 1928. Gruppi chiusi e gruppi aperti in alcuni fatti collettivi di combinazioni, *Bulletin de l'Institut international de statistique*, vol. 23(2), p. 362–383.

Bertaux Daniel, 1969. Sur l'analyse des tableaux de mobilité sociale, *Revue française de sociologie*, vol. 10(4), p. 448–490.

Bertaux Daniel, 1970. L'hérédité sociale en France, *Économie et statistique*, no. 9, p. 37–47.

Bertaux Daniel, 1977. *Destins personnels et structure de classe: pour une critique de l'anthroponomie politique*. Paris, Presses Universitaires de France, 323 p.

Bertaux Daniel and Thompson Paul, 1997. *Pathways to social class: a qualitative approach to social mobility*. Oxford, Clarendon Press, Oxford University Press, 334 p.

Billewicz Wladyslaw Z., 1955. Some remarks on the measurement of social mobility, *Population Studies*, vol. 9(1), p. 97–100.

Bishop Yvonne M., Fienberg Stephen E. and Holland Paul W., 1975. *Discrete multivariate analysis: theory and practice*. Cambridge, The Massachussetts Institute of Technology Press, 957 p. (2nd ed. 1989).

Blau Peter M. and Duncan Otis Dundley, 1967. *The American occupational structure*. New York, John Wiley & Sons, 520 p.

Blum Alain, de la Gorce Gilles and Thélot Claude, 1985. Mobilité sociale et mobilité géographique, *Population*, vol. 40(3), p. 397–434.

Boltanski Luc and Thévenot Laurent, 1987. *Les Économies de la grandeur*. Paris, Centre d'études de l'emploi and Presses Universitaires de France, 361 p.

Boudon Raymond, 1968. La mobilité sociale, *in*: *Encyclopædia Universalis*, vol. 11, p. 115–118. Paris, Éd. Encyclopædia Universalis (repr. in 1985 ed., vol. 12, p. 395–398, and in 1989 ed., vol. 15, p. 505–508).

Boudon Raymond, 1973a. *L'Inégalité des chances: la mobilité sociale dans les sociétés industrielles*. Paris, Armand Colin, 239 p. (repr. 1985, Eng. trans. 1974: *Education, opportunity and social inequality: changing prospects in Western society*. New York, John Wiley, 220 p.).

Boudon Raymond, 1973b. *Mathematical structures of social mobility*. Amsterdam, Elsevier, 168 p.

Bouglé Célestin, 1904. *La démocratie devant la science: études critiques sur l'hérédité, la concurrence et la différenciation*. Paris, Alcan, 312 p.

Bourdieu Pierre, 1966. La transmission de l'héritage culturel, *in:* Darras [group pseudonym] (ed.), *Le partage des bénéfices: expansion et inégalités en France*, p. 384–420. Paris, Éditions de Minuit, 447 p.

Bourdieu Pierre, 1979. *La distinction: critique sociale du jugement*. Paris, Éditions de Minuit, 672 p.

Bourdieu Pierre, Boltanski Luc and de Saint-Martin Monique, 1973. Les stratégies de reconversion: les classes sociales et le système d'enseignement, *Informations sur les sciences sociales*, vol. 12(5), p. 61–113.

Bourdieu Pierre and Passeron Jean-Claude, 1970. *La Reproduction: éléments pour une théorie du système d'enseignement*. Paris, Éditions de Minuit, 283 p.

Bozon Michel and Héran François, 1987. La découverte du conjoint. I. Évolution et morphologie des scènes de rencontre, *Population*, vol. 42(6), p. 943–986.

Bozon Michel and Héran François, 1988. La découverte du conjoint, II: les scènes de rencontre dans l'espace social, *Population*, vol. 43(1), p. 121–150.

Breiger Ronald L. (ed.), 1990. *Social mobility and social structure*. Cambridge, Cambridge University Press, 364 p.

Bresard Marcel, 1950. Mobilité sociale et dimension de la famille, *Population*, vol. 5(3), p. 533–566.

Capecchi Vittorio, 1967. Problèmes méthodologiques dans la mesure de la mobilité sociale, *Archives européennes de sociologie*, vol. 8(2), p. 292–300.

Carlsson Gösta, 1963. Sorokin's theory of social mobility, *in*: Philip J. Allen (ed.), *Pitirim A. Sorokin in review*, p. 122–139. Durham, Duke University Press, 527 p.

Carlsson Gösta, Glass David V. and Svalastoga Kaare (eds.), 1965. Social stratification and mobility, *Acta sociologica*, vol. 9(1–2), p. 1–192.

Cautres Bruno, 1992. Recherches récentes sur la mobilité sociale, *Revue française de science politique*, vol. 62(3), p. 441–479.

Chapman S.J. and Abbott W., 1913. The tendency of children to enter their fathers' trades, *Journal of the Royal Statistical Society*, vol. 76(6), p. 599–604.

Chauvel Louis, 1986. *Le destin des générations: structure sociale et cohortes en France au XX^e siècle*. Paris, Presses Universitaires de France, 386 p.

Chenu Alain, 1994. *Les employés*. Paris, La Découverte, 128 p.

Cherkaoui Mohamed, 1979. Les effets sociaux de l'école selon Paul Lapie, *Revue française de sociologie*, vol. 20(1), p. 239–255.

Cuin Charles-Henry, 1987. Durkheim et la mobilité sociale, *Revue française de sociologie*, vol. 28(1), p. 43–65.

Cuin Charles-Henry, 1988. Sorokin et le *Social Mobility* de 1927: naissance et mise en œuvre d'une problématique sociologique, *L'Année sociologique*, vol. 38, p. 275–308.

Cuin Charles-Henry, 1993. *Les Sociologues et la mobilité sociale*. Paris, Presses Universitaires de France, 304 p.

Cuin Charles-Henry, 1995. La sociologie et la mobilité sociale: les énigmes du cas français, *Revue française de sociologie*, vol. 36(1), p. 33–60.

de La Gorce Gilles, 1991. L'individu et la sociologie: soixante ans d'étude de la mobilité sociale, *Revue de synthèse*, vol. 112(2), p. 237–264.

de La Gorce Gilles, 1993. *Mobilité sociale et trajectoires individuelles: l'histoire d'une spécialité de la sociologie et l'étude du cas d'une génération de Parisiens*. Paris, École des Hautes Études en Sciences Sociales, 620 p. (Doctoral dissertation defended March 31, 1993).

Demaris Alfred, 1992. *Logit modeling: practical applications*. London, Sage Publications, 88 p.

Desabie Jacques, 1955. La mobilité sociale en France, *Bulletin de la SEDEIS*, Nov. 1 issue, (repr. in *Bulletin d'information de l'INSEE*, 1956, vol. 1, Jan., p. 25–63).

DesrosiËres Alain and Thévenot Laurent, 1988. *Les Catégories socio-professionnelles*. Paris, La Découverte, 128 p.

Dumont Arsène, 1890. *Dépopulation et civilisation*. *Étude démographique*. Paris, Éditions Lecrosnier et Babé, 520 p. (repr. by André Béjin, Paris, Economica, 1990).

Dumont Arsène, 1898. *Natalité et démographie: conférences à l'École d'anthropologie de Paris*. Paris, Librairie Cdt Reinwald et Schleicher Frères Éditeurs, 230 p.

Duriez Bruno, Ion Jacques, Pinçon Michel and Pinçon-Charlot Monique, 1991. Institutions statistiques et nomenclatures socioprofessionnelles: essai comparatif (Royaume-Uni, Espagne, France), *Revue française de sociologie*, vol. 32(1), p. 29–59.

Erikson Robert and Goldthorpe John H., 1987a. Commonality and variation in social fluidity in industrial nations, part I: a model for evaluating the 'FJH hypothesis', *European Sociological Review*, vol. 3(1), p. 54–77.

Erikson Robert and Goldthorpe John H., 1987b. Commonality and variation in social fluidity in industrial nations, part II: the model of core social fluidity applied, *European Sociological Review*, vol. 3(2), p. 415–441.

Erikson Robert and Goldthorpe John H., 1993. *The constant flux: a study of class mobility in industrial societies*. Oxford, Clarendon Press, xvi + 430 p.

Erikson Robert, Goldthorpe John H. and Portocarero Lucienne, 1983. Intergenerational mobility and the convergence thesis: England, France and Sweden, *British Journal of Sociology*, vol. 34(3), p. 303–343.

Euriat Michel and Thélot Claude, 1995. Le recrutement social de l'élite scolaire en France, *Revue française de sociologie*, vol. 36(3), p. 403–438.

Featherman David L. and Hauser Robert M., 1978. *Opportunity and change*. New York, Academic Press, 572 p.

Featherman David L., Jones Frank L. and Hause Robert M. R., 1975. Assumptions of social mobility research in the United States: the case of occupational status, *Social Science Research*, vol. 4(4), p. 329–330.

Forsé Michel, 1997. La diminution de l'inégalité des chances scolaires ne suffit pas à réduire l'inégalité des chances sociales, *Revue de l'OFCE*, no. 63, p. 229–241.

Galton Francis (Sir), 1869. *Hereditary genius: an inquiry into its laws and consequences*. London, Macmillan, 390 p. (2nd ed., rev.: 1892).

Garnier Maurice and Hazelrigg Lawrence E., 1974a. Father-to-son occupational mobility in France, *American Journal of Sociology*, vol. 81(2), p. 478–502.

Garnier Maurice and Hazelrigg Lawrence E., 1974b. La mobilité sociale en France comparée à celle d'autres pays, *Revue française de sociologie*, vol. 15(3), p. 363–378.

Gini Corrado, 1912. *I Fattori demografici dell'evoluzione delle nazioni*. Turin, Fratelli Bocca, 143 p.

Gini Corrado, 1914. Indici di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione, *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti*, vol. 74(2), p. 583–610.

Ginsberg Morris, 1929. Interchange between social classes, *The Economic Journal*, vol. 39(156), p. 554–565.

Girard Alain, 1961. *La Réussite sociale en France: ses caractères, ses lois, ses effets*. Paris, INED and PUF, 356 p.

Glass David L. (ed.), 1954. *Social mobility in Britain*. London, Routledge & Kegan Paul Ltd, 412 p.

Goldhamer Herbert, 1968. Social mobility, *in*: David L. Sills (ed.), *International Encyclopedia of the Social Sciences*, vol. 14, p. 429–438. New York, Macmillan and The Free Press.

Goldthorpe John H., 1982. On the service class: its formation and future, *in*: Anthony Giddens and Gavin Mackenzie (eds.), *Social class and the division of labour*, Cambridge, Cambridge University Press, 337 p.

Goldthorpe John H., 1995. Le noyau dur: fluidité sociale en Angleterre et en France dans les années soixante-dix et quatre-vingt, *Revue française de sociologie*, vol. 36(1), p. 61–79.

Goldthorpe John H., Llewellyn Catriona and Payne Clive, 1980. *Social mobility and class structure in modern Britain*. Oxford, Clarendon Press, 310 p. (2nd ed. 1987).

Goldthorpe John H. and Portocarero Lucienne, 1981. La mobilité sociale en France, 1953–1970: nouvel examen, *Revue française de sociologie*, vol. 22(2), p. 151–166.

Gollac Michel and Laulhé Pierre, 1987a. Les composantes de l'hérédité sociale: un capital économique et culturel à transmettre, *Économie et statistique*, no. 99–100, p. 95–105.

Gollac Michel and Laulhé Pierre, 1987b. Lignée paternelle, lignée maternelle: un rôle voisin dans l'hérédité sociale, *Économie et statistique*, no. 99–100, p. 107–113.

Gollac Michel and Laulhé Pierre, 1987c. La transmission du statut social: l'échelle et le fossé, *Économie et statistique*, no. 99–100, p. 85–93.

Gollac Michel, Laulhé Pierre and Soleilhavoup Jeanine, 1988. *Mobilité sociale: enquête Formation—qualification professionnelle de 1985*. Paris, INSEE, 203 p.

Goodman Leo A., 1979. Simple models for the analysis of association in cross-classification having ordered categories, *Journal of the American Statistical Association*, vol. 74(367), p. 537–552.

Goux Dominique and Maurin Éric, 1995. Origine sociale et destinée scolaire: l'inégalité des chances devant l'enseignement à travers les enquêtes Formation-Qualification professionnelle 1970, 1977, 1985 et 1993, *Revue française de sociologie*, vol. 36(1), p. 81–121.

Goux Dominique and Maurin Éric, 1997a. Démocratisation de l'école et persistance des inégalités, *Économie et statistique* (6), p. 27–39.

Goux Dominique and Maurin Éric, 1997b. Destinées sociales: le rôle de l'école et du milieu d'origine, *Économie et statistique* (6), p. 13–26.

Goux Dominique and Maurin Éric, 1997c. Meritocracy and social heredity in France: some aspects and trends, *European Sociological Review*, vol. 13(2), p. 159–177.

Gribaudi Maurizio and Blum Alain, 1990. L'espace social: des catégories aux liens individuels, *Annales (ESC)*, vol. 45(6), p. 1365–1402.

Hauser Robert M., 1978. A structural model of the mobility table, *Social Forces*, vol. 56(3), p. 919–953.

Heath Anthony, 1981. *Social Mobility*. London, Fontana, 303 p.

Héran François, 1984. La force du lien, *Autrement*, no. 61, p. 43–55.

Héran François, 1997. La catégorie socioprofessionnelle: réflexions sur le codage et l'interprétation, *in*: Élisabeth Dupoirier and Jean-Luc Parodi (eds.), *Les indicateurs socio-politiques aujourd'hui*, pp. 49–68. Paris, L'Harmattan, 364 p.

Hout Michael, 1983. *Mobility tables*. London, Sage Publications, 93 p.

Jones Frank Lancaster, 1985. New and (very) old mobility ratios: is there life after Benini?, *Social Forces*, vol. 63(3), p. 838–850.

Kahl Joseph A., 1957. *The American class structure*. New York, Rinehart & Co., 310 p.

Lapie Paul, 1899. *La justice par l'État: étude de morale sociale*. Paris, Ferdinand Alcan, 213 p.

Lapie Paul, 1904. Les effets sociaux de l'école, *Revue scientifique (Revue rose)*, vol. 41 (1–2), p. 6–12 and 42–46 (repr. in *L'École et la société*, Paris, Ferdinand Alcan, 1923, p. 111–147).

Lemel Yannick, 1988. Stratification et mobilité sociales, *in*: Henri Mendras and Michel Verret (eds.), *Les champs de la sociologie française*, p. 97–112. Paris, Armand Colin, 272 p.

Lemel Yannick, 1991. *Stratification et mobilité sociales*. Paris, Armand Colin, 226 p.

Lipset Seymour M. and Bendix Reinhardt, 1960. *Social mobility in industrial society*. Berkeley, University of California Press, 309 p.

Livi Livio, 1950. Sur la mesure de la mobilité sociale: résultats d'un sondage effectué sur la population italienne, *Population*, vol. 5(1), p. 65–76.

Luijkx Ruud, 1994. *Comparative loglinear analyses of social mobility and heterogamy*. Tilburg, Tilburg University Press, 326 p.

Merllié Dominique, 1990. Les classements professionnels des enquêtes de mobilité, *Annales (ESC)*, vol. 45(6), p. 1317–1333.

Merllié Dominique, 1994. *Les Enquêtes de mobilité sociale*. Paris, Presses Universitaires de France, 254 p.

Merllié Dominique, 1995. Les travaux empiriques sur la mobilité sociale avant la Première Guerre mondiale, *Revue française de sociologie*, vol. 36(1), p. 5–31.

Merllié Dominique and Prévot Jean, 1991. *La Mobilité sociale*. Paris, La Découverte, 128 p.

Pareto Vilfredo, 1916. *Trattato di sociologia generale*. Firenze. French trans. checked by the author: 1917–1919, *Traité de sociologie générale*. Paris, Payot, 2 vol., 1761 p. (repr. 1968, in *Œuvres complètes*, vol. 12, Geneva, Droz, 1822 p., preface by Raymond Aron).

Perrin Emily, 1904. On the contingency between occupation in the case of fathers and sons, *Biometrika*, vol. 3(4), p. 467–469.

Pohl Richard and Soleilhavoup Jeanine, 1982. La transmission du statut social sur deux ou trois générations, *Économie et statistique*, no. 144, p. 25–42.

Pohl Richard, Soleilhavoup Jeanine, and Ben Rezigue Joèlle, 1983. *Formation, mobilité sociale, salaires: enquête Formation—qualification professionnelle de 1977*. Paris, INSEE, 353 p.

Pohl Richard, Thélot Claude, and Jousset Marie-France, 1974. *L'enquête Formation—qualification professionnelle de 1970*. Paris, INSEE, 276 p.

Poulantzas Nicos, 1974. *Les classes sociales dans le capitalisme aujourd'hui*. Paris, Seuil, 367 p.

Praderie Michel, 1966. Héritage social et chances d'ascension, *in*: Darras [group pseudonym] (ed.), *Le partage des bénéfices: expansion et inégalités en France*, p. 329–348. Paris, Éditions de Minuit, 447 p.

Praderie Michel, Salais Robert and Passagez Monique, 1967. Une enquête sur la formation et la qualification des Français (1964). La mobilité sociale en France: liaison entre la formation reçue et l'activité professionnelle, *Études et conjoncture*, vol. 2,(10), p. 3–109.

Robertson Robert V. and Garnier Maurice, 1985. Class reproduction among men and women in France: reproduction theory on its home ground, *American Sociological Review*, vol. 91(4), p. 912–928.

Rogoff Natalie, 1950. Les recherches américaines sur la mobilité sociale, *Population*, vol. 5(4), p. 669–688.

Rogoff Natalie, 1953. *Recent trends in occupational mobility*. Glencoe, The Free Press, 131 p.

Schiltz Marie-Ange, 1986. Les modèles log-linéaires pour l'analyse des tableaux de contingence, *Bulletin de méthodologie statistique (BMS)*, no. 12, p. 4–21.

Shavit Yossi and Blossfeld Hans-Peter, 1993. *Persistent inequality, changing educational attainment in thirteen countries*. Boulder, Westview Press, 396 p.

Singly (de) François and Claude Thélot, 1986. Racines et profils des ouvriers et des cadres supérieurs, *Revue française de sociologie*, vol. 27(1), p. 47–86.

Singly (de) François and Claude Thélot, 1988. *Gens du privé, gens du public: la grande différence*. Paris, Dunod, 256 p.

Sorokin Pitirim A., 1927. *Social mobility*. New York, Harper & Brothers, 559 p. (expanded ed.: 1959, *Social and cultural mobility*, Glencoe, The Free Press, 645 p.).

Stevens Gillian, 1981. Social mobility and fertility: two effects in one, *American Sociological Review*, vol. 46(4), p. 573–585.

Strauss Anselm L., 1971. *The contexts of social mobility. Ideology and theory*. Chicago, Aldine Pub. Co., 263 p.

Tabard Nicole, 1983. Réflexions sur la relation fécondité—mobilité sociale, *Consommation. Revue de socio-économie*, vol. 30(3), p. 61–70.

Tabard Nicole, 1984. Mobilité sociale, fratrie et descendance, *Consommation. Revue de socio-économie*, vol. 31(3), p. 19–50.

Thélot Claude, 1979. Les fils de cadres qui deviennent ouvriers, *Revue française de sociologie*, vol. 20(2), p. 409–430.

Thélot Claude, 1982. *Tel père, tel fils? Origine familiale et position sociale*. Paris, Bordas and Dunod, 249 p.

Thélot Claude, 1983. L'évolution de la mobilité sociale dans chaque génération, *Économie et statistique*, no. 161, p. 3–21.

THÉVENOT Laurent, 1987. Les enquêtes Formation-qualification professionnelle et leurs ancêtres français, *in*: Joëlle Affichard (ed.), *Pour une histoire de la statistique, vol. 2: Matériaux*, p. 117–165. Paris, INSEE and Economica, 971 p.

THÉVENOT Laurent, 1990. La politique des statistiques: les origines sociales des enquêtes de mobilité sociale, *Annales (ESC)*, vol. 45(6), p. 1275–1300.

TOURAINE Alain, 1954. Rapport sur la préparation en France de l'enquête internationale sur la stratification et la mobilité sociale, *in*: Robert C. Angell (ed.), *Transactions of the Second World Congress of Sociology, Liège 1953*, vol. 2, p. 1–25. London, ISA, 450 p.

VACHER DE LAPOUGE Georges, 1896. *Les sélections sociales: cours libre de science politique professé à l'Unversité de Montpellier (1888–1889)*. Paris, A. Fontemoing, 503 p.

VALLET Louis-André, 1991. *La mobilité sociale des femmes en France: la participation des femmes aux processus de mobilité sociale intergénérationnelle*. Paris, Université de Paris-Sorbonne (Paris-V), 475 + 119 p. (Sociology dissertation).

VALLET Louis-André, 1992. La mobilité sociale des femmes en France: principaux résultats d'une recherche, *in*: Laurence Coutrot and Claude Dubar (eds.), *Cheminements professionnels et mobilités sociales*, p. 179–200. Paris, Centre d'Études et de Recherche sur les Qualifications, CNRS and La Documentation française, 312 p.

VALLET Louis-André, 1995. L'analyse de la mobilité sociale: conceptualisations sociologiques et modélisations statistiques, *La Lettre du CREST* no. 8, p. 1–3.

VALLET Louis-André, 1999. Quarante années de mobilité sociale en France, *Revue française de sociologie*, vol. 40(1), p. 5–64.

WESTERN Mark and WRIGHT Erik O., 1994. The permeability of class boundaries to intergenerational mobility among men in the United States, Canada, Norway and Sweden, *American Sociological Review*, vol. 59(4), p. 606–629.

WESTOFF Charles F., 1953. The changing focus of differential fertility research: the social mobility hypothesis, *The Milbank Memorial Fund Quarterly*, vol. 31(1), p. 24–38.

WRIGHT Erik Olin, 1989. The comparative project on class structure and class consciousness: an overview, *Acta sociologica*, vol. 32(1), p. 3–22.

WRIGHT Erik Olin, 1997. *Class counts: comparative studies in class analysis*. Cambridge, Cambridge University Press, 576 p.

YAMAGUCHI Kazuo, 1983. The structure of intergenerational occupational mobility: generality and specificity in resources, channels and barriers, *American Journal of Sociology*, vol. 88(4), p. 718–745.

YASUDA Saburo, 1964. A methodological inquiry into social mobility, *American Sociological Review*, vol. 29(1), p. 16–23.

ZARCA Bernard, 1993a. L'héritage de l'indépendance professionnelle selon les lignées, le sexe et le rang dans la fratrie, *Population*, vol. 48(2), p. 275–306.

ZARCA Bernard, 1993b. L'héritage de l'indépendance professionnelle: un ou plusieurs élus au sein de la fratrie?, *Population*, vol. 48(4), p. 1015–1042.

APPENDIX 92–1

A Log-Linear Model of the Association between Educational Attainment and Social Origin[11]

Let $F_{i,j,t}$ be the number of individuals aged 25 to 34 of social origin i and educational attainment j at date t. We can decompose $\mathrm{Log}\,F_{i,j,t}$ in a unique way (see Gouriéroux, 1984), as follows:

$$\mathrm{Log}\,F_{i,j,t} = \mu + \lambda_i^O + \lambda_j^D + \lambda_t^T + \lambda_{i,j}^{OD} + \lambda_{j,t}^{DT} + \lambda_{i,t}^{OT} + \lambda_{i,j,t}^{ODT} \quad (1)$$

where O denotes social origin, D educational attainment, and T the survey date. The γ coefficients meet the following identification constraints:

$$\begin{aligned}
\sum_i \lambda_i^O &= \sum_j \lambda_j^D = \sum_t \lambda_t^T = 0,\\
\sum_i \lambda_{i,j}^{OD} &= \sum_j \lambda_{i,j}^{OD} = \sum_j \lambda_{j,t}^{DT} = \sum_t \lambda_{j,t}^{DT}\\
&= \sum_i \lambda_{i,t}^{OT} = \sum_t \lambda_{i,t}^{OT} = 0,\\
\sum_i \lambda_{i,j,t}^{ODT} &= \sum_j \lambda_{i,j,t}^{ODT} = \sum_t {}_{i,j,t}^{ODT} = 0.
\end{aligned}$$

μ is a constant representing a parameter of scale. The λ_i^O, λ_j^D, and λ_t^T coefficients stand, respectively, for the effect of the distribution of individuals by social origin O, educational attainment D, and survey date T. The $\lambda_{i,j}^{OD}$, $\lambda_{i,t}^{OT}$, and $\lambda_{j,t}^{DT}$ coefficients stand, respectively, for the effect of the specific associations between social origin and educational attainment, between social origin and survey date, and between educational attainment and survey date. The $\lambda_{i,j,t}^{ODT}$ coefficients stand for the impact of the changes in the associations between social origin and educational attainment over time.

The advantage of this type of model is that it allows a simple test of the independence and conditional-independence hypotheses. In the study of social mobility, most of the key questions can be framed in terms of independence. For example, testing the stability of educational inequality over time is the same thing as testing the independence of the OT association and DT associations. This involves testing the simultaneous nullity of the $\lambda_{i,j,tt}^{ODT}$ parameters in Eq. 1, that is, verifying the statistical validity of the model:

$$\mathrm{Log}\,F_{i,j,t} = \mu + \lambda_i^O + \lambda_j^D + \lambda_t^T + \lambda_{i,j}^{OD} + \lambda_{j,t}^{DT} + \lambda_{i,t}^{OT} \quad (2)$$

An abridged representation of Eq. 2 is $(OD)(OT)(DT)$. This is consistent with the standard practice of describing log-linear models by means of a summary outline of the effects and interactions to which the model is confined.

To test the compatibility of Eq. 2 with the data, we can simply estimate the γ and μ parameters by using

[11] From Dominique Goux and Éric Maurin (1997a).

TABLE 92–A Change in Association between Education Attainment and Social Origin, 1970–1993: A Log-Linear Model

Models	Degrees of freedom	G^2	rG (%)	p
Benchmark model: *(OD)(OT)*	36	1,788	—	0.0001
Uniform democratization: *(OD)(OT)(DT)*	27	156	91.3	0.0001
Effect: *(DT)*	9	1,632		0.0001

1. T stands for the survey date. Four-category classifications are used for social origin (O) and educational attainment (D). The distance G^2, combined with the number of degrees of freedom, provides a measure of the difference between the observations and the model. The probability p measures the test's significance.

Interpretation: Between 1970 and 1993, 91.3% of the increase in educational attainment was uniformly distributed across all social strata.

Field: young adults aged 25 to 34 in 1993.

From Formation-Qualification Professionnelle surveys (INSEE, 1970, 1977, 1985, 1993).

TABLE 92–B Change in Association between Education Attainment and Social Origin, 1970–1993: A Log-Linear Model

Models	Degrees of freedom	G^2	rG (%)	p
Benchmark model: *(OD)(OT)*	60	870	—	0.0001
Uniform democratization: *(OD)(OT)(DT)*	50	104	88.0	0.0001
Effect: *(DT)*	10	766		0.0001

1. T stands for the survey date. Six-category classifications are used for social origin (O) and educational attainment (D). The distance G^2, combined with the number of degrees of freedom, provides a measure of the difference between the observations and the model. The probability p measures the test's significance.

Interpretation: Between 1970 and 1993, 91.3% of the increase in educational attainment was uniformly distributed across all social strata.

Field: young adults aged 25 to 34 in 1993.

From Formation-Qualification Professionnelle surveys (INSEE, 1970, 1977, 1985, 1993).

the maximum-likelihood method and then assess the distance between the observed data and the model forecasts by means of the likelihood statistic G^2.[12] Another method, which typically yields very similar results, consists in estimating the parameters by minimizing the distance of Pearson's chi-square.[13] Here, the two methods lead to the same results (Goux and Maurin, 1995).

Table 92–A summarizes the results. It reports a log-linear analysis of the mobility table whose elements are the $F_{i,j,t}$ values. These are obtained by combining education and social-origin classifications into four categories and by separating four dates. The $F_{i,j,t}$ vector studied has $4 \times 4 \times 4 = 64$ degrees of freedom.

Table 92–B shows the same analysis, but using the more detailed six-category classifications. The vector studied has $6 \times 6 \times 6 = 216$ degrees of freedom.

To study the dissemination of diplomas (i.e., school democratization), we chose the (*OD*)(*OT*) model as the benchmark. This model postulates that (1) inequality has remained stable (no *OT* or *DT* interaction), and (2) educational attainment has remained stable over time within each social category (no *DT* interaction). The model does not, of course, provide a statistically acceptable description of the changes in the distribution of individuals by social origin and education ($G^2 = 1788$ and the probability p that the omitted coefficients are null is a very small 0.0001). The distance between this model and the observed data is consistent with the full impact of school democratization on the shift in the mobility table over time. This is precisely the influence we are trying to isolate and analyze, and we have accordingly adopted this model as the benchmark in our study.

We next estimated the (*OD*)(*OT*)(*DT*) model. Here, only the $\lambda^{ODT}_{i,j,t}$ interactions have been constrained to zero. We kept 37 parameters (μ, (μ, $3\lambda^O_i$, $3\lambda^D_j$, $3\lambda^T_t$, $9\lambda^{OD}_{i,j}$, $9\lambda^{OT}_{i,t}$, and $9\lambda^{DT}_{j,t}$), leaving only 27 degrees of freedom. This model is based on the hypothesis of a uniform dissemination of diplomas, that is, no *OT* or *DT* interaction. The (*OD*)(*OT*)(*DT*) model is far more realistic than the previous one. It explains more than 90% of the differences between the frequencies predicted by the

[12] If $F^*_{i,j,t}$ is the estimate of the frequency $F_{i,j,t}$ by the log-linear model M, the quality of M's fit with the data can be estimated by

$$G^2 = 2\sum_{i,j,t} F^{Log}_{i,j,t}\left(F_{i,j,t}\middle/F^*_{i,j,t}\right)$$

This statistic displays additivity properties: if M' is a model nested in M, then

$$G^2(M') = G^2(M) + G^2(M,M').$$

[13] The chi-square statistic has the minor disadvantage of not being additive. The distance between two models is not quite identical to the difference between those models and observed data.

TABLE 92–C Main Differences between Observed Changes *(ODT)* and the Uniform Democratization Model *(OD)(OT)(DT)*

Educational attainment	1970	1977	1985	1993
1. Farmers' children[a]				
Higher education	0	0	0	0
Bacclaurêat or equivalent	–	0	0	+
CAP, BEP, BEPC	0	0	0	0
CEP or dropout	+	0	0	–
2. Manual or clerical workers' children[b]				
Higher education	+	0	–	0
Bacclaurêat or equivalent	+	0	–	0
CAP, BEP, BEPC	0	–	+	0
CEP or dropout	–	+	+	0

Bacclaurêat, senior-year high-school diploma; a prerequisite for admission to university; CAP, *certificat d'aptitudes professionnelles* (vocational training diploma); BEP, *brevet d'êtudes professionnelles* (vocational training diploma); BEPC, *brevet d'êtudes du premier cycle* (diploma at the end of fourth year of secondary school); CEP, certificat d'êtudes primaires (primary-school diploma, now abolished).

Field: young adults aged 25 to 34 in 1993.

[a]*Interpretation:* The estimates of the $\lambda^{O}_{i=farmer,}{}^{DT}_{j,t}$ coefficients that measure the differences between the *(OD)(OT)(DT)* model and the observed data *(ODT)*. Only the signs of the coefficients are given; 0 denotes a value not significantly different from zero. If democratization had been strictly uniform, there would have been more dropouts and fewer *Bacclaurêat* holders among farmers' children in 1993.

[b]*Interpretation:* The table reports the estimates of the $\lambda^{O}_{i=manual\ or\ clerical\ worker,}{}^{DT}_{j,t}$ coefficients that measure the differences between the *(OD)(OT)(DT)* model and the observed data *(ODT)*. Only the signs of the coefficients are given; 0 denotes a value not significantly different from zero. If democratization had been strictly uniform, there would have been fewer dropouts and more *Bacclaurêat* holders among manual and clerical workers' children in 1993.

From Formation-Qualification Professionnelle surveys (INSEE, 1970, 1977, 1985, 1993).

(*OD*)(*DT*) model and the observed frequencies (this is expressed in the formula $rG^2 = 91.3$). In other words, more than 90% of the dissemination of diplomas has occurred uniformly in each social category, without modifying inequality. Only the residual 8.7% is owing to changes in the association between education and social origin over time.

The differences between the (*OD*)(*OT*)(*DT*) model and the observed frequencies are weak but significant: in other words, the probability that the order-three interactions neglected by the model are all null is itself practically null. The analysis of the differences between the model and the observed frequencies allows us to locate the frequencies of the $\lambda^{ODT}_{i,j,t}$ interactions, whose nullity can be ruled out. Table 92–C reports the estimated values of these non-null interactions. They concern only the most modest social categories: farmers versus manual and clerical workers. For greater legibility, only the signs of the significantly non-null $\lambda^{O}_{i=farmer,}{}^{DT}_{j,t}$ and $\lambda^{O}_{i\,=\,manual\ or\ clerical\ worker,}{}^{DT}_{j,t}$ interaction are given; the others are equal to 0. Table 92–C thus shows that, if the children of manual and clerical workers had benefited as much as other children from school democratization, they would have included more holders of the *baccalauréat* and higher-education degrees, and fewer dropouts (Table 92–C-2). In contrast—again assuming strictly uniform democratization—we would now find more dropouts and fewer *baccalauréat*-holders among farmers' children (Table 92–C-1).

CHAPTER

93

Gender and Demography

ANTONELLA PINNELLI

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

INTRODUCTION

This chapter sets out to illustrate the main insights that can be expected from adding a gender perspective to demographic research and training. As has already happened with sociology, demography is shifting from often implicitly feminist-oriented studies on women to studies on gender, which looks more at male-female relationships and how they evolve over time. This is because changes in women's status clearly entail changes in that of men, so their respective agendas and conditions, and the interactions between them, must be examined to understand the trends and determinants of demographic behaviors.

This chapter first clarifies the definitions of some terms and expressions used in gender studies, which refer to social rather than biological differences between the sexes, followed by a brief background to the development of gender studies in demography. Recent findings in the key areas of demography—family, fertility and contraception, health and mortality, migration, aging—are then reviewed in a five-section survey of the literature on "gender and population," focusing on what are arguably the most promising lines of research. The gender system, or certain aspects of it, will in some cases be looked at as determinants of demographic behavior, whereas in other cases, the other side of the question will be examined: how do demographic behaviors modify gender systems? This is because the relation between the two aspects is not one-way but much more two-way, and it is important to highlight one or other of these two directions in the relation or their reciprocal linkages, as the case may be. The final section will consider the specific data and methods relating to gender and population research.

I. SOME DEFINITIONS

The word "gender" has only recently entered into demography. As Richard Udry (1994) points out, in a 12,000 title bibliography of the literature on marriage and the family from 1900–1964, the word gender is notable by its absence. In 1972, John Money suggested using the word *sex* for the biological distinction between males and females, and the word *gender* when discussing sex-specific behavioral differences (Money and Ehrhardt, 1972). Generally, the word gender is used to highlight the social development of relations between the sexes and to get away from the biological determinism implicit in the word "sex" or expressions such as "sex differences." The word "gender" also stresses the relational aspect: men and women are defined relative to one another, and neither can be understood other than by studying both (Watkins, 1993).

In social science terms, gender depends on three things: roles, socialization, opportunities. A *gender role* is a range of sex-specific and field-differentiated acceptable behaviors, perpetuated by *gender norms.* The bounds of acceptable behaviors differ for men and women: overstepping these bounds is penalized and, in any event, obstructed by the surrounding environment.

Where do sex-differentiated norms derive from? There are several theories, based variously on the influence of technologies; reproductive role differences, which in turn underlie the gendered division of labor; ideological theories; and, finally, those arguing for sex-differentiated behavior as simple historical accident (Bem, 1987). In the social sciences, gender is about differential social experiences that explain both between- and within-gender differential behavior. With these theories goes a biological theory of gender: sex hormones dictate sex-differentiated behavior and the development of the reproductive function among humans and animals, but in different ways. Biology and social experiences both have an influence, as shown by the undeniable historical changes in gender behaviors (Udry, 1994).

Karen Oppenheim Mason (1995) describes the socially constructed societal expectations for male and female behavior as a *gender system.* The expectations of a gender system prescribe a division of roles and responsibilities between men and women and grant different rights and obligations to them. The gender system comprises the complex of roles, rights, and statuses that surround being male versus female in a given society or culture. It has subcomponents, for example, it prescribes a division of labor and creates institutionalized inequality between male and female members of society to form a *gender stratification* that is measurable in terms of wealth, power, and prestige. The power dimension, for example, includes autonomy (the freedom to act as they choose rather than as others would have them act), control of material resources, and command of social honor or prestige.

But this is not to deny that sex-specific biological differences may shape men's and women's behaviors: a gender-neutral environment, if possible, would not produce a neutral society because the two sexes do not react in the same way to the same elements of socialization (Udry, 1994). Also, the individual behaviors of each sex (in variable form) are in great part owing to individual biological characteristics, which are also an explanatory factor of within- and between-gender variability (Udry, 1994). Generally, however, the gender system goes beyond the behavioral differences that can be expected of biological diversity alone and creates an injustice—that is, inequality between the sexes in power, autonomy, and well-being typically to the disadvantage of women, which change in time and space.

Improving women's status and reducing gender differences means increasing women's empowerment. The United Nations (UN) *Program of Action* produced by the *International Conference on Population and Development* (1994) lists actions to be undertaken in favor of women: to improve their political representation, education, health, participation in the labor market and social security systems; to eliminate poverty and discrimination of all kinds, including employment discrimination based on pregnancy status or the pledge not to have children, and stereotypes in the media and school textbooks; to make it possible to combine work and child-rearing; to give women full economic rights (e.g., to sell, buy, inherit, borrow, contract); to promote the setting-up of activist groups for women to foster women's awareness of their rights; to strengthen women's well-being, self-image, and self-esteem; and, finally, to provide consideration for women and children. All this involves male responsibilities and participation.

Women's increased empowerment is therefore measured in many and unaccustomed spheres, which open up new areas of research and require new data and new methods to evaluate the changes and their effects on demographic behaviors.

II. FROM WOMEN'S STUDIES TO GENDER STUDIES IN DEMOGRAPHY

From the 1960s, a handful of feminists (Blake, 1965; Ridley, 1968; Dixon, 1975; Germain, 1975) began showing that women's status has significant demographic implications. But it was chiefly from the 1980s that the issue entered the mainstream of demographic thought, although not yet a central variable of *transition theory* (see Chapters 68, 69). A series of studies have built up a clearer and more detailed picture of the significance of women's status as a multidimensional concept composed of autonomy, prestige, power, empowerment, access to and command of resources, etc. The framework for analysis devised by Christine Oppong and Katharine Abu (1986)—the so-called *seven roles* theory—clarified the different roles that women play (individual, conjugal, maternal, domestic, occupational, kin, and community) and helped shed light on the underlying dimensions of women's status and suggested methods of information collection and indicators that could turn the concepts into operational definitions. This activity of cerebration, theorizing, and development of working constructs was held back by the practice that had evolved in differential demographic analyses of using women's educational level and participation in the labor market as women's status indicators; this stemmed partly from practical considerations, some intellectual laziness, and inertia—seldom was the trouble taken to clarify the theoretical reference frame that set the limits of validity within which the selected data and indicators were usable for each study.

In each case, many of the concepts put forward implicitly referred to a male benchmark. Further, the concepts of women's status and social status were often used interchangeably, such that it gradually became seen as preferable to talk of *gender* rather than women's status, by explicitly including the comparison with men into the assessment of women's status.

The UN and its agencies, along with other international organizations involved in population and development policy, have made a big contribution to this process, which first revealed the importance of women's status as a determinant of demographic behavior and then clarified and developed the concept to accommodate gender systems. When the stages in this development and its gradual inclusion in the action programs drawn up by the UN for the three world population conferences in Bucharest, Mexico City, and Cairo are charted, it is striking to see the increased prominence given in them first to women (from 1974–1984) and then to gender, in 1994. This rising emphasis on women, finally resulting in a gender focus, comes chiefly from an awakening to the need to expand women's decision-making power over sexuality and reproduction, and to reduce male opposition to the idea, at the risk of seeing family planning programs collapse. The ideas, preferences, and expectations of women and men alike therefore had to be taken into account. This policy shift has had a significant influence on public opinions and research.

The International Union for the Scientific Study of Population (IUSSP) has played a key role in this development as an early adopter of issues around women's status, gender, and their relationships to demographic behavior (see Appendix 93–1). The IUSSP, first through the preparatory work for the Asker Conference and then through the activities of the scientific committees it appointed on the issue, has done much to raise demographers' awareness of gender issues, define and develop this field of research, devise theoretical frames of reference, and systematize empirical knowledge.

III. GENDER AND FAMILY

Marriage is the most common way to sanction the union between a man and a woman and is the origin of family formation. Historically, it has been used as means of creating wealth and power. The rules and norms governing marriage, divorce, conjugal relations, and the rights and duties of family members are important to each society. Very few countries anywhere in the world at present give the two sexes equal rights in relation to marriage, divorce, and family property. Put simply, institutionalized gender inequality still persists in most countries. There are still countries where the legal minimum age at marriage for women is 12 or 14, and millions of girl children are given in marriage in the world every year (UN, 1992). The spousal age difference, always to the man's advantage albeit to very varying degrees, is the simplest indicator of the unequal relationships within the union. These and other aspects of the gender disparity within the family have been extensively considered by Véronique Hertrich and Thérèse Locoh (1999) for developing countries, and by me for developed countries (Pinnelli, 1999).

One function of the family should be to protect its members, but in fact, violence and abuse are common within families, and the victims are invariably women. It may range from battering, burn injuries, and other types of assault on women (especially in connection with dowry disputes), through genital mutilations, child sexual abuse and psychological violence to spouse rape and murder (UN, 1989). Recent studies have shown that most family violence is perpetrated on the wife, invariably by the husband (Smyke, 1991), although mothers-in-law are also implicated to quite a high degree.

Beyond the male breadwinner–female housewife stereotype, gender relations within the family vary by family type (nuclear, cohabitation or long-distance union, step/blended, extended, kinship, tribal, polygamous) and life-cycle stage (courting, marriage or cohabitation, first childbirth, last child's home-leaving, labor market exit of aged partners, divorce or widowhood).

The developed countries have experienced big changes, especially in patterns of union formation: marriage is less frequent and more delayed, often after a period of cohabitation, and less stable. More children are born outside wedlock and to older mothers and increasingly often live with a social rather than a biological father or with a lone mother (Roussel, 1989; Pinnelli, 2001). This situation varies from country to country, with the economic and social situation, gender system, and institutional context. In southern Europe, for example, these trends are not yet particularly pronounced, because institutional support to families is limited (inadequate childcare provision, shorter maternity leave, tax system favoring legal families), because there is little assistance for young people made unemployed by the economic crisis, or because the gender system is more unfavorable to women (less autonomy, less work and more frequently interrupted careers, less political and administrative power) (Pinnelli, 1995, 1999). In contrast, changes in the family are more apparent where female labor force participation is more widespread and where the economic situation enables the formation of even nontraditional unions. That is not without adverse consequences for the life of women and children, however: out-of-wedlock children often have a lower birth weight and higher infant mortality than do in-wedlock children, notwithstanding that the rise in illegitimacy also reflects positive developments such as women's greater empowerment and autonomy. The increased instability of family situations has a negative influence on mothers' and children's physical and mental well-being (Pinnelli and Masocco, 1995).

Lone mother families resulting from union dissolution (break-up or divorce) are growing increasingly common, and women have difficulty juggling the children's needs with the need to work as well as assuming alone the physical, emotional, and economic family responsibilities in situations in which insufficient institutional support may be compounded by ex-husband or partner absence. Women and children bear the brunt of the increased instability of the new family patterns. But moving from studies on women to gender studies also means taking into account men's expectations and preferences to help understand and reconcile individual demands for personal happiness and responsibility to future generations.

Comparative analysis of male and female behaviors in the developed countries clearly reveals how women are torn among work, career, and family life and acknowledges the higher price they pay for the deinstitutionalization of marriage and increased instability of unions (Pinnelli *et al.*, 2003).

IV. GENDER, FERTILITY, AND CONTRACEPTION

1. Fertility

Karen Oppenheim Mason (1985) reviewed the literature on relations between gender systems and fertility in developing countries.

Little research has been done on how gender systems affect reproductive behavior, but there are concordant findings from case studies done in Bangladesh (Balk, 1994), Nepal (Morgan and Niraula, 1994), and India (Malhotra *et al.*, 1995) that both micro- and macrolevel changes in gender systems produce fluctuations in fertility and its intervening variables. The first of these used women's freedom of movement within the community, their authority within the home, the leniency of households toward their women, and their gender attitude as indicators of the gender system. The second considered women's freedom of movement and voice in household decisions on the use of contraception and fertility desires. The third took the relationship between female and male mortality, which preceding studies had shown to be a good proxy for *gender stratification*, and the proportion of economically active women. Further, various studies done in Bangladesh (Schuler and Hashemi, 1994) showed that involvement in programs to raise their standard of living enhanced women's literacy skills, earning abilities, and power and increased contraception use. However, the geographic and cultural coverage of these studies is limited, and it is not known whether identical findings would be made in less traditional contexts.

One intervening variable added to those designed to explain fertility patterns is son preference, which prompts "unlucky" parents to continue creating children until their sex preferences are fulfilled. A relationship has been observed between fertility decline and a lesser degree of son preference (Cain, 1993).

Studies based on evidence from the World Fertility Survey (WFS) and Demographic and Health Survey (DHS) analyzed the influence of women's education and employment on contraceptive behavior and fertility. The findings for education are robust and consistent with the hypothesis of an inverse relationship with fertility; those for employment are less so. Where employment fails to affect women's enhanced autonomy, power, or control of resources, fertility does not fall. Evidence from some studies, however, shows that individual-level measures of power and autonomy are unrelated either to fertility levels or to contraceptive use (Jejeebhoy, 1995).

In developed countries, with generally subreplacement level fertility, gender equality is associated with higher fertility (Pinnelli, 2001; Di Giulio and Pinnelli, 2003). This finding has surprised those who hypothesized that women's enhanced status led to reduced fertility. It suggests the converse—that women who acquire more autonomy and power more easily approach ideal fertility as measured by surveys on the number of children wanted, whereas others tend to have lower fertility (Pinnelli, 1995). In fact, the lowest fertility is found today in the southern European countries, where, as mentioned earlier, protracted economic crisis and youth unemployment are obstacles to first union formation and the lack of adequate institutional assistance is a disincentive to married women to have more than one child, notwithstanding the women's more traditional status in that location and its more pronounced gender differences.

2. Family Planning

Demographic research has long disregarded men's role in reproductive and contraceptive decisions; as a result, family planning programs in developing countries have been directed only toward women. One characteristic of contraceptive models in continuing very high fertility countries, therefore, is a lack of male involvement, evidenced by the infrequent use of male methods of birth control. This situation is only partly explained by the need for wholly reliable methods, because more medical research could have been directed toward male methods. The real problem is the failure to get men engaged with family planning programs. Too little attempt was made to encourage men and women to make joint decisions about their desired fertility, birth spacing, and the methods to be used to achieve that, or to tackle male abdication of responsibility or chauvinism.

A woman-centric approach may arguably be the most practical course, given that women do not want all the many pregnancies they have. But the evidence is that in fact women cannot solely take such key decisions for the family and society as family size: disempowered in so many spheres of both public and private life, they cannot act in this of their own volition. For instance, a study of women's and their partners' fertility and contraception demands and preferences in selected African countries revealed that men were less concerned than were women about fertility control, but that men's opinion prevailed over women's in moving from preferences to practice (Andro and Hertrich, 2001).

Another consequence of sidelining men from family planning programs is to have held back condom use by young people, women who cannot use the pill or intrauterine device (IUD) on health grounds, for example, or where there is a risk of sexually transmitted infection.

Getting men to engage more with contraception and family planning decisions also means involving them more in responsibility and care for children, prompting them to provide adequate economic resources for the needs of the large family to which they generally aspire, and, ultimately, inspiring them to limit their fertility.

In the long run, the success of family planning programs will depend on changing men's attitudes toward women, marriage, children, and the family, as well as on changing women's attitudes.

This means studying men's contraception and fertility desires, preferences, and behaviors in the same way as women's, as well as studying the interactions between the sexes. This is why fertility and birth control surveys, including, for example, the Family and Fertility Survey (FFS) and DHS, now systematically include a sample of men as well as the traditional sample of women, in recognition of the importance of recording data for both partners.

The contraceptive history of developed countries has taken a very different tack. Fertility control has been a male preserve, using such methods as condoms or coitus interruptus (a specifically male method), whereas women have contributed to reduce fertility through abortion (Dalla Zuanna, 1996). The 1960s, and especially the 1970s, with the spread of the pill, followed by IUDs and sterilization, signaled women's ownership of family planning. Falling abortion rates evidenced the gradual shift to more effective contraception methods. But also in the developed countries, the sudden emergence of AIDS put the merits of condom use and, more generally, negotiation with the partner for infection risk-free sexual activity back on the agenda. Involving men in contraceptive decisions has thus become as essential in developed as developing countries.

V. GENDER, HEALTH, AND MORTALITY

1. Child Health and Survival

Responsibility for the health of family members is traditionally left to women, but where there is a highly antiquated view of women's status and wide gender gaps, women are excluded from modern medicine, an act that has family-wide repercussions. It is in society's interests, therefore, that women should have access to modern medicine and the ability to get information and interact with the health system. A large body of research in fact shows that the gender system influences infant health and survival, both of which improve with mothers' empowerment. Islamic Middle Eastern countries, where women have little freedom of movement or autonomy, are those where, *ceteris paribus*, child health conditions are worst. This was reported by John Caldwell as long ago as 1986 and was subsequently confirmed by a series of studies (Oppenheim Mason, 1995) using micro- and macrolevel measures of the gender system.

Even the most approximate indicator—women's educational levels—bears out these findings, because the most well-educated women are more self-confident and resolute or are able to communicate with their partner. There is therefore a link between education and the intervening variables of infant survival, such as hygiene and utilization of health care provision, and that holds true even where socioeconomic conditions are controlled. The effect of women's economic participation is more contentious, however: although women's resources may help improve children's diet and health care, work may reduce their mother's available time and attention for them;

unsurprisingly therefore, the relation is found to be situation specific (Basu and Basu, 1991).

Excess girl child mortality has been observed in several countries, notwithstanding the greater biological weakness of boys (see Chapter 54). Some studies (Chen *et al*., 1981; Das Gupta, 1987) have evidenced girl child discrimination in relation to feeding and medical care. Excess female child mortality is a very robust proxy for sex discrimination against women.

2. Excess Male Mortality

As development has progressed, mortality generally has receded but excess male mortality has risen (see Chapter 54). Societies with no pronounced son preference are always found to have approximately 10% to 15% excess infant male mortality (Vallin, 1993). In developed countries, this excess male mortality then becomes significantly age specific, peaking at a very high level around the age of 65. Historically, however, and in developing country societies, age-specific excess female mortality has been observed, especially in infancy and at the reproductive ages (see Chapter 54).

Were lifelong excess male mortality to remain constant at its first-year level, gender-differential life expectancy would be approximately 2 years (Pressat, 1973). Arguably, therefore, a female survival advantage of 2 years could be seen as biologically normal, and a lower recorded difference in a given country could be attributed to women's disadvantage in treatment, diet, and economic participation, whereas a differential above 2 years reflects a male disadvantage.

Some of the male disadvantage stems from men's more aggressive, reckless, and less health-conscious behavior. But health and health care provision use surveys also evidence gender-differential attitudes: women are readier than men to acknowledge their unwellness and seek treatment, whereas men will seek treatment if they acknowledge their illness but are less ready to do so (Waldron, 1976, 1985). Arguably, that is a key explanatory variable of the sex-specific mortality differentials. But males are also more exposed to environmental risks (physical and psychological). A borderline case evidencing an exceptional male disadvantage is Russia, where the recent mortality crisis after the collapse of the communist system simply worsened the ongoing decline in the health situation since the mid-1960s (Shkolnikov *et al*., 1995a–d; Meslé *et al*., 1996, 1998). Even so, the female excess over male life expectancy at birth rose there from 9 years in 1965 to 14 years in 1995 (Meslé and Vallin, 1998).

A fascinating field of research straddling biology and sociology has opened up to encompass, beyond the often self-victimizing stereotypes of feminist studies, the highly meaningful causes of one of very few female advantages.

3. The Biological Role in Danger

Developing country infant and maternal mortality indicators bear an inverse relationship to care development and availability; conversely, in developed countries, where infant and maternal mortality rates are very low, no correlation is found with the intensity of care, which may be seen as excessive and unnecessary. There is in fact no evidence that the overmedicalization of pregnancy and childbirth produces more mother and child health and survival advantages. Women undergo very frequent checks and even risky analyses, labor is medically controlled, the frequency of caesarian sections is much higher than that which the World Health Organization (WHO, 1985) considers to be physiologically necessary, breast feeding has been reduced to a few weeks or abandoned altogether in many developed countries and is also declining in developing countries (for a short literature review, see Cesaroni and Pinnelli, 1999). Not only does this extreme overmedicalization of a function that is specific and exclusive to women bring no gain for either mother or child, it surely brings a psychological and emotional loss.

One factor that encourages overmedicalization of childbirth in developed countries is later motherhood: delayed first pregnancies are seen as medically delicate, so risk-avoidance controls and care are—consensually—increased. Birth deferral also exacerbates conception problems, evidenced by the steadily-growing recourse to assisted reproductive techniques, used increasingly—and increasingly earlier—when pregnancy does not rapidly follow the decision to procreate (WHO, 1985). This creates a general impression of a rise in sterility and infertility. But various studies (Mosher and Pratt, 1993; Beets *et al*., 1994) have shown that deferring the first birth until age 30 to 35 or even after, partly owing to women's rising educational levels and greater economic participation, is the structural cause of women's increasing procreation problems in developed countries. Whatever else, this difficulty helps to drive the market for medically assisted reproduction, with all its economic and psychological costs and the attendant risks of failure.

The inference from this example of the overmedicalization of birth is that greater gender equality makes it harder for women to accept and assume their biological role. They uncritically accept excess care provision as if pregnancy were a serious illness and not a natural fact of a woman's life, but seek to expedite a pregnancy that does not occur for being too long deferred (Leridon, 1992). They may need a greater ability at the right biological and social time in the life cycle to balance emotional fulfillment with social and career success, while preserving the priceless naturalness of motherhood as being a function specific to women.

VI. GENDER AND MIGRATION

There is scant information on gender and migration, because migration is still seen as a primarily male phenomenon. The stereotype is that women's migration follows that of their husbands, so most research is focused on male motivations and preferences. But women also increasingly frequently migrate of their own volition, especially never-married and very young women under 20 or older women aged around 45 to 50—especially those widowed, separated, or divorced. Historically, these movements were often thwarted by a range of circumstances: the gender division of agricultural labor assigned work requiring constant, ongoing care to women and more seasonal work to men, such that men were able, and women unable, to migrate. Furthermore, migration was seen as endangering their virtue, which would be thus placed beyond the control of parents and elders. Finally, women's lower average educational level reduced their employment opportunities and disadvantaged them against men (Makinwa-Adebusoye and Afolayan, 1995). But the situation has undergone rapid change, and an increasing number of women now migrate as free agents.

Going or staying is the outcome of a family survival strategy that decides who will migrate. It is important to consider this from the perspective of the microlevel, as well as at the macrolevel of social and cultural norms and employment policies that often disadvantage women. Women emigrate for education, which for many is the first step toward urbanization; as a permanent or occasional domestic worker or as a member in the sex trade, which is seldom permanent and less common than may be supposed; in farming, often in seasonal work, as is common in many regions of Latin America, Asia, and Africa; or in the act of joining their husband or a relative. Similar to men, women's migration is chiefly economic, especially when they are young, but less commonly is this the only reason. Women's children also accompany them, if not left with relatives. Personal and family considerations dictate the time and destination of migration (UN, 1994b). In international migrations, women suffer discrimination on three counts: class, race, and gender. Most work, usually in domestic service and childcare (by thus performing for reward some of the domestic and care duties of their female employers, they release these women to engage in other activities outside the home) and in shops. Working for lower pay and in less secure jobs, they nevertheless remit more money home than do men (Makinwa-Adebusoye and Afolayan, 1995). They have lower fertility than do nonmigrant women (Findley, 1999; Hugo, 1999).

Nonmigrant women are apt to be middle-aged mothers, resigned to their husband's emigration as a necessity. Where they already have some autonomy, it increases with the husband's emigration, and their situation may improve. Others, however, may find themselves in greater difficulty. Separation from the husband and the possible consequences are always stressful.

The personal and family lives, and the social and economic conditions, of women who migrate with or without husbands—but also nonmigrating women with migrant husbands—undergo major upheavals. The gender system changes in particular. Women who are significant wage-earners and also more reliable household breadwinners appreciably increase their value in the family. Migration changes the rules of care and attention provision and the models of intergenerational assistance. Migrant women have fewer children than do nonmigrant women and can play a leader role in fertility transition as models of innovation or sources of information on low fertility. This is true for migrant women and for the nonmigrant wives of migrant husbands: family structures change (Findley and Williams, 1991). Research has barely begun to consider migration from the migrant/nonmigrant woman viewpoint, and the subjects studied have only been touched on. It is clear that migration has many repercussions on the lives of women (work–life balance, well-being, education, social mobility opportunities, health, etc.), those for whom they are caregivers, and gender relations in the societies of departure and arrival. Research should inform all these aspects in the coming years.

Another important factor is changes in social structures and gender relations related to rapid urbanization. Gender structures shape women and men's involvement in management of the community, the extent of their representation in urban government, and how far their requirements inform urban development and organization. Structural adjustment programs, for example, have affected women's lifestyles more than men's: women have had to engage in paid work and reduce their family consumption. Women do most of the childcare and housework, but their demands on housing and essential services go unheeded. The house is often the center of productive activities. Women's greater physical and social vulnerability should inform the planning of community, housing, service, and transport development (UN, 1994b).

VII. GENDER AND AGING

Aging is a female topic in many ways: because at the older ages women outnumber men, because institutional support at the older ages is often deficient, because women are the main caregivers for older people, and because women's lower participation rates and lower-paid jobs expose them to a greater risk of old age poverty. Here, the gender perspective adds new areas for exploration, but research is barely at its beginnings. Consider a few examples of the gender-aging nexus. In some cultures, women are empowered by exiting reproductive age: they are then less subject to male authority and can engage in many activities prohibited to younger women. They no longer need to defer in the same way to men or dress or conduct themselves as modestly; they may disregard certain language taboos and speak more freely and with more authority; and they may move outside the home and travel much more freely than younger women. All this means that older women have the potential to challenge men in decision-making bodies, carry on commercial activities beyond their village, and work with others in organized activities. But, simultaneously, older women in traditional societies also often become protagonists in the transmission of gender norms that disadvantage females.

This contrasts with industrialized countries, where age tends to bring disempowerment for women. Granted, there are situations in which increased women's empowerment generally delivers greater well-being for older women: one example is care provision for older males, which falls to women through either tradition or demography (women are apt to care for their elderly spouses and males not, because generally in unions, the husband is the older and predeceases). When women are empowered in political decision making and local government, institutional assistance to families increases and this advantages all women (UN, 1993). This is a specific example of the mechanism of women's empowerment: here, it adjusts the balances in the gender system, enabling an improved quality of life for women without harming those needing care.

Gender role differentials may even decrease with age: retired men can help more with housework, whereas women can become more independent and outward-looking once the children have left home. This is borne out by various studies, although the results do not lend themselves to extrapolation (Arber and Ginn, 1995).

VIII. DATA AND METHODS FOR THE STUDY OF THE GENDER-DEMOGRAPHIC BEHAVIOR RELATIONSHIP

1. Stereotypes and Realities

Through an analysis of articles published in the first 30 years of the journal *Demography* since 1964, Susan Watkins (1993) showed how a preconceived and stereotyped view of the gender system underlies the bulk of the work. In a quarter of the literature references, this view is not expressly stated but implicitly shapes all demographic research from the type of data to be collected through to the choice of methods to be used, and the tables or more advanced analyses to be produced. The researcher's—even subconscious—assumption that the prevailing gender system model is the standard male worker/breadwinner with female housewife/homemaker-cum-child carer, will inform the type of data collected and analyses of it. For example, the fact that the information we have on economic activities is on formal for-gain work, and not on informal work or housework, although major contributors to the economy are normally assigned to women, is a direct product of the researchers' view of society and not mere chance or data-gathering issues.

2. Official Sources

The general population and household census is a fundamental source of information on total male and female populations in terms of many characteristics. Family surveys are good sources of national data, at least, and may cover the widest range of topics. For countries that systematically record vital events, these records are also a mine of information: many of the UN's social indicators are derived from them. Other statistics (e.g., tax, education, work, health, crime) can also be used. One problem with official statistics, however, is that women's status is a fairly low priority in the topics covered, and so often no useful information is collected on it. The way in which the collection of data on economic activity is informed by gender role stereotypes has already been mentioned. Moreover, in many cases, published data are not sex-disaggregated, which renders any gender analysis impossible.

Contextual analysis of a population's living conditions requires more than just collecting quantitative data. Defining a community's gender system means analyzing all the prevailing norms and values that may influence sex-specific roles, power, and value. That means referring to sources of information not normally used by demographers and bringing to bear legal, ethnological, and anthropological skills, in other words a firmly interdisciplinary effort.

3. Time Use Statistics

Time use statistics greatly inform the analysis of gender differences: a day comprises 24 hours, for men and women alike, and the use made of that time can provide much objective information about men and women's lifestyles, constraints, and preferences. Time use also clearly reveals the wide range of activities undertaken, especially all those carried out by women for the informal or family production of goods and services that do not enter into official statistics. The main issues with such statistics are the reliability of recording (daily time use logs are often vague), the categories of activity defined, the choice of reference period (risks of seasonal effects), time- and space-dependent comparability, and the cost of the surveys.

4. Wistat, a Database on Women

The year 2000 saw the publication of the fourth edition of *Wistat, Women's Indicators and Statistics Database* drawn up by the UN in conjunction with several of its specialized agencies.[1] This database provides a range of sex-specific or women-related information on specifically female matters (e.g., fertility) or other issues relevant to women's lives (e.g., childcare provision, environmental conditions that impact women's lives, like health service provision, availability of drinking water). The thinking is that women's status and the gender system cannot be studied separately from development. The database contains, for example, information on sex ratio at birth and at different ages; marital status; percentages of women ever married and never married, singulate mean age at marriage, and percentage ever married among persons aged 15 to 19. It also contains information about female household headship, contraceptive and unmet contraceptive demand; percentage of women requiring more than an hour travel time to the nearest facility providing family planning services or supplies; circumstances under which abortion is permitted; women and men's expressed ideal number of children for total, urban, and rural areas; age-specific fertility; and female population by age and number of children born alive. Other data include urban or rural residence, employment, foreign-born population, availability of safe drinking water and health services, life expectancy, infant mortality, child malnutrition indicators, smoking prevalence among persons 15 years of age and over, causes of death, HIV and AIDS, antenatal care, percentage of infants breastfed, percentage of births delivered by caesarian section, maternal mortality, learning and education, percentages of women teachers by level, women in international cooperation, youth work, commercial loans for women, unemployment, wages, maternity leave, years in which women acquired the right to vote and to stand for election, numbers of women ministers or deputy ministers, women in UN staff in posts subject to geographical distribution, date of ratification and entry into force of Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), prevalence of violence against women by an intimate partner, and deaths from external causes. The geographical and time coverage of these indicators is improving year by year.

[1] Wistat, http://unstats.un.org/unsd/demographic/gender/wistat/index.htm

5. How to Measure the Gender System?

The first requirement in studying the gender-demographic behavior nexus is to translate the gender system concept into indicators. This is no easy matter, because it is a multidimensional concept. Gender stratification aspects (power, wealth, and prestige) have been operationalized by measuring women's situation relative to that of men: this has, for example, included evaluating autonomy (Dyson and Moore, 1983), power and control of material resources (Cain *et al.*, 1979), and freedom of movement (Morgan and Niraula, 1994). Power has been evaluated through six specific components: (1) sense of self and vision of a future, (2) mobility and visibility, (3) economic security, (4) status and decision-making power within the household, (5) ability to interact effectively in the public sphere, and (6) participation in non-family groups. The information was gleaned from personal interviews with different questions on each component (Schuler and Hashemi, 1994). Richard Udry (1994) used 19 measures of gender behavior, including marriage with a man, number of live-born children, a sexual orientation index, importance of career, importance of children, domestic burden sharing, a scale of typical sex-specific activity, importance of marriage, an index of femininity, the list of main interests, pleasure taken in child care, percentage of women in employment, a socioeconomic index, percentage of women in the work unit, and personality type. The International Labor Office devised a series of recommendations for the collection of data on economic conditions to highlight women's contribution to the economy and the obstacles to their integration in development: employment, unemployment, income, sex segregation, gender differences in emigration, time use, access to credit, housing, physical mobility; support networks, and access to training (Anker, 1980). Christine Oppong and Katharine Abu (1986) proposed indicators at the community and individual level for each of the seven roles played by women (occupational, conjugal, domestic, maternal, community, kin, and individual), relating to activity, time use, knowledge and skill acquisition, resource use, power, and decision-making ability. The indicators are equally applicable to men and women and support an analysis of gender differences.

Many of the individual indicators referred to are drawn from face-to-face surveys or participant observations. Information of this type tends not to be extracted from WFS, DHS,[2] or FFS-type demographic surveys or official statistics, where indicators usable for an evaluation of gender systems are few and limited. For this reason, the demographic literature often uses information on education and learning, employment, and age at marriage, as *proxies* for gender stratification indicators (Sathar *et al*., 1988; Jejeeboy, 1991; Chowdhury and Trovato, 1994). Demographers have recently begun to assess the validity of these proxy measures. It had been found that in some cases, the proxies do not give a proper representation of gender stratification: in Islamic society, for example, better-educated women may be confined to *purdah*,[3] that is, have severely constrained mobility, whereas less educated women are forced to try and earn a living in the most difficult circumstances and so become more independent decision-makers. However, the fact of working does not necessarily imply that a woman has greater power and autonomy if, for example, she cannot freely use her earnings or they are earned to be spent on necessities for the family.

6. Synthetic Macro Indicators Sensitive to Gender Inequalities

The *Human Development Report* (1995) uses two synthetic measures of gender inequality, the Gender-related Development Index, a multiattribute measure of longevity, knowledge, and access to resources (life expectancy, education, and income) expressed as an equally distributed equivalent percentage for female and male shares, and the Gender Empowerment Measure, developed out of three variables for women's participation in political decision making, access to occupational opportunities, and earning power (female share of parliamentary seats; share of women as senior officials, managers, and professional and technical positions; share of female nonagricultural wages compared with that of men). These indicators have a limited value, because they make no allowance for many aspects of gender inequality, such as within the family or in community life, and do not lend themselves to an evaluation of the situation in rural areas; nevertheless, they are useful for an overall evaluation of the gender system at the macro level.

Pavel Sicherl, in a 1989 book, suggested simple and effective methods for evaluating gender disparities by using aggregate statistical data, providing worked examples for four fields: education, employment, wages, and time use (Sicherl, 1989). He stressed the importance of taking into consideration the absolute values for men and women, the relative values of gender disparities, as well as both static and dynamic measures. He introduced the time-distance index, that is, the number of years necessary for women to catch up with a given level reached by men (using historical data or projections), as a dynamic measure of gender differentials. He nevertheless cautioned that the male position should not always be regarded as an objective to be achieved, because value judgments change over time with the importance attached to different aspects of the quality-of-life evaluation. His proposed overall measure of gender disparities was a weighted combination of static and dynamic measures. He counseled that the analysis should always be set within the broader reference frame of development, because although development will not automatically eliminate gender inequalities, neither is the simple elimination of gender differences sufficient to improve women's standard of living and eradicate poverty.

7. Plan of Research

Karen Oppenheim Mason (1995) posited that the ideal research plan for analyzing the gender system-demographic behavior relationship should combine multivariate with multilevel analysis that embraces both the macro- and microlevels (collective factors related to development, culture, norms, and institutions with individual social condition, attitudinal, and behavioral factors), within a cohort approach. It should encompass external determinants, intervening

[2] Albeit the Demographic and Health Surveys done in Egypt in 1995–1996 and Haiti in 2000 contained a long series of questions on women's status, which yielded many indicators of the gender system and women's empowerment.

[3] Which forbids women to leave the house at all or only to a very limited extent.

variables, and individual characteristics and should cover successive cohorts. The analysis should bear on four levels at which the gender system acts: (1) through community norms and values, (2) labor market characteristics, (3) the structure and characteristics of the family, and (4) individual characteristics and attitudes (Sen and Batliwala, 2000).

The longitudinal approach is not really workable at the microlevel today, because it is difficult to obtain data on gender systems at the individual level owing to recall problems and recording distortions. Nor is there any easy way to collect long historical and geographical series, even at the aggregated level. Prospective research can be started now but will only produce results some years hence. Aggregated data alone could be more easily used, but only by foregoing any study of the linkages between gender systems and demographic behaviors at the individual level.

Period fluctuation analysis allows intercultural variations to be substituted for intertemporal variations and can include a multilevel strategy by incorporating both individual information and community data. One problem with intercultural studies, however, is that the meanings of questions may often be determined by context. But they are always useful and particularly advisable whenever a comparison of different realities can give insights into which factors are specific to a given culture and which have a more general relevance and influence.

8. Methods of Qualitative Research and Experience of Development Projects

Traditional research is often time-consuming and unsuited to the rapid evaluation of gender systems, and their interactions with demographic behaviors usually seen over 10-year periods. How can we prevent development projects from repeating the disastrous historical mistakes that flowed from having ignored this dimension? Methods have been developed for a rapid evaluation of gender systems and the potential impacts of targeted programs. Such studies aim to gain insights into gender relations that are object to the project, understand gender-differential perceptions of the proposed activities, evaluate whether the institutions involved are capable of addressing the gender issues implicit in the project, and evaluate the project's potential gender impact on various categories of people. They are short-run studies (a few months) using a range of qualitative information and data collection techniques; are highly practice-oriented; are based on interviews with women, men, and the organizations involved; and use lists of criteria such as the Organization for Economic Cooperation and Development's Development Assistance Committee/Women in Development (1992), the *Checklist for Development Projects* (Pezzullo, 1982), the Key Questions from Chapter 3 of the *Handbook for Social/Gender Analysis* (Moffat *et al*., 1991), or the *Harvard Analytical Framework* (Overholt *et al*., 1985). The gender system is analyzed chiefly by observing the division of labor, access to and control of resources and services, participation in decision-making and organizational opportunities, women's sense of self and self-perception, and men's and women's opinions and expectations of the project (Gianotten *et al*., 1994). These same research strategies can be applied to a gender system analysis of family planning, reproductive health, or any other kind of demographic issue. More generally, social anthropology methods can be used to see through the target population's eyes rather than through the often Western or Westernized filter of the observer. In addition to the *rapid rural appraisal* techniques, these methods include critical analysis of the historic reports and data, key anecdotal evidence, *sondeos* (an interdisciplinary team of men and women technicians and social science researchers spend several days working in permuted pairs such that each technician eventually works with each researcher, discussing their observations and interviews at the end of the day; this method fairly quickly introduces new individuals in a domain, and creates an environment for highly dynamic, interdisciplinary learning), mapping, and observation. Participatory research methods, in contrast, aim to involve stakeholders in decisions affecting them and in project implementation activities (e.g., family planning or health). They include such methods as group interviews, projective techniques, case studies, guided group discussions, and learning fora. Data collection itself is often not possible with traditional methods owing to poor literacy skills and low educational levels, lack of organizational ability, and cultural isolation. In such cases, group discussions, collection of life histories, projective techniques, and content analyses are often the most appropriate methods (CIDA, 1989).

Many research strategies proposed for genderized development projects can be useful to gender and population research. For instance, research on the relations between gender systems and population must enlist many different methodological skills from demography, sociology, statistics, social anthropology, etc.

CONCLUSION

Starting with some clarification of the vocabulary used in gender studies, this chapter set out to show what a gender perspective can add to demographic research and, conversely, how demographic analysis can inform gender studies. It was seen that research theories, data and strategies must adapt to this new perspective. It is apparent from this overview that there is an ongoing ferment of development in theory and practice in the field of gender studies, not merely through academic teaching and research but also by the needs of economic development and family planning programs. Gender studies in demography also have useful operational spin-offs.

But tomorrow's research depends on today's teaching, and it is increasingly essential for gender to be systematically mainstreamed across demography teaching to train researchers and practitioners that are sensitive to these issues and capable of reflecting them in their work. That means making a specific effort not only to devise new programs but also to identify the methods most apt to create this new sensitivity. Appendix 93–2 reviews the learning methods that could be explored in this field in the years and decades to come.

Acknowledgment

This chapter was translated from the French version by Glenn D. Robertson.

References

Andro Armelle and Hertrich Véronique, 2001. La demande contraceptive au Sahel: les attentes des hommes se rapprochent-elles de celles de leurs épouses? *Population*, vol. 56(5), p. 721–772.

Anker Richard, 1980. *Research on women's roles and demographic change: survey questionnaires for households, women, men and communities with background explanations.* Geneva, ILO, 314 p.

Arber Sara and Ginn Jay, 1995. *Connecting gender and ageing. A sociological approach.* Buckingham, Open University Press, 212 p.

Balk Deborah, 1994. Individual and community aspects of women's status and fertility in rural Bangladesh, *Population Studies*, vol. 45(1), p. 21–45.

Basu Alaka Malwade and Basu Kaushik, 1991. Women's economic roles and child survival: the case of India, *Health Transition Review*, vol. 1(1), p. 83–103.

Beets Gijs, te Velde Egbert, Verloove-Vanhorick Pauline, Merkus Hans and Bruinse Hein, 1994. Medical complications of aging fertility, *in*: Gijs Beets, Hans van den Brekel, Robert Cliquet, Gilbert Dooghe, and Jenny de Jong Gierveld (eds.), *Population and family in the low countries 1993. Late fertility and other current issues*, p. 3–23 Amsterdam, Lisse, Swets & Zeitlinger, 246 p.

Bem Sandra L., 1987. Gender schema theory and its implications for child development: raising gender-aschematic children in a gender-schematic society, *in*: Mary R. Walsh (ed.), *The Psychology of women: ongoing debate*, p. 206–225. New Haven, Yale University Press, 484 p.

Blake Judith, 1965. Demographic science and the redirection of population policy, *Journal of Chronic Diseases*, vol. 18, p. 1181–1200.

Cain Mead, 1993. Patriarchal structure and demographic change, *in*: Nora Federici, Karen Oppenheim Mason, and Solvi Sogner (eds.), *Women's position and demographic change*, p. 43–60. Oxford, Clarendon Press, 368 p.

Cain Mead, Khanam Syeda R. and Nahar Shansun, 1979. Class, patriarchy and women's work in Bangladesh, *Population and Development Review*, vol. 5(3), p. 405–438.

Caldwell John C., 1986. Routes to low mortality in poor countries, *Population and Development Review*, vol. 12(2), p. 171–220.

Canadian International Development Agency [CIDA], 1989. *A handbook for social/gender analysis.* Ottawa, Coady International Institute, Antigonish and Social Science and Human Resources Development Division of CIDA, 99 p.

Cesaroni Giulia and Pinnelli Antonella, 1999. L'allattamento materno, *in*: Paolo De Sandre, Antonella Pinnelli, and Antonio Santini (eds.), *Nuzialità e fecondità in trasformazione: percorsi e fattori del cambiamento*, p. 497–507. Bologne, Il Mulino, 837 p.

Chen Lincoln C., Huq Embadul and D'Souza Stan, 1981. Sex bias in the allocation of food and health care in rural Bangladesh, *Population and Development Review*, vol. 7(1), p. 55–70.

Chowdhury Fakhrul I. and Trovato Frank, 1994. The role and status of women and timing of marriage in five Asian countries, *Journal of Comparative Family Studies*, vol. 25(2), p. 143–159.

Cook Rebecca, 1999. *The human rights dimensions of maternal mortality.* Liège, IUSSP, 31 p.

Dalla Zuanna Gianpiero, 1996. Paesi ad antica transizione, *in*: Gianpiero Dalla Zuanna (ed.), *Contraccezione e aborto alle soglie del 2000. Paesi ricchi e paesi poveri a confronto*, p. 155–188. Rome, Dipartimento di Scienze Demografiche, Università di Roma "La Sapienza", 225 p.

Dalla Zuanna Gianpiero, 2000. *Material and method in gender and population research.* Liège, IUSSP, 38 p.

Das Gupta Monica, 1987. Selective discrimination against female children in rural Punjab, India, *Population and Development Review*, vol. 13(1), p. 77–100.

DeGraff Deborah and Anker Richard, 1999. *Gender, labour markets and women's work.* Liège, IUSSP, 41 p.

Di Giulio Paola and Pinnelli Antonella, 2003. Genere e demografia nei paesi sviluppati: evidenze micro e macro, *in*: Antonella Pinnelli, Filomena Racioppi, and Rosella Rettaroli (eds.), *Genere e demografia*, p. 3–35. Bologne, Il Mulino, 564 p.

Dixon Ruth B., 1975. *Women's rights and fertility.* New York, The Population Council, 20 p.

Dixon-Mueller Ruth B., 2000. *Women in the labour market in changing economies: demographic issues.* Liège, IUSSP, 28 p.

Dyson Tim and Moore Mick, 1983. On kinship structure, female autonomy, and demographic behabiour in India, *Population and Development Review*, vol. 9(1), p. 35–60.

Federici Nora, Caselli Graziella, Egidi Viviana and Fong Monica, 1985. The status of women, population and development, *IUSSP Newsletter*, no. 23–24, p. 77–98.

Federici Nora, Oppenheim Mason Karen and Sogner Solvi (eds.), 1993. *Women's position and demographic change.* Oxford, Clarendon Press, 368 p.

Findley Sally E., 1999. *Women on the move: perspectives on gender change in Latin America.* Liège, IUSSP, 48 p.

Findley Sally E. and Williams Linay, 1991. *Women who go and women who stay: reflection on family migration processes in a changing world.* Geneva, ILO, 95 p.

Garcia Brigida (ed.), 2000. *Women, poverty and demographic change.* Oxford, Clarendon Press, 311 p.

Garcia Brigida, Anker Richard and Pinnelli Antonella (eds.), 2003. *Women in the labour market in changing economies: demographic issues.* Oxford, Oxford University Press, 341 p.

Germain Adrienne, 1975. Status and roles of women as factors in fertility behaviour: a policy analysis, *Studies in Family Planning*, vol. 6(7), p. 192–200.

Gianotten Vera, Groverman Verona, Zuidberg Lida and van Walsum Edith, 1994. *Assessing the gender impact of development projects: case studies for Bolivia, Burkina Faso and India*. Amsterdam, Royal Tropical Institute of The Netherlands, 112 p.

Hertrich Véronique and Locoh Thérèse, 1999. *Rapports de genre, formation et dissolution des unions dans les pays en développement*. Liège, IUSSP, 46 p.

Hugo Graeme, 1999. *Gender and migration in Asian countries*. Liège, IUSSP, 45 p.

International Union for the Scientific Study of Population [IUSSP], 1993. *International Population Conference, Montreal 1993. Proceedings/Congrès international de la population, Montréal 1993. Proceedings*. Liège, IUSSP/UIESP, 4 vol., 2033 p.

International Union for the Scientific Study of Population [IUSSP], 1997. *International Population Conference, Beijing 1997. Proceedings/Congrès international de la population, Beijing 1997. Proceedings*. Liège, IUSSP/UIESP, 3 vol., 1532 p.

International Union for the Scientific Study of Population [IUSSP], 1998. *Seminar on men, family formation and reproduction, Buenos Aires, Argentina, 13–15 May 1998*. Liège and Buenos Aires, IUSSP Committee Gender and Population and Centro de Estudios de Poblacion, mult. pages.

International Union for the Scientific Study of Population [IUSSP], 2001. *IUSSP contributions to gender research*. Liège, IUSSP/UIESP, 187 p.

Jejeebhoy Shireen J., 1991. Women's status and fertility: successive cross sectional evidence from Tamil Nadu, India, 1970–1980, *Studies in Family Planning*, vol. 22(4), p. 217–231.

Jejeebhoy Shireen J., 1995. *Women's education, autonomy, and reproductive behaviour. Experience from developing countries*. Oxford, Clarendon Press, 306 p.

Leridon Henri, 1992. Sterility and subfecundity: From silence to impatience? *Population, An English Selection*, vol. 4(1), p. 35–54.

Locoh Thérèse, 2002. *Genre et population: quelles activités à promouvoir à l'UIESP?* Paris, INED, 21 p.

Makinwa-Adebusoye Paulina K. and Afolayan Adejumoke Alice, 1995. Migration and women's status in sub-Saharan Africa, *in*: Paulina Makinwa-Adebusoye and An-Magritt Jensen (eds.), *Women's position and demographic change in Sub-Saharan Africa*, p. 253–269. Liège, UIESP/IUSSP, 448 p.

Makinwa-Adebusoye Paulina and Jensen An-Magritt (eds.), 1995. *Women's position and demographic change in Sub-Saharan Africa*. Liège, IUSSP, 448 p.

Malhotra Anju, Vanneman Reeve and Kishor Sunita, 1995. Fertility, dimensions of patriarchy, and development in India, *Population and Development Review*, vol. 21(2), p. 281–305.

Meslé France, Shkolnikov Vladimir, Hertrich Véronique and Vallin Jacques, 1996. *Tendances récentes de la mortalité par cause en Russie, 1965–1994*. Paris, INED, 140 p. + 2 disks.

Meslé France and Vallin Jacques, 1998. Évolution et variations géographiques de la surmortalité masculine: du paradoxe français à la logique russe, *Population*, vol. 53(4–5), p. 1079–1102.

Meslé France, Vallin Jacques and Shkolnikov Vladimir, 1998. Reversal of mortality decline: the case of contemporary Russia, *World Health Statistics Quarterly/Rapport trimestriel de statistiques sanitaires*, vol. 51(2–4), p. 191–206.

Moffat Linda, Geadah Yolande and Stuart Rieky, 1991. *Two halves make a whole: balancing gender relations in development / Un autre genre de développement: un guide pratique sur les rapports hommes-femmes dans le développement*. Ottawa, Canadian Council for International Cooperation (CCIC), Association québécoise des organismes de coopération internationale (AQOCI) and MATCH International Center, 178 p.

Money John and Ehrhardt Anke A., 1972. *Man and woman, boy and girl: the differentiation and dimorphism of gender identity from conception to maturity*. Baltimore, Johns Hopkins University Press, 311 p.

Morgan S. Philip and Niraula Bhanu B., 1994. Gender inequality and fertility in two Nepali villages, *Population and Development Review*, vol. 21(4), p. 541–561.

Mosher William D. and Pratt William F., 1993. The demography of infertility in the United States, *in*: Richard H. Asch and John W. Studd (eds.), *Annual Progress in Reproductive Medicine 1993*, p. 37–43. Pearl River, Parthenon Publishing Group, 366 p.

Necchi, 1999. *Men, family formation and reproduction*. Liège, IUSSP, 30 p.

Oppenheim Mason Karen, 1985. *The status of women: a review of its relationships to fertility and mortality*. New York, The Rockefeller Foundation, 86 p.

Oppenheim Mason Karen, 1995. *Gender and demographic change: what do we know?* Liège, IUSSP, 32 p.

Oppenheim Mason Karen and Jensen An-Magritt. (eds.), 1995. *Gender and family change in industrialized countries*. Oxford, Clarendon Press, 329 p.

Oppong Christine and Abu Katharine, 1986. *A handbook for data collection and analysis on seven roles and statuses of women*. Geneva, ILO/OIT, 134 p.

Organization for Economic Cooperation and Development [OECD], 1992. *DAC principles for effective aid: development assistance manual*. Washington, OECD Publications and Information Centre et OCDE, 142 p.

Overholt Catherine, Anderson Mary, Cloud Kathleen and Austin James E. (eds.), 1985. *Gender roles in development projects: a case book*. West Hartford, Kumarian Press, 326 p.

Pezzullo Caroline, 1982. *Women and development: guidelines for programme and project planning*. Santiago, United Nations, CEPAL, 123 p.

Pierce Catherine, 1993. *Women's organizations/interventions for women: their impact on fertility and child mortality*. Liège, IUSSP.

Pinnelli Antonella, 1995. Women's condition, low fertility and emerging union patterns in Europe, *in*: Karen Oppenheim Mason and Ann-Magritt M. Jensen (eds.), *Gender and family change in industrialized countries*, p. 82–104. Oxford, Clarendon Press, 329 p.

Pinnelli Antonella, 1997. *Teaching gender and population*. Rome, DSD, 27 p.

Pinnelli Antonella, 1999. *Gender and the family in developed countries*. Liège, IUSSP, 52 p.

Pinnelli Antonella, 2001. Determinants of fertility in Europe: new family forms, context and individual characteristics, *in*: Antonella Pinnelli, Hans Joachim Hoffmann-Nowotny, and Beat Fux (eds.), *Fertility and new types of household and family formation in Europe*, p. 47–184. Strasbourg, Publications of the Council of Europe, 184 p.

Pinnelli Antonella and Ajello Anna-Maria, 1997. Genere e popolazione: idee per un curricolo, *Induzioni*, 14, p. 149–159.

Pinnelli Antonella and Masocco Maria, 1995. La diseguaglianza della nascita e dell'infanzia, *in*: Carlo Corchia, Dante Baronciani, and Vittorio Getti (eds.), *Epidemiologia della diseguaglianza nell' infanzia*, p. 55–80. Rome, Fondazione Smith & Kline and Istituto Italiano di Medicina Sociale, 346 p.

Pinnelli Antonella, Racioppi Filomena and Rettaroli Rosella (eds.), 2003. *Genere e demografia nei paesi sviluppati*. Bologne, Il Mulino, 564 p.

Pressat Roland, 1973. Surmortalité biologique et surmortalité sociale, *Revue francaise de Sociologie*, vol. 14, p. 103–110.

PRESSER Harriet, 1997. Fatherhood: recent changes and policy strategies, *in*: IUSSP (ed.), *International population conference, Beijing 1997/Congrès international de la population, Beijing 1997*, p. 829–866, vol. 2. Liège, IUSSP, 1531 p.
PRESSER Harriet B. and SEN Gita (eds.), 2000. *Women's empowerment and demographic processes. Moving beyond Cairo.* Oxford, Oxford University Press, 426 p.
RIDLEY Jeanne Clare, 1968. Demographic change and the roles and status of women, *Annals of the American Academy of Political and Social Science*, vol. 375, p. 15–25.
ROUSSEL Louis, 1989. *La famille incertaine.* Paris, Odile Jacob, 283 p.
SATHAR Zeba, 1993. The processes by which female education affects fertility and child survival, *in*: IUSSP (ed.), *International population conference, Montreal 1993/Congrès international de la population, Montréal 1993*, vol. 4, p. 227–292. Liège, IUSSP, 384 p.
SATHAR Zeba, CROOK Nigel, CALLUM Christine and KAZI Shahnaz, 1988. Women's status and fertility change in Pakistan, *Population and Development Review*, vol. 14(3), p. 415–432.
SCHULER Sidney R. and HASHEMI Syed M., 1994. Credit programs, women's empowerment and contraceptive use in rural Bangladesh, *Studies in Family Planning*, vol. 25(2), p. 65–76.
SEN Gita and BATLIWALA Srilatha, 2000. Empowering women for reproductive rights, *in*: Harriet Presser and Gita Sen (eds.), *Women's empowerment and demographic processes*, p. 15–36. Oxford, Oxford University Press, 426 p.
SICHERL Pavle, 1989. *Measuring disparity between men and women: a technical report.* Santo Domingo (Dominican Rep.), International Research and Training Institute for the Advancement of Women, 178 p.
SHKOLNIKOV Vladimir, MESLÉ France and VALLIN Jacques, 1995a. La crise sanitaire en Russie, I: tendances récentes de l'espérance de vie et des causes de décès de 1970 à 1993, *Population*, vol. 50(4–5), p. 907–943.
SHKOLNIKOV Vladimir, MESLÉ France and VALLIN Jacques, 1995b. La crise sanitaire en Russie, II: évolution des causes de décès: comparaison avec la France et l'Angleterre (1970–1993), *Population*, vol. 50(4–5), p. 945–982.
SHKOLNIKOV Vladimir, MESLÉ France and VALLIN Jacques, 1995c. Health crisis in Russia, I: recent trends in life expectancy and causes of death from 1970 to 1993, *Population, An English Selection*, vol. 8, p. 123–154.
SHKOLNIKOV Vladimir, MESLÉ France and VALLIN Jacques, 1995d. Health crisis in Russia, II: changes in causes of death: a comparison with France and England and Wales (1970 to 1993), *Population, An English Selection*, vol. 8, p. 155–190.
SMYKE Patricia, 1991. *Women and health.* London, Zed Books Ltd, 182 p.
UDRY Richard J., 1994. The nature of gender, *Demography*, vol. 31(4), p. 561–573.
UNITED NATIONS [UN], 1989. *Violence against women in the family.* New York, UN, Center for Social Development and Humanitarian Affairs, 120 p.
UNITED NATIONS [UN], 1992. *Family: forms and functions.* Vienna, UN, 21 p.
UNITED NATIONS [UN], 1993. *Older persons in the family: facets of empowerment.* Vienna, UN, 12 p.
UNITED NATIONS [UN], 1994a. *Programme of Action of the United Nations International Conference on Population and Development.* New York, UN, 100 p.
UNITED NATIONS [UN], 1994b. *A gender perspective on migration and urbanization*, p. 466–470. Vienna, UN.
United Nations Development Program [UNDP], 1995. *Human development report 1995. Gender and human development.* Oxford, Oxford University Press, 145 p.
VALLIN Jacques, 1993. Social change and mortality decline: women's advantage regained or achieved? *in*: Nora Federici, Karen Oppenheim Mason, and Solvi Sogner (eds.), *Women's position and demographic change in the course of development*, p. 190–212. Oxford, Clarendon Press, 368 p.
VALLIN Jacques, 1999. *Mortalité, sexe et genre.* Liège, IUSSP, 50 p.
WALDRON Ingrid, 1976. Why do women live longer than men? *Social Science and Medicine*, vol. 10(7–8), p. 349–362.
WALDRON Ingrid, 1985. Que savons-nous de la différentiation sexuelle dans la mortalité? Aperçu bibliographique, *Bulletin démographique des Nations Unies*, no. 18, p. 67–87.
WATKINS Suzan C., 1993. If all we knew about women was what we read in Demography, what would we know? *Demography*, vol. 30(4), p. 551–577.
World Health Organization [WHO], 1985. *Having a baby in Europe.* Geneva, WHO/OMS, 157 p.

APPENDIX 93–1

The IUSSP and Gender Issues

The IUSSP, the International Union for the Scientific Study of Population, has since the 1980s put a big focus on issues relating to women's status, gender, and their relationships to demographic behaviors. A series of activities on these topics, including publishing a literature review on "The Status of Women, Population and Development" (Federici *et al*., 1985) and holding an international conference on "Women's Position and Demographic Change in the Course of Development" (Asker, 1988), the proceedings of which were published in 1993 edited by Nora Federici, Karen Oppenheim Mason, and Solvi Sogner (1993), culminated in the IUSSP setting up a Scientific Committee on "Gender and Population"[4] in 1990.

Hitherto, the linkages between gender and demographic behavior had gone ignored in many fields of demographic research. The most pressing need was therefore to raise awareness on the issue. The committee first staged an international seminar on "Gender and Family Change in Industrialized Countries" in Rome in 1992 (Oppenheim-Mason and Jensen, 1995), followed by two sessions at the Montreal International Population Conference (IUSSP, 1993), one on "Women's Organizations/Interventions for Women:

[4] Cochaired by Karen Oppenheim Mason and Shireen Jejeebhoy, the committee's members also included Brigida Garcia, Ann-Magritt Jensen, Paulina Makinwa Adebusoye, and Catherine Pierce.

Their Impact on Fertility and Child Mortality" (Pierce, 1993) and the other on "The Process through which Female Education Affects Fertility and Child Survival" (Sathar, 1993), plus two further international seminars on "Women's Position and Demographic Change in Sub-Saharan Africa" (Dakar, 1993: Makinwa-Adebusoye and Jensen, 1995) and "Women, Poverty and Demographic Change" (Oaxaca, 1994: Garcia, 2000); it also contributed to a collective work on "Female Education, Women's Status, and Demographic Change in Developing Countries" edited by Shireen Jejeebhoy (1995).

In 1994, when the life of this first committee expired, the IUSSP set up a new Scientific Committee on "Gender and Population,"[5] which in turn programmed three international seminars on "Female Empowerment and Demographic Processes" (Lund, 1997, Presser and Sen, 2000), "Men, Family Formation, and Reproduction" in Buenos Aires in 1998 (IUSSP, 1998; Necchi, 1999), and "Women and the Labor Market in Changing Economies: Demographic Consequences" in Rome in 1999 (Dixon-Mueller, 2000; Garcia *et al.*, 2003), and organized two sessions at the 1997 Beijing International Population Conference (IUSSP, 1997), one on "Teaching Gender and Population" (Pinnelli, 1997) and the other on "Fatherhood: Recent Exchanges and Policy Strategies" (Presser, 1997); it also published a series of educational papers on "Gender in Population Studies."[6] Other IUSSP gender initiatives are also described in the book, *IUSSP Contributions to Gender Research* (IUSSP, 2001).

All these various activities have helped to define and develop this field of research, devise theoretical frames of reference, systematize empirical knowledge, and lay the foundations for a growing awareness of gender issues among researchers and teachers in demography. So successful did they prove that in 1998, the IUSSP Council queried whether a new committee was needed after the expiry of the second, compared with the benefits of mainstreaming the issue across all the traditional Scientific Committees. The IUSSP therefore tasked an exploratory mission[7] with producing an opinion on the matter. In light of its report, which stressed that although gender studies now enjoyed recognition in the scientific community, much remained to be done to ensure their continued presence on the future demographic research agenda, a new committee was set up in 2002.

[5] Cochaired by Brigida Garcia and Harriet Presser, its other members were Lawrence Adeokun, Richard Anker, Paulina Makinwa Adebusoye, Edith Pantelides, Antonella Pinnelli, Gita Sen, and Ferhunde Ozbay.

[6] A series of eight papers edited by Antonella Pinnelli on the labor market (De Graff and Anker, 1999), migration in Latin America (Findley, 1999) and in Asia (Hugo, 1999), mortality (Vallin, 1999), maternal mortality (Cook, 1999), the family in developing countries (Hertrich and Locoh, 1999) and in developed countries (Pinnelli, 1999), and materials and methods (Dalla Zuanna, 2000).

[7] Chaired by Therese Locoh, and comprising Maria-Eugenia Cosio-Zavala and Sally Findley (Locoh, 2002).

APPENDIX 93–2

Teaching Gender and Population

An International Objective

Raising awareness of gender issues is an internationally recognized objective. The *Program of Action* approved by the United Nations (UN) International Conference on Population and Development gives a central role to relations between gender systems, demographic behavior and sustainable development, in particular devoting a chapter to "Gender Equality, Equity, and Empowerment of Women." The UN's message is that starting or completing demographic transition and achieving sustainable economic development that is compatible with the survival of the planet are not possible without changes (among other things) to the gender division of roles and power relationships between men and women. Accordingly, the UN Program of Action places great importance on training. Chapter 11 on "Population, Development and Education" sets as an educational objective the promotion of a "greater responsibility and awareness on . . . gender equity."

Trainers

Who should provide demographer training to promote increased awareness of gender issues? Demographers are concerned with analyzing the determinants and consequences of demographic behaviors, and the gender system must be seen as both a determinant and a consequence of demographic behavior. This subject has only recently come onto the international scientific and political agenda, from its general absence from demographic training programs, which at best address the relations between women's status and demographic trends in very general terms.

To make it female-delivered provision (as was frequently done in research on relations between demographic behavior and women's status) not only risks ghetto-ization, but by no means ensures that the issues will be properly addressed, because it cannot be expected that female academics, whose gender experiences tend to differ widely from those of most other women, will have an awareness of the gender system that exists in actuality.

Objective

Genderizing demographic training is an immensely wide-ranging objective: the United Nations Program of Action talks of information on population and related issues being directed toward elected representatives at all levels: the scientific community; religious, political, traditional, and community leaders; nongovernmental organizations; parents' associations; social workers; women's groups; the private sector; qualified communication specialists and others in influential positions; age-appropriate education for young people, beginning in the home and continuing through all channels of formal and nonformal education; teachers; and all others involved in the planning of education and communication programs. The urgent need to address the demographic problem in developing countries is the reason for the importance attached to information and education, especially in these countries.

But, a culture of sustainable development and quality of life must also be spread in developed countries, and gender equity and population dynamics are a part of that. A first aim for a developed country could be academic training in the teaching of demography up to doctorate level and continuing training in statistics or demographic research bodies to increase *gender awareness* among statisticians and researchers, as well as in public enterprises, for public and private decision makers.

Examples of promoting continuing training have already been given, in the economic field, for appropriate reporting of women's economic participation both in the (formal and informal) market and out of the market, as well as in urban planning, to factor women's needs into community development (personal safety, as well as of children and older people, transport, socialization, etc.) in a way compatible with broader family needs. The same argument could be extended to other areas (e.g., work environment, services).

Education, both in the family and in the classroom, should remain a main focus of continuing training, because it fosters the gender socialization of the population. Women's enrolment and labor force participation are rising, whereas their socialization to domestic and family care roles is much less great. But men have become more socialized to these roles. The outcome will be dwindling numbers of adult men and women who have been educated to assume what were previously women's duties, with the foreseeable impact on child physical and emotional development, as well as the well-being of other family members, at an as-yet-unevaluated cost to society. Government education, health, and social affairs departments should put these changes at the top of their agendas and take the necessary measures to avoid a decline in personal and family life.

Contents

The topics that should feature in university-level "gender and population" training programs could broadly include:

Definitions and data at the macro and micro levels—*feminist studies*, (studies on women's status, for which there is already a long-established tradition) are often confused with *gender studies*, (i.e., the differences between men and women and the components of a gender system); the data available at the collective and individual levels determine the possible types of knowledge.

Qualitative methods of information-gathering on gender systems—participant observations, in-depth interviews, focus groups, etc.

Quantitative methods of analysis and synthetic gender system indicators—a strategic multivariate and multilevel analysis demands a wider array of methodological tools. Because a gender system is a multidimensional concept, composite indicators must be developed that account for the main dimensions of the system; that means using factor analysis, using multiple classification methods, and constructing latent variables (all methods that are little-used in many demographic research schools). Multilevel analysis methods address the need to consider several levels of analysis but are only in the earliest testing stages in demographic research.

Legal bases—for gender differential demographic behavior (e.g., family law norms where equal dignity is not granted to both sexes; different legal minimum ages at marriage, which is the first instance of the development of an unequal relationship).

Biological bases—of gender differences in demographic behavior.

Relations between gender systems and demographic phenomena—family; fertility, birth control, and reproductive health; health and survival; aging; and internal and international migrations.

Depending on the educational context in which the gender and population issue is set, the following related topics could be included:

Production and market—adding informal and domestic activities to formal work, looking at work patterns such as part-time and flex-time that are compatible with family care, factoring in men's and women's forms of work organization (individualistic, authoritarian and competitive for men, co-operative, interchangeable, flexible for women), taking account of institutional support to the family, etc.

Development programs—very often discussed only with men, the assumption being that women do not count; wrong decisions may therefore be taken simply by ignoring women's development input.

Environment—in developing countries, women manage the essential environmental resources such as water, wood, and food and spend many hours each day in getting them; they must therefore be involved in decisions concerning the use of these resources because wrong decisions can greatly add to their hardships.

These are some selected aspects of method and content. The latter vary widely with their reference historical, geographical, or social contexts. There is as yet no handbook that addresses all these issues—the series of papers published by the IUSSP (Pinnelli, 1999–2000) are merely a first step in that direction—but there is a wide variety of preparatory texts put out by international agencies and nongoverment organizations that have sought to genderize development programs or programs relating to the big areas of intervention, such as health, the environment, the economy, and agriculture. Many use case studies to illustrate the relations among gender, population, and development, because there is often too little research on which to base generalizations. Accumulated knowledge generally refers to developing countries and is especially valuable in identifying the obstacles to be overcome for a speedy completion of demographic transition.

Learning Strategies

Proposing new educational contents means deciding the most appropriate learning strategy. Teaching gender and population raises the issue of linkages with other subjects, because teachers and students alike have internalized from birth and throughout life ideas and stereotypes about the gender division of roles. If the aim of training is not just to transfer knowledge but also to raise awareness of the interrelationships between gender systems and population and to prompt empowerment to change them, the training method design must foster the challenging of deep-rooted ideas and stereotypes. This makes an educational psychology input vital to the development of discussions on the challenge and reconstruction of pre-existing knowledge.

The awareness of the difficulty that trainers have in this field is already clear to see in the training programs linked to the development projects run by United Nations agencies, in which the approaches are focused through the use of interactive and inclusive training strategies such as those based on case studies, video materials as a basis for discussion, and redesign, and role play.

The issue is couched in different terms in the academic environment, in which it is mainly about defining approaches by which to develop gender awareness among specialists training in population issues. One guideline could be a thorough review of academic curricula to genderize the contents. This would be the straight track to change mentalities and knowledge. But, of course, effective as that may be, there are objective limits to training on topics that are not about simple education-instruction learning but go to deep-rooted cultural issues.

The practical impossibility of this kind of approach (involving a paradigm shift that means overcoming fierce opposition from the considerable interests involved) has led to two other solutions: (1) introducing a specific curriculum comprising the most strongly gender-characterized topics, and (2) seeking out potential linkages between the different academic curricula as regards gender issues to devise systematic interfaces between the different areas of education. A credit-based system could address the evaluation issue.

This latter solution, which is a compromise to a complete refashioning of curricula, is arguably an acceptable short and medium term solution, because it would set in motion a valid process of reflection while avoiding ghetto-ization, which is the big risk of the alternative.

One possibility that springs to mind for academic education is to develop *cross-cutting training modules* as the common backbone for the different curricula concerned (history, geography, education theory, sociology, law, demography, economics, etc.), with a more specific focus on gender issues. Introducing elements of this kind could avoid resistance from educators with little knowledge of gender issues, because these modules would be taught by specialists.

Also highly relevant from the teaching viewpoint would be to bring in anthropological-cultural references to other cultures, those apt to promote an awareness that what may sometimes be absolutely unimaginable gender structures are a product of their cultures.

Obviously, from the learning viewpoint, the positive experiences of activities already carried out on the topic in developing countries should be brought in as these—as mentioned earlier—have highlighted the value of educational techniques that maximally involve those to whom they are addressed, specifically to facilitate an awakening to the powerful cultural connotation of gender structures (Pinnelli and Ajello, 1997).

CHAPTER

94

Economic Activity and Demographic Behaviors

GIUSEPPE GESANO

Istituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy

INTRODUCTION

There are multiple relationships between economic activity and demographic behaviors, and these relationships are obviously reciprocal. The importance that most people attach to economic activity puts it at the center of the fundamental choices that heavily influence demographic evolution, especially regarding reproductive behavior and geographic mobility. In the opposite direction, among the different options and strategies in the individual and family life cycle, marriage, procreation, and changes of residence are all decisive stages in the lives of persons and families, and they in turn influence the choices made regarding participation in economic activity, whether it concerns the types of work or the time consecrated to work.

These relationships, which are cardinal at the individual level or at the level of elementary social groups (family, clan, village, etc.), may also be interpreted at the level of populations by a dialectical relationship between demographic trends and the changes in the labor market and in the economic system. In the same way as at the microlevel, these relationships are reciprocal at the macrolevel, with population dynamics influencing the supply of labor, both in quantity and in quality (from a demographic viewpoint), and employment opportunities contributing in the opposite direction to determining demographic events and flows, especially regarding migration.

The principal objective of this chapter is to take account of the complexity of the linkages between demographic dynamics and social and economic conditions. Given the difficulties encountered in the search for cause-and-effect relationships, it is undoubtedly better to conceive the ambivalence and the circularity of linkages between these two components of the human condition, which it would be vain to reduce to simple one-way relationships. In sum, it is perfectly clear that these relationships between economics and demographics are always established at two levels: at the microlevel, for individuals and families, and at the macrolevel, for the aggregates of population and the labor market.

In the next section of this chapter, the relationships between demographic behaviors and economic activity, both at the microlevel and the macrolevel, and the pertinence of discussing them in a treatise on demography will be emphasized. Then several elements that, without being rigorously demographic, are nonetheless necessary to deal correctly with this type of subject, will be provided regarding sources of data on the labor force and methods for analysis of gender differences in economic activity. Finally, given that the effects of women's economic activity on fertility have already been largely considered in other chapters in this treatise (most notably in Chapter 37), the next to last section will focus on the interaction between economic activity and migration, which undoubtedly constitutes the most significant relationship in the area that we are interested in here.

I. IMPORTANCE OF ECONOMIC ACTIVITY IN THE STUDY OF DEMOGRAPHIC BEHAVIORS

The goods and services necessary for the survival and well-being of individuals, families, and collectivities come from work, either of the population to which they belong or of another population with which it is linked. This results from a continuous circular process of transformation of vital energy into work, seeking to produce goods. These goods, in turn, are necessary to reconstitute the expended physical energy and to take care of the entire patrimony of material and nonmaterial elements at the disposition of individuals, families, and collectivities.[1]

It follows from this approach that human activities to be considered as productive are not limited to those covered by classical economic categories and official statistical definitions. We must reintroduce not simply the activities contributing to the reproduction of the labor force (both with regard to its daily restoration and the more formally demographic reproduction of generations) and the activities of general training and technical apprenticeship, to the extent that they produce value added in terms of human capital, both individual and collective.[2]

The relationships between economic activity and population dynamics depend on the level of well-being and the degree of complexity attained by a given society. This concerns the primary relationships between the availability of food products and several demographic variables (in particular, mortality, nuptiality, and fertility) (Livi Bacci, 1987) and can be extended to the hypothesis of a reduction in fertility linked to new patterns of consumption characteristic of the abundant society (Ermisch, 1983). The mobility of the population can also find its motor force in geographic variations in the availability of goods. This latter factor, although it has taken many varied aspects, is nonetheless common to the great historical migrations induced by famines (as well as being still sometimes present, here or there, in the poorest regions of the world) and to that part of current migration from developing countries or from eastern Europe that is based on the illusion of ample availability of goods and services in Western countries.

In addition, economic activity itself constitutes the principal if not the only source of individual and family income, although carrying out this activity implies certain operating costs for everyone and, in many cases, specific investment costs, borne by the worker himself.

Decisions concerning entry to the labor force, the type and conditions of work, and leaving economic activity are taken on the basis of comparisons of costs and benefits, in time and space. Possible changes of residence that may be required by the exercising of the economic activity further complicate the assessment of the explicit and implicit costs and the effective advantages or those expected in the future. Before they are rationalized at the individual level, such choices often result from obligations. In the family context, intentions and individual and collective expectations are heavily influenced by the definition of the roles and functions of different members of the household. Nonetheless, whatever decisions are in fact made regarding participation in the labor market of the potential labor supply of family members, they are made, consciously or not, as a function of a comparison of costs and benefits. This includes, on the one hand, the explicit costs (such as the cost of a baby-sitter who replaces the mother while she is at work) and the opportunity costs (such as forgoing the possibility of improving one's education and consequently improving later one's remuneration level) and, on the other hand, the benefits, not simply the immediate (salary payment) or future (retirement pension) economic benefits but also the hope for a more important role in the family and in society.

In this case also, the linkages with certain demographic aspects are explicit, even if they are less direct. The size and composition of families constitutes a powerful regulator of the labor force participation of individuals who belong to those families, as a result of a complex evaluation of the relative costs and benefits. Moreover, the alternative that women often confront between economic activity and motherhood shows

[1] It is important to note the degree to which this process comes into play in the succession of generations and thus carries, including in its most nonmaterial aspects, a universal dimension. In reality, we benefit today from the fruits of the labor of very distant populations in time or space, including those that preceded us in the past as well as those from the four corners of the planet with which we trade, directly or indirectly. From a sustainable development perspective, we must succeed, by our own work, in bequeathing to future generations a patrimony that is at least equivalent.

[2] "Every action that tends, directly or indirectly, to make possible the satisfaction of individual or collective economic needs, present or future, is an act of production and the person who accomplishes it is economically active" ("*Ogni azione che tenda, direttamente o indirettamente, a rendere possibile il soddisfacimento di bisogni economici individuali o collettivi, presenti o futuri, è un atto di produzione, e chi la compie esercita un'attività economica*") (Mortara, 1960:50). It is worth noting that accepting this argument implies that we must take into consideration the set of domestic activities, especially of women, that statistics and national accounting do not consider as productive.

clearly that this comparison of costs and benefits includes some important noneconomic elements. In a wider framework, it appears that there are substantial consequences, for example, for the level of reproduction of a population, that may flow from these decisions. The choice with regard to mobility, in all of its various forms and types, rests in an even more explicit manner on a comparative evaluation of advantages to be obtained and costs to be borne. This is true as much at the level of individuals as, in a more complex manner, in the family context, with solutions that inevitably tend to optimize the different components.

In decisions pertaining to participation in productive economic activity, a "time balance" comes into play in considering the hours devoted to work and complementary nonwork time. Although time at work is largely imposed from outside the household, its regulation can become an important variable in a family's search for organizational equilibria when the labor market provides opportunities for flexible schedules or part-time activities. It is apparent that this may be mutually correlated with the decision-making spheres regarding the raising of children and family assistance to the ill or aged. It is worth noting as well that certain forms of occupational mobility impinge on time allocation, with more or less impact, between the time devoted to work and time available for rest.

But taking time into consideration also has a broader dimension, extending to the life cycle, during which periods of economic activity and periods of inactivity, either temporary (holidays, illness, maternity leaves, unemployment, etc.) or structural (childhood, schooling and training, retirement), are interspersed. In the daily, weekly, and annual cycle, economic development has substantially reduced the time devoted to productive work. Labor force activity, especially if understood in the strict economic sense, has thus undergone considerable reduction: on the one hand, in favor of periods of training and, on the other hand, in favor of leisure and retirement.

In economically less-developed societies, in the past or even still at present, life and work largely coincide, and age constitutes at most only a weak regulator of the intensity and the quality of productive activity at the two extremes of life. In advanced modern societies, human capital training requirements and the accomplishments of labor unions regarding restrictions on age and labor time increasingly disassociate economic activity and personal life. For quite some time, however, participation of the population in economically productive activities has been conditioned by the different roles given to individuals as a function of their sex, age, marital status, family size, their position within the family, etc.

The relationships between population and economic activity are pertinent both at the microlevel (individual or family) and at the macrolevel. They should be analyzed from the two classical demographic perspectives: cross-sectional and longitudinal.

1. Micro Aspects: Individuals and Families

At the individual level, the study of economically active life in the context of the life cycle is the best approach for identifying the turning points and the factors regulating individual decisions regarding participation in productive activity. In considering the different steps in successive order, the first key transition point is that separating the training phase from the work phase, that is, in modern contexts, the school-to-work transition.[3] The distinction between these two phases has been progressively reinforced to the extent that the objective of the training has gone from a simple apprenticeship for a specific occupation to the constitution of a more or less complete cultural baggage. At the same time, this passage is also taking place later and later in the life cycle, with the extension of compulsory schooling and the widening desire to reach the highest educational levels.

In this domain, the access of young women to higher education is important from a demographic perspective. Not only does this reduce the participation of very young women in economic activity, but it also delays the age at which the woman becomes part of a couple and, quite often, delays the age at which she has the birth of her first child. In societies in which fertility is still high, schooling provides women with the cultural tools for access to models of controlled fertility via the mastery of birth control techniques. In addition, development of schooling may also give a woman access to nonhousehold employment, and hence, the necessary status to discuss, as an equal partner within the couple, the use of contraceptive methods, as well as to be able, if necessary, to live an autonomous life in case of widowhood, divorce, or repudiation.

The prolongation of the schooling phase has, however, also been responsible for the phenomenon of delay in becoming an independent adult, even if this is not the only reason. In this regard, patterns of

[3] The continuous transformation of structures and types of production and the growth of occupational mobility have resulted in a growing frequency of occupational training and retraining during the working life, beyond the initial period of schooling and training. In general, however, these additional training activities do not entail real career interruptions and hence do not play an important demographic role except indirectly, via the increased income or changes in residence that may result from them.

TABLE 94–1 Percentage of European Young Adults Living with their Parents, 1986 and 1994, by Age Group and Sex

Country	Ages 20–24				Ages 25–29			
	Men		Women		Men		Women	
	1986	1994	1986	1994	1986	1994	1986	1994
Central Europe	59.9	61.2	37.9	41.3	23.1	24.7	9.4	11.4
France	56.9	61.8	36.4	41.6	19.3	22.5	8.4	10.3
Germany	64.8	64.6	42.8	44.6	27.4	28.8	11.0	12.7
United Kingdom	57.2	56.8	33.8	37.0	21.9	20.8	8.6	10.8
Southern Europe	87.1	90.9	71.1	81.3	51.3	65.3	28.8	44.3
Spain	88.1	91.5	76.1	84.3	53.2	64.8	35.3	47.6
Greece	76.5	79.3	52.3	62.3	53.8	62.6	23.8	32.1
Italy	87.8	92.2	70.4	82.4	49.6	66.0	25.5	44.1

From Fêrnandez Côrdon, 1997:580.

behavior vary considerably among the advanced societies (Moors and Palomba, 1995), because of both objective reasons and customary and psychological reasons. In the former case, the availability of adequate housing in terms of size and cost and the attainability of jobs accessible to youth and appropriately remunerated are key aspects. With respect to the latter reasons, one relevant aspect, for example, is the willingness of young workers to accept jobs, at least temporarily, for work that does not really correspond to the level of preparation acquired while waiting for eventual access to more appropriate employment.

The demographic and socioeconomic counterparts of this turning point are known: whether they are apparent, such as residential mobility linked to independence, or more indirect, such as the possible beginning of a marriage-like cohabitation. Consideration of these aspects can explain in part the small number of consensual unions among youth in the countries of southern Europe. These youth, in contrast to their northern European counterparts (Table 94–1), tend to stay in their family of origin until they obtain stable employment or get married.[4] Moreover, the idea remains widespread in these societies that economic independence stemming from stable and well-paid employment constitutes an indispensable precondition to the starting of a real family.

In rural societies, in which the extended family prevails and in which work is distributed among all members of the family as a function of their roles and their capabilities, the passage from the training phase to professional life is not as marked, and the creation of a new household depends above all on nonindividual factors such as agreements between families or the prosperity of that particular year. In modern societies with low fertility, in contrast, the reduced number of children per family and the considerable availability of resources that parents can transfer to their children may encourage the formation of new families even in the absence of an inheritance or an independent income that are deemed necessary for the formation of a new family.

It is also important to emphasize the profound modification of relationships between women's economic activity and family formation: initially, in relation to the length of the status of being an autonomous unmarried person and, then, in the relationship between the timing of these two phenomena. Formerly, the low schooling of women and the primacy of their domestic and reproductive functions after marriage resulted in a pattern of female labor force participation rates by age that peaked at a very early age, followed by a very rapid and sharp decline after age 20, closely linked to age at marriage or age at the birth of the first child. The prolongation of the training period and then the effort to conciliate labor force activity with family tasks and responsibilities have modified the pattern of female labor force participation rates, giving a profile much closer to that of men.

The relationships between women's work and their reproductive behavior have been studied a great deal since the 1960s, that is, since the study of fertility differences in the presence of voluntary birth control has highlighted the capital importance of women's level of education and their occupational situation as explanatory variables. In reality, this situation becomes important when economic activity becomes nonagricultural

[4] The phenomenon of adult children who reside with their parents well into adulthood was analyzed from a comparative international perspective (focused on Japan, the United States, France, and Italy) by Hiroyuki Takahashi and Jeanette Voss (2000). It is also worth looking at the special issue of the *Journal of Family Issues* devoted to this subject and edited by Andrew J. Cherlin, Eugenia Scabini, and Giovanna Rossi (1997).

and nondomestic, that is, in an economically developed context, and especially if the society is not organized so as to allow mothers to reconcile economic activity with the demands of motherhood and the raising of children[5] (Killingsworth, 1983; Brinkmann, 1987).

Below we will deal with the relationships between geographic mobility and occupational mobility more thoroughly. However, the disassociation between personal and family life and professional life has consequences, including in terms of the differentiation of the places in which they occur. As a first effect, this implies very short distance mobility, notably in the form of daily commuting between home and work. But more substantial and more definitive moves are also linked to the evolution of one's career or to changes in place of work, enterprise, or occupation, and the beginning and the end of professional life, in particular, can bring about important changes of residence. On one hand, there is migration in search of new employment or at the beginning of one's career, and on the other hand, there are moves related to leaving employment, both to the place of origin (return migration) and to agreeable and warm localities, perhaps even to foreign countries, to spend long years of retirement there.

The relationship between economic activity and demographic behaviors becomes much more complex at the family level because of the juxtaposition and, sometimes, the opposition of different interests of family members. Thus, family choices regarding the work of the two spouses, with the resulting family situation influenced by whether there is one income or two, certainly have a demographic impact. This in turn affects the assessment of costs and benefits, time use, and the mutual division of roles and duties: in societies in which birth control is prevalent, strategies relative to the number and timing of births in the family life cycle derive in large measure from this set of elements.

Even more important are the family strategies pertaining to the duration and type of schooling to provide to each child (or, alternatively, whether the child should enter the labor market early), notably as a function of the child's sex and birth order. Patterns differ considerably according to the culture and inheritance rules. In traditional economies, girls and the youngest children are often sacrificed as part of this choice. In the current developed societies, certain difficulties encountered by youth when entering the labor market can, on the contrary, lead families to try to offer to their reduced number of children the option of continuing their studies, while hardly taking account of either their capabilities and their actual chances of success or of the strategic opportunity of the training that is undertaken.

It is absolutely necessary to take the family into account when considering occupational mobility strategies, taking into consideration the demands and desires of different family members. Choices thus often result from mediation among the interests of the principal actors, and in case of incompatibility, they can lead to conflicts and renunciations, including those pertaining to economic activity, on the part of the weakest members of the family.[6]

However, especially in societies in which unemployment and underemployment are high, temporary labor migration by one or several members of a family may constitute a survival strategy or one of relative enrichment in relation to the initial situation. After the great European emigrations of the first part of the 20th century (see Chapter 58), these strategies have henceforth been largely adopted in Third World countries, both at the level of the extended family or clan, via emigration of their most capable and entrepreneurial members, and at the level of nuclear families, via the temporary migration of the husband or wife, based on the employment opportunities that the host countries may offer (Piore, 1979).

2. Macro Aspects: Cross-Sectional and by Cohorts

The complexity at the microlevel of individual and family relationships between economic activity and demographic aspects is partially simplified at the macrolevel, but there are some specific issues that come into play when one takes a global approach. In particular, it is necessary to analyze the reciprocal influences that exist between, on the one hand, the size

[5] Beyond the existence of an adequate number of daycare centers and nursery schools that are economically accessible, the compatibility of their annual and daily schedules (as well as those of all schools) with work time is capital (Del Boca and Repetto, 2002). However, strong family structures and the proximity of residence of relatives or in-laws and economic inactivity of at least some of them can also facilitate work in the labor market by mothers, even in societies that are not well-equipped and well-organized in this domain.

[6] In the most intense and chaotic phase of urbanization in certain Western societies, abandoning the rural residence and activities may have temporarily penalized woman's explicit economic activity, putting her in the petit bourgeois role that made her the "angel of the hearth." Moreover, the type of activity and work schedules characteristic of early industrialization were hardly compatible with the traditional tasks of a family mother who, among other things, found that urbanization and family nucleation reduced the assistance that, in the countryside, she could receive from a much more extended family network.

and structure of a population and, on the other hand, the labor market and labor migration.

Thus, the vicissitudes that modify the magnitude and the distribution by age and sex of the fraction of the population more interested in carrying out a productive economic activity have effects on the size and on the types of labor supply, which, along with other conditions, appear in the labor market. In addition, precisely because it is a market, the reactions of the market and the autonomous changes in the demand for labor will have, in turn, consequences for employment, for labor supply, and probably—that which is of particular interest here—for certain demographic behaviors such as the formation of new families and choices regarding fertility and migration.

It is apparent that the study of this domain requires the combined expertise of a demographer and a labor economist. The former, among other things, not only provides the basic elements for the static and dynamic analyses of the populations in question but also makes available extremely useful analytical instruments for understanding the phenomena and their relationships. Among these instruments, notably, is the perspective provided by longitudinal analysis, which allows one to follow the continuous process of renewal of cohorts and the linking up of their experiences.

Starting from this postulate, Richard Easterlin (1987) developed, beginning in the 1960s, a hypothesis of differential success of cohorts, based precisely on the relationships that they have with the labor market as a consequence of their relative size. In very summary terms, the members of relatively small cohorts will face less internal competition when they enter the labor market and then during their careers, and they will thus experience greater and more rapid success than do members of relatively large cohorts. This success will allow them to begin their own families sooner and to have more children, thus resulting in larger subsequent cohorts. The inverse results for larger cohorts, and in the long run, this phenomenon should tend to smooth out demographic waves resulting from the alternation of more or less numerous cohorts of parents.[7]

This theory found a rather general verification in the American and north European baby boom that took place after World War II. But the profound restructuring of productive systems that occurred after 1973 undermined one of its implicit hypotheses, which rested on a demand for labor that was at least stable, if not growing. Moreover, the continual and prolonged reduction of European fertility since the middle of the 1960s led to an empirical contradiction of the theoretical result of a new cyclical quantitative equilibrium between the generations of parents and those of children. However, two aspects of the theory remain valid: the generational postulate and the hypothesis that there are reciprocal effects between the labor market and the population. In particular, the sudden structural and cyclical contractions of the demand for labor in the developed countries beginning in the 1960s may very well have been owing in part to the evolution of the level and timing of fertility. Indeed, the difficulties of entry to the world of work have increased for youth, and moreover, those entering the labor market, in their effort to begin their careers, come up against the crowding created by the larger generations that preceded them and that, by virtue of seniority, occupy and will continue to occupy for some time the best jobs in a productive system in which there has sometimes been a slowing of the growth of the demand for labor.

Cross-sectional differences between situations that occur simultaneously in different labor markets often result in labor migration. Demographic disequilibria and differences in the pace of economic development among European countries and, in their midst, among the different regions, effectively created after World War II different situations in their respective labor markets, situations that in numerous cases were brought back into equilibrium by temporary or permanent mobility of workers. At the theoretical level, a purely quantitative first approach to these phenomena has given way to sophisticated considerations concerning the qualitative characteristics of jobs most frequently offered to the immigrant labor force and to the theory of the "dual labor market" one reserved for native workers and the other available to immigrants (Stalker, 1995).

A hypothesis relevant here is that there is an ongoing process whereby employments at the top of the ladder first go to native workers and then to immigrants who came long ago, and this process leaves to recent immigrants the dirty and difficult jobs that are refused by workers already well-positioned in the labor market (Böhning, 1988). However, without fundamentally contradicting this hierarchy of employments, the most recent international migrations have demonstrated the inadequacy of this paradigm of the labor market in explaining the direct flows toward the west from countries of the south of the world and those of eastern Europe. Underlying this phenomenon is a strong and growing pressure of unemployment and underemployment in the countries of origin, as well as a general propensity in the destination countries to exploit immigrant labor, especially when that

[7] For a critical review of the Easterlin hypothesis, see Macunovich (1997).

labor is in the precarious situation of being irregular or clandestine.

II. THE LABOR FORCE: PRINCIPAL SOURCES OF DATA AND SOME ANALYTICAL INSTRUMENTS

The principle of the market on which economic activity is supposed to be based and regulated implies the existence of a demand and a supply, potential or effective. Statistical treatment of these concepts requires, however, precise definitions that vary in time and space as a function of economic structures, systems of production, and behaviors in use by the population.

To define in a universal way *potential supply* confronts the strong dependence that links the labor market to the cultural context, upstream of the social and economic spheres in which they are operating. A definition from the potential population reservoir in which the effective supply is very often formed seems, however, possible: its definitional terms change more slowly as a function of legal norms, usage, and actual behaviors of the population in its relationship to economic activity and with respect to the duration of the life-cycle stages that precede or follow the productive period.

With regard to legal aspects, there are numerous rules that establish the minimum age of entry into different economic activities, forbidding child labor and providing a legal framework for work by minors. For reasons of coherence, the minimum working age often corresponds to the maximum age of compulsory schooling, but special requirements exist for older ages for certain specific occupations. Conversely, there is generally no legal upper age limit to economic activity other than in the inverse form of a minimum age at retirement, which is very variable from one country to another and from one epoch to another. In addition, the minimum retirement age may be a function of the type of employment, and it can be exceeded by workers who wish to remain economically active beyond that age.

Empirically, the recent history of the most economically developed societies shows that there is a progressive raising of the age at entry to the labor force and a reduction in the age at exit. The delay at entry is obviously linked to the diffusion of higher education and its lengthening. The earlier exit is a consequence of both the nearly total coverage of workers by social security systems and the specific demands of restructuring of production systems. Briefly, the crisis of social accounts that is affecting the Western countries will in all likelihood impose drastic changes with regard to the minimum age of retirement and, thus, an important delay in the exit from economic activity by workers.

It is common in international comparisons to treat as the potential supply of labor a *population of working age,* often fixed between ages 15 and 65. For the economically developed countries, it would be more realistic to set the limits around ages 20 and 60, if account is taken of recent behavior of most workers. In the near future, however, it is very likely that the upper limit should be set at least at age 65 in these countries also, as a consequence of the more severe rules that appear to be imminent regarding retirement payments.

For a precise definition of the aggregates that interact in the labor market, texts of economic statistics or publications of the International Labor Organization (ILO) on sources and methods of labor statistics may be consulted.[8] It should not be forgotten, however, that even if the concepts can be clearly defined, it is still difficult to translate them into precise practical parameters, both because of their variability across societies and over time and because it is rarely possible to translate them into objective measurements.

Simplifying greatly, one can define the *(economically) active population* as that which, in a given time period, provides the supply of labor for the production of the goods and services that are taken into consideration in the system of national accounting approved by the United Nations, which includes own consumption of goods and services produced as well as market production of goods and services. A distinction may then be made between the *usually active population,* the population observed as such for a comparatively long time period via a general question on the individuals' usual work status, and the *currently active population,* more frequently referred to as the *labor force,* observed with reference to a short period close to the date of observation.

The *active population,* notably that which forms the *current labor force,* includes the *employed,* who were able to demonstrate the existence, during the reference period, of a relationship with an economic activity (whether or not this activity entails working for someone else, is currently ongoing or temporarily suspended for whatever reason) and the *unemployed (in the broad sense),* that is, those who did not have work during the reference period but were nonetheless

[8] In particular, see the Web site of the International Labor Office in Geneva. The definitions to be found there were adopted at the 13th International Conference on Labor Statistics at Geneva in 1982 and are still in effect today. See the following addresses: http://www.ilo.org/public/english/bureau/stat/res/ecacpop.htm; http://laborsta.ilo.org/appl/data/c3e.html.

available for work and had undertaken some action to find a job.[9] Among these *persons in search of a job,* there is usually a distinction made between *the unemployed in a strict sense,* that is, those who lost a previous job, and *persons seeking a first job.* Finally, persons of labor force age who are not in the labor force are asked if they would be willing to work under certain conditions (regarding place, position, schedule, salary, etc.), in order to identify the entire *potential labor force,* which, in addition to the actual labor force, includes *discouraged workers,* that is, the undeclared supply of labor that remains out of the labor force owing to current market conditions.

These definitions allow one to identify the corresponding categories in population censuses and specific surveys. During the operation of these data collection, the employment status of the interviewee is in general defined with reference to a precise period of time (often the week preceding the survey), close to when the data collection takes place, so as to avoid ambiguities and reporting errors. However, differences in data collection methodology can explain apparent inconsistencies between different sources. Employment survey data gathered by direct interview have rather good reliability. Conversely, censuses, which generally are based on self-administered questionnaires, tend rather to identify the occupational situation judged to be "normal" by the person in question (or the person providing the information for him) and thus runs the risk of providing less reliable information, notably with regard to transitory situations such as unemployment.

In many developed countries, employment surveys are systematically organized at regular intervals. This is notably the case in the European Union countries, for which Eurostat has made a major effort to ensure that statistics produced by different member countries are comparable. In general, these surveys are repeated several times per year (in the European countries, they are often quarterly) in order to take account both of current economic conditions and structural characteristics of the labor market. This entails sample surveys, which are implemented by using partially rotating *panels* to take into account the changes (and hence the flows) between the individuals' economic statuses observed from one survey to the next (or to the survey taking place at the same time the following year, to obtain annual flows).

A partial or complete description of the working life cycle can, however, only be obtained via follow-up analyses. Although there only exist some very partial examples, such analyses could be developed by relying, for example, on the databanks of social security institutions or else via retrospective surveys, with all of the relevant problems of reliability and extrapolation to the universe from which the surveys are drawn. In particular, the French "3B" survey (family biography, migration biography, and economic activity biography) sought to gather retrospectively the relations among events, both individual and family, for the three indicated areas. Unfortunately, difficulties in analyzing the data have limited the results to a few of the very many variables considered (Courgeau and Lelièvre, 1993).

Even if the *absolute quantities* describing the labor market and its evolution (such as the number of unemployed workers) often give an interesting view of economic activity, comparisons in space and time generally require construction of adequate *relative measures.* Moreover, these phenomena vary greatly as a function of the age and sex of those being considered, and because of this, it is necessary to differentiate the intensity measures. Hence, most often relative measures are used.

The *activity rate* (or *labor force participation rate*) is defined as the ratio of the active population to the total population. These rates are then calculated separately for each sex, now limiting the denominator to the population most susceptible to belong to the active population, that is, the *population of working age.* The *unemployment rate* (a measure more economic than demographic) is defined as the ratio of the sum of those seeking a first job plus the unemployed in the strict sense (those who, having lost a previous job, are actively in search of new employment) to the corresponding total number of economically active persons. With regard to employment, two measures may be calculated: either the rate complementary to the unemployment rate or the employment-to-population rate (the ratio of the employed to the total population at the same ages). It is thus important to specify what definition is used.

The profile of rates by age and sex will be described below. We should, however, note here that in addition to the obvious importance of the number of active individuals (employed or unemployed) per 100 residents of the same age, this indicator may also be considered as giving an image of the portion of the year spent by each individual of this age in the corresponding labor force status. In this specific interpretation, the sum of rates by age, cross-sectionally or better yet longitudinally, from the beginning of the working life until its end measures the average duration of the total working life. Table 94–2 presents an international comparison of these durations calculated from the age-

[9] The reference period for such actions is generally much longer than is the reference period for the definition of employment status.

specific activity rates estimated by the ILO (1997). It is interesting to note the differences between men and women, not only in terms of durations but also for the contrast between men who are shortening the duration of their working lives and women who are lengthening theirs. Moreover, the diversity observed across countries is growing over time.

This duration, in years and fractions of a year, can then be compared to the analogous durations spent in other economic condition or outside of the labor market, thus synthesizing the working life cycle in the more general context of the unfolding of life in its entirety, in its different socioeconomic contexts. For example, Fig. 94–1 shows, for each sex, the total durations of different economic statuses, calculated from the age-specific rates recorded in the 2001 Albanian census for those aged 15 to 64. This schema is interesting not only for the substantive results it shows but also as an example of application of a demographic approach to economic questions.

TABLE 94–2 Duration of Working Life by Sex in Some Countries (Total Activity Rate for Ages 15 to 64, in Years)

Country	Men			Women		
	1950	1980	2000	1950	1980	2000
France	46.2	40.5	36.3	20.0	27.1	29.7
Italy	48.1	40.5	38.0	14.3	19.1	23.9
United Kingdom	48.4	44.9	41.3	18.3	28.4	32.6
Sweden	45.7	42.7	41.4	16.8	34.7	40.0
United States	44.8	42.1	39.8	17.6	28.7	33.9
Japan	45.3	42.9	41.9	25.3	25.9	30.3

Calculated by the author using the International Labor Organization estimates (ILO, 1997).

One kind of analysis that is especially useful in macroeconomic studies is that of *stocks* of labor market aggregates and the *flows* into and out of the labor market that change these stocks. Both the basic principle and the principal components of such an analysis have obvious relationships with demography. However, the underlying logic often remains rather simplistic, unable to describe the internal complexity of the processes at work, because the aggregates cannot be adequately refined. Age is the key element that is usually ignored, both in the calculation of stocks and in the calculation of flows. Moreover, the data needed for this type of analysis are not always available (*panel* data are required or, in their absence, retrospective data). This means that for one of the flows (e.g., where the demographic component, which can be taken into account using more reliable data and recommended techniques, has the greatest impact), we

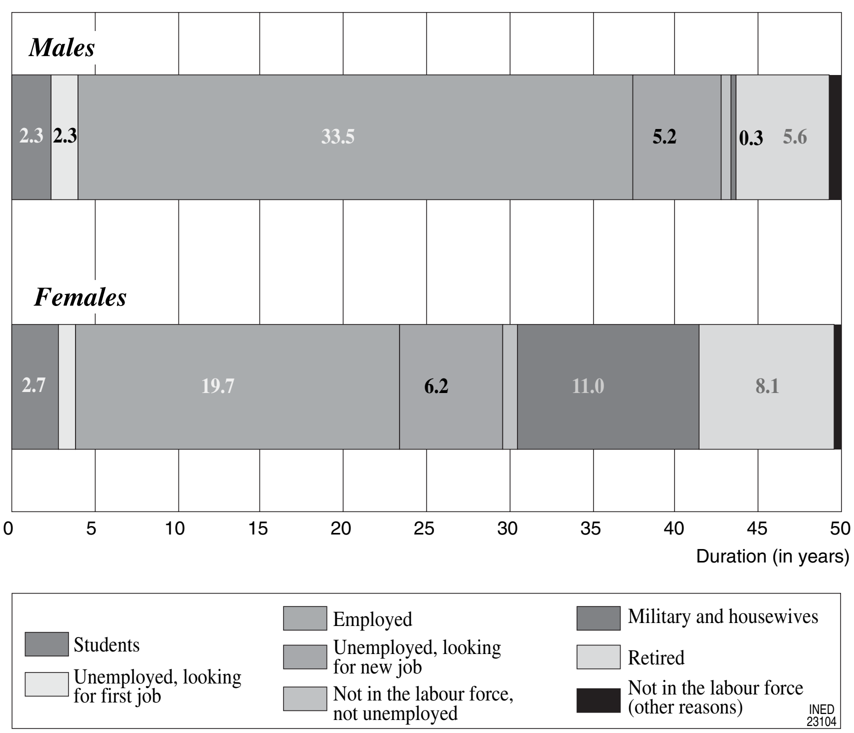


FIGURE 94–1 Total number of years between ages 15 and 64 spent in different economic statuses calculated from the Albanian Census data of 2001.

are left with competing pertinent hypotheses, in order to obtain the other flow as a residual, comparing the stocks over time.

If the necessary data are available, a more rigorous approach would consist of applying the method of multistate life tables to employment situations, thereby allowing one to take account of the articulation between economic activity and demographic parameters (Lalu, 1992). Thus, by use of panel or retrospective survey data, transition probabilities from one economic status to another may be calculated, according to age, sex, and, of course, the initial employment status. The concrete and theoretical problems to resolve when using this multistate life table method are well known: scarcity of the required basic data; arbitrary nature of the estimated probabilities, based on surveys that are generally limited and hence always representative of a specific reality; and hypothesis of stability over time of transition probabilities (even though it is known that the phenomena being studied may vary considerably with economic conditions). In addition, the process is assumed to be a Markov process, with no memory regarding previous moves, whereas it is known that certain states, such as unemployment, for example, tend on the contrary to repeat themselves with greater likelihood among certain individuals and largely bypass others.

Complete retrospective life-history surveys are theoretically much more appropriate for discerning reality in these cases. But these surveys present problems in data analysis and in the construction of indicators, problems that will not be discussed here (see Chapter 23 of Volume I). The advantage of such surveys compared with other sources of data is in their capacity to link events that occur in relation to different aspects of individual and family life. This greatly enriches the analysis. In particular, it becomes theoretically possible to link a person's working life history with his geographic mobility and his fertility and family history, and to identify interesting causal relationships. But this comes at the cost of difficult complications to overcome, as we saw regarding the French 3B survey discussed above (Courgeau and Najim, 1995).[10]

Given what we saw in second section, it should not be surprising that behavior regarding economic activity varies considerably by age and sex. It is also quite evident that the nature of these variations depends largely on productive and economic structures, on rules and norms of social organization, and on the acknowledged and recognized roles of individuals as a function of their sex, their marital status, or their age. It should, however, be made explicit that this argument makes sense only if one considers just the economic activities that are officially recognized and, hence, statistically observed. Otherwise, nearly all of the life cycle can be considered more or less as active for both sexes. In addition, and for the same reasons, it is only in the context of modern economies that activity profiles that are clearly differentiated by age and sex can be readily identified. In this context, an individual's economically active life is more clearly distinct from other roles within the life cycle.

Thus, in particular, males present a typical activity profile by age that has a unimodal curve and that over a large intermediate age range remains very close to 100% of the corresponding population. This central phase is surrounded by initial and final phases of entry and exit, respectively, that are more or less rapid and characterized by portions of logistic curves with two asymptotes, with growth that is initially positive and then negative. Outside of this curve there is, of course, first the training phase and then that of retirement: the rapidity of entry to and exit from work depends on the rules and on the actual behavior that guides these transitions between these two extreme phases and the intermediate phase characterized by a massive presence in the labor market. With the development of higher education and the growth of publicly funded retirement programs, recent changes have generally reduced the age interval during which economic activity rates remain at the highest levels and, on the contrary, have made the phases of entry and exit more extended and flexible.

Women's economic activity profiles were more varied in their form, depending on the social and economic contexts, and they have changed more than have those of men in recent decades. Fig. 94–2, which, by choice, refers to an already dated period (the beginning of the 1980s) so as to illustrate the different patterns of female labor force participation that were apparent then, shows how different the Finnish, Japanese, and Spanish patterns were. Since that time, in the advanced societies this diversity of women's situations has given way to a double homogenization, with female economic activity profiles becoming more alike among different developed countries and also becoming more like the male ones. Although there are more or less marked differences that still exist among women's activity levels, it is also the case that these levels reach more than 80% in certain countries of Northern Europe. In the less advanced societies, women's participation in economic activity showed a peak at young adult ages (especially when schooling

[10] Another example, although limited to aspects of employment status mobility, is given by Miguel Malo and Fernando Muñoz-Bullón (2003), who used data from the British Household Panel.

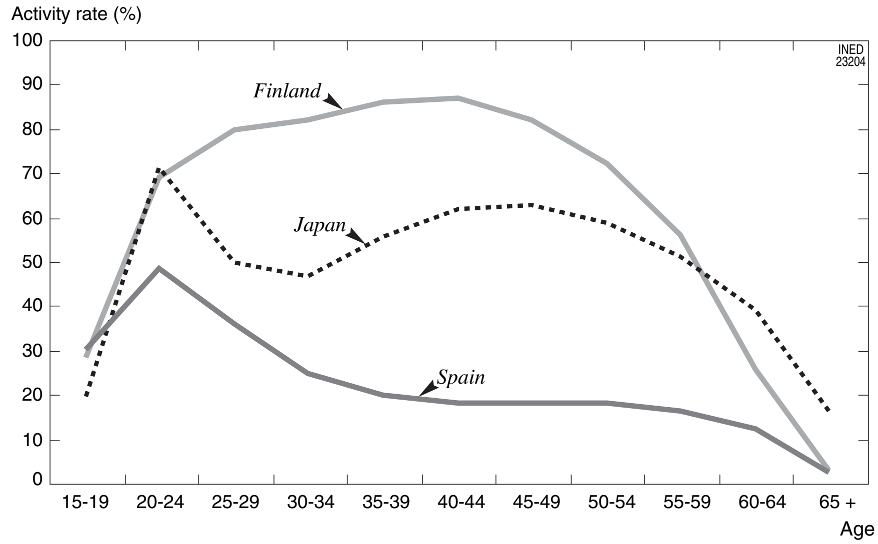


FIGURE 94–2 Women's economic activity rates by age group in three developed countries around 1980.

retarded in extending to these ages) before diminishing, after age 25, at the ages of marriage and childbearing, and as a function of the still quite distinct differentiation between the roles of wives and husbands. Likewise, normative and cultural factors resulted in variations in the right-hand side of the curves, showing women's activity rates depending on whether the exits from the labor force after age 25 were permanent (at least from the official point of view) or temporary, with a return to the labor market once past the life-cycle stage most encumbered by childbearing and the raising of preschool-age children. In the first case, the activity curve stagnated at a very low level up to the age of retirement (as in Spain during the 1980s), whereas in the latter case, a second maximum would be evident between ages 40 and 50 (as in Japan around 1980).

III. ECONOMIC ACTIVITY AND GEOGRAPHIC MOBILITY

In view of studying the numerous relationships between geographic mobility and socioprofessional mobility, the problem is to successfully identify which is the primary phenomenon and which is the induced phenomenon. Only in a few cases is the answer very clear: for example, it is easy to say that forced displacements and movement of refugees are indubitably the primary factor in relation to work changes that may flow from them. Retirement migration is, conversely, the immediate consequence of an important change in one's socioprofessional situation. In other cases, however, one should be able to analyze, case by case, the set of relationships between these two types of mobility.

Nora Federici (1974) made a nice distinction in differentiating between "physiological migration" and "pathological migration," according to whether geographic moves were pursuant to the normal evolution of social and professional life, or changes in socioprofessional status were consequences of a decision to migrate, respectively. Today, in economically advanced societies, it seems that physiological migration has prevailed, with socioprofessional mobility normally being the primary factor in relationship to possible geographic moves. On the contrary, in the relations between developing countries or regions and developed countries or regions, the determining factor regarding geographic mobility tends to be the need for a new equilibrium between population and resources. Thus, especially in periods of reduced labor demand, such movements are above all determined by push factors in the countries of origin, opening the way to typically pathological migration. In these cases, the decision to migrate precedes the possible change in socioeconomic and work status, which is at best but the principal objective of the move and an objective that is often far from being realized.

Table 94–3 shows a typology of the most common forms of geographic mobility as a function of the type of occupational and work mobility. It is ordered on a decreasing scale of importance and complexity of the move undertaken. However, especially for geographic moves, this does not constitute the only possible articulation, as there are numerous criteria that one can think of to classify the different forms of geographic mobility (see Golini, 1987). Likewise, the modalities

TABLE 94–3 Typology of Geographic Mobility as a Function of Occupational and Work Mobility

Occupational and work mobility	Geographic mobility					
	With change of place of residence				Without change of place of residence	
	International	Interregional	Intraregional	Local	Back-and-forth	Nonsystematic
Entry to the labor market for a permanent job	Permanent or long-term migration	Permanent or long-term migration	Permanent or long-term migration	Change of residence pursuant to change of status	Beginning of back-and-forth work movements	Beginning of nonsystematic work mobility
Entry to the labor market for a temporary or seasonal job	Temporary or seasonal migration	Temporary or seasonal migration	[unlikely combination]	[unlikely combination]	Beginning of back-and-forth work movements	Beginning of nonsystematic work mobility
Entry to the labor market for a precarious job	Migration "of fortune" (e.g., irregulars)	Migration "of fortune" (e.g., nomads)	Rural-urban and "fortune" moves	[unlikely combination]	Beginning of back-and-forth work movements	Beginning of nonsystematic work mobility
Permanent exit from the labor market	Different from of permanent return or of new migration	Different forms of permanent return or of new migration	Different forms of permanent return or of new migration	Change of residence pursuant to change of status	End of back-and-forth work movements	End of nonsystematic work mobility
Temporary exit from the labor market (nonemployment, long vacation, etc.)	Temporary or seasonal returns or new migration	Temporary or seasonal returns or new migration	Temporary or seasonal returns or new migration	[unlikely combination]	End of back-and-forth work movements	End of nonsystematic work mobility
Change of sector of economic activity	[rather frequent combinations]	[rather frequent combinations]	Rural-urban moves	Drawing closer of place of residence and new place of work	Change of destination and distance	Change of destination and distance
Change of work situation (employee/ independent)	[likely combination]	[likely combination]	[likely combination]	Drawing closer of place of residence and new place of work	Change of destination and distance	Change of destination and distance
Change of employer	[likely combination]	[likely combination]	[rather frequent combinations]	Drawing closer of place of residence and new place of work	Change of destination and distance	Change of destination and distance
Change of place of work (implantation, office, etc.)	Mobility in international enterprises or firms	Mobility in enterprises with multiple locations	Mobility in enterprises with multiple locations	Drawing closer of place of residence and new place of work	Change of destination and distance	Change of destination and distance
Change of occupational position (career)	[unlikely combination]	[unlikely combination]	Change of residence pursuant to change of status	Change of residence pursuant to change of status	[unlikely combination]	Change of destination and distance

given in each of the cells certainly do not claim to exhaustively describe the very large spectrum of individual situations that may in fact fit together with one or another combination of geographic and professional mobility. The modalities described in the table relate primarily to situations most often prevalent currently in developed economies.

Despite the impossibility of defining in a general way fixed relationships in the complete and logical sequence of linkages between the determining events of geographic and professional mobility, one can try to identify some typical cases, corresponding to mass phenomena and implicating large groups of individuals.

In the first place, the professional life cycle articulates itself in several fundamental steps that often include one form of geographic mobility or another. Thus, it is not only that labor market entry and exit can activate (or deactivate) systematic mobility (commuting and back-and-forth movements) or nonsystematic mobility precisely owing to the change in economic status or in the kind of work, but they create new social relations and new psychological and economic conditions as well as precise requirements for localization of residence that may induce more or less permanent and more or less distant geographic moves.

Only specific surveys can provide pertinent information on this particular linkage between these two types of mobility. The considerable internal migration flows and, in particular, the phenomenon of massive urbanization that, in the west, accompanied and marked the economic transformations of the Industrial Revolution, can, however, indirectly tell us about the amplitude of the link between these two phenomena. Abandoning the sector of activity of the family of origin—especially when it is the agricultural sector, where nearly always the place of residence and place of work are the same—constitutes in these cases the sign of a likely migratory move subsequent to entry to the labor market.[11]

From this perspective, something may have changed or is changing in the most developed countries. It is certainly the case that the great agro-industrial transformation in these countries has already been accomplished, and the next stage, in which the tertiary sector of the economy is dominant, tends to take place in the urban environment without any substantial new population movements. Further, the phenomena of decentralization of production and geographic redistribution of services reduce, if not reverse, the previous trends toward the concentration of workers in a small number of locations. As a consequence, entry into working life in sectors or industries are different from those where the family of origin may already be found, or subsequently, changes in one's sector during the working life less and less imply, by themselves, the need for a change of residence. In addition, the development of means of mass communication and transportation has no doubt been a factor in the sense of reducing permanent moves in favor of temporary and recurring moves.

Developing countries have not yet reached this stage. Not only do they remain characterized by ongoing urbanization, but almost always this urbanization is driven by the push of rural populations toward the cities. In the urban destinations, there is not a real demand for labor corresponding to the level of pressure brought to bear by the in-migrants. It is thus not the case in these countries that entry to the labor market motivates such large migratory moves: people are only prompted to migrate by an often false hope of securing a higher income via obtaining a job and by the illusion of being able, in the urban setting, to have opportunities that do not exist in the countryside. Migration precedes a possible access to economic activity and hence establishment in the urban labor market, but these remain uncertain as a consequence of the limits and precariousness of labor demand. Thus, one can see, from rural to urban areas, *migrations of fortune* (that is, people who migrate as an adventure), a phenomenon currently typical in many Third World countries, even if in the past it also, at least in part, marked the history of the growth of western cities and the depopulation of rural places.

Migrations of fortune have, moreover, always constituted an important share of international mobility, in which it is difficult to have a prior agreement between local employers and workers coming from more or less distant foreign countries. Often, in this case, the worker migrates essentially on the basis of some simple general indications regarding the possibilities of finding some kind of employment in the destination country.[12] Such possibilities may become reality, but only a posteriori, in the meantime exposing the immigrant to the risks of long periods of unemployment and serious problems of getting established in the new society.

[11] Catherine Bonvalet and Éva Lelièvre presented some interesting results on this subject in two articles published in *Population* in 1989 and 1991. Based on sample surveys, the 3B survey already mentioned and the survey *Peuplement et dépeuplement de Paris*, these results confirm the relationships between the socioprofessional mobility of children in relation to their parents and geographic mobility, particularly if the mobility involved Paris.

[12] In these processes an important role is traditionally played by the so-called migration chains that circulate information and descriptions, sometimes illusory, about the destination society in the places of origin, via those who have migrated previously.

When these phenomena occur very frequently and in the absence of policies or rules pertaining to immigration on the part of the receiving country, the end result is illegal immigration. However, whatever the temporal logic between geographic mobility and professional mobility may be, migration between different countries and between continents has lost its character of conquest, and colonization reflected by migration in past centuries and has become today the primary instrument for re-equilibrating labor markets on an international scale. As a consequence, the amplitude of the flows, their direction and even the means of mobility are, more or less explicitly, increasingly dictated by the requirements of labor demand. Repeat migration stems not only from an adjustment of the duration of the displacement to the duration of the demand for labor but also from the rules established by host countries that are opposed to the permanent settlement of foreign workers. Seasonal migration may seem to be outdated, in that it is primarily linked to agricultural work, but today it is re-emerging in relation to the typically seasonal nature of certain commercial activities (peddlers) and services (e.g., in the tourist sector).

Recently, economic, social, and political push factors have become more important in causing international migration flows that often go beyond the requirements and the rules of the destination countries. These countries, in turn, increasingly evaluate the need for accepting the immigrants, both as a function of an overall evaluation of their international relations and as a function of their own internal manpower needs. As a consequence, there is a growing difficulty in clearly distinguishing from among the total international moves those for work reasons and in singling the economic immigrants out from the total number of foreigners who live in a country.

In all of these cases, such an important move, in terms of both sociocultural distance and geographic distance, as it is implied by international migration (and this is also the case, to a slightly lesser degree, for many long-distance internal or inter-regional moves, especially when they involve moves from depressed regions to developing regions) is almost always accompanied by an even more important modification of the nature of the individual's economic activity. Often an entry to the labor market (stable, temporary, or precarious) is observed, accompanied by changes in economic activity (sector or industry, occupation or employment) that can be attributed to new entrance to a labor market that is totally different from that which prevailed at the place of origin.

But if the relationship between work and location is in these cases rather straightforward, strict, and exact, one must also ask what happens, in general, when for whatever reason the work stops. Of course, the reactions will be different according to the nature of the work interruption and the degree to which the migrant has been able to establish himself since immigrating. When the migrant's situation is precarious, a brief period of unemployment is sufficient to provoke a return to the place of origin (returning in failure) or a departure to another destination (stepwise or circular migration). However, systematic interruptions can be foreseen by the nature of the activity or by the legal framework in effect: in this case the result is a return flow of seasonal or temporary and repeat migration.

It is, however, the permanent exit from the world of work that often causes, in this situation, a definitive return move for retirement. At this stage, preference can be given to the linkages and interests maintained at the place of origin, especially in the case where local socioeconomic conditions have in the meantime improved thanks to the emigrants themselves, both by their monetary remittances and by their contribution in specific know-how or simply as part of the general evolution of standards of living.

Along with the natural and foreseeable consequences of retirement for the migration/work tandem, the favorable economic prospects and the longer life expectancy of the newly retired in our advanced societies has given birth to a new kind of migratory move, not yet massive but probably more selective, both in terms of age structure and social group as well as the destinations chosen. A first stage of this phenomenon can be identified in the American Sun-belt or on the Mediterranean coasts and islands, inhabited permanently or used systematically by number of retirees, for a tourism extended to the entire year.[13] Moreover, the process will evolve very rapidly, at least when the degree of aging of the populations of origin will be accentuated, if the favorable economic conditions for the elderly are maintained. In these cases, the magnitude of the move, far from being reduced, will be increasingly motivated by the desire to find a new place to settle as different and distant as possible from the place of residence during the working life, especially when that place is not particularly pleasant. This provokes an inter-regional or even international movement that cannot really be considered as labor migration, but which nevertheless is part of the considerable mobility of the population following changes in the relationships between individuals and the world of work.

[13] For an analysis of the literature on this theme from a European perspective, see Casado-Díaz *et al*. (2002), available at *http://www.shef.ac.uk/sisa/esf/EW_Bibliography.shtml*. One can also refer to King *et al*. (1998), as well as to the complete issue of the *International Journal of Population Geography* in which this article is found, because it is entirely devoted to international retirement migration in Europe, notably that coming from Great Britain.

Also at the boundaries of real migratory movements as such may be found long-distance mobility determined by the work itself or by certain modifications of aspects of its being carried out. At the lowest level, or even in the absence of occupational mobility, specialized labor services can be found which by their nature or the location where the work must take place induce a more or less prolonged and/or systematic dislocation between the places of work and of usual residence, very distant from one another. But stimulus to geographic mobility stems also from the type of economic structure in which the work is accomplished: being part of organizations or multinational corporations entails invitations and obligations to make international moves as a function of changes in one's duties or even transfers to the same job at a site situated in another country. Moreover, the same can be said, all things considered, in the case of belonging to firms or entities organized into decentralized units within a given national territory. In this case, there is a high probability that a work-related move will be positively correlated with the level of professional qualifications attained and will represent in addition an indicator of success for career progress.

Along with these long-distance moves, there is a lesser mobility linked to work and the labor market that develops. This mobility is lesser not in terms of the number of people involved, because at the lowest level it tends to be the normal situation for nearly all work activity. Rather, it is lesser in terms of the distances covered and especially the effects that are felt by the persons and social environments that are involved.

In this perspective, the most important event is the relocation of own residence as a function of the evolution of work. The drawing closer of the residence to the place of work or relocating as part of the process of adapting to a new standard of living pursuant to permanent entry, definitive exit, or promotion to a better and higher-paying job are the most frequent cases. Whether the magnitude of the move is regional or only local becomes less important with the development of means of transportation and communication on a large geographic scale. This in turn results in a disconnection, at least within certain geographic limits, of the choice of a place of residence from the location of the job, and the ultimate choice of residence may in the end result from factors that, in large part, have nothing to do with the place of work.

In such cases, even if the *living space* (*espace de vie*, following a nice definition by Daniel Courgeau, 1970) of the individuals involved only changes a little, we are nonetheless considering a change of residence, an infrequent and important event in the life course of a person or family. But beyond this almost rare mobility, one considers as a widespread and normal condition, owing to its ineluctable and habitual character, all recurring mobility, whether of a systematic or nonsystematic nature, that is closely linked to work and its changes: occupations that by their nature are undertaken in variable places, random mobility stemming from ongoing economic activity, and commuting between home and work.

From the same perspective, it may be said that with the declining role of agriculture or the disappearance of artisan workshops and also with the progressive exemption from the assembly line and from great industrial concentrations, economic activity is increasingly carried out as forms of communication. At the current stage, we are witnessing an exaltation of the physical mobility of persons. Very soon, progress and the diffusion of computerized telecommunications will perhaps allow information flows to increasingly attenuate the direct effects on location and on the geographic mobility of persons, thereby reducing mobility in relation to economic activity.

OVERVIEW AND CONCLUSIONS

This chapter sought to illustrate the complex and multiple relationships that link two essential aspects in the lives of individuals: *demographic behaviors* (as reflected in matrimonial, reproductive, migratory, and residential choices) and *economic activity,* notably *occupational* activity, seen from different angles: (1) means of production of the goods and services necessary for the survival and the well-being of individuals and their families, (2) major source of income, and (3) importance for time allocation (notably for certain stages of the life cycle). Because of the importance that each of these two aspects assumes in the lives of most people, the choices cannot be made on one side without taking the other side into account, so a compromise is struck, as a consequence of the attainments and the opportunity costs in each domain.

The study of the modifications of the life cycle and, under the circumstances, of the working life cycle, can provide interesting keys for understanding the changes that are taking place at present, regarding both the intensity and timing of demographic events and the time allocation and types of participation of individuals in economic activity. The full importance of the cultural and socioeconomic context is thus understood, especially the fundamental role played by the life stages that precede and follow the working life, that is, the phases of training and those which, in the advanced societies with some sort of social security system, coincide with retirement.

If, on the one hand, the family allows for a compensation between the capacities and aspirations of the individuals who are in it, thereby opening the way to more flexible and complex strategies with regard to demographic behaviors and participation in the labor market, the family is also, on the other hand, the place where the choices and the compromises between individual projects or plans are struck. Thus, it is essentially the compensation function that prevails in intergenerational relations, notably in favor of children during the training period (which in general also precedes their exit from the family of origin and formation of a new, autonomous family unit). But it is the dialectical confrontation and the function of interindividual compromise that govern survival strategies of families (and even of clans), notably via migration and residential choices.

Although the contradiction and the arbitrage between demographic behaviors and economic choices are evidently situated at the microlevel of individuals and families, there is also a confrontation between demographics and economic activity at the macrolevel: for a number of different aspects, the structure and the dynamics of the population interact with the state of the labor market and the conditions under which it is functioning. This interaction may be analyzed in terms of cohorts, for example, to inquire into or verify the hypothesis of Richard Easterlin of an influence of the relative size of successive cohorts and the resulting labor market competition on fertility behaviors. That interaction can also be seen cross-sectionally to measure quantitative disequilibria and qualitative disequilibria that demographic dynamics can provoke between supply and demand for labor and migration, which such disequilibria can induce.

The macro perspective also allows for better definition of the aggregates and measures relative to the labor market. In particular, it permits analysis of the different economic behaviors associated with differing economic conditions, as a function of the evolution of the cultural and socioeconomic environment, notably in the domain of the roles borne by women.

Finally, we saw, under numerous aspects, the importance of the relationships that the state of the labor market and its variations maintain with the different forms of geographic mobility of the population. There is a wide variety of relationships via which the evolution of productive structures and migration strategies both bring about substantial changes.

In conclusion, because economic activity and work form a specific part of the life of individuals and families, one which is distinct and circumscribed in time, the study of the relationships between demographic behaviors and economic activity constitutes a fundamental element of any population study seeking to identify different behaviors and to determine their causes. The joint study of demographic dynamics and of the evolution of the labor market is especially important in order to understand their mutual interactions and to foresee possible disequilibria as well as their likely consequences in terms of migration flows.

We may expect that in the future changes in the nature and timing of economic activity may modify, here or there, the nature or intensity of this relationship. However, so long as work remains a vital economic function, it will play a valuable social role and will remain a major means of success for individuals, thereby confirming the key position occupied by the relationships that it maintains with this other aspect of life, even more responsible for social worthiness and individual accomplishment, where demographic behaviors are defined.

Acknowledgment

This chapter was translated from the French version by David Shapiro.

References

Böhning W. Roger, 1984. *Studies in international labor migration.* London, Macmillan, 318 p.

Bonvalet Catherine and Lelièvre Éva, 1989. La mobilité résidentielle en France et à Paris depuis 1945. I. Bilan résidentiel d'une génération, *Population,* vol. 44(3), p. 531–560.

Brinkmann Christan, 1987. *Demographic aspects of the labor force and employment.* Strasbourg, Council of Europe, 126 p.

Casado-Díaz María Angeles, Lundh Ulla, and Warnes Tony (eds.), 2002. *European dimensions of retirement; older migrants in Europe: projects and sources.* Strasbourg, European Science Foundation, 39 p.

Cherlin Andrew J., Scabini Eugenia, and Rossi Giovanna (eds.), 1997. Still in the nest: delayed home-leaving in Europe and the United States, *Journal of Family Issues,* vol. 18(6). (Thematic issue).

Courgeau Daniel, 1970. Les champs migratoires en France (1) et (2), Paris, Presses Universitaires de France, (Cahier de Travaux et documents, no. 58, INED), 158 p.

Courgeau Daniel and Lelièvre Éva, 1993. *Vers une analyse biographique des ménages*. Paris, INED, 17 p.

Courgeau Daniel and Najim Jamal, 1995. Analyse des biographies fragmentaires, *Population,* vol. 50(1), p. 149–168. (Also published in English: Interval-censored event history analysis, *Population, an English Selection,* vol. 8, 1996, p. 191–208.)

Del Boca Daniela and Repetto Alaia Margherita (eds.), 2002. *Women's work, the family and social policy: focus on Italy in a European perspective*. Berne, Peter Lang, 196 p.

Easterlin Richard A., 1987. *Birth and fortune: the impact of numbers on personal welfare*. Chicago, University of Chicago Press, 206 p.

Ermish John, 1983. *The political economy of demographic change.* London, Heinemann, 317 p.

Federici Nora, 1974. Migrazioni fisiologiche e migrazioni patologiche, *in:* Istituto di Statistica Economica (ed.), *Scritti in onore di Guglielmo Tagliacarne,* p. 181–202. Rome, Istituto di Statistica Economica, Università di Roma, 571 p.

Férnandez Córdon Juan Antonio, 1997. Youth residential independence and autonomy: a comparative study, *Journal of Family Issues,* vol. 18, p. 576–607.

Golini Antonio, 1987. Population movements: typology and data collection, trends, policies, *in:* IUSSP/UIESP, EAPS and FINNCO (eds.), *Plenaries of the European population conference 1987: issues and prospects,* p. 113–200. Helsinki, Central Statistical Office of Finland, 434 p.

International Labor Office [ILO], 1997. *Economically active population 1950–2010.* Geneva, ILO.

Killingsworth Mark R., 1983. *Labor supply.* Cambridge, Cambridge University Press, 493 p.

King Russel, Warnes Anthony M., and Williams Allan M., 1998. International retirement migration in Europe, *International Journal of Population Geography,* vol. 4(2), p. 91–111.

Lalu Nirannanilathu M., 1992. Multistate life tables and its application to the analysis of labor force, *in:* Parameswara Krishan, Tuan Chi-Hsien, and Kuttan Mahadevan (eds.), *Readings in population research: policy, methods and perspectives,* p. 145–164. Delhi, B.R. Publishing, 586 p.

Livi Bacci Massimo, 1987. *Popolazione e alimentazione: saggio sulla storia demografica europea.* Bologna, Il Mulino, 173 p.

Macunovich Diane, 1998. Fertility and the Easterlin hypothesis: an assessment of the literature, *Journal of Population Economics,* vol. 11(1), p. 53–111.

Muñoz-Bullón Fernando and Malo Miguel A., 2003. Employment status mobility from a life-cycle perspective: a sequence analysis of work-histories in the BHPS, *Demographic Research,* vol. 9(7), p. 119–161.

Moors Heins G. and Palomba Rossella (eds.), 1995. *Population, family and welfare: a comparative survey of European attitudes.* Oxford, Clarendon Press, 297 p.

Mortara Giorgio, 1960. *Economia della popolazione. Analisi delle relazioni tra fenomeni economici e fenomeni demografici.* Turin, Utet, 514 p.

Piore Michael J., 1979. *Birds of passage. Migrant labor and industrial societies.* Cambridge, Cambridge University Press, 229 p.

Stalker Peter, 1995. *Travailleurs immigrés: étude des migrations internationales de main- d'oeuvre.* Geneva, International Labor Office, 346 p.

Takahashi Hiroyuki and Voss Jeanette, 2000. *"Parasite singles"; a uniquely Japanese phenomenon?* Tokyo, Japanese Economic Institute.

CHAPTER

95

Demography and Ethnicity

An Ambiguous Relationship

JEAN-LOUIS RALLU, VICTOR PICHÉ, AND PATRICK SIMON

Institut national d'études démographiques (INED), Paris, France

Département de démographie, Université de Montréal, Montreal, Canada

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

One of the most striking features of the end of the 20th century was the resurgence of the ethnic question in public debates, both in developing and in developed countries. Between conflicts interpreted from an ethnic perspective (the Balkans, central Africa), nationalist struggles (the Basque country, Quebec), and demands for ethnic identity by new minorities resulting from immigration, every country is currently affected by what is commonly known as cultural pluralism (Dieckhoff, 2000). Similar to all social sciences, demography is affected by the debate and is sometimes even directly involved, if only in its use of official and administrative categories.

Ethnicity is not easy to define. There are several definitions that refer to numerous concepts and theories of ethnicity (Poutgnat and Streiff-Fenart, 1995:93–133). Here we will retain the definition borrowed from Martin Bulmer (1996:35) that an ethnic group is "a collectivity within a larger population having real or putative common ancestry, memories of a shared past, and a cultural focus upon one or more symbolic elements which define the group's identity, such as kinship, religion, language, shared territory, nationality, or physical appearance." It is clear that race, defined by the criterion of "physical appearance," is included in this definition, and this would appear to be increasingly common practice in the field of ethnic relations. The concept of ethnicity emerged after World War II, whereas previously the concept of race was employed to include ethnic group and skin color (Kertzer and Arel, 2002b:12). Even though some authors denounce the confusion of race with ethnicity (see Oomen, 1997:58), we will retain Bulmer's definition as the issues associated with ethnicity (e.g., discrimination) also apply to race (racism).

There is extensive literature on ethnicity as a phenomenon deserving analyses and explanations in its own right, and producing theories that are often contradictory, swinging between essentialism (ethnic group as an innate reference) and constructivism (ethnic group as a social construction) (for some reviews of the theories, see Glazer *et al.*, 1975; Portes, 1984; Martiniello, 1995; Poutignat and Streiff-Fenart, 1995). Until recently, demography was hardly present in these theoretical constructions. Without reviewing the publications, we can quote three basic premises from recent works: (1) ethnicity is a determining social marker, in the same way as are social class and sex (Simon, 1983); (2) the essentialist definition is to a great extent rejected (Poutignat and Streiff-Fenart, 1995) in favor of a definition based on social and historical construction, and involving changing boundaries (Barth, 1969); and (3) despite the real difficulty of empirically measuring the concept of ethnicity, a difficulty that provokes caustic criticisms that go so far as to suggest all attempts in this direction should be rejected, it is increasingly acknowledged that greater efforts should be made in experimental formalizations to construct and mobilize categories closer to social and economic reality.

It is in connection with this function attributed to statistics (Desrosières, 1993) that demography has become implicated in research on ethnicity. Indeed, demographers have participated and contributed to the development of "ethnic statistics," defined here as the production and analysis of official categories such as are those provided by censuses and large surveys carried out by public statistical bodies. Even in descriptive works, it is obvious that the development of conceptual tools and categories at the source of these analyses are produced from a more or less explicit theoretical formalization. In this sense, the spread of theories of ethnicity in national social sciences has had a direct effect on the degree of acclimatization in the demographic research of ethnic categories in each country.

The use of ethnic categories is not without problems, and several works discuss the issue of the social and political significance of such categories (see Anderson and Fienberg, 1999; Zuberi, 2001). We will steer the analyses that follow in this perspective, by placing ourselves at four different levels: (1) attempting, by examining demoethnic analyses, to answer the question "why are demographers interested in ethnicity?"; (2) giving a brief overview of the diversity in the political and social use of statistical categories; (3) illustrating our words with the study of three cases where ethnic data have been used in the United States, Canada, and Great Britain; and (4) looking critically at the production of ethnic data in censuses.

I. WHY DEMOGRAPHERS ARE INTERESTED IN ETHNICITY

The answer to this question is not univocal or obvious insofar as demography is not fond of epistemological and theoretical reflection. The examination of demographic publications (particularly Anglo-Saxon) show that the ethnic variable is analyzed in association with population components supplied by censuses. Moreover, demographic manuals deal with the ethnic question in this perspective (see Beaujot, 1991:297). This type of analysis aims at studying the relative size of ethnic groups, their growth over time, and their socioeconomic characteristics. Research into integration has also drawn on this type of data in order to compare the economic performance of different ethnic groups, or that of the native population with that of immigrant populations (North American literature is very extensive: see Lieberson and Waters, 1988; Borjas, 1994; Halli and Driedger, 1999; Pendakur, 2000; for Europe: Coleman, 1994). The concept of "residual discrimination," which is defined as the persistence of economic differences even after human capital factors have been taken into account, could be empirically verified in these data (Renaud *et al.*, 1997).

Beyond the immediate interest for each variable that characterizes population, the theoretical interest in the ethnic variable for population sciences comes from the conception of the population system. To reproduce (or renew) itself, a population can only count on a limited number of mechanisms: the replacement of generations through natural growth or net migration. When the ethnic (or racial) composition of a population is a sociopolitical issue, the population dynamics of subgroups becomes a wide-spread field of research, as is the case in Canada and the United States (Bean and Frisbie, 1978). From a methodological point of view, the ethnic group constitutes a dependent variable that varies according to demographic variables. It is an essentially descriptive approach that, at first sight appears nontheoretical, but behind these analyses there is always a certain reference to a desired or desirable demoethnic balance. Population projections often serve as a warning to adopt more selective or restrictive immigration policies. Beyond these sociopolitical considerations, this type of analysis causes huge classification problems, as we will show later, as it is based on categories that are assumed to be fixed and invariable, a fiction that is contradicted by the fluidity of ethnic identities (Waters, 1990; Hout and Goldstein, 1994).

However, another response to the question "why is demography interested in ethnicity?" is not strictly demographic but more generally linked to social science concepts, which consider ethnicity as a factor in behavioral variations. The ethnic category is raised to the level of an explanatory variable of demographic phenomena. For a long time, the descriptive approach was preferred when ethnicity was used in the analysis of differential demography in the same way as sex, age, marital status, family structure, and social categories. Recently and increasingly, research in this area has been using statistics from multivariate analyses. Once again, despite the empirical character of the research, the interpretations most frequently borrow from culturalist theories (Piché and Poirier, 1995). Beyond the general criticisms that call the adequacy of these approaches into question (Poutignat and Streiff-Fenart, 1995:120–123), a more radical criticism suggests that ethnic categories are inappropriate for causal analysis and are carriers of a racist ideology if they are not placed in their sociopolitical context (Zuberi, 2001).

Migration is a distinctive case and deserves to be considered separately. For a long time, researchers were obliged to choose between a Marxist approach that rejects the pertinence of the ethnic variable and a functionalist, neo-liberal approach that considers the ethnic variable as a proxy for cultural factors. In the first case, migration determinants are of a macrostructural nature in relation to capitalist development. When Marxists deign to speak of ethnicity, it is considered as a smokescreen or as a secondary contradiction in relation to the war of the classes (Bonachich, 1999). In the second case, the accent is placed on the individual (microscopic approach), and ethnicity is most often presented as an obstacle to mobility (a retention factor) and an impediment to integration. For some time, authors have been trying to reconcile the two micro-macro approaches (Wood, 1982; Massey *et al.*, 1998:chapter 2) and introduce a communal level (the "meso" level) at which ethnicity is a strategic link. Thus, ethnicity acts as a networking factor in he field of migration at the point of origin, by participating in the organization of migration, and at the destination, by constituting itself as a channel of information for the other members who have remained in the country of origin and as an interdependent group within the community.[1] The concept of ethnicity has therefore become central to all contemporary theories on migration.

In any case, the importance of ethnic categories will always remain ambiguous. Ethnic categories can be useful to dominant groups (those who have the power to nominate) as classification tools in order to maintain the privileges and systems of racial segregation (Zuberi, 2001). Classification therefore aims to determine hierarchical statuses that stipulate who has or does not have political and civil rights, as was the case for the apartheid system in South Africa, racial segregation in the United States, and ethnic selectivity in Canadian immigration policies. Censuses are therefore used to count, classify, organize into a hierarchy, and control.

With the adoption by various countries of rights and freedom charters, programs for equal access to employment, and antidiscrimination measures (at the beginning of the 1960's), the role of censuses changed radically, as we shall see in the next section. The principle of these charters is that no specific group has intrinsic, normative, or cultural superiority over another. As a matter of principle, therefore, there is only one category of citizens. But in reality (and this is the role of the analysis of ethnicity), there really is an ethnic stratification. Classification therefore aims at determining which groups are disadvantaged (less equal) and which mechanisms are the basis of ethnic stratification. In this approach, the census is used to enumerate in order to highlight the stratification and to identify the target groups toward which equal opportunity programs should be aimed. Here the criteria are presumed to be objective (color, national origin, ancestral origin, language), and exogenous (defined by the statistical system or the researcher). Very recently with self-identification, the criteria have become subjective and endogenous. But regardless of whether the criteria are exogenous or endogenous, the ethnic categories can be used by minority groups to demand equal access to economic, social, and political resources.

In short, the intrinsically ambivalent nature of ethnic data must be emphasized, in the sense that, as we stated previously, they can confirm hierarchical organization, as well as be used in aiming for equality. Undeniably, ethnic inequality is present in every country of the world and ethnicity is a major principle for division within human society. It therefore seems difficult to analyze societies, or transform them (which is not, strictly speaking, the purpose of demography, but is also what it is used for), without taking their organizational principles into account.

In demography, the analysis of ethnicity is nearly always based on official data provided by censuses. An analysis of the types of data collected (or not collected) is very revealing of subjacent conceptions. The next section presents three case studies (United States, Canada, Great Britain) that illustrate the issues behind the statistics on origins. The last section presents a more technical discussion of the production of ethnic data.

II. OFFICIAL CATEGORIES AND THEIR POLITICAL UTILIZATION: A GREAT DIVERSITY

The linkages between political views and projects and the statistical categories that support them are present in every society (Nobles, 2000). In this respect, censuses are a special place in which views on race and ethnicity are confronted by official statistics. In this sense, censuses do more than reflect social realities, they also participate in the construction of these realities (Kertzer and Arel, 2002b:1). However, statistical constructions around ethnic categories cover a wide diversity, which makes any attempt at generalization difficult. This is essentially owing to historical and social specificities and the various procedures that are at the origin of racial and/or ethnic stratification. Conquests, annexations, redefinition of borders, or migration have placed certain groups in minority positions, whether they are old or recent minorities (Striff-Fenart,

[1] For a more detailed analysis of this concept, see Piché's contribution in this treatise (Chapter 84).

1997:3). To simplify, four cases could summarize past and present experiences regarding ethnic statistics. The first case *(counting to dominate)* has characterized a major part of historical experiences associated with colonialism. Although very widespread in the past, it still exists in some newly independent countries. The second case *(not counting in the name of national integration)* is associated with the republican perspective for which national identity is unique. Ethnic differences are therefore denied and should disappear through an unavoidable assimilation process. The third case *(counting or not counting in the name of multiculturalism)* is associated with counting or not counting but for reasons that have little to do with racist views or national unification, but rather with the increased appreciation of cultural mixing. The fourth case *(counting to justify positive action)* has appeared recently and implies a complete reversal of the racist and discriminatory perspective of the first case. Although the first three cases currently apply to the majority of countries, there is increasing pressure to recognize minorities and the necessity of ethnic statistics in order to document discriminatory practices.

1. Counting to Dominate and Exclude

Historically, this type of categorization was closely associated with the colonial situation, which was impregnated with racist ideologies and the issue of superiority/inferiority. According to colonial logic, the *others* should be collectively identified, and censuses (or administrative surveys in the case of the majority of African countries) are the most visible and most politically important tools for the state's statistical needs (Kertzer and Arel, 2002b:3). This approach characterized all the regions under colonial domination, particularly Africa and Asia but also eastern Europe. Generally, the approach employed categories that reflected racial and ethnic differences considered to be scientifically apparent and administratively pertinent (Uvin, 2002). Such a stratification essentially aims to recognize the rights of the dominant and to exclude the dominated from citizenship. In some extreme cases, racial or ethnic markers are recorded on identification papers: this was notably the case in *apartheid* in South Africa (Kertzer and Arel, 2002b:3), in Rwanda (Chrétien, 1997:18), and in Soviet controlled countries (Haug, 1998:13).

Regarding the postcolonial period, the paths taken by the newly independent countries were extremely diverse despite their very similar colonial situations. The case of Africa illustrates this diversity well. Although after independence, as we will see later, the majority of African countries opted to negate ethnicity in the name of national integration, some of them adopted an ethnicist approach by using ethnic categories as tools for domination. The comparison of Burundi and Rwanda is an exemplary illustration of the political stakes of statistical data. In the case of Rwanda, ethnic categories served as administrative markers in order to exclude the Tutsis from power, thus leading to a policy of numbers that tends to cover discriminatory practices (Chrétien, 1997:37). In Burundi, the reverse strategy was adopted: ethnicity is denied (therefore no questions in the census) in order to mask the unequal power structure and the system of mass exclusion (Uvin, 2002:166). In other cases, censuses are simply abandoned (Nigeria), or the results are not published (Mauritania) (Kertzer and Arel, 2002b:25).

Eastern Europe's experience should also be placed under the title of "counting to dominate". The Soviet experience regarding nationality is all the more interesting, as many authors see the similarity between the concept of nationality and that of ethnicity (Courbage, 1998:27; Arel, 2002b:802). Historically, the former Soviet Union was presented as a federal structure based on ethnic and territorial foundations (Abramson, 2002b:176). Nationality or membership of an ethnic group was recorded on each citizen's identity card (Haug, 1998:13). The Soviet empire's constitution is a unique historical case given the huge challenge of managing a multinational empire. Ethnic statistics are indeed used for imperial management here and follow from the colonialist logic of territorial extension (Blum and Gousseff, 1997:49). The colonial nature of the Russian territory raised the issue of the administrative and legal status of the *conquered people*, and social distinctions based on geographic origin (national or ethnic) become necessary in order not to give to the inhabitants of these regions the same attributes as those of the central regions. Censuses have played a key role in the establishment of precise nomenclatures enabling territories and nationalities to be associated, initially on the basis of ethnography (between 1897 and 1937) and then on a purely official, administrative, and legal basis (1839–1989) (Blum and Goussef, 1997:55–61).

With the collapse of the Soviet empire, the policy of numbers continued to dominate the debates regarding official census categories, but this time in the context of independent States. In post-Soviet censuses, the association between nationality and territory is maintained everywhere as the question of nationality continues to be linked with that of national sovereignty. Indeed, the legitimacy of independence is perceived as being based on the fact that the country constitutes the land of origin of the *titular nation*, that is, the ethnic nationality that has given its name to the state, hence the necessity of creating ethnic majorities via census statistics, which provokes struggles by different groups for recognition on the list of nationalities (Arel, 2002b:804–805). "Post-Soviet elites use the census to legitimize the right of titular nations to rule over their homeland, in perfect continuity with Soviet practice," according to Arel (2002:817). The case of Uzbekistan illustrates this continuity in the practice of national and ethnic categorization in which the policy of numbers persists even after independence (Abramson, 2002). As we will see later, taking into account the fact that these new states are emerging within a new global system of nation states, the necessary recognition of ethnic diversity will transform the enumeration practices of censuses (Abramson, 2002b:196). Moreover, these pressures also apply in Russia (Blum and Gousseff, 1997; Haug, 1998).

This said, it must be stressed that the approach of "counting to dominate and exclude" did not develop only in colonial situations. As we will see in greater detail, this issue has been present in countries of immigration such as the United States and Canada for a long time. For example, in Canada during the 19th and at the beginning of the 20th century, the racial issue was included in an immigration policy that aimed at excluding certain groups such as blacks and Asians. These racist criteria were not eliminated until the 1960s (Piché, 2003). However, although the racist issue has lost momentum or even disappeared in the official statistics of these countries, racial categories continue to be included in censuses, even though ethnic categories are dominant, but for reasons that are diametrically opposed and associated with the struggle against discrimination.

2. Not Counting to Unify and Assimilate

In contrast to "counting to dominate," many countries have opted not to count populations based on ethnic or national criteria. The justifications are not all identical, but in all cases, ethnicity as a basis for social stratification is rejected, either in the name of national integration, as is presently the case in many African countries, or in the name of the republican principle of national unity, as is the case in several countries in Western Europe. The case of west Africa is particularly interesting insofar as the scale of ethnic analysis, which was omnipresent during the colonial period, has been abandoned in the postcolonial period in favor of a nation-building ideology. Indeed, they have moved from a situation in which ethnic classification, invented by colonial administrators and ethnologists (Amselle, 1990:22), was used as a basis for social stratification to a situation in which the effort of developing and building a national identity is monopolized by the postcolonial state, which is involved in a modernization process that should result in the disappearance of ethnic awareness (Otayek, 2000:87, 90). In this context, ethnic analysis becomes taboo owing to the weakness of the state, which is trying to build a nation within borders drawn by the colonizers, borders that rarely take ethnic geography into consideration. For west Africa, analysis of ethnicity is all the more lacking as vast circular migratory movements have turned the region into an integrated economic space (Adepoju, 1988:60; Cordell *et al*., 1996:13). Membership of an ethic group is of little importance in this context, as long as labor needs remain high, and the concept of citizenship is of little importance in societies in which access to *social security* remains essentially within the family sphere and outside of state-controlled legislation.

Therefore, during this period, demographers ignored the ethnic variable in their analyses of censuses and surveys, even when the variable had been explicitly included in the questionnaires. It is only recently, notably in Côte d'Ivoire, that the concept of citizenship has come into effect and is used for exclusion purposes (Blion and Bredeloup, 1997; Dozon, 1997). It goes without saying that even if ethnic categories are rejected in the name of national integration, this does not mean that ethnicity is not included in political practices (Otayek, 2000:55). The situation is at the very least paradoxical: on the one hand, ethnic data are absent in order to avoid claims of ethnic identity and to promote national integration, and on the other hand, ethnic based policies are omnipresent.

The refusal to include ethnic categories in official statistics characterizes nearly all the countries of western Europe. According to a recent inventory, the reasons that these countries refuse to include a question regarding ethnic groups are mainly political, constitutional, and legal: this is the case notably of France, Germany, Spain, Belgium, Denmark, and Italy. It would seem that this information is not pertinent for Sweden, Switzerland, and Ireland and that there is no social demand for these types of data (Courbage, 1998:55).[2] The case of France deserves particular attention insofar as the debates there are particularly virulent. Indeed, in France, the question of using ethnicity as an analytical or simply descriptive category is far from being resolved as is confirmed by debates that are ideological and controversial in nature (Taguieff, 1991; Lorcerie, 1994). The frequently violent arguments between demographers testify to the difficulty of

[2] Other unitary states exist, such as Greece and Turkey, which, although proclaiming the homogeneity of the nation state, would not recognize the existence of minority groups on the territory or official statistics on these groups (Haug, 1998:13).

introducing the ethnic variable into the analysis of social reality. For some, the study of ethnicity is dangerous because of the risk of making it essential—and, in so doing, reducing social issues to biology (Blum, 1998)—and playing into the hands of the extreme right (Le Bras, 1998). For others, taking ethnic categories into account is necessary if we wish to understand the mechanisms at work in the exclusion phenomena based on national or ethnic origins (Tribalat, 1995). The rise in the assertion of identity highlights the inadequacy of traditional categories (French–foreign) in the study of the future of immigrants and their descendants, particularly with regard to discrimination (Simon, 1998).

It is necessary here to note the double standard of metropolitan practices in the colonies. Thus, France, Great Britain, and Belgium, which have refused to include questions regarding cultural or ethnic identity because of its incompatibility with their conception of a nation state, did not hesitate to do so in the colonies (Kertzer and Arel, 2002b:10). In the specific case of France, colonization, which creates a system of ethnic classification, constitutes a denial of republican universalism (Otayek, 2000:24–25).

3. Counting or Not Counting in the Name of Multiculturalism

The Latin-American experience is different from the majority of historical experiences discussed thus far. In this region, racial mixing is acknowledged in political and ideological views as a positive value and should therefore not cause a problem. Within this ideological context, there are two distinct statistical practices. On the one hand, in the majority of countries (e.g., Venezuela, Colombia, Cuba, Dominican Republic) the census does not enumerate by race or color (Nobles, 2000:66). On the other hand, and this is the case in Brazil, the use of the census in racial attitudes is particularly well developed. Thus, racial views are in a way used to promote multiculturalism, insisting on harmonious race relations, and statistical data serve to measure the population's degree of whitening (Nobles, 2000:43, 62). However, since the 1980s, numerous surveys have shown socioeconomic differences between the categories of *color*: we thus move from the idea of a radically democratic and nondiscriminatory society to an image of distinct groups that are discriminated against (Nobles, 2000:65).

4. Counting for Positive Action

All ideological constructions regarding notions of race and ethnicity are currently under pressure and being called into question. In particular, countries using statistical categories in order to dominate are being challenged by the minority rights movement and antiracist and antidiscriminatory ideologies. Although Africa seems so far to have avoided this movement (e.g., Côte d'Ivoire), the recognition of pluralism is imposing an increasing number of new statistical practices and risks transforming the enumeration practices of censuses. This is particularly true in the case of eastern Europe, where the issue of pluralism has appeared (Blum and Gousseff, 1997; Abramson, 2002; Arel, 2002b). Even countries such as France, which are characterized by the absence of ethnic statistics, are confronted by the increase in social demand for data requesting information concerning the integration of immigrants beyond the first generations (Striff-Fenart, 1997). A new issue is appearing in the countries of the European Union, supported by the Council of Europe, which expresses the need for reliable statistical data "to encourage peaceful intercultural relations and ensure the protection of national minorities" (Haug, 1998:11). Pressure in favor of positive action appears as much in Latin America, for example, in Colombia (Barbary, 2001) and in Brazil (Nobles, 2000), as in Asia, for example, in China (Courbage, 2002). We are progressing from the diversity of past situations to a pluralist model that tends to replace ethnic nationalism by a more civic nationalism (Dieckhoff, 2000:191).

Three countries deserve particular attention for having made this transition toward a pluralist model, they are the United States, Canada, and Great Britain. In these countries, ethnic diversity resulting from immigration has necessitated calling official views and statistical categories into question. Here we witness a complete reversal in views and policies, which now sees ethnic categories of censuses as tools in the fight against discrimination. The following section will present these three case studies.

III. STATISTICS OF ORIGINS: UNITED STATES, CANADA, AND GREAT BRITAIN

Countries that have introduced ethnic or racial classifications[3] in their statistical systems have all, in one

[3] Despite the scientific invalidation of "race" as a biological concept, it continues to be used as a descriptive category of populations in social practices and has been maintained in numerous international texts, the most famous being the 1948 Universal Declaration of Human Rights, and UNESCO's 1950 "Experts' Declaration on Race" (see Gayon, 2001). Based more on social perception than on improbable genetic variations, the "racial" categories des-

way or another, been confronted with the experience of population diversity associated with a political connection: slavery, immigration (postcolonial in particular), or the registration of minorities within the political borders of a nation state. The definition of the categories, the methods used for recording them, and the use made of them vary considerably among the countries, which underlines the weight of historical-national formations and the difficulty of providing a description and an analysis of classification methods independently of their political contexts (taken in the broadest sense). The experiences of the United States, Canada, and Great Britain make it possible to highlight certain constants in the models developed by immigration countries.[4] The omnipresence of the ethnic and racial reference not only concerns legal or social policy bodies, administrative bodies, the media, and collective representations, but it clearly distinguishes the demographic (and social science) research of these countries where the principal divisions place race and ethnic group at the same level as sex, age, and social category.

1. The United States: From Slavery to Multiple Race

The extraordinary development in the use of racial and ethnic categories in the United States is part of a historical continuity with origins that are in slavery and immigration. These two major facts are part of the United States' constitution. The new nation's first census, in 1790, used the categories that had already been used in colonial censuses. It includes a distinction of the population according to race and condition, with regard to slavery. This founding distinction continued until well after the abolition of slavery and the federation's 10-year censuses continuously included questions regarding the "race" of the territory's inhabitants (Anderson and Fienberg, 1999). To this fundamental distinction was quickly added differentiation according to ethnic origin, whose characterization was continually being refined during the development of mass migration and then during the period of assimilation in the years between the two wars (Schor, 1999), and after the success of the civil rights movement and the emergence of ethnic revival affecting the descendants of European immigrants of the beginning of the 20th century. In the extension of the civil rights movement that obtained the suppression of segregation policies (civil rights acts) and the establishment of affirmative action, a revision of the categories included in censuses became necessary (Nobles, 2000).

Indeed, affirmative action programs require a numerical comparison of the distribution of racial groups in the economy, housing, or education. The disparities within an equal representation serve as indicators of the size of the phenomenon and its eventual decrease or increase. However, the list of groups benefiting from these programs was officially set in 1977 by "Federal Statistical Directive 15" at four minorities: "Blacks," "Indians," "Hispanics," and "Asians." As a result, these four groups are subject to special treatment in censuses, either through a question regarding "race,"[5] or by a special formulation for "Hispanics."[6] The naming of target groups in affirmative action programs therefore reproduces the American society's racialization, as it aims to quash the negative effects of the circulation of racial prejudices. By doing this, it reinforces the racialization of the statistical system.

Jointly with the consolidation of racial taxonomy, a re-evaluation of categories used to classify immigrants and their descendants was undertaken. The combing operated by the question on foreign parentage was indeed of no use in describing the situation of immigrants who had arrived between 1880 and 1920. To study the identity demands of the *hyphenated Americans* in more depth and to re-establish genealogical links, another approach was required.[7] The concept of ancestry appeared within this context. Contrary to the double classification that resulted from individuals' place of birth and that of their parents, ancestry records the origin with which the respondent identifies, with reference to the nationality or the country of origin of the last ancestor born outside the United States. The U.S. statistical office carried out several

ignate phenotypic, physical, or chromatic markers that serve to distinguish individuals and outline populations. Some countries, such as the United States as we will see later (but also Brazil), officially use a racial classification. Others use proxies, such as "visible minorities" in Canada, or "ethnic group" in Great Britain. The sociological literature of ethnic and racial studies extensively uses the concept of race, thus conferring social or cultural acceptance on it. Simply put, race is what is produced by racism, dealing with the effects of racism therefore presupposes speaking of race, while being aware that this is not based on any objective distinction between humans. In a later section of this chapter, we will refer to ethnic and racial categories, without quotation marks, to evoke the classification used by the statistical bodies of the countries in question.

[4] For further developments, see Simon, 1997.

[5] With the following formulation in 2000: "What is this person's race? Mark one or more races to indicate what this person considers himself/herself to be" (the underlining is ours). The possibility of selecting more than one race appeared for the first time in the 2000 census. The expression "considers to be" emphasizes the subjective nature of the response.

[6] For background on the "Hispanic question," see Cholsin (1986).

[7] On the links between ethnicity and race in the United States' statistical classification, see Schor (1999).

surveys in 1969 and 1979 (Current Population Survey) that tested the questions on ancestry, and the question was introduced into the census in 1980.

From then on, ancestry was used as an elementary parameter for pinpointing ethnic groups in the United States. Although such groups do not benefit from differential treatment programs, their recognition through the census consolidates the development of the ethnic pluralism present in American society since the mid-1960s (Petersen, 1997). From a methodological point of view, the recording of declarations on individuals' origins is an innovation in the census. Beyond an objective evaluation of ethnicity, self-classification is based on the expression of identity. Indeed, by asking individuals to choose an origin or even several origins, the question presupposes a selection among the plurality of possible origins, according to criteria that largely elude statistical services. In other words, the responses to the question on origin indicate the *ethnic options* of individuals, to quote the title of Mary Waters' famous publication (1990) devoted to the analysis of data on ethnic origins.

The significance of data on race and ethnic origins poses a real epistemological problem for statistics and, in particular, for official statistics whose purpose is to authenticate the numerical size of the groups that make up the country's population. The variations in the declarations of ethnic origin in each census demonstrate the instability of identities and their sensitivity to current events, whether they are political, cultural, or social (Skerry, 2000). The interweaving of the collection of racial and ethnic data and affirmative action programs mainly provoke debates on the categories. Their legitimacy is rarely called into question as they have become almost natural in American society. The division of society into a racial pentagon is conceived as consubstantial and is never questioned (Lacorne, 2001).

In contrast, the debate surrounding the policy of affirmative action has weakened the statistical base and provoked changes in the method of handling racial groups. The introduction of the possibility of multiple choices in declaring race in the 2000 census formalized the change in society toward greater racial diversity. Indeed, the results (fewer than 7 million declarations of several races, 2.4% of the population) seem to demonstrate the resistance of racial boundaries, if not in reality, at least in the representations of individuals who favor an uncombined genealogy in their declarations (Grieco and Cassidy, 2001). Multiple choice for race has, however, opened the Pandora's box of the reliability of the declarations and definitively damaged the pertinence of the questioning (Perlmann and Waters, 2002). Although some consider that the low rate of racial mixing declared reinforces the pertinence of affirmative action, others predict that the convention of racial self identification will be called into question (Goldstein and Mourning, 2002).

2. Canada and the Visible Minorities

Since the beginning of the 1970s, the Canadian government has been professing a multiculturalism based on recognition of the population's ethnic peculiarities, associated with the exercise of specific rights for minorities within the confederacy. However, this conversion should not make one forget the particularly assimilation policies of successive governments faced with foreign immigration. The multicultural opening exists within the conflict between the two "founding peoples," France and Great Britain, since the promulgation of the 1867 Act of British North America that created the dominion of Canada.

A regular and detailed count of inhabitants according to their origins and first or usual language occupies a strategic position in the formation of Canadian identity and, in more concrete terms, in confederate political management. More recently, debates surrounding the emergence of a Canadianism that could be expressed in the same way as that of other ethnic groups, or be defined as a priority, is developing among the two founding peoples as a result of the pressure from large-scale foreign immigration.[8] The recording of origins in censuses has experienced numerous changes since the end of the 19th century. As well as nationality and place of birth, data on *ethnic group* and *racial origin* were collected until after World War II, when there was a return to *ethnic origins*. Opposition between a geographical-political definition of origin and a definition of ethnicity through *linguistic ancestry* has been seen in Canadian classifications from 1871 up to the last census in 1996 (White *et al.*, 1993).

The definition of ethnic origin developed dramatically between 1951 and 1991. Although the 1951 census was the first to offer a geographical definition of origin, the 1971 census introduced a series of significant modifications in the conceptualization and the formulation of ethnic origin. In particular, it marked the passage to self-counting. Origin is characterized as the "ethnic or cultural group [to which] the ancestor on the male side belonged upon arrival on the continent" (*Census Guide*, quoted in White *et al.*, 1993). The *Census Guide* recommends referring to the language spoken by this ancestor. In 1981, the reference to the ancestor on the male side was abandoned and replaced by a wider

[8] The concept of "foreign immigration" is a Canadian specificity and indicates all origins other than French and British.

meaning of ancestry. The concept of roots replaced that of descent in the guide's instructions. The use of a list of names, organized hierarchically according to the number of times they were used in the preceding census, was proposed in several censuses. This organization method in the form of origins listed in descending order resulted in competition between different groups to be recognized in the listing and to obtain, thanks to multicultural policy, a significant numerical representation. The pressure exerted by ethnic lobbies to modulate the formulation of the questions or to include one name or another started to multiply.

The multiplication of demands from the first nations led to the removal of the reference to the first ancestor to arrive on the continent. This step toward a broader concept of the Canadian population's origins also resulted in the acceptance of multiple responses in even the wording of the question. The result did not take long: multiple responses represented 11% of all responses in 1981, 28% in 1986, and 38% in 2001. The affirmative action programs included in the 1982 Charter, then in the Employment Equity Act of 1986, modified in 1996 (Equity Act, 1996), and the Multiculturalism Act of 1988 also had a significant effect on the data collected in the census. The inclusion of a "Black" category in 1986 responded to the Employment Equity Act, without it being really sufficient to cover the range of minorities that were likely to be discriminated against. Working on the same principle as the affirmative action programs, the act established quotas or numerical objectives to achieve equal opportunity. However, the Canadian census, contrary to the U.S. census, has not included a reference to race since 1951. The evaluation of the population concerned by the Equity Act revealed itself to be delicate, and the act introduced the concept of "visible minorities." By visible minorities, the concept of "nonwhite"[9] people is understood, that is, an all-inclusive list combining all individuals according to their country of origin but referring implicitly to visible attributes such as skin color or appearance.[10] In 1981, all visible minorities combined made up nearly 5% of the total population, by 2001 they had increased to 13%. This increase is essentially owing to high, mainly Asian, immigration in the 1990s.

After nearly a century of evolution, the categorization of ethnic groups has changed considerably. Origin is now clearly distinguished from language, it has lost its reference to the paternal line (awareness of gender inequality in Canadian society), and its reference to the first arrival on the continent (increase in autochthonous claims). Henceforth, it is turning toward a more subjective definition, closer to identification than ancestry, which in any case is very hypothetical. Taking into account multiple origins puts into perspective claims to reconstitute ethnically homogenous lines from the migration period up to the census date. The high levels of mixed marriages between ethnic groups should favor a more subjective definition of origin, because the identification of a specific ancestry depends on a grid of complex decisions that are specific to each respondent.

Increasing the relativity of origin as the expression of ancestry, the emergence of a *Canadian* category disturbs the classification usually used. For a long time regarded as inadequate, claims of Canadian origin have increased significantly. The introduction of a Canadian category in the list of ethnic groups was tested in preparation for the 1991 census. Nearly a third of respondents chose it to describe the origins of their parents or grandparents, and half chose it to describe their own ethnic identity (Pryor *et al*., 1992). Quite obviously, a process of indigenization consecutive to the date of the initial migration (more than 2 centuries for the founders) has greatly affected the Canadian population. This refusal to be included in the old ethnic distinctions by affirming a national identity can be explained as a reaction to the multicultural policy. The 1996 and 2001 censuses explicitly included the Canadian category in the section on origins. Nearly one-quarter of the population indicated it as their only ethnic origin in 2001, whereas 16% declared it in combination with another origin. A postenumerations survey on ethnic identity is carried out in conjunction with the census in order to provide an evaluation of the population's declaration and membership logic, beyond a national feeling that obviously destabilizes the statistical systems and public policies toward minorities.

3. British Postcolonialism

Although questions regarding race and language have been authorized by the Census Act since 1920, data on ethnic or racial ancestry have only been collected in the United Kingdom census since 1991. Up to this date, censuses had only recorded data on nationality and place of birth. These were used to establish statistics on ethnic groups with the short-lived introduction in 1971 of a question on parents' place of birth based on the model of the American census.

[9] More precisely "persons, other than natives, who are not of white race or who do not have white skin."

[10] The list of groups considered as "visible minorities" indicates visible perceptions as clearly as the terminology: "White, Chinese, South Asian, Black, Arab/West Asian, South East Asian, Latin-American, Japanese, Korean, Other."

However, the categories thus obtained did not correspond to the groups studied by the sociology of *ethnic and racial studies*, or to the population targeted by campaigns against discrimination. Created by the Race Relations Act (RRA) of 1976, the *Commission for Racial Equality* was given a mission "to work towards the elimination of discrimination; to promote equality of opportunity; to influence policy; and to develop and encourage research in the field of race relations" (RRA). From this point of view, public action and the development of tools of knowledge and surveillance appear closely related.[11] However, the development in the fight against ethnic and racial discrimination encounters many problems, the least of which is not the complete lack of quantitative information on the phenomenon. The concept of a minority, the usage of which spreads meteorically in public view, does not have a statistical basis. In the 1971 census, a question on the parents' country of birth, similar to that of the American census, was introduced. It was used as a basis for identifying minorities concerned with the RRA. However, the information thus collected is not really satisfactory, and from the point of view of the 1981 census, reflection on taking ethnicity into account began. A virulent argument began between those who support pinpointing individuals' ethnic and racial origin and those for whom such a process is not only prejudicial to individual freedom but also springs from scientific trickery. To the detractors of a question on ethnicity, Martin Bulmer (1980) responded by pointing out the relative imperfection of all measurement of social phenomena. By placing the specific issue of race in a more general perspective, he neutralized the controversy regarding the legitimacy of racial representation and only dealt with the technical issue.[12]

In parallel with the epistemological and ethical debate on the racialization of statistics, the Office of Population Census and Survey carried out a series of tests to develop the questionnaire on ethnic and racial origins. In the face of resistance, the idea was abandoned by the government, and the 1981 census did not include any questions on ethnic origin or parents' place of birth. After this first failure, the gaps in the information provided by the census on the situation of ethnic minorities caused major problems for the Commission for Racial Equality. The adjustment of local positive action policies required very accurate information that the census could not provide. As a result, the government asked the Office of Population Census and Survey to begin exploratory work again. This time, the changes in the census did not meet with any major opposition. The 1991 census therefore included a question regarding a person's ethnic origin (Coleman and Salt, 1996). The taxonomy used combined racial and ethnic elements, such as *White*, *Black-Caribbean*, or *Pakistani*. This confusion reflects the ambiguity in the definition of minorities in British legal texts.

In Great Britain, more so than in the United States, the borders between race and ethnic group appear to be extremely blurred. In most cases, these two terms overlap or are used one for the other. Thus, the RRA of 1976 defines discrimination as unfavorable treatment of a person on the basis of "racial grounds." The text uses an extensive definition of racial grounds: " 'racial grounds' [refers] to 'color, race, nationality (including citizenship) or ethnic or national origin.' " With criteria combining objective (place of birth, language of socialization, religion, culture, nationality, skin color) and subjective elements (identification of a real or imagined ancestry, feeling of belonging to a group), such a liability is hardly surprising. It reproduces the broad outline of the divisions inherited from the colonial experience and the construction of the British empire, while striving to match the most common representations in British society. This is why the "ethnic origin" category is biological, cultural, political, and social (Ballard, 1997).

Such a characterization indicates a continuity between British colonial classifications and those used for minorities settled in the United Kindgom. The fact that the majority of the immigrant populations came from former colonies of the empire reinforces the transposition of taxonomies. The colonial heritage[13] greatly modifies the management and position of immigrants in former empires. As Didier Lapeyronnie (1993:132) points out, the immigrant of colonial origin is a victim of a specific rejection, more racist than xenophobic, which "is born from the difficulty that is specific to universalistic and modern cultures of incorporating the particular experience of the colonized." The progression from colonialism to racism rises from a connection between the immigrant and the host society that precedes the migration. In this sense, the racialization of immigration statistics in Great Britain is situated in the extension of the classification formed

[11] Maire Ni Brochlain (1990:546–550) lists no less than seven reasons justifying the presence of an "ethnic question" in the census. These arguments all have, with no exceptions, an operational range. Here, knowledge is subordinate to action, it starts with a recognition of multiple "racial" discriminations in all aspects of social life, then the statement directs the initiatives taken by the services in charge of "ethnic minorities."

[12] A very "positivist" position, which he reiterated in the Office of Population Census and Survey publication that followed the 1991 census (Bulmer, 1996).

[13] On this point, British and French experiences show many similarities.

within the framework of colonial management and reactivated in immigration control policies.

The ambiguity of such a determination corresponds to the broad spectrum covered by the laws on race relations. Social intervention programs, in placing the fight against racial and ethnic discrimination at the centre of their concerns, have encouraged the racialization of statistics while polarizing social issues on ethnic-racial problems. The British experience once again illustrates the close links between racism and antiracism in their interdependent development. How can the terrible contradiction that is attached to any reparation be avoided? By refining the methods of pinpointing exposed minorities, the classification locks them into an autoreferential dynamic, excluding de facto other media of identification (notably socioeconomic).[14] In the same way as the United States—which began to take multiple declarations into consideration, opening the way to an apprehension of multiculturalism or intermixing—the census in 2001 also introduced a mixed position in a formulation that was even more explicit than the U.S. example by introducing a special heading. A question on religion was also added. However, there are no longer any questions on citizenship.

IV. THE VERY DIFFICULT USE OF ETHNIC DATA OR THE BIAS OF ETHNIC STATISTICS

The ideological and ethical aspects considered so far should not blind us to the simply methodological problems. It is necessary to ask about the quality of ethnic data and the specific methods of analysis that are associated to them, which conditions the legitimacy of their use and on which the image of the components of society depends.

Here we will only discuss the statistical aspects: data collection; the manner in which respondents can be situated in the categories, notably with the question of people of mixed race (for convenience, all persons of multiple or mixed ancestry will be so called, but the terms should be understood in the social and not the biological sense); and the presentation of information by statisticians who regroup and classify the data once more. The importance of data collection for these types of data is such that it is necessary to consider this stage in sufficient detail before moving on to the issue of analysis. Then we will consider analysis problems at a given time: estimates of data quality and common bias in these types of data. Last, we will look at the problems of studying population trends through ethnic statistics.[15]

1. Collecting Ethnic Statistics

The statistician's approach wishes to be objective (Goldman and McKenney, 1993; Goldmann 1997) by ensuring that the ethnic categories reflect the sociocultural reality of the various components of society. The statistician is involved twice: during the design of the questionnaire and during analysis of the data collected.

a. The Wording of the Question on Ethnic Origin

Before the questionnaire is written, the issue of the method of declaration of ethnic origin comes up: self-declaration by the respondents, declaration by the interviewer, or from a document (e.g., identity card, passport) that includes the origin, which was very common in the Soviet Union and still is today in countries that were part of it. Self-declaration conforms best to the freedom of the individual but is still not universally applied. The declaration method obviously affects the results (Telles and Lim, 1998).

The question on ethnic origin is usually presented as a list, closed or open. Experience has shown that not only do these lists differ from the terms used by anthropologists, but that they do not fulfill scientific criteria and that they mix different dimensions: racial, cultural, and national. Thus, one often comes across terms referring to race (skin color), language, and nationality with the names of ethnic groups. The United States divide the question into three: that is race (e.g., "White," "Black," "Amerindian," "Chinese"), Hispanic origin ("no/yes Mexican, Porto Rican"), and ancestral origin ("ancestry") of an ethnic-cultural nature (including nationalities and ethnic groups). In Great Britain (1991 census), a single question includes racial aspects and nationality in the terminology of the categories (e.g., "Black African," "Black Caribbean," "Indian," "Pakistani,") (Bulmer, 1996). Australia uses a question on ethnic origin in order to distinguish the aborigines and only introduced a question on ancestral origin in 1986 (and more recently in 2001). New

[14] Criticism aimed at ethnic and racial classification in the census and at the entire statistical machinery remains relatively widespread in Great Britain. It also comes from users who work in the field of discrimination, and who, although challenging the production of ethnic statistics, also set up systems for evaluation. For an example of this position, see Moore (1997).

[15] Although the issue of people's origins has been called many different names—ethnic origin, race, and ancestry in Anglophone countries—we always use the term ethnic origin for more clarity and because the problems of collection and analysis are the same regardless of the term used.

Zealand uses one ethnic question that can include nationalities ("ancestry") in the open category, and people of mixed race can tick several boxes.

At the level of the ethnic groups themselves, various inaccuracies are frequent: a group can appear on the list at the same time as one of its subgroups, particularly when the subgroup is sufficiently represented in the population for its size to be of interest. Certain names are more popular than scientific, and they often correspond to the manner in which the dominant group has named the other ethnic groups rather than the manner in which the ethnic group calls itself, a situation that was common in colonial statistics. Last, certain very minority ethnic groups do not appear on the list and can be specified in the "other" category if space has been provided. The list is sometimes very short and aims to group into large categories as is the case in Great Britain.

The considerations that determine the establishment of the list of categories proposed are not strictly anthropological but are associated with the demographic and social significance of the groups. However, they tend to carry the majority's a priori vision, represented by the administration and including the statisticians, and can make it difficult for respondents to accept the suggested categories and to identify themselves therein. It is therefore desirable that the list be prepared in collaboration with representatives of the different components of the population.

Because of the existence of various levels (race, nationality, ethnicity) of groups and subgroups in the proposed list, the order of the categories has a nonnegligible influence on the results. Alphabetical order can place very poorly represented groups at the top of the list, whereas ordering according to numerical importance is subject to the quality of prior estimates and variations with time. For this reason, it is frequently contested by various minorities. Statistics Canada's recent tendency has been to remove the list altogether and replace it with an open question consisting of four blank spaces (Canada, 1996 census: Goldman 1997:45).[16]

b. Mixed-Race and Multigroup Membership

The individuals who make up the population can belong to several ethnic groups at the same time owing to intermixing. Statisticians may wish to ignore this fact and insist that an individual be attached to one and only one group. Ethnic membership is therefore considered as a social (we attach ourselves socially to a community) rather than a genealogical phenomenon. Transmission of origin can therefore follow a patrilineal or a matrilineal tradition, or another method: social, emotional, etc. Statisticians can also decide at which degree of intermixing a person belongs to a community; the U.S. practice of *one drop blood,* according to which with 1/16 of black blood a person is black, has been the object of much criticism. The choice of single membership within a social context is not contradictory to the definition of ethnic origin as membership of a sociocultural group but has implications for analysis. Whereas a characteristic fixed at birth such as sex is (nearly) unchangeable, a social characteristic is subject to variation and to various interactions. The example of the 2000 U.S. census illustrates these aspects clearly. Few mixed-race individuals identified themselves as such and the declaration of ethnicity is therefore social rather than biological (African-Americans would seem to make up a social class). There is nothing to ensure that some of the people of mixed-race will not choose a different category in the next census.

If the declaration of mixed-race is allowed, it can be done in many different ways: ticking as many boxes in the list as necessary, describing in the "other" category, having several boxes in an open list as in the 1996 Canadian census, or dividing the question into two questions on parents' origins (with a double answer possible for each parent such as in the 1988 French Polynesian census). The approach of two questions on parents would be capable of reflecting the genealogical reality of the parents if they had at the most two origins, otherwise a question on each grandparent would be necessary. It is however not a biological racial reality (whose existence has been definitively invalidated) that we wish to identify, but sociocultural membership, a fact that justifies taking simplified representations into account.

c. Respondents' Answers

The respondents' answers represent the subjective aspect of ethnic membership. Choosing imposed declarations would not solve all the problems. In the absence of a document, the interviewer must decide on the origin of the persons and his/her opinion varies according to the interviewer's own origin—a member of a majority group, of the interviewed minority, or another (Telles and Lim, 1998)—and would influence the results (Davis, 1997). Coercive practices can cause an outright refusal to reply, either by avoiding the survey or by giving inaccurate information, which results in under-representation, a high rate of nonresponse, and a reduction in data quality.

The principal difficulties undoubtedly stem from the existence of groups that are divided into subgroups and people of mixed-race. Thus when two groups that

[16] The purists claim that even the few examples quoted in the title of the question can influence the results (Goldmann, 1997).

are in a relation of inclusion are present on the list, an individual might not read the list entirely. Some members of a subgroup not present on the list might wish to distinguish themselves from the main group that is listed and place themselves in the "other" group, clearly mentioning their origins or not. Persons of mixed-race have the choice among their various origins, which they can only mention if the questionnaire offers the possibility. If this is not the case, in order to show a dominant origin, they are obliged to omit their other origins. Last, because origins are of a social nature, sociodemographic factors such as length of stay, nationality, social situation, and cultural integration can influence the declaration.

In the field of ethnic categories, the collection method used is particularly susceptible to ambiguity between categories in which respondents can place themselves according to their perceived origins, and objective categories that reflect society as viewed a priori by statisticians. The trend is to give more freedom to the respondents and to move away from single answers (the United States recorded multiple answers in the 2000 census, see above) and to replace a list of hotchpotch categories that reflect the concerns of the majority with open responses.

2. Presentation of the Results

Although a table usually presents all responses, possibly including the most common responses to the open question and the different types of multiple origins when they can be declared as such, tables that compare different variables should, for practical reasons, group the origins in fewer categories. The groupings are usually done by geographical region or by cultural groups, unless particular origins are of specific significance because of their number or social position. If the questionnaire is designed for this purpose, one can still choose to continue separating people of mixed-race into one or more categories or integrating them into monoethnic groups at this level.[17]

[17] Only the method of dividing the question on origin into two questions on the parents makes it possible to calculate the dominant origin of mixed-race individuals: persons with three-fourths of an origin or one-half *a*, one-half *b*, and one-fourth *c*; however, if only two origins are required for each of the parents there will be ambiguity at this level. The question on individual origin does not make it possible to weight the different origins to which a person is attached, a one-half *a*, one-half *b*, and one-fourth *c* person appears to belong equally to three ethnic groups from ticks on a list or the counting of his/her origins in an open question. Placing mixed-race individuals into one single group is not mathematically correct and is based on the choice of hierarchical organization (or attribution priorities). Usually, high priorities are attributed to minority groups, which results in placing mixed-race individuals with them, but this choice does affect the results of the analysis.

Since 1991, New Zealand has been using a particularly interesting double analysis of ethnic categories associated with the existence of people of mixed-race. An initial analysis provides additive numbers. Statistics New Zealand has used, according to priority, a maximum of three ethnic groups mentioned by mixed-race individuals and published a table according to the number of origins: single and mixed-race by two or by three or more origins with the details of these origins (Statistics New Zealand, 1993). This implies category changes when grouped at a wider level: an English-Greek mixed-race individual is no longer of mixed-race if European origins are grouped. Another analysis of the data produced tables in which an individual is included in a category as soon as he/she mentions it. This method makes it possible for each component of the population to include all its elements, as well as those who share their origins with others, but the results are not additive.

As soon as the results are presented, the frequency tables of ethnicity and the analysis of open answers make it possible to consider changes to the list. As with any changes in nomenclature, it is important to preserve the comparability of the information with that collected in the past. However, poorly represented groups are sometimes withdrawn from the list, which causes problems in ascertaining to which group they will attach themselves, or sometimes groups that seemed well represented in the open question are added to the list. It would seem preferable to avoid withdrawing groups from the list, but adding also has an effect on the evolution of other groups. Individuals whose group does not appear on the list can change their declaration, detailing their origins in the open question or attaching themselves to a different group in various censuses. Individuals thus move from one category to another simply as a result of the imprecision of the categories (a phenomenon that should be distinguished from voluntary changes, see above), which results in population transfers between groups. However, the absence of a list obviously avoids any such changes.

In summary, even though the large number of ethnic groups cause coding and grouping problems for the presentation of the results, it would seem preferable, from an essentially theoretical point of view, to abandon the lists and to ask about ethnic origin in a completely open question. This would avoid the problems associated with the imprecision of terms and the inevitable changes to the list that this then entails. The second conclusion is more imperative: it is necessary to give mixed-race individuals the possibility of declaring themselves as such and to present them separately in the results.

a. Data Quality and Variations of Declarations of Origin

The populations that we wish to study through the question on ethnic origin are old, historic, or autochthonous minorities or are new minorities, resulting from migration. The phenomena studied are the socioeconomic and sociocultural (e.g., language, religion) characteristics and more generally social integration, which presupposes an analysis over time but which can be carried out through comparison between diverse components of a population by cross-section. This presupposes reflection on data quality.

Comparison of data on ethnic origin with other variables makes a relative verification of the quality of the answers possible. The most interesting variables are place of birth and original nationality. However, there are hardly any monoethnic countries, and multisecular migrations, including colonization, add to this confusion, although analysis by age or period does sometimes make it possible to distinguish old migrants (e.g., returning colonists). Religion and first language in particular also make verification possible; however, these concepts do not always overlap. Declaration of religion varies according to practice, and persons born in the host country or who migrated at a young age may not have spoken their parents' language enough to consider it as their first language.

Whereas up to now we have mainly considered that respondents declare their ethnic origin in its unchanging aspect (as an origin) and the objective aspect (in the sense of ethnic-cultural filiations reflecting social reality), we will now consider the cases in which this declaration is affected by other criteria and can change over time.

Individuals can refuse to be identified with a group of origin for numerous personal, social, or political reasons and attach themselves to a group that they believe is closer. A high social position can result in an individual attaching to a more favored group: artists, for example, tend to attach themselves to groups of exotic origins. This phenomenon particularly affects mixed-race people but is not exclusive to them. Thus, in Australia and Canada, people of all origins have started to declare themselves to be Australians or Canadians, respectively, in censuses (Pryor *et al.*, 1991, 1992). These phenomena generally tend to lower the socio-occupational levels of minorities. Measuring social distance, and socioeconomic characteristics in general is obviously biased if the definition of the population in which the characteristics are being measured is in itself social, in other words, if there is interaction between the social situation and the declaration of origin. More generally, a variant characteristic is measured, for example, socio-occupational situation, against an invariant characteristic, such as birth year, place of birth, and sex, but it is far more risky to do it in comparison to a characteristic that is itself variant, such as the declaration of a person's ethnic group.

The socioeconomic position of people of mixed-race is especially sensitive to this type of measurement. Their attribution to one or another of these groups in the context of a single category will affect the results of the comparisons. Thus, the tendency, through the interplay of priorities, is to include them in the visible minorities; by so doing, the intergroup distance in the measurement is reduced, as the mixed population would be more favored than the minority groups. The same is true if we consider that persons with faint kinship ties to an ethnic group do not belong to this group, whereas visible characteristics still distinguish them from the majority group and are therefore the cause of discrimination. The process is therefore reversed: underprivileged people are included within the privileged population. To measure a social phenomenon such as discrimination, it is necessary for the affected group to be fully counted, in other words that there be adequacy between the division of the population by the statistician and that caused by the discrimination itself.

With the existence of intermixing, ethnic groups take us back to set theory in mathematics. Mixed-race people belong to several groups, and comparison between subsets should include every individual from each group or the separate analysis of each intersection. Therefore, the solution lies either in including mixed-race people in all the groups to which they belong (this has been the case in New Zealand since 1991), which avoids the above mentioned bias (but a difference in proportion effect of mixed-race people remains in some groups), or in analyzing them separately.

Another source of inaccuracy is linked to the fact that the ethnic origin variable combines several generations whose degree of integration undoubtedly varies. It is therefore necessary, in the case of recent migration, to compare it with the place of birth to distinguish the migrant generation (first generation) from that of their descendants. If the parents' place of birth is available, it is also possible to distinguish migrants' children (second generation) from the following generations.[18] However, particular components of migra-

[18] One might also believe that discrimination is easier to measure shortly after the migrant's arrival, with compensatory phenomena coming into force after a while in sectoral or ethnic labor markets (Renaud *et al.*, 1997); it would therefore be more interesting to study recent migrants in a survey or from a question on residence a year before the census.

tion (such as the return of colonists) cannot easily be distinguished from nationality. Although nothing prevents the place of birth and ethnic group from being used simultaneously, when these variables are combined (or cross-tabulated) biases linked to ethnicity are found.

In summary, using monoethnic categories leads to the isolation of nonhomogenous populations, which are made up of people with double (or multiple) membership, and ethnic origin alone does not make it possible to distinguish different generations, which, added to the social aspect of the declaration of origin, can bias the estimate of social distance. Simultaneous approaches by variables not susceptible to biases linked to the social nature of origin appear necessary, and it would be useful to have information on parents' birthplace for populations resulting from recent migration.

3. Longitudinal Observation

It is also necessary to consider the difficulties in measuring the evolution of population components in the context of simultaneous variations in the quality of enumeration,[19] of natural growth, migration, and changes in declarations of origin. Declaration of ethnic origin can vary over short periods. In Australia, between the 1986 census and the postcensus survey, 15% of changes in the declarations of aborigines were observed (Howe, 1994). How can longer-term changes be measured?

The evolution of the population by ethnic group, all ages combined is first of all affected by the different composition of births and deaths with regard to the proportion of mixed-race people, which is interpreted differently according to the method of classifying them. A population that becomes mixed can lose, according to the method of classification used, some of its descendants, which can reduce and even reverse its growth.

The analysis of the evolution of an ethnic group is more precise if it is carried out by birth cohort through the method of expected population (by survival between t and t').

$$P^{t'}_{x+a} = P^{t}_{x} * S_{x,x+a}$$

where S = population's probability of survival (P);

x = age in t;
a = interval t,t'.

However, estimates of mortality by ethnic group often give controversial results, owing to the bias associated with the declaration of different origins for the same person in censuses (denominators) and in civil registration data (numerators), because the declaration is made by the relatives or neighbors of the deceased. This is a first illustration of the problems in using ethnic data for measuring demographic phenomena such as mortality and natality in order to measure demographic trends and socioeconomic characteristics. Indeed, in relation to a stable attribute (e.g., sex, place of birth . . .), a cohort increases or decreases (apart from mortality) only as a result of migration, but it is not the case with an unstable variable susceptible to changes in declaration such as ethnic origin. It is therefore difficult to decide whether the changes are due to migration or changes in declaration, which react one against the other as interference. The ethnic origin variable alone is not sufficient to measure migration. Because migration is frequently ascertained from other sources, we can therefore attempt, by difference, to estimate the changes in the declaration of origin. Ideally, migration data according to the same categories as those of the census or the survey should be available.[20] The use of ethnic group, in censuses, for example, seems to call for its use once more in other statistical fields: civil registration, migration, etc.

In the absence of a direct measurement of migration, changes according to place of birth or the nationality of origin provide an estimate of migration[21] when there is a good match between an ethnic group and these variables. In the case of censuses, migration can also be estimated with a question on former residence, directly for the arrivals and from the expected population for the departures (the difference between the population expected by survival between t and $t'$and population observed in t' who declared residence in the country in the question on former residence in t). The growth by birth cohorts, of an ethnic group that is higher than migration or higher than the growth of another variable that enables the size of this population to be estimated without ambiguity represents an increase in the declaration of this ethnic group in these birth cohorts, and vice versa, lower growth indicates a decrease in declaration.

Thus, in the absence of immigration, an increase in the cohorts of Australian aborigines has been observed since 1981, after a reduction, in the absence of emigration, between 1976 and 1981. In Canada, American Indians have recently experienced an increase owing to changes in the declaration of origin. In Yugoslavia,

[19] The methods of verifying the completeness of the enumeration are usually postenumeration surveys and comparison with other, for example, administrative, sources.

[20] Provided that that the categories are the same and that the declaration of origin is identical and stable.

[21] Provided that the statistical operations are similarly complete.

the Yugoslavs[22] increased between 1961, 1971, and 1981, but they decreased significantly between 1981 and 1991. These variations were often more than 10%. In Australia, the increase resulted from the adoption of laws that are more favorable to the aborigines. In Yugoslavia, the reduction resulted from an increase in competition between the nationalities, with political parties demanding each person to attach themselves to their origins (Mrdjen, 1992). Between the French Polynesian censuses (Baudchon and Rallu, 1993), significant variations are observed around the age of 20, when individuals move out of family homes and complete the questionnaire themselves without always taking into account the declaration that their former household head made when he/she completed the questionnaire on their behalf. The changes should also be considered in relation to migration, which is considerable at these ages.

The evolution of the population's ethnic components is interesting not only at a national level but also at regional and local levels in order to measure phenomena such as confinement and ghetto-ization. The problems that appear at national level can also be found at these levels (evaluation of enumeration quality, migration, and changes in declaration of origin) and necessitate, when the information is available, the same comparisons with stable attributes as at the national level.

Taking into consideration the difficulties of following the trend in numbers and the social nature of origins (e.g., susceptible to vary with the occupational situation), measuring the evolution of the socioeconomic characteristics of the population's ethnic components seems particularly hazardous. Admittedly, rarely are huge changes in declaration observed, but, for example, variations of 5% in the total number concentrated on certain social categories can be enough to make the assessment of the trend in levels of qualifications or occupation impossible.

Finally, the use of ethnic origin, which is not a stable attribute even though it is often considered as such, does not make it possible to separate the factors of change of ethnic groups. As in cross-sectional analysis, it is necessary to use together (or simultaneously) other approaches based on stable attributes in order to confirm the changes.

In conclusion, evaluating the ethnic origin variable is particularly delicate as it is based on different points of view regarding the composition of society, the statistician's objective viewpoint, and the more subjective viewpoint of individuals: that of the majority group and that of the minorities. It represents different realities, social for the statistician and genealogical or biological for others. The fact that the ethnic origin variable includes nonhomogenous populations, including individuals with double (or multiple) origins and of different generations, makes its use difficult, resulting in considerable biases in the measurement of social distance. Last, as an unstable variable, it does not allow a precise measurement of changes, particularly of migration. With regard to minorities resulting from recent migration, the individual's place of birth and that of the parents can be used at the same time as ethnic origin or preferably apart to improve the homogeneity of the populations studied and enable change factors to be analyzed.

It is surprising that current practice most often deviates from the appropriate methods: lists composed of terms referring to various dimensions, frequent changes in the list, imprecise treatment of people of mixed-race, delimitation of nonhomogenous populations, and approximate evaluation of changes without distinguishing the various factors; all of this leads to measurement of social distance and its evolution without the sources of bias even being mentioned. Compared with analyses of stable attributes, the analysis of ethnic origin can be an interesting complement or a source of bias, more especially so as the methods of collection and analysis are approximate. It remains, however, the only variable that enables autochthonous minorities or those whose migration dates back to three generations or more to be studied.

CONCLUSION

Demography has played a preponderant role in ethnic statistics. In so doing, it participates and has always participated in the *power of naming*. For this reason, complicated problems exist in the demographic analysis of ethnicity. The ambivalence still remains as, beyond the numerous problems of production and analysis of statistical categories such as described above, the essentialist and biologizing drift is still possible. Moreover, the increasing ethnic pluralism of modern societies, in association with the migratory flows of the past 30 years are a major challenge in that they question the old definitions of identity based on the idea of an ethnic nation and necessitate new approaches to citizenship (Bauböck and Rundell, 1998).

One of the major questions raised by the emergence of new ethnic minorities resulting from immigration concerns their position within the society and the

[22] Individuals declaring themselves to be of Yugoslav nationality, rather than Serb, Croat, or Muslim, etc., which should not to be confused with Yugoslav citizenship.

nation. In particular, the issue of economic integration finds its meaning in the fact that minority groups have to contend with discriminatory and racial obstacles in the labor market. The numerous antidiscrimination and antiracism measures have created considerable social demand for data providing information on the experience of immigrants and their descendants, and data that identify the groups who suffer most from the mechanisms behind ethnic stratification. Censuses continue to provide the best opportunity for producing such data, insofar as the so-called ethnic categories reflect, if only in an approximate way, new migratory realities. The trend toward self-identification in censuses and the evolution toward subjective and endogenous definitions must be interpreted in this sense.

The evolution of ethnic categories calls into question two traditions of demographic research concerning the use of these categories. Their changing and fluid nature decreases their causal value considerably and therefore makes it difficult to interpret the variations in behavior according to these categories. It is traditional, differential demography that is called into question here. But more fundamentally, the subpopulation approach, which is so dear to demography, must be abandoned as it becomes difficult, if not impossible, to follow ethnic groups from one census to another. Expressed in more technical terms, the comparability of ethnic categories is seriously disrupted by the new definitions. Nevertheless, that which is lost in comparability is gained in legitimacy and validity insofar as minority groups participate in the definition of the categories. Where demography can play a strategic role in studying ethnic groups is in its *statistical power* of information on ethnic inequality. The collection of ethnic data is only legitimate if it is not used to the disadvantage of minority groups (Pool, 1991:16) but serves to reinforce the policies and measures aimed at eliminating racist and discriminatory practices.

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Abramson David, 2002. Identity counts: the Soviet legacy and the census in Uzbekistan, *in:* David I. Kertzer and Dominique Arel (eds.), *Census and identity: the politics of race, ethnicity, and language in national censuses*, p. 176–201. Cambridge, Cambridge University Press, 210 p.

Adepoju Aderanti, 1988. International migration in Africa south of the Sahara, *in:* Reginald T. Appleyard (ed.), *International migration today*, p. 17–88. Paris, UNESCO, 382 p.

Amselle Jean-Loup, 1990. *Logiques métisses: anthropologie de l'identitéen Afrique et ailleurs*. Paris, Payot, 257 p.

Anderson Margo and Fienberg Stephen, 1999. *Who counts? The politics of census-taking in contemporary America*. New York, Russell Sage Foundation, 319 p.

Arel Dominique, 2002. Démographie et politique dans les premiers recensements post-soviétiques: méfiance envers l'État, identités en question, *Population*, vol. 57(6), p. 791–820.

Ballard Roger, 1997. The construction of a conceptual vision: `Ethnic groups' and the 1991 UK census, *Ethnic and Racial Studies*, vol. 20(1), p. 182–194.

Barbary Olivier, 2001. Mesures et réalités de la segmentation socio-raciale à Cali: une enquête sur les ménages afro-colombiens, *Population*, vol. 56(5), p. 773–810.

Barth, Fredrik, 1969. Introduction, *in:* Fredrik Barth (ed.), *Ethnic groups and boundaries: the social organization of culture differences*, p. 9–38. Boston, Little Brown. 153 p.

Baubock Rainer and Rundell John, (eds.), 1998. *Blurred boundaries: migration, ethnicity, citizenship*. Vienna, European Centre, Ashgate, 356 p.

Baudchon Gérard and Rallu Jean-Louis, 1993. Évolution démographique récente dans les TOM du Pacifique 1970–1990, *Population*, vol. 48(4), p. 885–918.

Bean Frank D. and Frisbie W. Parker, 1978. Some issues in the demographic study of racial and ethnic populations, *in:* Frank D. Bean and W. Parker Frisbie (eds.), *The demography of racial and ethnic groups*, p. 1–11. New York, Academic Press, 321 p.

Beaujot Roderic, 1991. *Population change in Canada: the challenge of policy adaptation*. Toronto, McClelland & Stewart, 379 p.

Blion Reynald and Bredeloup Sylvie, 1997. La Côte d'Ivoire dans les stratégies migratoires des Burkinabèet des Sénégalais, *in:* Bernard Constamin and Harris Memel-Fotê (eds.), *Le modèle ivoirien en questions*, p. 707–738. Paris, Karthala and ORSTOM, 802 p.

Blum Alain, 1998. Comment décrire les immigrés? A propos de quelques recherches sur l'immigration, *Population*, vol. 53(3), p. 569–588.

Blum Alain and Gousseff Catherine, 1997. Nationalité, groupes ethniques, peuples: la représentation des nationalités en Russie, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 49–72. Paris, John Libbey and INED, 397 p.

Bonachich Edna, 1999. The site of class, *in:* Rodolfo D. Torres, Louis F. Miron, and Jonathan Xavier Inda (eds.), *Race, Identity, and citizenship*, p. 297–303. Oxford, Blackwell Publishers, 452 p.

Borjas George J., 1994. The economics of immigration, *Journal of Economic Literature*, vol. 32(4), p. 1667–1717.

Bulmer Martin, 1980. On the feasibility of identifying "race" and "ethnicity" in censuses and surveys, *New Community*, vol. 8(1–2), p. 3–16.

Bulmer Martin, 1996. The ethnic group question in the 1991 census of population, *in:* David Coleman and John Salt (eds.), *Ethnicity in the 1991 census*, vol. 1, p. 33–62. London, OPCS and HMSO, 290 p.

Cholsin Harvey M, 1986. Statistics and politics: the "Hispanic Issue" in the 1980 census, *Demography*, vol. 23(3), p. 403–418.

Chrétien Jean-Pierre, 1997. *Le défi de l'ethnisme.* Paris, Karthala, 400 p.

Coleman David, 1994. International migrants in Europe: adjustment and integration processes and policies, *in:* Miroslav Macura and David Coleman (eds.), *International migration: regional processes and responses*, p. 41–76. New York, United Nations, 200 p.

Coleman David and Salt John, 1996. The ethnic group question in the 1991 Census: a new landmark in British social statistics, *in:* David Coleman and John Salt (eds.), *Ethnicity in the 1991 census*, vol. 1, p. 1–32. London, OPCS and HMSO, 290 p.

Cordell Dennis, Gregory Joel W., and Piché Victor, 1996. *Hoe and wage: a social history of a circular migration system in West Africa*. Boulder, Westview Press, 384 p.

Courbage Youssef, 1998. Enquête sur les sources statistiques relatives à la religion, la/les langue(s), le groupe national et le groupe ethnique en Europe, *in:* Werner Haug, Youssef Courbage, and Paul Compton (eds.), *Les caractéristiques démographiques des minorités nationales dans certains États européens*, p. 27–82. Strasbourg, Council of Europe.

Courbage Youssef, 2002. Les minorités nationales en Chine: l'exception démographique?, *in:* Isabelle Attané (ed.), *La Chine au seuil du XXIe siècle*, p. 215–254. Paris, INED and PUF, 602 p.

Davis Darren W., 1997. The direction of race of interviewer effect among African-American: donning the black mask, *American Journal of Political Science*, vol. 41(1), p. 309–322.

Desrosières Alain, 1993. *La politique des grands nombres. Histoire de la raison statistique*. Paris, La Découverte, 456 p.

Dieckhoff Alain, 2000. *La nation dans tous ses États*. Paris, Flammarion, 355 p.

Dozon Jean-Pierre, 1997. L'étranger et l'allochtone en Côte-d'Ivoire, *in:* Bernard Constamin and Harris Memel-Fotê (eds.), *Le modèle ivoirien en questions*, p. 779–798. Paris, Karthala and ORSTOM, 802 p.

Gayon Jean, 2001. Faut-il proscrire en biologie l'expression "races humaines" (UNESCO 1950–51)? *L'aventure humaine*, no. 12, p. 9–40.

Glazer Nathan, Moynihan Daniel P., and Schelling Corinne S., (eds.), 1975. *Ethnicity, theory and experience*. Cambridge, Harvard University Press, 531 p.

Goldman Gustave J., 1997. La mesure de l'ethnicité au Canada: évolution et débats actuels, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities*, p. 31–48. London, John Libbey Eurotext, 397 p.

Goldman Gustave J. and McKenney Nampeo (eds.), 1993. *Les défis que pose la mesure de l'origine ethnique: science, politique et réalité*. Washington, U.S. Government Printing Office, 593 p.

Goldstein Joshua and Morning Ann, 2002. Back in the box: the dilemma of using multiple-race data for single-race law, *in:* Joel Perlmann and Mary C. Waters (eds.), *The new race question: how the census counts multiracial individuals*, p. 119–136. New York and Annandale-on-Hudson, Russell Sage Foundation Press and Levy Economics Institute of Bard College, 398 p.

Grieco Elisabeth M. and Cassidy Rachel C., 2001. *Overview of race and Hispanic origin*. Washington, U.S. Census Bureau, 11 p.

Halli Shiva S. and Driedger Leo (eds.), 1999. *Immigrant Canada: demographic, economic, and social challenges* Toronto, University of Toronto Press, 358 p.

Haug Werner, 1998. Introduction: la statistique des minorités prise entre la science et la politique, *in:* Werner Haug, Youssef Courbage, and Paul Compton (eds.), *Les caractéristiques démographiques des minorités nationales dans certains États européens*, p. 11–26. Strasbourg, Council of Europe.

Hout Michael and Goldstein Josuah R., 1994. How 4.5 million Irish immigrants became 40 million Irish Americans: demographic and subjective aspects of the ethnic composition of white Americans, *American Sociological Review*, vol. 59(1), p. 64–82.

Howe Andrew, 1994. *Experimental estimates for the Aboriginal and Torres Strait islander population 1986–1991*. Canberra, Australian Bureau of Statistics, Demography Section, 28 p.

Kertzer David I. and Arel Dominique (eds.), 2002a. *Census and identity. The politics of race, ethnicity and language in national censuses*. Cambridge, Cambridge University Press, 240 p.

Kertzer David I, and Arel Dominique, 2002b. Censuses, identity formation, and the struggle for political power, *in:* David I. Kertzer and Dominique Arel (eds.), *Census and identity: the politics of race, ethnicity, and language in national censuses*, p. 1–42. Cambridge, Cambridge University Press, 210 p.

Lacorne Denis, 2001. Je n'y crois pas, mais quand même. Mesure et malmesure de l'ethnicité aux États-Unis, *in:* Michel Wieviorka and Jocelyne Ohana (eds.), *La différence culturelle: une reformulation des débats (Colloque de Cergy)*, p. 381–392. Paris, Balland, 479 p.

Lapeyronnie Didier, 1993. *L'individu et les minorités: la France et la Grande-Bretagne face à leurs immigrés*. Paris, PUF, 362 p.

Le Bras Hervé, 1998. *Le démon des origines: démographie et extrême droite*. Paris, Édition de l'Aube, 260 p.

Lieberson Stanley and Waters Mary C., 1988. *From many strands: ethnic and racial groups in contemporary America*. New York, Russell Sage Foundation, 289 p.

Lorcerie Françoise, 1994. Les sciences sociales au service de l'identité nationale, *in:* Denis-Constant Martin (ed.), *Cartes d'identité. Comment dit-on "nous" en politique?* p. 245–263. Paris, Presses de la FNSP, 304 p.

Martiniello Marco, 1995. *L'ethnicité dans les sciences sociales contemporaines*. Paris, PUF, 128 p.

Massey Douglas S., Arango Joaquin, Hugo Graeme, Kouaouci Ali, Pellegrino Adela, and Taylor J. Edward, 1998. *Worlds in motion: understanding international migration at the end of the millennium*. Oxford, Clarendon Press, 362 p.

Moore Robert, 1997. *Positive action in action: equal opportunities and declining opportunities on Merseyside*. Aldershot, Ashgate, 129 p.

Mrdjen Sznejana, 1992. Qui sont les Yougoslaves? *in:* Michel, Roux (ed.), *Nations, États et territoires en Europe de l'Est et URSS*, p. 259–266. Paris, L'Harmattan, 294 p.

Ni Brochlain Maire, 1990. The ethnicity question for the 1991 census: background and issues, *Ethnic and Racial Studies*, vol. 13(4), p. 542–567.

Nobles Melissa, 2000. *Shades of citizenship: race and the census in modern politics*. Stanford, Stanford University Press, 256 p.

Oommen T. K., 1997. *Citizenship, nationality and ethnicity*. Cambridge, Polity Press, 270 p.

Otayek René, 2000. *Identité et démocratie dans un monde global*. Paris, Presses de Sciences Po, 228 p.

Pendakur Ravi, 2000. *Immigrants and the labour force*. Montreal, McGill-Queen's University Press, 246 p.

Perlmann Joel and Waters Mary C. (eds.), 2002. *The new race question: how the census counts multi-racial individuals*. New York, Russell Sage Foundation Press and Levy Economics Institute of Bard College, 398 p.

Petersen William, 1997. *Ethnicity counts*. New Brunswick, Transaction Publishers, 323 p.

Piché Victor, 2002. Immigration, diversity and ethnic relations in Québec, *Canadian Ethnic Studies Journal*, vol. 34(3), p. 5–27.

Piché Victor and Poirier Jean, 1995. Divergences et convergences dans les discours et théories de la transition démographique, *in:* Dominique Tabutin, Thierry Eggerickx, and Catherine Gourbin (eds.), *Transitions démographiques et sociétés*, p. 111–132. Louvain-la-Neuve and Paris, Académia/L'Harmattan, 692 p.

Pool Ian, 1991. *Te Iwi Maori: a New Zealand population past, present & projected*. Auckland, Auckland University Press, 271 p.

Portes Alejandro, 1984. The rise of ethnicity, *American Sociological Review*, vol. 49(3), p. 383–397.

Poutignat Philippe and Streiff-Fenart Jocelyne, 1995. *Théories de l'ethnicité*. Paris, PUF, 270 p.

Pryor Edward T., Goldmann Gustave J., and Royce Donald A., 1991. Future issues for the Census of Canada, *International Migration Review*, vol. 25(93), p. 167–175.

Pryor Edward T., Goldmann Gustave J., Sheridan Michael, and White Pamela, 1992. Measuring ethnicity: is "Canadian" an evolving indigenous category? *Ethnic and Racial Studies*, vol. 15(2), p. 214–235.

Renaud Jean, Piché Victor, and Gingras Lucie, 1997. Immigration et intégration économique à Montréal, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds), *Old and new minorities/Anciennes et nouvelles minorités*, p. 89–111. Paris, INED and John Libbey, 398 p.

Schor Paul, 1999. Les "immigrés," objets et sujets des statistiques: la négociation des catégories de l'origine dans le recensement américain, 1860–1940, *Raisons Politiques*, no. 2, p. 27–47.

Simon Patrick, 1997. La statistique des origines. "Race" et ethnicité dans les recensements aux Etats-Unis, Canada et Grande-Bretagne, *Sociétés Contemporaines*, no. 26, p. 11–44.

Simon Patrick, 1998. Nationalité et origine dans la statistique française: les catégories ambiguës, *Population*, vol. 53(3), p. 541–567.

Simon Pierre-Jean, 1983. Le sociologue et les minorités: connaissance et idéologie, *Sociologie et sociétés*, vol. 15(2), p. 9–22.

Skerry Peter, 2000. *Counting on the census? Race, group identity, and the evasion of politics*. Washington, Brookings Institution Press, 261 p.

Statistics New Zealand, 1993. *New Zealand's multicultural society: 1991 census of population and dwellings*. Wellington, Department of Statistics, 138 p.

Steph Thernstrom (ed.), 1980. *Harvard encyclopedia of American ethnic groups*. Cambridge, The Belknap Press of Harvard University Press, 1076 p.

Streiff-Fenart Jocelyne, 1997. Présentation, *in:* Jean-Louis Rallu, Youssef Courbage, and Victor Piché (eds.), *Old and new minorities/Anciennes et nouvelles minorités*, p. 3–8. Paris, John Libbey and INED, 397 p.

Taguieff Pierre-André, 1991. Le nationalisme des nationalistes. Un problème pour l'histoire des idées politiques en France, *in:* Gil Delannoi and Pierre-André, Taguieff (eds.), *Théories du nationalisme: nation, nationalité, ethnicité*, p. 47–124. Paris, Kimé, 324 p.

Telles Edward E. and Lim Nelson, 1998. Does it matter who answers the race question? Racial classification and income in Brazil, *Demography*, vol. 35(4), p. 465–474.

Tribalat Michelle, 1995. *Faire France: une enquête sur les immigrés et leurs enfants*. Paris, La Découverte, 232 p.

Uvin Peter, 2002. On counting, categorizing and violence in Burundi and Rwanda, *in:* David I. Kertzer and Dominique Arel (eds.), *Census and identity: the politics of race, ethnicity, and language in national censuses*, p. 148–175. Cambridge, Cambridge University Press, 210 p.

Waters Mary C., 1990. *Ethnic options: choosing identities in America*. Berkeley, University of California, 197 p.

White Pamela M., Badets Jane, and Renaud Viviane, 1993. Mesure de l'origine ethnique dans les recensements du Canada, *in:* Gustave Goldman and Nampeo Mckenney (eds.), *Les défis que pose la mesure de l'origine ethnique: science politique et réalité*, p. 259–293. Washington, U.S. Government Printing Office, 593 p.

Wood Charles H., 1982. Equilibrium and historical-structural perspectives on migration, *International Migration Review*, vol. 16(2), p. 298–319.

Zuberi Tukufu, 2001. *Thicker than blood: how racial statistics lie*. Minneapolis and London, University of Minnesota Press, 193 p.

CHAPTER

96

Genetics and Demography

ENZO LUCCHETTI AND LAMBERTO SOLIANI

Dipartimento di Genetica Antropologia Evoluzione, Università degli Studi di Parma, Parma, Italy

Dipartimento de Scienze Ambientali, Università degli Studi di Parma, Parma, Italy

INTRODUCTION

The collective genetic heritage (*gene pool*) of a population is made up of all of the hereditary information (*genes*) carried by individuals in the population. The goals of human population genetics are to study the structure of the gene pool, to identify the modes and mechanisms that control the transmission of genes from one generation to the next, and to track the underlying variations and their causes.

The intersection between demography and population genetics is clear: both fields study the structure of human groups and their evolution, distribution in space, and perpetuation over time. However, they differ in their approach to the topic. Demographers are interested in individuals and their behavior (e.g., the population structure by age, sex, marital status, survival or mortality, nuptiality, fertility, mobility), whereas population genetics focuses on the allelic structure of the gene pool and the probability of change over time in the relative frequency of alleles. Interaction between the disciplines is intrinsic to the phenomena themselves, especially with reference to the mechanisms and modes by which a population is perpetuated over time. On the one hand, genes are not independent entities; their dynamic can only be studied in relation to the characteristics and behaviors (notably demographic) of their carriers. On the other hand, demographic phenomena, such as individual behaviors, are partly determined by individuals' genetic heritage.

In the first section of this chapter, we review the main components of the genetic structure of a population. We then examine and assess the principal models used in population genetics. Next, we study the evolutionary mechanisms of population genetic structure, using both theory and examples. We conclude the chapter by discussing the important role that the study of family names can play in the intersection of genetics and demography.

I. THE STRUCTURE OF A POPULATION'S GENE POOL

The characterization of the collective genetic heritage of a population is based on the observation of *phenotypes*, manifestations of hereditary characteristics that can be discerned by the use of our senses or using available instruments. For example, *eye color* can have the *blue* or *brown* phenotype; for the characteristic *ABO blood type*, we distinguish the phenotypes *type A*, *type B*, *type AB*, and *type O.*

The phenotypic manifestation of hereditary characteristics is determined by the *genotype*, the set of genes that the individual received from his or her parents. For each autosomal characteristic[1] the individual has two genes, one inherited from the mother and one from the father. These two genes form the individual's genotype for this characteristic.

Alleles are different forms of the same gene or different hereditary information for the same characteristic. For the characteristic *eye color*, genetic information

[1] For clarity, this discussion will be limited to autosomal characteristics, characteristics that do not differ by sex.

TABLE 96–1 Characteristics, Phenotypes, Genotypes, and Alleles

Characteristic: Eye color		Characteristic: MN blood group		Characteristic: ABO blood group	
Alleles: C, c		Alleles: M, N		Alleles: I^A I^B, i	
Relationship: C, dominant; c, recessive		Relationship: codominant		Relationship: I^A and I^B, codominant; i, recessive	
Phenotype	Genotype	Phenotype	Genotype	Phenotype	Genotype
Brown eyes	CC, Cc	Type M	MM	Type A	$I^A I^A$, $I^A i$
Blue eyes	cc	Type MN	MN	Type B	$I^B I^B$, $I^B i$
		Type N	NN	Type AB	$I^A I^B$
				Type O	$i\ i$

might be *blue eyes* (*c* allele) or *brown eyes* (*C* allele); for the characteristic *blood type* there are three principal types of allele, designated by the symbols I^A, I^B, and i. An individual who carries two identical alleles for a given trait is called a *homozygote*; an individual whose genotype consists of two different alleles is called a *heterozygote*.

Different alleles have different capacities to determine phenotypes. An allele is called *recessive* when the information it carries is only manifested if the other allele is identical (condition of homozygosity); it is called *dominant* if its manifestation is independent of the other allele; and two alleles are *codominant* if, when present in the same individual, they interact to produce a phenotype different from that produced by either of the two alleles in homozygous condition (Table 96–1).

Studies of population genetics start by observing phenotypes (phenotypic frequencies) in order to deduce genotypic and allelic frequencies (Appendix 96–A). Allelic frequencies are the base parameter. A population is characterized by the relative frequency of the alleles in its gene pool, and studying the genetic evolution of a population means analyzing the internal dynamics that determine changes in allelic frequencies from one generation to another.

II. MODELS OF POPULATION GENETICS

Population genetics uses explanatory models to study changes in allelic frequencies. Even the most refined models are merely theoretical instruments and can never achieve a complete description of reality. However, models have been proven to be generally effective in identifying the forces and mechanisms of genetic evolution and analyzing their modes of action.

1. Panmixia

The simplest model, *panmixia*, assumes that evolution of the gene pool over generations is solely determined by allelic frequencies. The birth of a new individual (or the formation of a new genotype) is equivalent to two independent selections from the gene pool of the parent generation. If p and q are the frequencies of the A and a alleles of an autosomal bi-allelic system ($p + q = 1$), the probability that a new individual has the genotype AA is p^2, $2pq$ is the probability that the individual is an Aa heterozygote, and q^2 is the probability of aa homozygosity. The Hardy–Weinberg theorem, named after its inventors, describes the relationship between allelic frequencies (p, q) and genotypic frequencies, given by $(p + q)^2$. In the new generation, a fraction p^2 of individuals will have the AA genotype, a fraction $2pq$ will have the Aa genotype, and a fraction q^2 will have the aa genotype.

The model is stable: allelic frequencies remain constant across generations, and the underlying distribution of genotypes and phenotypes is also unchanging (panmictic equilibrium). A population that is initially unbalanced (i.e., with a significant difference between the observed genotypic frequencies and the frequencies expected based on the model) reaches equilibrium in the first generation for which the model's hypothesis holds. Equilibrium is then maintained as long as the hypothesis holds (Table 96–2).

2. Random-Mating Model

Populations perpetuate themselves via sexual reproduction, which requires the formation of couples.

TABLE 96–2 Panmictic Model

Allelic frequencies in the parent generation	Genotypic composition of the filial generation	Allelic frequencies in the filial generation
Allele *A*: frequency = p	Genotype *AA*: $p*p = p^2$	p_f = $(2p^2 + 2pq)/2 = \mathrm{p}$
Allele *a*: frequency = q	Genotype *Aa*: $2*p*q = 2pq$	
	Genotype *aa*: $q*q = q^2$	$q_f = \mathrm{x}$ $(2p^2 + 2pq)/2 = \mathrm{q}$

TABLE 96–3 Random-mating Model

Type of pairing	Frequency	Distribution of offspring according to genotype		
		AA	***Aa***	***Aa***
AA **AA*	p^4	p^4		
*AA** *Aa*	$4p^3q$	$2p^3q$	$2p^3q$	
*AA** *aa*	$2p^2q^2$		$2p^2q^2$	
*Aa** *Aa*	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
Aa aa*	$4pq^3$		$2pq^3$	$2pq^3$
aa aa*	q^4			q^4
Total	1	p^2	$2pq$	q^2

The reference model (random mating) assumes that couples are formed randomly with respect to genotypes (or phenotypes) and that their contribution to the perpetuation of the group does not vary systematically. Table 96–3 describes the model: the possible types of pairings (in terms of genotypes), their relative frequency, and the distribution by genotype of the progeny of each couple. It is clear that the condition of *panmixia* (random extraction of pairs of alleles) coincides with that of *random mating* (random formation of couples).

III. MECHANISMS OF VARIATION OF GENETIC STRUCTURE

These models assume that the population is large (theoretically infinite) and that individuals (classed by phenotype or genotype) all contribute equally to the formation of the filial generation, with each individual having an equal probability of coupling and no differences in fertility or mortality. They also assume that there is no *mutation* (passage from one allelic form to another), so allelic forms remain unchanged and that the populations are not affected by migration (or at least that there is no differential migration by phenotype or genotype). When these conditions are met, the population is said to be in panmictic equilibrium; allelic frequency (and genotypic and phenotypic frequency) are stable across generations.

There are several situations in which these hypotheses are not met.

1. Assortative Mating or Nonrandom Formation of Couples

We speak of *assortative mating* when the probability of couple formation is not independent of the phenotype (or genotype) of the partners. Assortative mating can act in a positive direction (if a couple is more likely to form when the partners are alike than when they are different) or in a negative direction (in the reverse case), and can act either completely or partially.

Positive assortative mating (when partnerships are more likely to form—or only form—between bearers of the same trait) does not cause variation in allelic frequencies, but it does imply a different distribution of genotypes and phenotypes than that expected under the panmictic model. In effect, the population is divided into several subgroups; couples form between individuals of the same subgroup. In the case of *complete genotypic sorting*, for example, the population is divided into three groups, each made up of individuals with the same genotype—*AA*, *Aa*, and *aa*, respectively. All children of *AA*AA* couples and *aa*aa* couples belong to the same groups as their parents, whereas children born to *Aa*Aa* couples are distributed among the three groups in a 1 : 2 : 1 ratio. In each generation, the frequency of heterozygotes declines, whereas the frequency of homozygotes rises. The *AA* and *aa* groups of homozygotes each increase by half the amount by which heterozygotes were reduced. Allelic frequencies, however, remain constant. After many generations, heterozygotes approach extinction, and almost all individuals are of *AA* or *aa* genotypes. The relative frequencies of these two genotypes approach p and q. At that point, the two remaining groups are completely separated, forming two distinct populations between which there is no exchange of genes.

Consanguinity, marriages between individuals with common ancestors, can be seen as a particular form of assortative mating: unions are formed between individuals carrying a higher than average proportion of identical alleles, with the proportion being higher the more closely the two individuals are related. These identical alleles have not only the same effect but also the same origin, because they are pairs of alleles that were present in the couple's common ancestors. Children born to consanguine marriages show a higher frequency of homozygous *loci* and a lower proportion of heterozygous *loci* than do other children. The probability that recessive alleles will be manifested at the phenotypic level increases, which explains the higher frequency of hereditary diseases among children of consanguine parents.

Negative sorting means that couples are more likely to form between individuals with different characteristics. Sex is the classic example, although the example is not quite correct: the two chromosomes X and Y, which jointly determine sex (XX for women, XY for men), can be seen as two allelic forms of a gene. Pairings only take place between individuals of the opposite sex. From these pairings, individuals of both sexes are born, in a 1 : 1 ratio. The system is stable, with the frequency of the two alleles constant at 0.75 and 0.25. Only two genotypes (or phenotypes) are possible (XX and XY), male or female, each with a frequency of 0.5.

2. Selection: Differences in Contribution to Subsequent Generations

Selection occurs when individuals' contributions to subsequent generations differ as a function of the genetic information they carry. Survival and fecundity, the two primary components of the biological capacity for reproduction, are largely determined by genetic heritage, although fertility also depends on many other factors, with socioeconomic and cultural factors playing a growing role.

The relative effective contribution to the formation of subsequent generations is measured using the *index of fitness.* A *fitness* of one is assigned to the individuals (or couples) who have the largest number of *successful offspring*, that is, children who survive and reproduce in turn. The value of the index of fitness for other groups is calculated by dividing their average number of successful offspring by that of the group with fitness 1. The *index of selection* describes the phenomenon from the opposite perspective and is calculated by subtracting the index of fitness from one. Selection is set at zero for the group whose contribution is largest and is equal to one for the group with zero contribution (fitness equal to zero).

Table 96–4 gives an example in which the average number of successful offspring varies by genotype and shows the relative values of fitness and selection for each group. For each successful child produced by an *AA* individual, *Aa* and *aa* individuals produce 0.75 and 0.5 children, respectively. It does not matter for the purpose of this example whether all *Aa* and *aa* individuals have lower effective fertility do than *AA* individuals or whether, for example, 75% of *Aa* individuals reproduce as do *AA* individuals and 25% do not reproduce at all, or the survival of children born to *Aa* individuals is 75% that of children born to *AA.*

TABLE 96–4 Indices of Fitness and Selection

Genotype	Average no. of successful offspring	Indices of	
		Fitness	Selection
AA	4	1	0
Aa	3	0.75	0.25
aa	2	0.50	0.50

Selection can act at different levels: in the formation of gametes, during the course of the definition of the genotype, or in the expression of the phenotype. It can act through mortality or differential survival, through differences in fecundity or fertility, or even through differences in the capacity to raise children after their birth. In all cases it influences the overall results through differential contributions to the perpetuation of the population. Because of the large reduction in infant mortality, the number of children born alive now provides a good estimate of fitness.

Categories (gametes, genotypes, phenotypes) that are associated with reduced fitness are submitted to *negative selection*; we speak of *positive selection* for groups with higher fitness. To indicate the object and the direction of selection, expressions such as *selection against the recessive allele, against the recessive phenotype, selection in favor of*, or *against heterozygotes*, are used. Appendix 96–B explains in more detail the effects of different types of selection on the genetic structure of a population.

Over the course of generations, the continuous action of selection reduces the frequency (or even completely eliminates) of the alleles being selected against and the genotypes and phenotypes formed by these alleles. Selections against the heterozygote constitute a special case, because this genotype contains different allelic forms (Appendix 96–B). *Selection against the heterozygote*, when it has a lower fitness than do the two homozygous genotypes, leads to fixation of the allele with a higher initial frequency. *Selection in favor of the heterozygote* (heterozygote with higher fitness than homozygotes), in contrast, leads to a stable equilibrium, determined by the intensity of selection against the two homozygotes independently of their initial allelic frequencies.

The direction of selection is not always constant. The eradication of malaria is a classic example of a process that has changed the direction of genetic selection. When a population is subject to both thalassemia and malaria, selection favors the heterozygote, to the detriment of the two homozygous forms. However, with the disappearance of malaria, selection acts against the thalassemia homozygote and, to a lesser degree, against the heterozygote.

3. Mutations

The expression *genetic mutation* signifies the passage from one allelic form to another. Mutations occur naturally in the course of the DNA replication that precedes the formation of gametes. Certain environmental conditions can influence the frequency of mutations or *mutation rate.* The most common effect of a mutation is the loss of a function: for example, an error in replication might cause an allele to lose the information necessary for the synthesis of an enzyme. Such a mutation is equivalent to the passage from the dominant allelic form to a recessive form: individuals that are homozygous with respect to the new allele are not able to synthesize the enzyme, and the reaction that requires its presence cannot take place in these individuals. Mutation in the opposite direction (*retro-mutation*), which confers a previously missing functional capacity on an allele, is less frequent but still possible.

Mutations change allelic frequencies. Given a rate of mutation μ from A to a, in each generation a number $p\mu$ of A alleles mutate to a. In subsequent generations, the allelic frequency p of A varies according to the equations:

$$p_1 = p_0 - \mu p_0 = p_0\,(1-\mu)$$

$$p_2 = p_1 - \mu p_1 = p_1(1-\mu) = p_0\,(1-\mu)(1-\mu) = p_0\,(1-\mu)^2$$

$$p_n = p(1-\mu)^n.$$

If, in addition to the mutations from A to a, retro-mutations also take place at a rate ν, in each generation μp A alleles mutate to a and νq alleles mutate from a to A. If the two quantities are equal, the allelic frequencies do not change. Equilibrium is thus determined by the values of the mutation and retro-mutation rates, not by the initial values of the allelic frequencies. In an equilibrium state

$$\hat{p} = \frac{\nu}{\mu+\nu} \quad \text{and} \quad \hat{q} = \frac{\mu}{\mu+\nu}$$

Mutation-Selection Interaction

A mutation that causes the loss of a function, such as the capacity to synthesize an enzyme, can lead to the phenotypic manifestation of defects or disease in the homozygote. In this case, there will be selection against the recessive phenotype. Selection tends to eliminate harmful alleles, whereas mutations continue to reform them. If s is the intensity of the action of selection and q is the allelic frequency of a, sq^2 individuals will be eliminated in each generation as aa homozygotes are struck by selection. Mutations will simultaneously recreate $\mu p \cong \mu$ new a alleles. Equilibrium conditions are defined by

$$\hat{q} = \sqrt{\mu/s}.$$

If there is complete selection ($s = 1$), the equilibrium frequency of a becomes:

$$\hat{q} = \sqrt{\mu}.$$

4. Individual Mobility and Migration

When individuals migrate, they take their genes with them. Depending on the characteristics of the sending and receiving populations and the intensity and direction of migration, migratory flows can modify the genetic structure of the populations involved. Models are used for reference in studying the effects of migration. Migration-induced variations are generally evaluated based on allelic frequencies in the gene pool of the receiving population as a whole (Appendix 96–C). Migration often leads to the constitution of subgroups within the receiving population; couples form preferentially within each subgroup (assortative mating based on origin). Usually, it requires many years, or even many generations, before groups of immigrants can be considered to be integrated into the receiving population to the point at which the hypotheses of the random-mating model can be applied to the population as a whole.

Migration may also influence fertility. The average number of children in immigrant families is often different from that of native-born families. Migrants may have higher fertility owing to the conservation of traditions specific to the population of origin or, fertility may be lowered by difficulties caused by migration itself.

Appendix 96–C illustrates some of the effects of migration on the genetic structure of populations. These effects can vary depending on the direction and strength of migratory flows, allelic frequencies in the populations in question, and, finally, the subgroup structure that develops in the host population. Models address one phenomenon at a time; they can be combined and integrated to describe interaction between different processes taking place in the same population.

Because migrants are not a random genetic sample of the population of origin, migration can also induce changes in the sending population. In the models considered here, migratory behavior is hypothesized to be independent of the genotype or phenotype of the individual, but it is nevertheless possible that genetic characteristics could affect the propensity to migrate. For example, migration by groups (family groups, groups of relatives, or inhabitants of the same village) can produce concentrations of specific genetic markers, which differentiate migrants based on their population of origin.

5. Genetic Drift

Genetic drift describes the effect of random variations occurring in a finite population with an intensity inversely proportional to the size of the group. The action of drift, which accumulates from generation to generation, can lead to fixation of one of the initial allelic forms. It is impossible to predict which allelic form will be fixed. In general, the allelic form with the highest frequency has the highest probability of being fixed; if several populations with the same initial allelic frequencies are considered (p for allele A and q for a), the most likely outcome is that allele A becomes fixed in a proportion p of the populations and a in a proportion q. However, this distribution is only the most probable distribution; the outcome in a given population cannot be specified a priori.

The *fixation rate* is a useful parameter for describing the dynamics of finite populations: it indicates the proportion of alleles established in each generation in a population, that is, the proportion of the population in which an allele is fixed.

Genetic drift reduces the internal variation in a population, thus limiting the possibility of adaptation; that is, in evolutionary terms, a population's capacity to survive in the face of environmental changes. At the extreme, when all individuals in a population have the same phenotype, environmental variation incompatible with this phenotype can lead to the extinction of the entire population. However, genetic drift increases differences between populations, by causing the fixation of different alleles in initially similar populations. One or more of these populations might contain an allele more capable of responding to environmental variation. Only these populations would survive, possibly occupying vacated ecological niches. Drift cannot be defined a priori as having positive or negative effects on a population's survival capacity; it produces an advantage in some cases and a disadvantage in others.

Migration counteracts drift by increasing internal variability and reducing divergence between populations. In the presence of migration, even at low rates, the index of fixation by drift decreases rapidly. Mutations also work in opposition to drift, but because mutation rates are generally much lower than are migration rates, their effects are even weaker.

IV. ANALYSES AND APPLICATIONS

The development of electrophoresis techniques has led to great advances in the observation of allelic frequencies. These techniques make it possible to observe polymorphism at the proteic level; today, DNA sequencing techniques have opened the way for new and more radical means of observing genetic variability, based on the sources of information as well as it manifestations.

1. Geographic Distribution of Allelic Frequencies

Graphing techniques can be used to establish the spatial distribution of allelic frequencies as a function of the geographic location of populations. Lines of equal frequency, analogous to level curves uniting points at the same altitude on a relief map, mark zones where populations share the same allelic frequency, and color and shadings distinguish geographic areas with different frequencies.

When these techniques are applied to the different regions of Italy (Fig. 96–1), zones with strong or weak diffusion of an allele are immediately visible, as is the existence of a gradient. In this situation, the distribution of allelic frequencies is the result of specific patterns of diffusion produced by both migration and factors of isolation. Environmental characteristics can also influence the frequency distribution. For instance, because the hemoglobin allele S confers a relative resistance against malaria, the distribution of the allele is associated with the presence of the disease.

The works of Arthur Mourant (Mourant, 1954; Mourant *et al.*, 1976) provide a remarkable set of observations on this subject. The monumental work of Luigi Cavalli-Sforza *et al.* (1994) is more up-to-date both in the data collected and in the methods used.

2. Evolutionary Reconstructions and the Relationship between Populations

The relationships between populations, and their genetic similarity (or distance), can provide information about their biological evolution. Clearly, two populations that have recently descended from the same

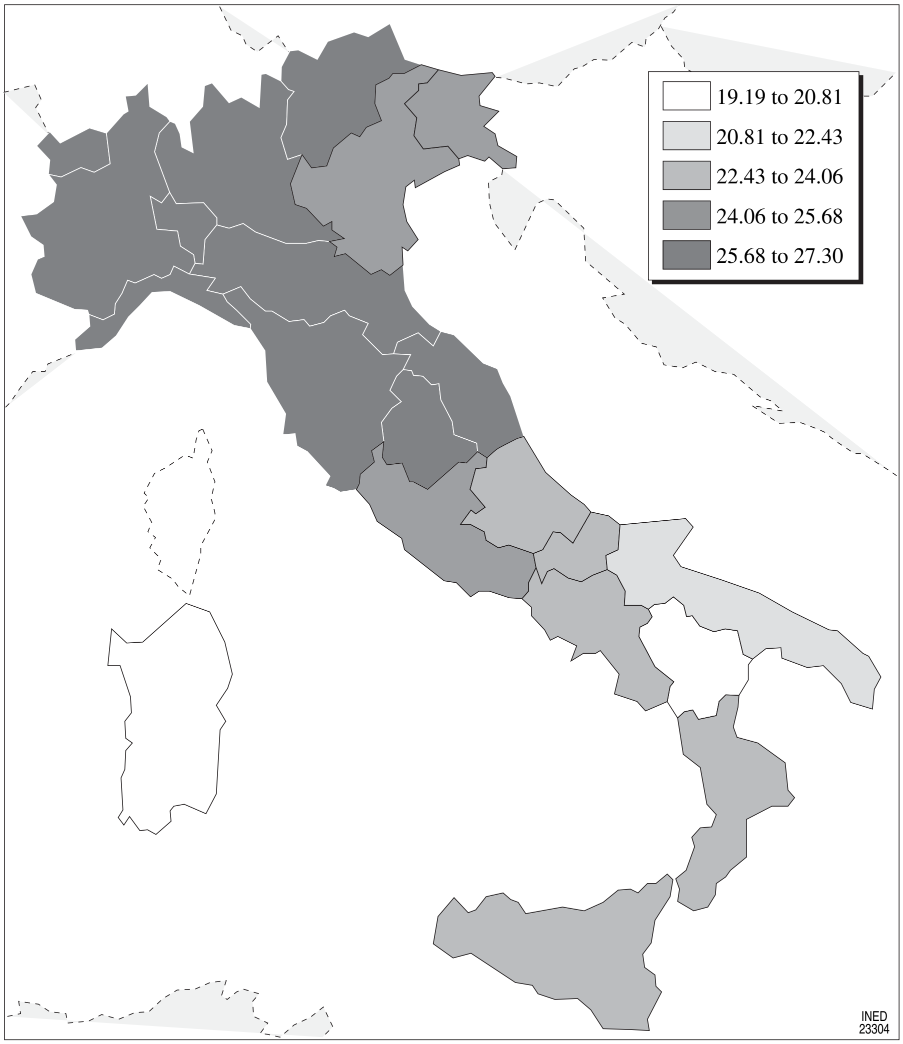


FIGURE 96–1 Average frequencies of I^A alleles (ABO system) in regions of Italy. (From Soliani *et al*., 1981.)

ancestral population will have similar allelic frequencies. However, populations that have been evolving separately for a long period will have more different distributions of allelic frequencies. It is even possible to estimate the time elapsed since the separation of two populations by hypothesizing a relationship between the degree of differentiation between the populations and the time necessary to develop this difference. That is, comparing the gene structures of populations makes it possible to formulate hypotheses about their evolution and the relationships between them. Conclusions should be accepted with caution; genetic similarity can result if populations are submitted to analogous factors of selection and does not necessarily imply a common evolutionary path. In addition, the relationship between genetic difference and time elapsed is not always linear.

3. Genetic Distance, Distance Matrices, and Methods of Representation

Indices of genetic distance provide a synthetic measure of divergence between the gene pools of different populations. There are numerous methods for calculating the genetic distance between populations; Jorde (1980) and Lalouel (1980) provide overviews of these methods that are both detailed and wide in scope.

One of the most commonly used measures of genetic diversity is Masatoshi Nei's *D* value (1972, 1978). Nei's *D* is calculated by using the probability of having identical alleles on the same locus chosen randomly from the gene pool of the two populations being compared, weighted by the probability of equality within each population. If p_{Ai} and p_{Bi} are the frequencies of the allele i in populations A and B, respectively, then:

$$I_{AB} = \Sigma p_{Ai} p_{Bi},\ I_{AA} = \Sigma p_{Ai}^{2} \text{ and } I_{BB} = \Sigma p_{Bi}^{2};$$

the normalized identity *I* is given by the ratio

$$I = I_{AB} / \sqrt{I_{AA} I_{BB}}$$

and the standard genetic distance is

$$D = -\ln(I).$$

Anthony Edwards and Luigi Cavalli-Sforza (1965) proposed a geometrically based method. Populations are represented by points whose coordinates correspond to the square roots of the allelic frequencies. Because the sum of the allelic frequencies of each system is equal to one, these points are located on the surface of a hypersphere with radius 1 and axes given by the square root of the frequencies. In a bi-allelic system ($p + q = 1$), the population-points are positioned on the circumference with the equation $x^2 + y^2 = 1$, with $x = \sqrt{p}$ pand $y = \sqrt{q}$. In a tri-allelic system $p + q + r = 1$, they are on the surface of a sphere. The distance between two populations can be defined as a function of the arc defined by the two points or of the length of the chord that joins them. The chord provides a linear distance and makes it possible to combine distances calculated from independent genetic systems (characteristics carried on different chromosomes) geometrically.

Researchers argue over the usage of genetic distances, the theoretical conditions required for validity of the measures, and the interpretation of results. However, both theoretical and empirical examples show a strong correlation between the different methods of calculating genetic distance, generally between 0.8 and 0.95 (Hedrick, 1975).

4. Graphical Representations

Distance matrices, constructed from the indices of diversity between populations taken in pairs, can be used to produce graphic representations which are easier to read and interpret than distance measures. The most common approaches are tree diagrams (or *dendrograms*) and topological maps.

The *dendrogram* describes the evolutionary relationships between populations. It assumes that currently existing groups are derived from the same ancestral population submitted to a series of splits, and that the cumulated genetic difference between two populations is a function of the time elapsed since their separation. Starting with the distance matrix, it is possible to retrospectively reconstruct the order of splits and the probable evolutionary paths (*phyletic trees*) traced by the populations. Different methods often produce different results; specialized techniques, such as bootstrap resampling (Efron, 1982), can be used to assess the credibility of the trees produced and to identify the most likely tree.

At the conceptual level, one problem with the dendrogram is that splits are generally not followed by complete separations between populations, and genetic flows, in the form of individual migrations, continue between groups belonging to different branches. Satisfactory methods taking into account these postseparation genetic flows do not yet exist.

Take, for example, the relationships between Italian regions, each characterized by allelic frequencies of the classic blood groups (ABO, MNSs, Rh haplotypes). We first recorded the average frequency of each allele in each region, then constructed the between-region similarity matrix. From there, we used the program Mega2 (Kumar *et al.*, 2001) to produce the dendrogram in Fig. 96–2, which describes the relationships between the gene pools of different regions. The uniqueness of Sardinia is immediately evident; a second division separates the southern regions from the center and north regions. These groupings are in turn progressively subdivided into distinct subgroups. The dendrogram as a whole reflects the geographic dispersion of the regions, their history, and the relationships that they have maintained over the past centuries fairly well.

Topological maps are based on methods such as principal components analysis and multidimensional scaling. In general, they provide good descriptions of the relationships between groups and are more informative than are tree diagrams when there are clear divisions among the groups or when the goal is to compare results obtained by using different data sets. Meaning is assigned to the axes afterward, and is often based on geographic position or ethnic origin, cultural identity and historic relations, territorial characteristics that impede or encourage relationships between groups, or migratory dynamics. Based on the organization of the population under study, topological maps can provide information on evolution. They do not necessarily assume that evolution has taken place in the form of a series of splits, with interruption of all genetic flows between the groups that result.

The topological representation in Fig. 96–3 is based on the same data used to construct the dendrogram in Fig. 96–2. The values of the first dimension of a multidimensional scaling analysis are plotted on the horizontal axis, whereas the second dimension is on the vertical axis. The first dimension can be easily interpreted as the north–south direction. Similar to the

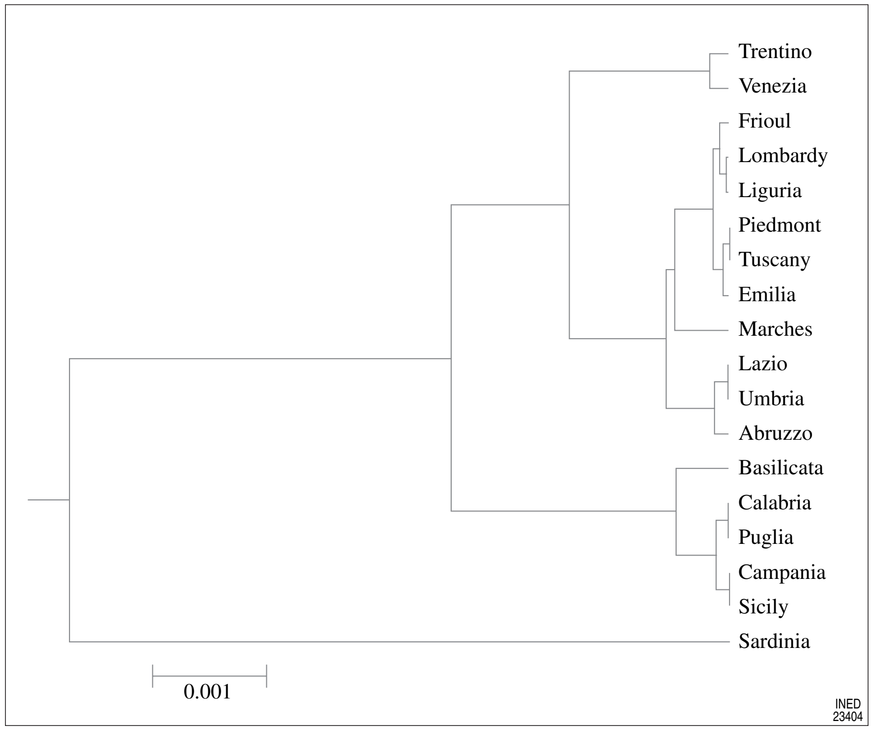


FIGURE 96–2 Dendrogram of relationships between Italian regions based on allelic frequencies of ABO, MNSs, and Rh blood groups. (From Soliani *et al*., 1982.)

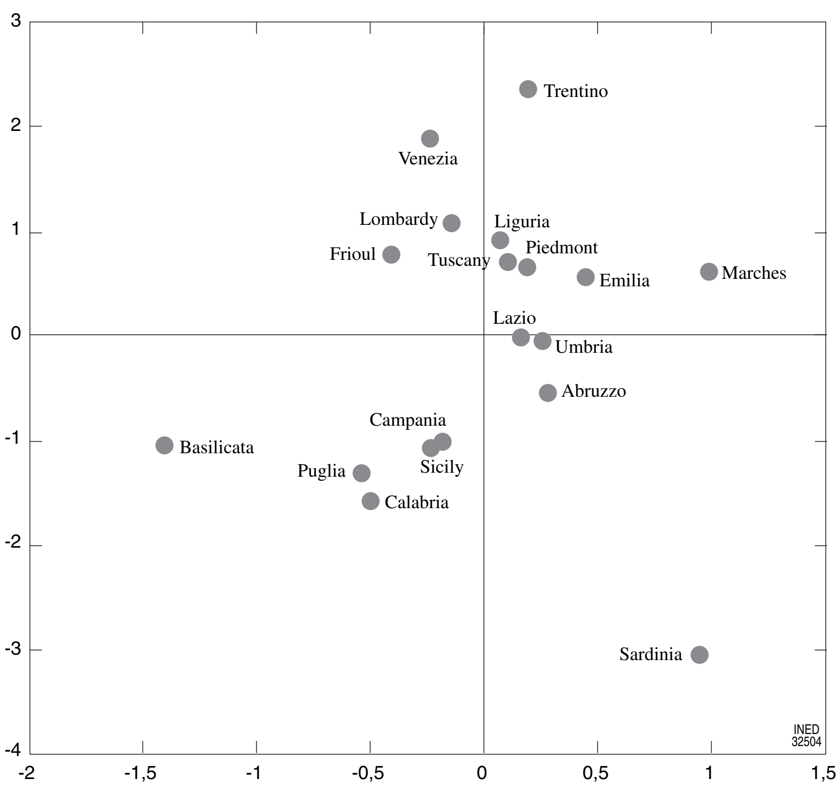


FIGURE 96–3 Topological representation of the relationships between Italian regions based on the similarity matrix of allelic frequencies (ABO, MNSs, and Rh blood groups). (From Soliani *et al*., 1982.)

dendrogram, the topological representation reflects the geographic distribution and the history of the regions. However, it describes the relationships solely based on the proximity of the points, without having to assign each region to a particular branch. Thus, for example, Liguria and Piedmont are closely linked both geographically and historically, but under the complete taxonomy of the dendrogram they were assigned to different subgroups.

V. BIODEMOGRAPHIC RESEARCH

When it comes to genetic evolution, human population biology shares many interests with the study of demographic evolution—the study of relationships between isolates and the groups that make up populations, and of variation over time and space. Scientists can study distributions of genic frequencies, by using erythrocytic and proteic markers and recent improvements allowing the observation of the sequencing of nuclear and mitochondrial DNA. This understanding makes it possible to make deductions about the origin and the evolution of human groups, and even to reach some conclusions about the origins of modern humans (Cann *et al.*, 1987; Cavalli-Sforza *et al.*, 1994).

Research on micro-evolutionary processes—the concrete mechanisms by which groups are transformed over time—as a function of sociodemographic behaviors is also very useful. In particular, it is necessary to consider changes in fertility, nuptiality, mortality, and migration, which can vary across populations and have reached new levels over the past few centuries.

An early area of research, which is now fairly old, studies demographic behavior in order to identify the mechanisms causing the evolution of genetic and genotypic frequencies: selection via differential mortality and fertility, migration, genetic drift, preferential mating, and subdivision of the population into isolates or subgroups. These biological and genetic transformations can be clarified and quantified by measuring evolution and demographic variation.

Consanguinity is one of the themes that has long been embraced by biodemography (Cavalli-Sforza and Bodmer, 1971; Conterio and Moroni, 1974). The type and degree of consanguinity as well as its incidence are tied to the age structure and the natural and migratory dynamics of the population. These factors in turn depend on social organization such as division into classes or castes, economic parameters such inheritance systems, religion, laws, traditions, and familial and collective customs. Studying the historical evolution of consanguinity and its variation over time and space, as well as differences in behavior within communities, makes it possible to evaluate the specific variations in the genetic structure of populations. Consanguinity is also important from a medical point of view, because it encourages the appearance of pathologies caused by harmful or lethal recessive genes; morbidity, disability, and infant mortality, in addition to morphological characteristics such as size, are closely correlated with consanguinity.

By studying union formation and modes of choosing spouses, it becomes possible to demonstrate that assortative mating by phenotypic characteristics or even social or cultural characteristics is taking place. Analyzing marital mobility, both within and between groups or populations, allows researchers to estimate a community's degree of isolation and to evaluate the effects of isolation on the evolution of genetic structure.

Migration has a strong influence on the evolution of the gene pool in a given geographic area. In genetics, the cumulative effect of continuous migration is important, even aside from the short-term effects of migration. Entries and exits have different impacts on the genetic pool. Emigrants are not a representative sample of the sending population, either socially or biologically. For the receiving population, the consequences for genetic heritage vary based on the life trajectories of immigrants, especially in terms of reproduction. Economic and social factors such as occupation, education level, marital status, and income affect both reproductive behavior and the likelihood of eventual integration into the receiving society. Thus, migratory flows must be analyzed at a fairly fine level, at minimum separately by sex and age group.

The nominative reconstruction of families, extended to entire populations or wide geographic areas, furnishes the most interesting, most complete, and most complex examples of interaction between demography and genetics or population biology. When it is extended to large territories and over several generations, it provides geneticists and population biologists with real-world experience ("real" populations whose outcomes can be followed over time) in which evolutionary forces and mechanisms, as well as their results, can be observed concretely.

This reconstruction takes the form of what we call "useful history" of a population, integrating demographic history, social history, and genealogical and genetic history.

The pioneering research in this domain was carried out on the population of Quebec in the 19th and 20th centuries (Charbonneau *et al.*, 1987; Desjardins, 1988; Bideau *et al.*, 1995). It represents a significant example of this type of approach because of the impressive number of individual records; the extent of the historical field covered; and the quantity of biological, medical, social, and economic characteristics considered. This research assessed the consequences of founder effects, the choice of spouse, the number of children per family, migratory flows, consanguinity, endogamy, and the evolution of isolates.

Because this analysis included information on hereditary diseases (Bideau *et al.*, 1979), it was possible to construct a map illustrating their spatial diffusion. "It is evident that the eastern regions are affected by rare diseases, mostly characteristic of the French Canadian population (tyrosinemia, spastic ataxia of the Charlevoix-Saguenay, intestinal atresia, agenesis of corpus callosum, myotonic dystrophy . . .). On the other hand, in the center and west regions, we find for the most part more common diseases such as Duchenne's disease, phenylketonuria, Friedreich ataxia, thalassemia. . . . Diseases including cystic fibrosis and fragile X syndrome are spread over the entire population" (Bouchard, 1993).

Demographic analyses, both aggregate-level and based on individual reconstructions, have shown that the region's population was produced by three waves of migration (from three founder effects?), which took place one after another in the three regions of Quebec between the 17th and 19th centuries. These immigrants, who were likely not a random sample of the French population of origin, gave birth to a population that was genetically different from the population of origin. The contribution of pioneering immigrants to the genetic pool is also clearly recognizable in reconstructions of ascending genealogies of the bearers of hereditary diseases: "A study of the first couples to arrive in the Saguenay (before 1842) demonstrated that after six generations, more than 20% of the first immigrants no longer had any descendants in the regional population, whereas 10% of them had more than 3000 descendants" (Bouchard, 1993).

Overall, the unique characteristics in the gene pool of the population under study can be attributed to (1) natural fertility that led to population doubling every 25 years, (2) distinctive modes of familial reproduction that encouraged the establishment of descendants of longstanding family lines, and (3) a relatively endogamous marriage market, all accentuated by (4) highly homogenous immigration.

Contrary to what is generally believed, studies have shown that closely consanguine parents were not significantly more common among carriers of rare hereditary pathologies. Because of differential reproduction patterns by founders within the population being considered, the risk of hereditary disease was diffused throughout the majority of the population rather than being concentrated in certain easily identifiable families or groups of descendants. Close consanguinity between spouses thus adds relatively little to the general risk of a given couple producing a child affected by pathologies owing to rare genetic mutations (Bouchard, 1993).

Research on reconstructed populations has slowed considerably since the 1980s because of the numerous and serious practical difficulties in organizing studies. Research is based on the collection of a large mass of data and requires costly long-term investigations and precise organization of data. In addition, it takes a long time to arrive at concrete results using these data. Genealogical reconstructions also demand a high level of competence in different scientific domains, leading to difficulties in collaboration and integration, and the substantial financial resources necessary for these studies are difficult to mobilize in a single field. Finally, because reconstructions necessarily apply to specific populations and periods, the results obtained describe specific situations and are not always easily generalized. Once carried out, however, this type of study can be highly productive, as evidenced by the example described above. Developments in computing and the expansion of available archives and computerized databases have opened new possibilities for the application of these methods, including in large present-day populations.

VI. THE ANALYSIS OF SURNAMES

In patrilinear societies, family names can be treated like allelic forms of a gene situated on the Y chromosome and transmitted through the male line. The origins of surnames are cultural. Since the codification of surnames and the establishment of rigorous norms of recording, however, the diffusion and population structure of surnames has been a product of the reproductive and migratory behavior of their bearers. (In many Italian regions and in a variety of European countries, family names were widely used starting in the 17th century.) Currently, the modes of diffusion of family names are analogous to those that prevail for genetic characteristics.

Surnames, which are inherited from the family of origin, define an individual identity that is both biological and cultural. At the population level, the distribution of surnames can help define the biological and cultural identity of different groups, as well as provide insight into the structural characteristics of the population itself.

The correspondence between family names and genes is obviously not complete. Problems arise from the polyphyletic origin of surnames: the same family name could form independently in different regions, and having the same family name does not always mean sharing the same origin. Illegitimacy, adoption, transcription errors, and legal changes, as well as the evolution of the language, introduce variations at the individual level, although they do not generally compromise the ability to characterize a population on the aggregate level. Finally, the length of time that is observable is not ideal: the history of family names is fairly brief compared with the long evolutionary period of allelic frequencies. Still, from the rich bibliography already available, it is clear that surnames and genes follow similar dynamics: conclusions derived from analyses using the two types of data are generally consistent (Lucchetti and Soliani, 1989; Zei *et al*., 1993).

Family names have the practical advantage of being easy, quick, and inexpensive to record and categorize. The record of surnames can be extended to cover the whole population and can often be directly obtained from lists that have already been established in various forms by various public and private institutions. Many of these lists now exist, and some are even computerized. Different sources of names can be used depending on the phenomenon being studied: stock data (censuses, population registers, lists of residents) or vital data (births, marriages, deaths) of people who are native to a certain area or have experienced certain events (marriage, birth, parenthood, death) in that area. By using historical registers (religious or civil), the record of family names can be extended into the past as well. A population's evolution, at least for the past few centuries, can be traced completely by observing actual movements in detail. It is thus possible to overcome the fundamental ambiguity of attempting to reconstruct the past from the present when experimental verification is not possible.

Frequencies of surnames form a system with high polymorphism, with a large capacity for characterization and differentiation. Systems of surnames can even be used to distinguish between neighboring populations, in which biological characteristics are generally not sufficient to identify significant differences (Lucchetti and Soliani, 1989).

1. Isonymy, Consanguinity, Matrimonial Structure

The use of surnames to analyze consanguinity is generally traced to George Darwin (1875), who estimated the frequency of marriage between cousins based on the incidence of isonymic marriages, or marriages between people with the same family name. Analyses of this type were carried out in numerous studies, based on a precise quantification of the relationship between consanguinity and isonymy (Crow and Mange, 1965). More recent applications address the repetition of surname pairs among spouses, thus extending the analysis to the marital behavior of the population as a whole. To uncover and assess population subgroups within which marriages take place preferentially, Gabriel Lasker and Bernice Kaplan (1985) calculate the number RP (for "repeated pairs") of repetitions in associations of spouses' last names according to the following formula:

$$RP = \frac{\sum [S_{ij}(S_{ij}-1)]}{N(N-1)}$$

where S_{ij} represents the number of couples in which the husband has the surname i and the wife the surname j and N is the total number of couples. The index RP varies between zero (each pair of last names appears only once) and one (all couples have the same combination of family names). The expected value $E(RP)$ for a random combination of family names is calculated by using the method proposed by Ranakut Chakraborty (1985). Comparing observed and expected values for a random sample makes it possible to measure the *nonrandom* component and thus to evaluate the influence of preferential choice on the formation of couples. Preferential choice is interpreted as an expression of subdivisions within the population (Mascie-Taylor *et al*., 1987). Applications to historical populations in Massachusetts (Relethford, 1992) have shown that the random component of RP is negatively correlated with the size of the population and with exogamy. The relative excess of RP is positively correlated with the size of the population, demonstrating a positive relationship between population size and internal stratification.

2. Family Names and the Internal Structure of the Population

The typology of the distribution of surnames supplies information on the internal structure of a population. Intuitively, there are two extreme situations: a population with few family names, all with a high degree of repetition, and a population with many family names, all rare. The first suggests a high degree of isolation, with no immigration of individuals from the outside: the survival of the population depends on local families, and the number of surnames declines over time, with the remaining names increasing in frequency. In the second type of population, the greater variety of surnames suggests the presence of immigration introducing new family names with low occurrence in the population.

The number S of different surnames in relation to the number N of individuals is a measure of the internal homogeneity (or variability) of a population. For populations of equal size, differences in the number S of patronymic forms stem from characteristics of the community, in particular on its level of isolation or openness.

The ratio S/N is a crude but simple indicator of a population's richness in surnames; a higher value implies a greater variety of family names. At the extremes, a ratio equal to one means that every individual in the population has a different last name, whereas values close to zero indicate that there are few patronymic forms and each is carried by a relatively large number of individuals. In the first case, the family name has great power to identify the individual within the population but can only rarely distinguish the population from other groups. In the second case, the family name cannot identify the individual within the group but is strongly predictive of the community that the individual belongs to.

The ratio S/N is defined based only on the diversity of patronymic forms in a population and does not always distinguish between different populations with the same level of variation. For instance, the ratio has the same value when there are S last names appearing with equal frequency in the population and when some of the S surnames are common whereas others are rare or unique.

Family names that appear only once (family names with frequency one) have a strong influence. In historical populations, especially isolated populations, the modes of transmission of family names and patrilocal customs meant that surnames with frequency one were mostly carried by women who had married in to the population. In present-day populations, however, family names with incidence one, or more generally with low frequency, are primarily associated with the mobility of single people, essentially men, who migrate for work reasons.

Numerous indices originally used in analyses of biological variability (genetic or ecological biodiversity) can be applied to the distribution of family names in order to describe the structural characteristics of a population. The probability that two randomly chosen individuals carry the same family name (Relethford, 1992) defines the random component of the internal isonymy of a group. As James Crow and Arthur Mange (1965) have shown, there is a specific relationship between this measure and the random component of the consanguinity in a population.

The frequency and distribution of patronymic forms can be used to evaluate a population's richness in surnames and even to estimate the importance of migratory flows. Ronald Fisher's α index (Fisher *et al.*, 1943), developed to assess an ecosystem's richness in species, measures the abundance of surnames in a population. The ν index developed by Samuel Karlin and James McGregor (1967) measures the rate at which new alleles appear in a population. When applied to surnames, it is interpreted as a measure of the rate of entrance of new names and used to estimate immigration.

Applications using historical data from the Parma Valley (Yasuda *et al.*, 1974) and Sardinia (Zei *et al.*, 1983), as well as on recent Italian (Piazza *et al.*, 1987), Sicilian (Scapoli *et al.*, 1997), and French (Darlu and Ruffie, 1992) data, confirm the credibility of this usage of family names. The results of this research are in agreement with observations based on demographic data.

3. Migration Matrices

By observing the distribution of surnames at different periods in the same group of populations, researchers can analyze changes in these populations as a function of migratory and reproductive behavior. For instance, Ellen Wijsmann *et al.* (1984) proposed an interesting method and applied it to historical data of Sardinia. The patronymic structure of several populations can be described using a matrix $\mathbf{N}$ of dimension $P \times S$, where P is the number of populations being studied and S is the number of names. Each row of the matrix corresponds to a population, and each column to a surname. Each cell element n_{pk} represents the number of individuals in the population p with family name k. The patronymic structure of the P populations observed at different times can be described using two matrices, $\mathbf{N_1}$ and $\mathbf{N_2}$, indicating times $t = 1$ and $t = 2$, respectively. If the P populations represent a relatively isolated system, $\mathbf{N_2}$ can be expressed as a function of $\mathbf{N_1}$ according to the relationship

$$\mathbf{N_2} = \mathbf{B} \times \mathbf{N_1} + \mathbf{E}$$

where the matrix $\mathbf{E}$ represents effects due to chance, or at least to causes other than migration and fertility. The matrix $\mathbf{B}$, a square matrix with dimensions $P \times P$, combines the joint effects of migratory and reproductive behavior. $\mathbf{B}$ can be decomposed into two matrices $\mathbf{R}$ and $\mathbf{M}$ such that $\mathbf{B} = \mathbf{R} \times \mathbf{M}$. $\mathbf{R}$ is a diagonal matrix: each cell r_{ij} describes the reproductive behavior in the jth population. $\mathbf{M}$ is a square matrix, also $P \times P$, in which each element m_{jk} indicates the proportion of individuals migrating from population k to population *j*. The elements of the main diagonal estimate the proportion of native born residents.

Ignoring **E** and assuming that only migration and reproduction are at work, we can write

$$\mathbf{N_2} = \mathbf{R} \times \mathbf{M} \times \mathbf{N_1}$$

where the value of each cell n_{pk} in $\mathbf{N}_2$ (the number of individuals in population p with surname k at time $t = 2$) is calculated by adding the number individuals with the kth name at time $t = 1$, reproducing at the rate of their population of origin, who migrate from each population to population p (including the quota m_{pp} of people originating in population p who stay in the same population). Ellen Wijsmann suggested a way of deriving **B** from the matrices of isonymy between the populations and then deriving **M**, which describes the probabilities of migration between the populations being considered. When applied to historical data from Sardinia, the method has produced results that are in good agreement with what is already known about migratory movements in the region.

4. Relations between Populations

Models and methods similar to those used for allelic frequencies can be applied to the study of family names and can shed light on the relationships between two or more populations. Indices of distance or of similarity can be calculated based on the probabilities that two individuals taken at random from the two populations have the same family name.

One of the most commonly used is the R_{ij} index developed by Gabriel Lasker (1977, 1985):

$$R_{ij} = \frac{\sum(S_{ik} \cdot S_{jk})}{2 \cdot N_i \cdot N_j}$$

where S_{ik} and S_{jk} are the number of individuals with surname k in populations i and j, respectively, and N_i and N_j are the total number of individuals in the two populations. R_{ij} is an index of similarity (of relationship): it is high when the two populations have a significant proportion of patronymic forms in common, zero when they have no patronymic forms in common.

To facilitate interpretation and comparison of values, other indices impose a constraint that the index of similarity of a population with itself be one. One of the most prominent of these indices is the index of "standardized isonymy" proposed by Kuan Ho Chen and Luigi Cavalli-Sforza (1983), which is calculated from the relative frequencies p_{ik} and p_{jk} of surname k in populations i and j:

$$R_{ij} = \frac{\sum(p_{ik} \cdot p_{jk})}{\sqrt{(\sum p_{ij}^2 \cdot \sum p_{jk}^2)}}$$

The methods of representation described for matrices of genetic distance can also be used for matrices of similarity obtained from family names. In general, results agree with those obtained on the basis of allelic frequencies. As noted above, these methods can also be used to compare neighboring populations with negligible genetic differences.

The diffusion of family names, in accordance with the geographic dispersion of their bearers, provides information on the history of populations. In particular, family names can shed light on the presence of groups linked by a common cultural-historical origin, or groups that are culturally differentiated owing to different ethnic origins or long periods of isolation. The application to the Italian population (Zei *et al*., 1993) has clearly shown correlations between the history of different regional populations, the influence of immigrants of different origins, and historic separations (Barbujani *et al*., 1991, 1992). Analyses carried out based on surnames are in good agreement with maps produced based on allelic frequencies (Piazza, 1988), as is shown in the examples described here.

These analytic methods have been applied to distributions of surnames in Italian regions (Lucchetti *et al*., 1996), constructed by using lists of telephone customers (Seat, 1993). The results of this analysis are presented in a multidimensional scaling representation (Fig. 96–4). The spatial distribution of regions in this figure echoes their geographic location, if the first dimension is graphed on the vertical axis (to identify the north–south direction) and the second dimension on the horizontal axis (to identify the east–west direction). The uniqueness of Sardinia is clearly visible—the region is characterized by very specific patronymic forms. Trentino and the Valle d'Aosta, where there are names of German and French origin respectively, are also distinctive. The southern regions are clustered in the lower part of the figure, whereas the northern regions are gathered in the upper section and Lazio, Abruzzo, and Molise are in the middle. On the other hand, Piedmont, Lombardy, and Liguria occupy unexpected positions based on their geographic location: the position of the regions of the industrial triangle demonstrates the effect of the powerful processes of migration coming from the southern regions.

The observation that north–south migratory flows are largely directed toward the industrial cities inspires an examination of the patronymic structure of the five principal metropolitan areas (Turin, Milan, Genoa, Rome, and Naples) separately from their

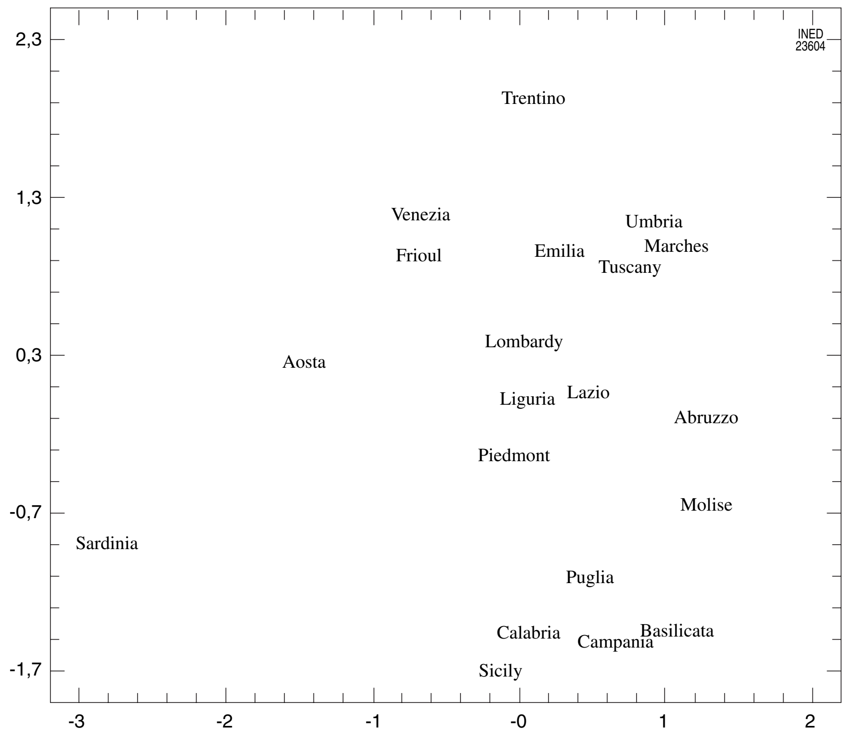


FIGURE 96–4 Division of Italian regions based on the distribution of surnames in 1993. (From Lucchetti *et al*., 1996.)

regions (Fig. 96–5). By using this refinement, the remaining parts of Lombardy and Piedmont rejoin the group of northern regions, whereas Milan and Turin fall even lower on the vertical axis than do their regions as whole in Fig. 96–4. On the other hand, Genoa, because of the greater dispersion of immigrants over the Ligurian territory, and Rome, toward which immigrants converge from all over the country, move very little in relation to their location in Fig. 96–4. Finally Naples, the arrival city for the flows of immigrants from the other regions of the south of Italy, is differentiated from Campania by moving even further toward the bottom of the figure.

The study of communal populations scattered over the Apennines (Parma and Piacenza in the north; Massa Carrara, La Spezia, and Genoa in the south), belonging to different provinces and linked to each other by various means of communication (railroad, roads, highways), reveals the direction of mobility and the contrasting effects of separation by the Apennines and connection by communication routes (Lucchetti *et al*., 1990).

Because data can be collected on the patronymic structure of a population in past centuries, analyses can be extended fairly far back in time. Research on the evolution of patronymic structures in parishes of the high Parma Valley (Lucchetti and Soliani, 1989) shows moments of rapid change alternating with periods of relative stability and highlights the effects of particular events, such as mortality crises or periods of war, and periods of marked isolation or rapid population growth.

5. Multidisciplinary Approaches: From Genetics and Anthropology to Demography and Sociology

"Current studies on family names, far from being limited to genealogy, are now a full part of the research domain. They facilitate an original decoding of both the historical and anthropological conditions that spurred the process of identification of individuals at different times in various parts of Europe." "Studying in turn the origins and the history of family names, their geographic distribution and the ties between genetic markers and family names . . . anthropologists, family historians, demographers, specialists in population genetics, and sociologists show the extent to which a multidisciplinary approach is necessary in order to explain the phenomenon of the family name." The project *Le patronyme. Histoire, anthropologie, société* (*The surname: History, anthropology, society*), published by the Editions du CNRS (Brunet *et al*., 2001), on whose

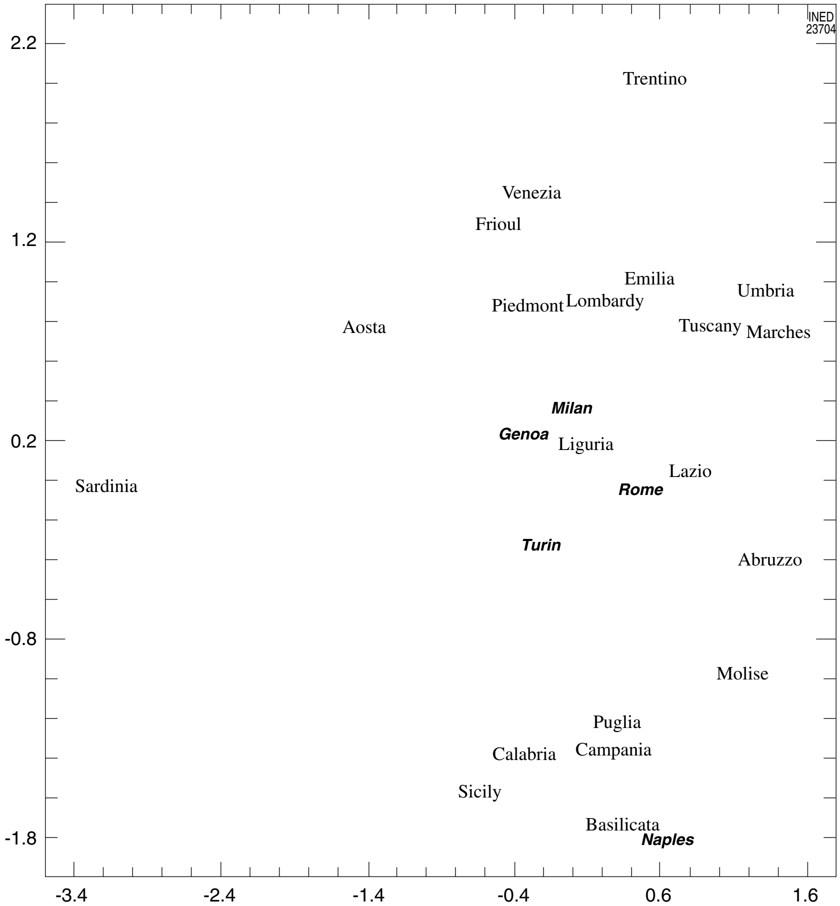


FIGURE 96–5 Division of Italian regions and five principal cities (Milan, Turin, Genoa, Rome, and Naples) based on the distribution of surnames in 1993. (From Lucchetti *et al*., 1996.)

cover these quotes are displayed, presents a large number of studies on the subject: "Based on concrete examples, from the Baltic provinces to Sardinia or the Moroccan Atlas, from the Flemish of France to the French of Quebec, this work offers a perspective without equivalent on the question of the origins and the evolution of the surname."

We take one of the diverse collection of articles as an example, both for the breadth of the field that it covers and for its demographic interest: "Immigration flamande en France au XIX^e^ siècle et au début du XX^e^ siècle."[2] In this article, European specialists from different disciplines show how characteristic family names of eastern and western Flanders diffused in the three French departments that were the closest, geographically and linguistically, to Flanders: the Nord, the Pas-de-Calais, and the Aisne. During the 19th century, there was preferential migration toward centers of labor, such as Paris, the Moselle, the Meurthe-et-Moselle, and Nantes. In addition to tracing the destination of migrants, the methods used make it possible to quantify their presence in specific zones and their dispersion over the course of the 20th century (Degioanni *et al*., 2001).

For the past few years, the major demography journals, including *Population*, have turned their attention to the usage of family names as an instrument of analysis, in particular for the study of migration. Italian migration to France has been analyzed in the framework of a collaboration between French- and Italian-speaking geneticists and demographers (Degioanni *et al*., 1996). They first reconstructed the distribution of family names in France and in Italy, using births from the vital registration system collected by National Institute for Statistics and Economic Studies for the 36,500 French communes, grouped by department, and marriage dispensations from the Vatican, grouped by diocese, for Italy. They then compared variations in the two series of distributions for the 50-year period

[2] "Flemish immigration to France in the 19th century and the early 20th century."

1891–1940 in order to estimate migratory flows from Italy to France. Despite the technical difficulties stemming from using these sources, the authors show that migration can be quantified using these methods, although in approximate fashion. Above all, these analyses clarify a number of points concerning the geographic origins of migrants, the direction of flows, and the time of migration, information that cannot generally be derived from census data. Residents of Piedmont and Liguria were the first to migrate to the south of France; starting in the early 20th century, they were followed by Venetians, many of whom moved toward the Garonne and along the Nantes-Rennes axis. Venetians and residents of Lombardy were also numerous in the Lorraine, and residents of Lombardy in Franche-Comté. Numerous groups from all over Italy, but especially from Emilia-Romagna, migrated to Paris. The zone around Marseilles attracted Italians from the center-south, especially from Lazio, Sicily, Sardinia, and Tuscany.

European demographic history societies occasionally follow along with the development of this new discipline, which applies easily accessible information to in-depth studies of the behavior of human populations, often impossible to analyze otherwise. As noted above, the analyses address the internal structure of specific communities and their demographic dynamics; the size of the marriage market and the modes of formation of couples; relations between populations; and the presence, intensity, and direction of migratory flows (Pizzetti *et al*., 2001). Further development of these research methods invented by anthropologists and geneticists, and their application, could be greatly encouraged by their diffusion among demographers and sociologists.

The evolution of human populations is the global result of a complex network of interactions between a number of different factors, often defined and analyzed in the context of different disciplines. Reciprocal forces overlap, cancel each other out, or reinforce each other in multiple and varied combinations (Livi Bacci, 1986). Evolution cannot be fully reconstituted through reductive approaches. When it comes to evolutionary and historical reconstructions, the phenomena under analysis can not be studied in a perfectly controlled manner or be subject to repeatable experiments in a laboratory. Studying these phenomena becomes a scientific challenge, and the credibility of the conclusions is rarely comparable to that of those obtained in the experimental sciences. The trust placed in these reconstructions depends on the convergence of multiple sources of evidence offering independent support to a given interpretation. The accumulation of new sources of evidence from varied research fields plays a vital role in creating a tight network of information to verify a proposed reconstruction (Ammerman and Cavalli-Sforza, 1986).

Acknowledgment

This chapter was translated from the French version by Sarah Hayford.

References

Ammerman Albert J. and Cavalli-Sforza Luigi Luca, 1986. *La transizione neolitica e la genetica di popolazioni in Europa*. Turin, Boringhieri, 210 p.

Barbujani Guido and Sokal Robert, 1991. Genetic population structure of Italy, II; physical and cultural barriers to gene flow, *American Journal of Human Genetics*, vol. 48(2), p. 398–411.

Barbujani Guido, Vlan Paolo, and Fabbris Luigi, 1992. Cultural barriers associated with large gene frequency differences among Italian populations, *Human Biology*, no. 64, p. 479–495.

Barrai Italo, 1978. *Introduzione alla genetica di popolazioni*. Milano, Isedi, 275 p.

Bideau Alain, Brunet Guy, Desjardins Bertrand, and Prost Michel, 1995. La reproduction de la population aux XVII^{e}, XVIII^{e} et XIX^{e} siècles: exemples français et québécois, *Annales de démographie historique*, p. 137–148.

Bideau Alain, Plauchu Henry, and Jacquard Albert, 1979. Démographie historique et génétique des populations: la concentration géographique d'une maladie héréditaire rare, *Annales, Economies, Société et Civilisation*, p. 85–105.

Bouchard Gérard, 1993. Population et génétique: une nouvelle frontière pour les sciences sociales, *Annales de démographie historique*, p. 397–409.

Brunet Guy, Darlu Pierre, and Zei Gianna, 2001. *Le patronyme: histoire, anthropologie, société*. Paris, Éditions du CNRS, 419 p.

Cann Rebecca L., Stoneking Mark, and Wilson Allan C., 1987. Mitochondrial DNA and human evolution, *Nature*, vol. 325(1), p. 31–36.

Cavalli-Sforza Luigi L. and Bodmer Walter F., 1971. *The genetics of human populations*. San Francisco, W.H. Freeman, 965 p.

Cavalli-Sforza Luigi L. and Edwards Anthony W. F., 1964. Analysis of human evolution, *in: Proceedings of the 11th Congress of Genetics*, vol. 3, p. 923–933. New York, Pergamon Press.

Cavalli-Sforza Luigi Luca, Menozzi Paolo, and Piazza Alberto, 1994. *The history and geography of the human genes*. Princeton, Princeton University Press, 518 p.

Chakraborty Ranajit, 1985. A note on the calculation of random RP and its sampling variance, *Human Biolology*, vol. 57(4), p. 713–777.

Charbonneau Hubert, Desjardins Bertrand, Guillemette André, Landry Yves, Legaré Jacques, Nault François, Bates Real, and Boleda Mario, 1987. *Naissance d'une population: les Français établis au Canada au XVII^{e} siècle*. Paris and Montreal, Institut national d'études démographiques and Presses de l'Université de Montréal, 232 p.

Chen Kuang-Ho and Cavalli-Sforza Luigi L., 1983. Surnames in Taiwan: interpretation based on geography and history, *Human Biology*, vol. 55(2), p. 367–374.

Conterio Franco and Moroni Antonio, 1974. Démographie et génétique, *Population et famille*, vol. 31(1), p. 64–115.

Crow James F. and Mange Arthur P., 1965. Measurement of inbreeding from the frequency of marriages between persons of the same surname, *Eugenics Quarterly*, vol. 12(4), p. 199–203.

Darlu Pierre and Ruffie Jacques, 1992. L'immigration dans les départements français étudiée par la méthode des patronymes, *Population*, vol. 47(3), p. 719–734.

Darwin George, 1875. Marriages between first cousins in England and their effects, *Journal of the Statistical Society*, vol. 38(2), p. 153–184.

Degioanni Anna, Antonella Lisa, Zei Gianna, and Darlu Pierre, 1996. Patronymes italiens et migrations italiennes en France entre 1891 et 1940, *Population*, vol. 51(6), p. 1153–1180.

Degioanni Anna, Darlu Pierre, Poulain Michel, and Foulon Michel, 2001. Immigration flamande en France au XIX[e] et au début du XX[e] siècle, *in:* Guy Brunet, Pierre Darlu, and Gianna Zei (eds.), *Le patronyme: histoire, anthropologie, société*, p. 189–201. Paris, CNRS Editions, 419 p.

Desjardins Bertrand, 1998. Le registre de la population du Québec ancien, *Annales de démographie historique*, no. 2, p. 215–226.

Efron Bradley, 1982. *The Jackknife, the bootstrap and other resampling plans.* Philadelphia, Society for Industrial and Applied Mathematics, 38 p.

Fisher Ronald A., Corbet Steven A., and Williams C.B., 1943. The relation between the number of species and the number of individuals in a random sample of animal population, *Journal of Animal Ecology*, vol. 12(1), p. 42–58.

Hartl Daniel L. and Clark Andrew G., 1989. *Principles of population Genetics.* Sunderland, Sinuaer Associates, 542 p.

Hedrick Philip W., 1975. Genetic similarity and distance: comments and comparisons, *Evolution*, vol. 29, p. 362–366.

Jacquard Albert, 1970. *Structures génétiques des populations.* Paris, Masson, 399 p.

Jorde Lynn B., 1980. The genetic structure of subdivided human populations: a review, *in:* James H. Mielke and Michael H. Crawford (eds), *Current developments in anthropological genetics, vol. 1*, p. 135–208. New York, Plenum.

Karlin Samuel and McGregor James, 1967. The number of mutant forms maintained in a population, *in: Fifth Berkeley Symposium of Mathematics, Statistics and Probabilities, Vol. 4: biology and problems of health*, p. 415–438. Berkeley, University of California Press, 934 p.

Kumar Sudhir, Tamura Koichiro, Jakobsen Ingrid J., and Nei Masatoshi, 2001. MEGA2: Molecular Evolutionary Genetics Analysis software, *Bioinformatics [Arizona State University, Tempe, Arizona]*, vol. 17(12), p. 1244–1254.

Lalouel Jean M., 1980. Distance analysis and multidimensional scaling, *in:* James H. Mielke and Michael H. Crawford (eds.), *Current developments in anthropological genetics*, p. 209–250. New York, Plenum, 436 p.

Lasker Gabriel W., 1977. A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations, *Human Biology*, vol. 49(3), p. 489–493.

Lasker Gabriel W., 1985. *Surnames and genetic structure.* Cambridge, Cambridge University Press, 148 p.

Lasker Gabriel W. and Kaplan Bernice A., 1985. Surnames and genetic structure: repetition of the same pairs of names of married couples: a measure of subdivision of the population, *Human Biology*, vol. 57, p. 431–440.

Livi Bacci Massimo, 1986. *Popolazione e alimentazione.* Bologne, Il Mulino, 192 p.

Lucchetti Enzo, Battista Carla N., Ghisolfi Giovanna, and Soliani Lamberto, 1990. Delimitation and aggregation in human populations examined by surname structure, *International Journal of Anthropology*, vol. 5(1), p. 49–62.

Lucchetti Enzo, Siri Enzo, and Soliani Lamberto, 1996. Internal migrations and relationships between Italian regional populations: an analysis based on surname distributions, *Rivista di Antropologia (Roma)*, vol. 74, p. 171–185.

Lucchetti Enzo and Soliani Lamberto, 1989. Similarità tra popolazioni esaminate mediante i cognomi, *Rivista di Antropologia*, vol. 67, p. 181–198.

Mascie-Taylor Nick C. G., Lasker Gabriel W., and Boyce Anthony J., 1987. Repetition of the same surnames in different marriages as an indication of the structure of the population of Sanday Island, Orkney Islands, *Human Biology*, vol. 59(1), p. 97–102.

Mourant Arthur E., 1954. *The distribution of the human blood groups.* Oxford, Blackwell Scientific, 438 p.

Mourant Arthur E., Kopec Ada C., and Domaniewska-Sobczak Kazimiera, 1976. *The distribution of the human blood groups and other polymorphisms.* London, Oxford University Press, 1055 p.

Nei Masatoshi, 1972. Genetic distance between populations, *American Naturalist*, vol. 106, p. 283–292.

Nei Masatoshi, 1978. The theory of genetic distance and evolution of human races, *Japanese Journal of Human Genetics*, vol. 23, p. 341–369.

Piazza Alberto, Cappello Nazario, Olivetti Eleonora, and Rendine Sabina, 1988. A genetic history of Italy, *Annal of Human Genetics*, vol. 52(3), p. 203–213.

Piazza Alberto, Rendine Sabina, Zei Gianna, Moroni Antonio, and Cavalli-Sforza Luigi L., 1987. Migration rates of human populations from surname distributions, *Nature*, vol. 329, 714–716

Pizzetti Paola, Lucchetti Enzo, and Soliani Lamberto, 2001. L'uso dei cognomi nella ricerca biodemografica, *Popolazione e Storia*, no. 1, p. 107–136.

Relethford John H., 1992. Analysis of marital structure in Massachusetts using repeating pairs of surnames, *Human Biology*, vol. 64(1), p. 25–33.

Scapoli Chiara, Rodriguez-Larralde Alvaro, Beretta Maria, Nesti Chiara, Lucchetti Enzo, and Barrai Italo, 1997. Correlations between isonymy parameters. *International Journal of Anthropology*, vol. 12,(1), p. 17–37.

Soliani Lamberto, De Iasio Sergio, Lucchetti Enzo, and Conterio Franco, 1981. Frequenze alleliche nelle popolazioni regionali italiane, I: le distribuzioni regionali dei sistemi Ab0, Rh, MN, Kell e delle aptoglobine, *Antropologia Contemporanea*, vol. 4(2–3), p. 283–297.

Soliani Lamberto, De Iasio Sergio, Lucchetti Enzo, and Conterio Franco, 1982. Freqquenze alleliche nelle popolazioni regionali italiane, II: distanze genetiche, *Antropologia Contemporanea*, vol. 5(3–4), p. 201–217.

Wijsmann Ellen, Zei Gianna, Moroni Antonio, and Cavalli-Sforza Luigi L., 1984. Surnames in Sardinia II: computation of migration matrices from surname distributions in different periods, *Annals of Human Genetics*, vol. 48, n° 1, p. 65–78.

Wright Sewall, 1968–1984. *Evolution and the genetics of populations.* Chicago, University of Chicago Press, 4 vol., 2173 p.

Yasuda N, Cavalli-Sforza Luigi L., Skolnick Mark, and Moroni Antonio, 1974. The evolution of surnames: an analysis of their distribution and extinction, *Theoretical Population Biology*, vol. 5(1), p. 123–142.

Zei Gianna, Barbujani Guido, Lisa Antonella, Fiorani Ornella, Menozzi Paolo, Siri Enzo, and Cavalli-Sforza Luigi L., 1993. Barriers to gene flow estimated by surname distribution in Italy, *Annals of Human Genetics*, vol. 57(2), p. 123–140.

Zei Gianna, Guglielmino Carmela, Siri Enzo, Moroni Antonio, and Cavalli-Sforza Luigi L., 1983. Surnames in Sardinia, I: fit of frequency distribution for neutral alleles and genetic population structure, *Annals of Human Genetics*, vol. 47(4), p. 329–352.

APPENDIX 96–A

Calculating Allelic Frequencies

Codominant Alleles

Calculating allelic frequencies is trivial when there are codominant alleles: because there is a one-to-one correspondence between genotype and phenotype, the alleles of each type can be directly enumerated and the relative frequencies derived from them. For example, for a sample of 1000 individuals (T = 1000), the following distribution of MN blood groups is observed:

TABLE 96–A

Group	Genotype	No. of individuals
M	MM	380
MN	MN	490
N	NN	130
Total		1000

The individuals in the sample carry a total of 2000 alleles (1000 × 2). The 380 individuals with phenotype M carry 380 × 2 M alleles, the 130 individuals with phenotype N carry 260 type N alleles, and the 490 individuals with phenotype MN each carry one M allele and one N allele. Of 2000 alleles in total, 1250 (380 × 2 + 490) are of type M and 750 (130 × 2 + 490) are type N. The allelic frequencies, generally represented by lower case p and q, are

$$p_{(\mathrm{M})} = (380 \times 2 + 490)/2000 = 0.625$$

$$q_{(\mathrm{N})} = (130 \times 2 + 490)/2000 = 0.375$$

More generally,

$$p_= \frac{2(MM \text{ homozygotes}) + MN \text{ heterozygotes}}{2{*}\mathrm{T}}$$

and

$$q_= \frac{2(NN \text{ homozygotes}) + MN \text{ heterozygotes}}{2{*}\mathrm{T}}$$

Because only two alleles are involved in the determination of the characteristic studied here, by definition $p + q = 1$.

The error involved in the estimation of the frequencies is equal to $\sigma_p = \sigma_q = \sqrt{pq/2T}$.

The existence of the Hardy–Weinberg equilibrium can be verified by comparing the observed genotypic frequencies to the frequencies that would expected given the relationship

$$\text{MM } (p^2);\ \text{MN } (2pq);\ \text{NN } (q^2)$$

Dominant Alleles

In the case of dominance, there is no longer one-to-one correspondence between phenotype and genotype. The direct counting method proposed above is no longer feasible, because homozygotes for the dominant allele cannot be distinguished from heterozygotes at the phenotypic level. We therefore assume that the population is in panmictic equilibrium, and the frequency of the recessive phenotype, that is, of homozygotes for the recessive allele, is assumed to be equal to q^2.

Take as an example a sample of 1000 (T = 1000) people composed of 350 people with blue eyes and 650 people with brown eyes. The allelic frequency q of c, the recessive allele determining blue eyes, is estimated by assuming

$$q^2 = \text{frequency of recessive homozygotes}$$

or, in other words,

$$q^2 = (350/1000) \quad q = \sqrt{0.035} = 0.187.$$

The frequency p of the dominant allele C is therefore given by the difference:

$$p = 1 - q = 1 - 0.187 = 0.813.$$

The error in this estimation is equal to $\sigma_q = \sqrt{(1-q^2)/4T}$.

Hardy–Weinberg equilibrium cannot be verified, because the hypothesis is taken as a given.

Calculating Allelic Frequencies using Maximum Likelihood Methods

Take the ABO blood type system as an example. Label the allelic frequencies of I^A, I^B, and i alleles p, q, and r, respectively ($p + q + r = 1$), and let n_1, n_2, n_3, and n_4 be their distributions into phenotypes A, B, AB, and O observed in a sample. Knowing the means of determination of phenotypes and the expected probabilities for each of them, we construct the probability function Λ, in which p, q, and r ($p + q + r = 1$) represent the unknowns. The values that maximize Λ are the most likely values for the allelic frequencies. The method also provides the estimation error.

TABLE 96–B

Phenotype	Genotype	Expected probability	Observed frequency of phenotype
A	I^AI^A and I^Ai	$p^2 + 2pr$	n_1
B	I^BI^B and I^Bi	$q^2 + 2qr$	n_2
AB	I^AI^B	$2pq$	n_3
0	Ii	r^2	n_4

$\Lambda = (p^2 + 2pr)^{n1}\ (q^2 + 2qr)^{n2}\ (2pq)^{n3}\ (r^2)^{n4}$

APPENDIX 96–B

Selection Models

Selection against the Recessive Phenotype

Selection against the recessive phenotype operates in the presence of diseases that are caused by a recessive allele *a* and are manifested in homozygotes. If selection is total, *aa* individuals do not contribute at all to the reproduction of the group, and population increase only comes from healthy individuals—*AA* homozygotes and *Aa* heterozygotes. However, individuals with the *aa* genotype continue to be produced as offspring of pairings between *Aa***Aa* heterozygotes as long as the population contains heterozygotes, who are healthy but carry *a* alleles.

Take a population in panmictic equilibrium, with the standard distribution of genotypes *AA*: p^2, *Aa*: $2pq$, and *aa*: q^2, where p and q are the allelic frequencies of *A* and *a*. When this population is submitted to a regime of total selection against the recessive phenotype, only *AA* and *Aa* individuals, a proportion of the population equal to $p^2+ 2pq = p(p + 2q) = p(1 + q)$, participate in reproduction.

The allelic frequencies within this fraction of the population are $p_1 = 1/(1 + q)$ and $q_1 = q/(1 + q)$: these become the allelic frequencies in the generation following selection. Clearly, $p_1 > p$ and $q_1 < q$: the frequency of allele *A* increasses, whereas that of *a* decreases. Similarly, in the next generation we have $q_2 = q/(1 + 2q)$, and after n generations of selection, $q_n = q/(1 + nq)$.

It is possible to evaluate the decrease in q at each generation or after n generations and the corresponding decrease in homozygous individuals *aa* carrying the disease caused by the recessive allele. It is also possible to calculate the number of generations necessary to reduce the initial frequency of the allele *a* by a given quantity given the regime of selection observed.

Selection with Respect to Heterozygotes

Selection against Heterozygotes

Selection against heterozygotes occurs when the contribution of heterozygotes to the formation of new generations is lower than that of the two types of homozygotes. To simplify, suppose that selection is total and that the two homozygotes have the same fitness:

TABLE 96–C

Genotype	Frequency	Fitness	Selection	Successful portion of the transmission
AA	p^2	1	0	p^2
Aa	$2pq$	0	1	
aa	q^2	1	0	q^2

Only *AA* and *aa* individuals ($p^2 + q^2 = 1 - 2pq$) participate in reproduction: within this group allelic frequencies are $q_1 = q^2/(1 - 2pq)$ and $p_1 = p^2/(1 - 2pq)$. In each generation the change in the allelic frequency q is given by

$$\Delta q = q_1 - q = [q^2/(1 - 2pq)] - q = [-pq(1 - 2q)]/(1 - 2pq).$$

When $\Delta q = 0$, the system is in equilibrium, with no changes in the frequency q over the passage of generations. The change Δq is zero if either p or q is equal to zero (if $q = 0$ and $p = 1$, all individuals are *AA*, whereas if $q = 1$ and $p = 0$, all are *aa*; in either case, there are no heterozygotes) or if the quantity $(1 - 2q)$ is zero, that is, if $p = q = 0.5$. The first two equilibrium conditions are stable, but the third is unstable. If $q > 0.5$, Δq is positive, that is, q increases from one generation to the next; if $q < 0.5$, Δq is negative and q decreases. Small random variations from the condition $q = 0.5$ are sufficient to spur evolution toward one or the other allele. If the two alleles in question have different frequencies when selection against the heterozygote begins, the allele with the higher frequency will reinforce its position. If the equilibrium condition, $p = q = 0.5$, is disturbed by random genetic variation, one of the two alleles will be reinforced. If selection against heterozygotes is introduced in several populations where the equality of frequency $p = q = 0.5$ initially was in effect, the two alternative alleles will each become fixed in approximately half of the populations; but the destiny of any given population cannot be predicted a priori.

Selection in Favor of the Heterozygote

Favorable selection toward the heterozygote takes place when heterozygotes have higher fitness than the two homozygotes.

The classic case of selection in favor of the heterozygote is that of thalassemia in a malarial environment. Thalassemia is a genetically determined disease that is less severe among individuals who are heterozygous with respect to the *t* allele responsible for the disease (thalassemia minor) and more severe among homozygotes (thalassemia major or Cooley's anemia). Selection takes place in a malarial environment because healthy individuals are more susceptible to malaria. Healthy individuals are subject to selection by malaria, and *tt* homozygotes are subject to selection by thalassemia. As a result, heterozygotes contribute most to the formation of the next generation.

TABLE 96–D

Genotype	Phenotype	Initial frequency	Agent of selection	Fitness	Selection	Successful portion of the transmission
TT	healthy	p^2	malaria	$1{-}s_1$	s_1	$p^2\,(1{-}s_1)$
Tt	thalassemia minor	$2pq$		1	0	$2pq$
Tt	thalassemia major	q^2	thalassemia	$1{-}s_2$	s_2	$q^2\,(1{-}s_2)$

Individuals $2pq$, $p^2\,(1 - s_1)$, and $q^2\,(1 - s_2)$ participate in reproduction; the frequency q of t alleles in the next generation will be:

$$q_1 = [(pq + q^2(1 - s_2)]/(1 - p^2s_1 - q^2s_2),$$

with

$$\Delta q = q_1 - q = -\,pq(qs_2 - ps_1)/(1 - p^2s_1 - q^2s_2).$$

The conditions where $\Delta q = 0$ correspond to the values $q = 0$ ($p = 1$), $q = 1$ ($p = 0$), and $= s_1/(s_1 + s_2)$. Selection in favor of the heterozygote produces a stable equilibrium condition (when $qs_2 - ps_1 = 0$), that is, when $q = s_1/[s_1 + s_2]$), which depends on the intensity of selection against the two homozygous genotypes. The equilibrium is stable, because when $q > s_1/(s_1 + s_2)$, Δq is negative and when $q < s_1/(s_1 + s_2)$, Δq is positive: in both cases the forces of selection work to maintain equilibrium, and there is no possibility of moving away from it.

APPENDIX 96–C

Migration and its Effects on the Genetic Structure of Human Populations

Island Model Migration

The island model describes the effects of constant unidirectional migration from a large population (a continent) toward a small population (an island). The model analyses the evolution of allelic frequencies in the smaller (receiving) population by assuming that the large population, because of its large size and the negligible influence exerted by the out migrations, does not experience significant change.

Let there be two populations, that of an island (i) and that of a continent (c), initially in internal equilibrium for a bi-allelic characteristic, with q_{i0} and q_{c0} being the respective initial frequencies for an allele a and m the proportion of individuals migrating from the continent to the island in each generation. The allelic frequency on the island becomes

After the first generation with migration: $q_{i1} = (1 - m)\,q_{i0} + mq_{c0}$;
After the second generation: $q_{i2} = (1 - m)q_{i1} + mq_{c0} = \ldots = (1 - m)^2q_{i0} + [1 - (1 - m)^2]q_{c0}$;
And, finally, after n generations: $q_{in} = (1 - m)^nq_{i0} + [1 - (1 - \mathrm{m})^n]q_{c0}$.

The proportion of natives among the residents of the island decreases with each generation. Over time, the island population will become more like that of the continent, and once the equation $q_{in} = q_{c0}$ holds the island population will no longer change from one generation to the next. The change Δ from the equilibrium value, initially $\Delta_0 = q_{i0} - q_{c0}$, becomes $\Delta_1 = (1 - m)(q_{i0} - q_{c0})$, $\Delta_2 = (1 - m)^2(q_{i0} - q_{c0})$, . . . , $\Delta_n = (1 - m)^{\mathrm{n}}(q_{i0} - q_{c0})$, declining by a fraction equal to $(1 - m)$ at each generation.

If the migration rate is known, allelic frequencies in the island population can be predicted. Reciprocally, if allelic frequencies at different times are known, they can be used to estimate the rate of genetic flow from the continent toward the population of the island.

This model has been applied to the North American population of African origin. This population is taken as the island population, subject to a steady arrival of genes from the much larger white population; movement in the opposite direction has very little influence on the white population. Typical allelic frequencies in the white and African-American populations q_{c0} and q_{i0} are known, so it is possible to estimate the average magnitude of genetic flows by generation, and from there to calculate the time necessary to reduce the remaining difference by half.

The R° allele (in the Rh system) has a frequency of 0.63 in the African population and a frequency of 0.03 in white populations; its frequency in current North American populations is estimated to be equal to 0.45. The length of the period of migration of genes toward the African population living in North America is estimated at 10 generations (approximately 300 years).

When the model is applied, the rate m of passage of genes from whites toward African-Americans is estimated to be 0.033 per generation; if this rate were held constant, another 10 generations would be required to reduce the initial difference between the two populations by half (Barrai, 1978).

The model can thus be applied to populations that live on the same territory and are separated by social, economic, or cultural factors rather than geography. In these cases, the two populations could be wholly or partly integrated at some future time.

Archipelago Model

The archipelago model accounts for migratory flows between several populations, analogous to a series of islands of an archipelago with migratory exchanges taking place in all directions. The model can be described by a matrix **M** of migrations, a square matrix with dimensions equal to the number of populations being studied, where each element m_{ij} represents the population of island i that immigrated during the previous generation from island j. The elements of the main diagonal $m_{ii} = 1 - \Sigma_{k \neq i} m_{ik}$ represent the number residents of island i who stayed on the same island during the next generation. A vector **q** describes the values of the allelic frequencies on the different islands, and the allelic frequencies of the next generation are calculated from the product of the vector **q** and the matrix **M**.

Because migration flows go in all directions, it is expected that the system will move toward an equilibrium at which allelic frequencies are the same on all islands and equal to $\bar{q} = \sum q_i / N$ (or a weighted average if the population sizes are different).

The model can be used to analyze changes in allelic frequencies in each population, the relative distance from equilibrium position, and the rate of reduction of the distance.

Migration, Isolation: Effects on Genotypic and Phenotypic Structures

The population that occupies a given geographic area is rarely a homogenous whole. Most of the time populations are divided into more or less separate subgroups. The problem is further complicated by large migratory flows.

Subdivision into groups influences the genotypic and phenotypic structures of the population: there are more homozygotes in populations subdivided into nuclei that are partially or totally separated from each other than in populations without subdivisions. Take, for example, a population formed from two isolates. Assume for simplicity that the populations are of equal size, and that both are in panmictic equilibrium with respective frequencies q_1 and q_2 for allele a. Within the two isolates the frequencies of homozygotes are q_{12} and q_{22} respectively; in the total population the frequency is $(q_{12} + q_{22})/2$. This frequency is greater than that expected when the population is taken as a homogenous whole: the frequency of the allele a in the whole population is the average of q_1 and q_2, and the expected frequency of aa homozygotes is $[(q_1 + q_2)/2]^2$.

Thus, fusing the two populations into a homogenous whole reduces the frequency of homozygotes and increases the frequency of heterozygotes. Conversely, the subdivision of the population into distinct groups is associated with an increase in the frequency of homozygotes and a reduction of the frequency of heterozygotes. The same phenomenon is observed in positive preferential mating. Specific statistics (Wright, 1968) are used to study the phenomena of division into subgroups (Whalund effect), both to detect the presence and intensity of structure and to predict its consequences. These methods focus in particular on migration, and the extent to which immigrants constitute partially distinct isolates within the receiving population rather than being integrated completely and rapidly into it.

Index

B

G

I

M

Demography: Analysis and Synthesis

Demography: Analysis and Synthesis

Graziella CASELLI, Jacques VALLIN, and Guillaume WUNSCH

with contributions by

Daniel COURGEAU, Nico KEILMAN, Eva LELIÈVRE, James VAUPEL, Anatoli YASHIN, and John WILMOTH

VOLUME IV

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data
Caselli, Graziella.
[Dimographie. English]
Demography : analysis and synthesis : a treatise in population studies / Graziella Caselli, Jacques Vallin, and Guillaume Wunsch ; with contributions by Daniel Courgeau . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0-12-765660-X (set : alk. paper)—ISBN 0-12-765661-8 (volume 1 : alk. paper)—ISBN 0-12-765662-6 (volume 2 : alk. paper)—ISBN 0-12-765663-4 (volume 3 : alk. paper)—ISBN 0-12-765664-2 (volume 4 : alk. paper) 1. Demography. I. Vallin, Jacques. II. Wunsch, Guillaume J. III. Title.
HB871.C37513 2005
304.6—dc22 2005019413

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 13: 978-0-12-765660-1
ISBN 10: 0-12-765660-X

For all information on all Elsevier Academic Press publications visit our Web site at www.books.elsevier.com.

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Contents
Volume IV

Contents of Other Volumes

Volume I Contents

Volume II Contents

Volume III Contents

Contributors

Amoakon Anoh, Ècole nationale supérirure de statistique et d'économie appliqué (ENSEA), Abidjan, Côte d'Ivoire.

Philippe Antoine, Institut pur la recherché et le developpement (IRD), Dakar, Senegal.

Isabelle Attané, Institut national d'études démographiques (INED), Paris, France.

Alexandre Avdeev, Université de Moscou, Moscow, Russia, and Université Marc Bloch, Strasbourg, France.

Maria Avdeeva, Departement de bibliographie et de l'information, Centre d'etudes demographiques, Faculte d'economie, Université de Moscou, Moscow, Russia.

Brigitte Baccaïni, Institut national d'études démographiques (INED), Paris, France.

Stefano Baldi, Permanent Mission of Italy to the United Nations, United Nations Plaza, New York, United States.

Magali Barbieri, Institut national d'études démographiques (INED), Paris, France.

Odo Barsotti, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Françoise Bartiaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Gijs Beets, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Giovanni Berlinguer, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Elza Berquo, Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil.

Francesco C. Billari, Istituto di Metodi Quantitativi, Università Bocconi, and Innocenzo Gasparini Institute for Economic Research, Milan, Italy.

Jean-Noël Biraben, Institut national d'études démographiques (INED), Paris, France.

Anna-Maria Birindelli, Dipartimento di Statstica, Università degli Studi Milano-Bicocca, Milan, Italy.

Alberto Bonaguidi, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Corrado Bonifazi, Instituto di Ricerche sulla Popolazione e le Politiche Sociali, Rome, Italy.

Marco Bottai, Dipartimento di Statistica e Matematica applicata all' Economia, Università di Pisa, Pisa, Italy.

Michel Bozon, Institut national d'études démographiques (INED), Paris, France.

Vittoria Buratta, Istituto Nazionale di Statistica (ISTAT), Rome, Italy.

Raimondo Cagiano de Azevedo, Facoltà di Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Gérard Calot, Institut national d'études démographiques (INED), Paris, France.†

Emmanuelle Cambois, Institut national d'études démographiques (INED), Paris, France.

Graziella Caselli, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Jean-Claude Chasteland, Institut national d'études démographiques (INED), Paris, France.

Jean-Claude Chesnais, Institut national d'études démographiques (INED), Paris, France.

Luciano Ciucci, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

†Deceased

Robert Cliquet, Population and Social Policy Consultants, Brussels, Belgium.

David Coleman, Department of Social Policy and Social Work, University of Oxford, United Kingdom.

Maria-Eugenia Cosio-Zavala, Universite de Paris X-Nanterre, Nanterre Cedex, France.

Daniel Courgeau, Institut national d'études démographiques (INED), Paris, France.

Jean Coussy, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Gianpiero Dalla Zuanna, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy.

Patricia David, Harvard School of Public Health, Boston, United States.

Lorenzo Del Panta, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna, Italy.

Bart de Bruijn, Netherland Interdisciplinary Demographic Institute (NIDI), Den Haag (La Haye), Pays Bas, Netherlands.

Jean-Michel Decroly, Laboratoire de Géographie Humaine, Université Libre de Brussels, Brussels, Belgium.

Ainhoa de Federico de la Rúa, Institut Federatif de Recherche sur les Economies et les Societes Industrielles (IRESI), Universite des Sciences et Technologies de Lille 1, Lille, France.

Arna Dellis, University of Hawaii, Manoa, Hawaii, United States.

Paul Demeny, Population Council, New York, United States.

Alessandra De Rose, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'Analisi Regionale, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Paolo De Sandre, Departimento de Scienze Statistiche, Università di Padova, Padova, Italy.

Gustavo De Santis, Facoltà di Scienze Politiche, Università di Messina, Messina, Italy.

Martine Deville, Institut national d'études démographiques (INED), Paris, France.

Manon Domingues Dos Santos, Centre de recherché en economie et statistique (CREST), Malakoff, France.

Josianne Duchêne, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Viviana Egidi, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Vincent Fauveau, United Nations Population Fund, New York, United States.

Alexis Ferrand, Institut de Sociologies et d'anthropologie, Universite des Sciences et Technologies de Lille, Villeneuve d'Ascq, France.

Patrick Festy, Institut national d'études démographiques (INED), Paris, France.

Judith A. Fortney, Family Health International, Research Triangle Park, North Carolina, United States.

Luisa Frova, Instituto Nationale di Statistica (ISTAT), Rome, Italy.

Alexia Fürnkranz-Prskawetz, Vienna Institute of Demography, Vienna, Austria.

Hubert Gérard, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Giuseppe Gesano, Instituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS), Consiglio Nazionale delle Ricerche (CNR), Rome, Italy.

Piero Giorgi, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Valérie Golaz, Institut national d'études démographiques (INED), Paris, France.

Antonio Golini, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Rosa Gomez-Redondo, Universidad Nacional de Educacion a Distancia (UNED), Facultad de CCPP y Sociologia, Madrid, Spain.

Catherine Gourbin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Carl Haub, Population Reference Bureau, Washington, D.C., United States.

François Héran, Institut national d'études démographiques (INED), Paris, France.

Kenneth Hill, Department of Population and Family Health Sciences, Johns Hopkins University, Maryland, United States.

Serguey Ivanov, Population Division, United Nations, New York, United States.

Carol Jagger, Department of Epidemiology and Public Health, University of Leicester, United Kingdom.

Heather Joshi, Centre for Longitudinal Studies, Institute of Education, University of London, London, United Kingdom.

Nico Keilman, Department of Economics, University of Oslo, Oslo, Norway.

Shigemi Kono, Faculty of International Economics, Reitaku University, Chiba-ken, Japan.

Marlène Lamy, Institut de demographie de l'Universite de Paris (IDUP), Paris, France.

Jacques Légaré, Department of Demography, University of Montreal, Montreal, Canada.

Luc Legoux, Institut de démographie de l'université de Paris (IDUP), Paris, France.

Eva Lelièvre, Institut national d'études démographiques (INED), Paris, France.

Henri Leridon, Institut national d'études démographiques (INED), Paris, France.

Ron Lesthaeghe, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Thérèse Locoh, Institut national d'études démographiques (INED), Paris, France.

Michel Loriaux, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Enzo Lucchetti, Dipartimento di Biologia Evolutiva, Università degli Studi di Parma, Parma, Italy.

Dionisia Maffioli, Università degli Studi di Bari, Bari, Italy.

Godelieve Masuy-Stroobant, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

France Meslé, Institut national d'études démographiques (INED), Paris, France.

Alain Monnier, Institut national d'études démographiques (INED), Paris, France.

Karel Neels, Interuniversity Program in Demography, Vrije Universiteit Brussels, Brussels, Belgium.

Annunziata Nobile, Dipartimento di Istituzioni politiche e Scienze sociali, Università degli Studi Roma Tre, Rome, Italy.

Alberto Palloni, Center for Demography and Ecology, University of Wisconsin, United States.

Sophie Pennec, Institut national d'études démographiques (INED), Paris, France.

Pierre Pestieau, Université de Liège, Liège, Belgium and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Victor Piché, Inter-university Centre for Demographic Studies, University of Montreal, Montreal, Canada.

Marc Pilon, Institut de recherché pour le developpement (IRD), Piagadpigpi, Burkina Faso.

Gilles Pison, Institut national d'études démographiques (INED), Paris, France.

Antonella Pinnelli, Department of Demographic Science, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Denise Pumain, Institut national d'études démographiques (INED), Paris, France.

S. Irudaya Rajan, Centre for Development Studies (CDS), Kerala, India.

Jean-Louis Rallu, Institut national d'études démographiques (INED), Paris, France.

Rosella Rettaroli, Dipartimento di Scienze Statistiche, Università de Bologna, Bologna Italy.

Jean-Marie Robine, Démographie et Santé, Institut national de la sante et de la recherché medicale (INSERM), Montpellier, France.

Paul-André Rosental, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris, France.

Silvana Salvini, Dipartimento di Statistiche, Università degli Studi di Firenze, Florence, Italy.

Antonio Santini, Dipartimento di Statistiche, Univrsità degli Studi di Firenze, Florence, Italy.

Francis Sartor, Departement d'epidemiologie-toxicologie, Institut scientifique de la Sante Publique, Brussels, Belgium.[†]

Bruno Schoumaker, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Catherine Sermet, Institut national d'études démographiques (INED), Paris, France.

Jolande Siebenga, Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas.

Patrick Simon, Institut national d'études démographiques (INED), Paris, France.

Lamberto Soliani, Dipartimento de Scienze Ambientali, Università degli Studi di Roma "La Sapienza," Rome, Italy.

Salvatore Strozza, Dipartimento di Scienze Statische, Università di Napoli Federico II, Napoli, Italy.

Pierre Surault, Groupe d'études démographiques, economiques et socials (GEDES), Universite de Poitiers, Poitiers, France.

Dominique Tabutin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michael Teitelbaum, Alfred P. Sloan Foundation, New York, United States.

†Deceased

Marc Termote, National Institut of Scientific Research, University of Quebec, and Department of Demography, University of Montreal, Canada.

Laurent Toulemon, Institut national d'études démographiques (INED), Paris, France.

Tapani Valkonen, Département de Sociologie, Université de Helsinki, Helsinki, Finland.

Jacques Vallin, Institut national d'études démographiques (INED), Paris, France.

Celine Vandermeersch, Institut national d'études démographiques (INED), Paris, France.

Jean-Pascal van Ypersele, Institut d'astronomie et de geophysique G. Lemaitre, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

James Vaupel, Max Planck Institute for Demographic Research, Rostock, Germany.

Jacques Véron, Institut national d'études démographiques (INED), Paris, France.

Éric Vilquin, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Paolo Vineis, Dipartimento di Scienze Biomediche e Oncologia Umana, Università di Torino, Turin, Italy.

Anatoly Vishnevsky, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Tania Vishnievskaia, Institut national d'études démographiques (INED), Paris, France.

Carolyne Wanja Njue, Population Council, Nairobi I, Kenya.

Christine Wattelar, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Michel Willems, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

John Wilmoth, Department of Demography, University of California, Berkeley, California.

Guillaume Wunsch, Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Anatoli Yashin, Max Planck Institute for Demographic Research, Rostock, Germany.

Sergei Zakharov, Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia.

Hania Zlotnik, Population Division, United Nations, New York, United States.

S E C T I O N

I

HISTORY OF POPULATION THOUGHT AND POLICIES

Introduction to Volume IV

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Volume I looked at the central study of the discipline—the built-in mechanisms of population dynamics. Volumes I and II respectively addressed the determinants of the three big forces of population change (fertility, mortality, and migration). Volume III drew on all this groundwork to paint a broad panorama of the historical and geographical forces behind population processes and sketch out prospective future medium- and long-term trends. Volume III examined the biological, economic, social, and political consequences of population changes and, more generally, looked at the big issues facing society where these changes intersect.

But, should and can society as a whole try and shape the course of all of these processes? This is the central question in this fourth volume on population thought and policies. Volume IV will then go on to offer some insights into data collection methods and tools for demographic analysis.

Section I of Volume IV will address this issue of population thought and policies on three different levels, the first being the abstract level of ideas, the other two the concrete level of political realities, seen from two complementary angles: the broad concerns of a population policy and individual country case studies.

Before moving onto population policies as such, it was thought necessary to look first at the dominant thinking that shaped historical and contemporary population theory and informed the concerns of demographers, naturally, but more especially of philosophers, politicians, and the public. This short overview in two chapters is followed by a third on ethics as it affects demography. This first part concludes with an introduction to population policies proper. Because most of this work focuses largely on post-World War II developments, a historical review is given of pre-1940 population policies, while a few very basic reference points on the principles and context behind contemporary population policies are outlined in a very short chapter.

The major concerns of population policies (fertility, health, migration, age structures) are then considered in three parts, beginning with a world overview that takes in as far as possible the wide range of situations and policy approaches. This is followed by a comparative analysis of family and fertility policies pursued since World War II. Health policy is dealt with in the same way in the third part. The fourth part looks first at migration policies, then at policies concerned with population age structure.

But for all that it enables the topic to be considered in depth from a broadly comparative analysis standpoint, splitting the field up into these different subject chapters risked obscuring the fact that population change comes about as a result of the interplay between all these elements, and government attempts to shape population trends cannot ignore that: what they do (or do not do) in a sphere will affect population change only as it connects with the other factors in population change and hence with what they do (or do not do) in those other spheres. It was therefore thought necessary to amplify the subject approach with a fifth part containing country case studies to give an overview of all policies that have demographic consequences across all spheres. Eight chapters are therefore given over to as comprehensive as possible an analysis of the population policies pursued, purposefully or otherwise, over some 50 years in eight major countries, concluding with a survey of policies conducted at the international level to illustrate the unprecedented and at times quite decisive role played by the international community since World War II in the establishment of population policies.

P A R T

I

HISTORY OF POPULATION THOUGHT AND SHORT INTRODUCTION TO POPULATION POLICIES

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Although the word *demography* was a 19th-century invention (coined by Achille Guillard, 1855), population issues were a very early concern for human societies. The need to count people is lost in the mists of time. A major legacy has come down to us from the ancient world in this area, and no analysis of the population policies of our era can be undertaken without some mention of the history of ideas developed in different cultures down the centuries. The belief that births and deaths were ordained directly by a god, whose will alone determined their equilibrium or otherwise, left little scope for human population policies or even any real thinking on the matter, even when the mere fact of counting people was not itself taboo.[1] But philosophers from antiquity onward reflected on the theory, while administrators took censuses and, over time, the thinkers developed their ideas, explored the pros and cons of population growth, suggested to rulers policies or measures that might work toward the desired ends.[2] This is considered by Éric Vilquin in Chapter 97, in a concise but wide-ranging review of over 2000 years of history of population thought leading to the situation as it stood on the eve of World War II.

[1] The taboo on counting has religious connotations and is found in the Bible, but in both ancient and more recent history, has also often served as a cover for political strategies.

[2] A concern central to the work of the 18th-century Pastor Johann Peter Süssmilch ([1741], 1998), for whom "God's order" in no way precluded human intervention, provided it was consistent with what Providence ordained: the increased peopling of the Earth (Rohrbasser, 2001).

The difficulty of condensing into a single chapter so much history from so many different civilizations may be exceeded only by that of outlining the history of population thought in the few decades that followed, for at least three reasons. One is the indisputable difficulty a contemporary author, however learned, faces in objectively reviewing the development of fellow specialists' thought in light of a largely yet-unwritten history. But this standard difficulty is compounded by two things. Another is that this is among the most eventful periods in the demographic history of humankind—the most crucial phase of worldwide demographic transition, where population growth peaked. But it also witnessed a real growth spurt in the still very young discipline of demography, and a proliferation of new ideas and theories of all kinds. Paul Demeny tackles this in Chapter 98, on the trend in ideas on population since 1940, starting with an extensive look at the tools and materials readers can use to explore the matter in greater depth, and moving on to a more specific review of selected major concerns in contemporary demographic thought.

One consequence of the development of population science has been the increase in ethical issues surrounding the work of demographers, and the use to which the results of their research may be put, or perhaps even more so, the scientific or other use that may be made of the investigative techniques they employ. This sensitive issue clearly seemed to fit here, following on from the first two chapters of consideration on population thought, and as a prelude to the following chapters on population policies. François Héran therefore considers ethics and population in Chapter 99.

Part One concludes with two chapters that touch briefly on population policies proper. Plainly, these chapters concentrate mostly on developments in recent decades, seldom harking back to before World War II. But even modern population policies have historical roots, and a rapid review of population policies pre-1940 from Michael Teitelbaum in Chapter 100 informs the situation. Without preempting what follows, Stefano Baldi and Raimondo Cagiano de Azevedo sketch out in a concluding Chapter 101, some of the fundamentals of the principles and contexts on which today's population policies are based.

References

Guillard Achille, 1855. *Éléments de statistique humaine, ou démographie comparée*. Paris, Guillaumin, 376 p.

Rohrbasser Jean-Marc, 2001. Sussmilch et Frédéric II: le sceptique et le ministre de la providence, *L'expertise, Annales littéraires de l'Université de Franche-Comté*, p. 197–224. Besançon, France, Presses universitaires fran-comtoise.

Süssmilch Johann Peter, 1741. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortplflanzung desselben erwiesen.* Berlin, Kultur-Verlag, 360 p. (Reprint of the original edition by J. C. Spener, Berlin, 1741).

Süssmilch Johann Peter, 1998. *L'ordre divin dans les changements de l'espèce humaine démontrés par la naissance, la mort et la propagation de celle-ci.* Paris, INED, PUF, 352 p. (Unabridged version of the 1741 edition, translated by Jean-Marc Rohrbasser).

CHAPTER

97

History of Population Thought

ÉRIC VILQUIN

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

INTRODUCTION

Once a group of women and men develops the awareness to form a population which has a future of its own that transcends the overlaying succession of individual futures, some of its members will inevitably start thinking about how to manage this common history. Having perceived that the group has so far outlived the passing of individuals, they reflect on how best to ensure the same continuity in the future. Self-perpetuation is the aim of all populations! The mechanism of continual replacement that enables a population to endure over time (i.e., the combination of entries, births, and immigrations with exits, deaths, and emigrations) assuredly contains varying elements of inevitability or laws of nature that are relatively difficult to gauge and cannot but be endured. But humans would not be humans without a belief in their ability to shape a population, at least to some degree, as they have thought to do with so many other mechanisms in life and in the world; it is this specifically human conceit that is the source of progress. Our forebears were no more wearily resigned than we to the trends and vagaries of the components of demographic changes.

Population sustainability requires that the individuals comprising the population remain sufficiently healthy and survive long enough to contribute to population replacement. This in particular involves a sufficient proportion of the population managing to feed itself, protect itself day by day from the fatal stresses and strains of its environment, be sufficiently fertile, and be sufficiently untempted to emigrate. At what stage in the hominization process these basic considerations may have become expressly imprinted into the consciousness of some human being is not known and matters little. Suffice it to acknowledge that, long before writing and long before history, our remote ancestors certainly possessed the ability to conceive of population.

In order to influence the factors that shape their destiny, humans must assess their present situation, choose a goal to achieve or a course to follow, and frame an appropriate strategy. Where the future of the population is concerned, the point is to define a demographic ideal, measure how far one is from it, and decide how best to close the gap. The ideal population will be that whose absolute size, growth rate, and composition—judged by various criteria—are most apt to ensure both its long-term sustainability and its happiness; the strategy will be formed of a coherent set of effective, realistic, and legitimate ways of steering the population dynamics of the current population toward the ideal population characteristics. This means demographers, philosophers, and politicians—often one and the same—must work in conjunction.

If the aim of population sustainability and the strategic criteria of effectiveness and realism come down on balance to just plain common sense and calculation (sound demographic analysis), the same cannot be said of the aim of happiness and the criterion of legitimacy. What, within the bounds of the sustainability requirement, is a desirable population? Large or small? Dense or sparse? Increasing, stationary, or decreasing? Young or old? Homogeneous or heterogeneous (ethnically, socially, linguistically, politically, etc.)? In addition, which of the many effective and realistic policy approaches for achieving the chosen ideal characteristics should be selected? There is no scientific answer to this host of questions. Every society is deemed to have its own answers that square with the ethical choices and definition of happiness it has settled on; each individual, guided by his or her own conceptions of happiness and ethics, can formulate and express an opinion as to which population theory they would like to see the community follow. The history of population thought is a descriptive and comparative survey of theories that are, in the final analysis, only the demographic elements (where they exist and are overtly expressed) of belief systems of a much broader compass.

Advances in scientific knowledge have demographic consequences that must be taken into consideration by all subsequent population theories. So strong may be the influence of the underlying belief system, however, that all theories may not invariably do so. Also, demographic science has scarcely offered a description and explanation for all phenomena (which, indeed, no observer of human behavior can do, being unable to set up chemistry laboratory-style experiments), such that the gaps and uncertainties in knowledge leave a vast clear field for subjective and ideological argument, not to say theorists' own philosophies and personal psychologies.

Attempts have been made to apply different classification criteria to population theories, but none has become the standard. All told, a chronological classification is no more artificial than any other and suits the purpose here, barring the occasional liberty taken with the timeline. Also, rather than striving for a comprehensive survey—an equally hopeless and unfeasible endeavor—it seemed preferable to sketch out the history of population thought through an (inevitably personal) selection from among the most original and influential theories. Let me also stress that the main reason why Europe dominates this chapter is simply because European population theories have attracted significantly more analysis than those of other continents.

I. SOME OF THE EARLIEST KNOWN POPULATION THEORIES

We can ascribe to prehistoric populations a minimal implicit population theory: because the survival of the group depends on numbers,[1] it is essential to keep watch, and, if possible, act on, mortality, fertility, and migration; the pursuit of a more sophisticated social ideal than collective survival alone can only heighten nuance and add complexity to this type of concern. But it would be a mistake to conceive of a simple, coherent primitive population theory. The different situations and philosophies of the different peoples necessarily colored their thinking in this as in other fields: nomadic and sedentary communities were probably inclined to different demographic ideals; some may have been driven by the fear of underpopulation, others by overpopulation; all necessarily made their demographic choices agree with their religion, moral code, political organization, technology, psychology, and so forth.

[1] Volume of population, but also the specific absolute numbers of certain key subgroups: women (for reproduction), defenders (for war), producers (for the economy), and so forth.

Very few peoples in the ancient world gave expression to their population theories, at least in writing, which is the only sufficiently detailed and precise form. The interpretation of the archaeological remains left behind by preliterate peoples is too uncertain and approximate on this: the existence of (female and earth) fertility divinities in almost all nature-worshipping or primitive religions testifies only to an awareness of the importance of this factor for life and survival.

What theories are known, moreover, tend to be formulated by thinkers who may have been expounding only their personal ideas (e.g., Plato's theory is not universal for ancient Greece), although some admittedly influenced the behavior of their people. Care must be taken, therefore, not to generalize to all the great civilizations in the ancient world the smatterings of population theory that have survived for us in the works of their most renowned writers.

It would be impossible to survey all the religious, legal, philosophical, historical, or literary writings that fall within the ambit of demographic thought because they inform us about peoples' norms in relation to marriage, divorce, homicide, infanticide, sexuality, abortion, contraception, or their demographic agendas. What follows are merely examples of ancient theories which, in addition to their inherent merits, are deserving of special interest for the influence they may have on more recent theorists.

1. China: Confucius

Confucius (555–479 BC) codified and unified elements of long pre-existing theories, thereby investing Chinese civilization with a system of moral and political philosophy that left a powerful imprint on the rest of imperial history, permeating institutions and behaviors (Confucius, 1981; Gay-Sterboul, 1974).

The main aspects of Confucianism that affect the population are moral laws. They are not demographic in intent, but have demographic implications.

All Chinese—from beggar to emperor—are subject to the moral duty of filial piety based on the precept that the souls of the dead can only be kept alive (and favorably disposed to the living) in the afterlife by regular ancestral worship rites performed by their direct male descendants. For the man, therefore, reproduction is both a duty to his ancestors and a personal imperative. As an unbroken male line must be preserved at all costs, children must be produced in sufficient numbers to ensure multiple sons, and, in light of early mortality, ensure the survival of some children long enough to produce further children before dying. Given the extreme uncertainty of individual fertility and mortality, the only reasonable strategy is to aim for maximum fertility at almost any cost. Institutions and morality are therefore all ordered to that end. Its principal purpose being reproduction, marriage, arranged by families, sometimes before the birth of the young parties directly interested, is entered into only when the engaged couple have not only attained puberty, but are mature enough to bear and raise children. To further secure the birth of male offspring, a man may have concubines in addition to his wife. Repudiation of the wife is permitted and easily done; infertility is a normal reason for repudiation. A woman's life is justified only by bearing several male children to her husband. Bearing a female child is a failure, and it is understandable that, for the poor, a girl should be regarded as a superfluous burden that it is not unthinkable to rid oneself of. Remarriage by widows is frowned on, but that of widowers is a matter of course. Paradoxically, the art of loving, including contraceptive techniques, seems to have been cultivated by the elite in Chinese society, but it was strictly an indulgence of the nobility.

The emperor has duties specific to his function, which includes a duty of benevolence toward his subjects in return for their absolute obedience. He must ensure harmony and equilibrium in the working of the administrative and economic machinery of society through a realistic and fair policy. To this end, Confucius advises him to count the population in close detail: sex, age, social class, trade, and health are data that will enable the imperial government to accurately assess needs and resources. In fact, regular census taking was standard practice for the Chinese administration, so much so that it introduced for the purpose, at various eras and in different forms, hierarchically organized canvassing systems for the entire population, which at certain times provided more or less the same statistical services as a permanent population register.

Confucianism is a populationist theory. If the population is large, its work will ensure the welfare of all, the wealth of the country, and the emperor's power, provided there is a fair distribution of land and resources and the emperor treats his subjects benevolently. The success of this policy will be proved by the immigration of foreigners, further reflecting on the emperor's glory.

Confucius seems not to have expressly considered the demographic content of his moral precepts. He sees them as so absolute as to render any inquiry into the point or value of their consequences superfluous. He professes the optimistic belief that if all people, emperor and subjects alike, keep their place, perform their duties, and abide by the rules of moral conduct, the harmony of society and the world is assured.

Some of Confucius's followers glimpsed the possibility of a discrepancy between population growth and the finite character of natural resources: land that is farmed becomes exhausted, uncultivated lands decrease, people's needs increase, and so on. But they believed that overpopulation would result only from poor resource management, and the problem would be cured by a return to a sounder economic policy of low consumption, saving, and providence. There is therefore an optimum relationship between population and resources toward which the emperor's wisdom must steer the ship of state.

2. Greece: Plato and Aristotle

Two of ancient Greece's foremost thinkers developed somewhat similar population theories within the all-encompassing framework of a total political utopia. Plato (Vilquin, 1982; Plato, 1997) and Aristotle conceived an organization of society that was radically new in the Greek geopolitical context of their time: the city, a small territory centered on a single town, in which the citizens (free men) attended to public civil and military business while women and slaves did the work necessary for general life. They also formed part of a line of thought that questioned the traditional values of Greek civilization, beginning with religion and morality. The old morality based on ancestor worship, which made reproduction a religious duty to the family line, was in disarray, but was supplanted by a civic morality that made reproduction a civic duty to the State.

In his *Republic* and *Laws*, Plato (428–348 BC) conceives of a city whose organization and legislation strictly regiment each person's life so that no individual whim risks upsetting the eternal static equilibrium that constitutes the Platonic ideal.

Here, he develops the odd theory of the two fertilities formulated in his other philosophical works (in particular, the *Banquet*). Fertility of the soul is an exchange of thoughts between the minds of men. Its perfect form is the love between master and disciple, which is the vehicle for the transmission of science, philosophy, and the sense of beauty. It is Platonic love, with no sexual overtones. The other fertility is reproduction, reduced to a simple biological and social necessity. All the worst individual and social disorders, the attendant passions, crimes, lunacies, and rebellions are the result of commingling the two fertilities which should remain utterly discrete. Plato also sees the interplay of love and sexuality as the gravest threat to the harmony of his perfect city. So high are the stakes that the dissociation of the two fertilities must be inculcated in children at all costs, and any means by which to ensure that individuals respect it are acceptable.

Because reproduction is an entirely civic function, it is for the State alone to organize it. The State must regulate not only the institution of marriage, but also sexual relations, the "selection of sons" (Republic, III, 415 a–c)[2] (eugenics), and the education and guidance of children toward the different functional specializations that society needs. The State's power in these matters must be absolute and no personal, family, or other interest must compete with or challenge it. These premises therefore lead Plato into constructing an unwieldy, pettifogging, sometimes monstrous framework of demographic laws.

The only exception to the Platonic ideal of moderation, "right measure," even nonmovement, is the quest for physical (eugenics) and spiritual (philosophy) perfection on which no bounds can be set. The city must be neither rich nor poor, the population neither too sparse nor too numerous.

The *Republic*, the first description of Plato's great political utopia, is the product of this obsessional quest for "right measure"/moderation and eternal static equilibrium. Like all other aspects of the city (which should be "neither large nor small" V, 460 a), the population absolutely must be stationary. This obviously cannot be achieved if reproduction is left to the free will of individuals and families. All the strings of this infinitely delicate biosocial function must be completely in the hands of the State, that is, the aristocracy of the magistrates. This elite of scholars and sages knows and lays down the conditions for the happiness of the individual and the community, resulting in their absolute authority and the total obedience by all other citizens. The State's power is totalitarian: the State may compel, mislead, or brainwash. For the system cannot suffer the slightest discordance; total law and order cannot tolerate the slightest measure of individual fancy or initiative. This leads logically to Plato's abolition of property and the family, which could only stand in the way of the magistrates' authority. The objectives of a stationary population and eugenics dictate the laws.

There may be a community of property among all citizens, but it is not possible to talk of a community of women, since total separation of the sexes is the general rule, with marriage reduced to simple cohabitation ordained by the magistrates for a specific period. The magistrates must be perfect demographers: using calculations, which Plato does not explain, they must have point-in-time information on the number of newborns the city needs to ensure a constant, sustainable absolute size of population and its qualitative advancement. For that, the city ordains a given number of marriages. The couple will be designated by the drawing of lots (so that only fortune or the gods are to blame for dissatisfaction with the partner they acquire), but the draw will have been deftly manipulated to ensure that the best of men are partnered with the best of women—for the good citizen is physically healthy and attractive, politically docile, and psychologically predisposed to his or her allotted social function. To limit the production of weak or malformed children, Plato recommends that reproduction be authorized only within specified age limits (that vary somewhat from chapter to chapter). Like physically deficient babies, children conceived without the magistrates' authorization must also be eliminated (the technique of infanticide as practiced in Greece and mentioned by Plato was to expose children to the elements). The others will be raised collectively and conditioned to perform allotted duties to society and the economy. They will never know their parents or siblings, which will create problems for the eugenicists!

Plato is well aware of the difficulties that his magistrates will face in making the necessary prospective mortality and fertility calculations, and his chief fear is that a forecasting error may irreparably damage the entire edifice.

In the *Laws*, written 40 years after the *Republic*, Plato endeavored—frankly, with little success—to add a touch of realism to his great dream. Population stationarity and genetic selection remain the two main focuses, but there is a lighter touch to the social organization and laws. There is less duress and more incentive (rewards, penalties) and social pressure (taboos). Instead of abolishing the family and property, Plato neutralizes them: they will remain, but every care will be taken to see that they do not thwart the magistrates' social engineering. The problem of land distribution

[2] Like the Bible, Plato's writings have references independent of the page numbering, which are common to all editions and translations.

between citizens' families, which regularly exercised the Greek cities (sometimes to the point of civil war), is settled once and for all when Plato ordains that the city will only ever have 5,040 families. The land is fairly divided into 5,040 indivisible and nontransferrable lots, one per family, and this distribution is immutable and never to be challenged. Each lot must therefore have not only one owner, but also only one heir (male citizens only). This gives a new justification to the demographic laws: the magistrates must evaluate and manage the fertility of unions such that the total number of families neither increases nor decreases, and that each lot has one male heir only. Plato bitterly regrets the impossibility of ensuring that each union produces exactly one male and one female child that are fit to survive and reproduce, and so robust measures are called for: large families will surrender children to infertile unions, while supernumerary children will be taken out of society (by infanticide or, if adults, banishment). The magistrates, in charge of stationarity and eugenics, must gauge the city's short-, medium-, and long-term population requirements, and, on the basis of their calculations, issue mandatory orders: if the city is likely to suffer a shortage of children, they will enjoin reproduction for all for a specific period; in the event of a likely surplus, they will prohibit all sexual intercourse for a specific period. To enforce their orders, they will use a range of measures increasing in stringency from exhortation, through rewards and threats to compulsion. In the final degree of coercion, "nuptial inspectors" (i.e., matrons with authority over marriage) will have the power to enter homes to ensure that couples are obeying orders (!), exacting punishment on offenders.

Plato allows migration only as a last resort. Immigration is ruled out because of the danger of allowing into the city individuals not inculcated with its specific moral precepts from birth. Emigration (in fact expulsion under the guise of establishing settlements) is a means of ridding the city of undesirables of all kinds: individuals whose upbringing had not produced the desired outcome, antisocial elements, troublemakers or, quite simply, supernumerary individuals too old to be discreetly put to death as a baby would be.

Only in the *Laws* does Plato set the total number of the stationary population. Specifically, he sets the number of family heads at 5,040. Why this figure, rather than, say, 5,000? Because 5,040 is the product of the first seven whole numbers, which gives it precisely 60 divisors, including the first ten whole numbers and 12. For a philosopher fascinated by the arithmetical properties of numbers, the figure of 5,040 is that of the numbers of the same order of magnitude (wherein Plato probably situates "right measure") which best enables the population to be exactly subdivided into numerically equal groups and subgroups.

Averaged over time, each head of family must have a wife, a son (heir), and a daughter (who will marry a neighboring heir), so at all times there will be just over 20,000 free citizens. Strange as it may seem, at no point does Plato concern himself with the total population size of the city. He refers to slaves for the functions they perform, but considers neither their numbers nor their population dynamics. It was probably obvious to him that family heads, having absolute power over their slaves, would be capable of managing them like livestock in their best economic interests through a judicious and easily readjustable mix of commerce and breeding. The State had no need to concern itself.

The optimum population thus devised by Plato is not an economic optimum. There is no question of balancing population with land size, production, or consumption. The ideal absolute number is that which best enables the State to control the population, ensures that all the city's political institutions work to the best effect, and delivers a high level of satisfaction to citizens themselves in the fulfillment of their political offices (pledge of social stability). Overpopulation is a purely numerical concept: there is point-in-time overpopulation of the city when there are more than 5,040 male heirs.

Aristotle (384–322 BC) espouses Plato's theories in his *Politics* (Books II and IV). Certain aspects are nuanced and tempered, but the fundamental principles are no different (absolutism, stationarity). He fears population growth, but only because distributing the land and its produce among a growing number of individuals would impoverish some of the people, which can only result in social unrest. But he fears underpopulation no less, since it would deprive society of the economic advantages of the division of labor. He is no less authoritarian than Plato, for example in considering that the State could compel women to terminate their pregnancies if the number of children were deemed to be excessive.

3. Rome: Augustus

The Romans saw the issue couched in very different terms. Rome was the center of a sprawling empire that the Latin people had conquered by arms: Rome's view of the benefits of a large population was mainly from the standpoint of military might.

Augustus' predecessors had already noted the waning popularity of marriage and reproduction among Roman citizens. In the first century BC, the citizen class were finding the idea of a wife and children common. Spending one's wealth on festivals and banquets was a greater sign of status than keeping a family. As the literature, drama, and biographies of famous citizens attest, the unmarried state was fashionable among this class of the population, and there was no certainty of its replacement. The chief ambition of a young Roman citizen was to be adopted by a rich, aging bachelor and inherit his fortune. Augustus (63 BC–AD 14) thus found it increasingly hard to recruit young Romans of noble birth into the officer class of the army and to administer a rapidly expanding empire. Having clearly identified the sources of the problem, he promulgated a highly sophisticated set of laws to promote marriage and reproduction, and tackle what was hindering them. These laws (*lex Julia*, 17 BC, and *lex Papia Poppaea*, 9 AD, are the main ones) lay down a mix of generally financial deterrents against the declining morality of Roman society which lay behind a falling birth rate and the undermining of civic virtues. Adultery was severely punished. Cohabitation was automatically legalized. The financial advantages of being unmarried were scrapped. The ability to benefit under a will was radically reduced and in some cases abolished for single people, childless couples, women with fewer than three children, un-remarried widowers and divorcees. The confiscated portion legacies were allotted to large families. Parents with more than three children were highly favored.

These first laws failed to remedy matters, so Augustus introduced a 100-day delay between death and the vesting of the inheritance to give unmarried heirs the time to marry in order to claim their portion. But this final "tweak" seems to have added nothing to the effectiveness of this major body of demographic legislation. Circumventions and exemptions were quick to develop. Although they remained formally in force for several centuries, Augustus' laws failed to stem the decline of the citizen class. This may well be one reason why Rome found it increasingly hard to enforce its presence and authority at the far-flung borders of its vast empire. The fall of the Roman Empire may in part be due to the loss of demographic dynamism among its nobility.

II. THE DEMOGRAPHIC DICTATES OF SOME MAJOR RELIGIONS

1. Judaism and Christianity

For the Jewish people, the injunction in Genesis to "Be fruitful and multiply, and replenish (fill) the earth" (Gen.1.28 and 9.1.7), is both a commandment and a promise: let Israel be confident in its reproduction; God guarantees its fertility and continuance. God said to Abraham, "I will make thy seed as the dust of the earth [. . .] tell the stars if thou be able to number them [. . .]: so shall thy seed be [. . .] I will make thee exceedingly fruitful [. . .] I will bless thee and [. . .] will multiply thy seed as the stars of the heaven and as the sand which is upon the sea shore" (Gen.13.16; 15. 5; 17.6; 22.17), and He would promise the same to Hagar the Egyptian and her son Ismaël (the forebears of the Arabs), and to Isaac and his son Jacob (the forebears of the Israelites). Fertility thus has a dual ontology: human duty and divine blessing. God approves and favors the man who creates many children, but also, it is the man with whom He is well-pleased to whom He grants a large progeny.

Populationism was an obvious imperative for this broad mix of tribes threatened (and often vanquished) by powerful neighbors. Being unmarried was an unthinkable state; the patriarchs' polygamy found justification in the need to increase their cumulative fertility; infecundity was among the worst of misfortunes. As a particularly striking and spectacular demonstration of His intervention in the story of humankind, God makes an infertile woman fertile.

The Bible has many verses, often in the form of aphorisms, with clear populationist overtones—"In the multitude of people is the king's honour, but in the want of people is the destruction of the prince" (Prov.14.28); "Children's children are the crown of old men" (Prov.17.6)—which would be formalized into shibboleths by 17th-century populationist–mercantilists like Bossuet.

The Old Testament also gives an account of three population counts done in Israel (Num.1, Num.26, and 2Sam.24) which, as well as their obvious purpose of measuring the nation's military and economic might, have a sacred nature whose ambiguity long prevented any census being taken in the Judaeo-Christian world. This is because although, like Moses before him (Num.1, Num.26), King David was ordered by God to number the people (2Sam.24.1),[3] he was conscious of committing a sin entailing harsh retribution. One possible interpretation of this paradox would be that knowing the size of the flock is the exclusive prerogative of the owner: the king is only the shepherd; God alone is the Master. Perhaps also the multitude of descendants promised by God to Abraham is beyond any attempt at measurement. It is accepted that the sta-

[3] Another later account of this episode attributes the initiating of the operation to Satan (1Chr.21, 1–27). The census is nonetheless an impious act.

tistics presented as the findings of these censuses are purely symbolic figures.

Early Christianity expresses no views on population. It is chiefly about ethical matters, without regard for the strictly demographic implications of its moral precepts. Christian populationism is therefore only a logical consequence of the Christian attitude toward life, the individual, and love.

Respect for life, the prospect of resurrection and eternal happiness, trust in Providence, and condemnation of material gain are all implicit encouragements for individuals to reproduce. Christian morality is opposed to polygamy, divorce, abortion, and infanticide. While extolling the virtues of indissoluble, monogamist marriage, the apostle Paul saw the unmarried state and chastity as morally superior, but he was concerned with a small elite, not the population as a whole.

2. Islam

For Muslim believers, the Koran and the example of the Prophet are the sources of precepts that must be followed and are not open to interpretation. Islamic populationism has the same expansionist overtones as biblical populationism. Children are repeatedly described as a blessing from Allah and many precepts encourage men to marry and reproduce. Marriage is the normal state of the believer; to remain voluntarily unmarried is offensive. Early marriage is encouraged (the traditional Arab practice of prepubertal marriage is not prohibited), as is remarriage. Spouses are forbidden to live apart for long periods. Marital sexual intercourse is expressly praised. Traditional Arab polygamy is maintained, but now confined within strict limits: a man may not have more than four wives at once, and must treat them equally.

The natalist hue of these promarriage and -reproduction precepts is somewhat qualified by other tenets that specify a large number of periods of sexual abstinence (during menstruation, after widowhood, during periods of fasting and pilgrimage). Protracted breast feeding is strongly urged. Repudiation of a wife is easily done; a wife can also obtain a divorce, but with great difficulty.

The issue of contraception is not clearly addressed by the holy writings. It is therefore a field in which different schools have put forward different theories. For popular religion, permeated with pre-Islamic superstitions, contraception is impious. Infanticide is expressly proscribed, but abortion is not mentioned and is therefore also the subject of varied doctrinal positions.

III. THE MIDDLE AGES AND THE RENAISSANCE

1. The Fathers of the Church. The Reformation

In their treatises on morality and the documents of councils convened to condemn various heresies (Manicheism, Gnosticism), theologians from the early centuries AD up to the Middle Ages expounded on certain aspects of Christianity that merely reinforced its populationist hue: humankind must populate the Earth and Heaven, so fertility is the primary purpose of marriage; any sexual activity other than for reproduction is unacceptable, nonmarital sexual intercourse is unlawful, and abortion equals infanticide. The medieval Church exhorted monks and nuns to chastity while propounding fertility for married couples; it imposed celibacy on priests and made dissolution of marriage almost impossible. Most Christian writers followed St. Paul[4] and St. Augustine in believing that the ideal state for human beings is pure celibacy; the alternative for those unable to achieve or not called to it is fertility within marriage.

Saint Thomas Aquinas (13th century), perhaps the only doctor of the church to have thought in specifically demographic terms, was a firm advocate of a large and growing population as the foundation of the glory and prosperity of terrestrial kingdoms.

The Reformation, while not calling biblical populationism into question, did relax some of the unyielding rules of Catholic morality: divorce and remarriage were accepted subject to conditions, reproduction was not the main purpose of marital relations, and in the 20th century, contraception would come to be regarded as morally acceptable between spouses not acting out of pure self-interest.

2. Ibn Khaldoun

The Arab historian and statesman Ibn Khaldoun (1332–1406) conceived the idea of writing a vast universal history, starting with a methodological introduction: *Al Muqaddima* ("Prolegomena" or "Introduction," 1968). It is in this first volume of his major opus that he develops a cyclical theory of history, which he calls the "science of human sociability" and is the first tentative nontheological sociology. The dialectical relationship between wealth and population, taxation and psychology, plays a key role in it.

[4] See in particular the First Epistle to the Corinthians (1Cor.6, 13–20, 7, 1–9 and 7, 25–27) and the First Epistle to the Thessalonians (1Thes.4, 1–5).

Wealth and population are interdependent, but the substance of their relationship depends on the sociopolitical context. In the earliest stages of a civilization (settlement), the population grows. With increasing population density, division of labor lightens the burden on all, bringing a rising standard of living and increased efficiency of political, military, and economic institutions. The leaders follow the Koran, so taxation is minimal. Farmers therefore produce great wealth because they keep the profit of their labor. The social climate is naturally optimistic and highly conducive to rapid population growth. Economic and population growth enrich the State; governments are caught in the trap of excess and avarice and raise taxes. This is the start of the downturn in the cycle. Rising taxation gradually erodes producers' interest in working, and production falters, prompting the leaders to further raise taxes in order to maintain revenues. The population feels oppressed, and its original optimism fades into disenchantment and rebellion, which are not conducive to reproduction. Fertility drops while mortality and emigration are temporarily exacerbated by civil war, which brings the nation to its knees. The field is now open for a new and uncorrupt people to embark on a new cycle where growth will again be followed by the decline of civilization, wealth, and population.

IV. MERCANTILISTS AND PHYSIOCRATS

The implicit populationism of church doctrine seems not to have been called into question in the pre-Renaissance Christian world. But, with the Renaissance, a rediscovery of the thinkers of the ancient world, the development of centralized states and state-to-state relations, the discovery of the New World, the birth of capitalism, and many other innovations prompted intellectuals to revisit the dominant theories in all fields. By overlaying political and economic considerations onto the moral precepts, traditional populationism would then be given a more explicit and voluntaristic gloss by some and would be challenged by others. The old biblical aphorism that the glory and strength of the king are made by the number of his subjects became formal doctrine and would go virtually unchallenged for three centuries.

1. Mercantilist Populationism (16th–18th Centuries)

For theorists of monarchical absolutism, like Niccolò Machiavelli (1469–1527) (*Florentine Histories*, 1520–1526), the purpose of the political and social organization of a nation, the supreme goal of government of a people, is the power and wealth of the State (incarnated in the person of the monarch). Any individual concern for welfare comes second to this ideal of national greatness, which means that it is often sacrificed to it with glib fatalism. The population is an essential factor of the leader's political and military might: his advantage therefore lies in the population being large, and he must actively promote its growth.

The mercantilists developed additional, relatively new populationist arguments out of their economic theories. They argued that a nation's wealth lay in the quantity of gold and silver it possesses, and anything that draws these precious metals into the royal coffers is good for the State. The discovery and systematic conquest of America having brought a sudden influx of gold and silver into the Iberian kingdoms, the rest of Europe feverishly sought ways of siphoning off some of this windfall. International trade became the favored means of accumulating wealth. Each country was to produce the agricultural foodstuffs and industrial products necessary for the subsistence of its inhabitants as well as a surplus of manufactured goods to be sold at the highest possible price abroad.[5] Why this emphasis on manufactured goods? Because, on the international market, they are far more advantageous than foodstuffs: industrial activity is much more profitable than agricultural activity, costs less, and brings greater returns. Some even went so far as to see agriculture as being governed by the law of diminishing returns, while industrial returns increase thanks to the division of labor.

Where does population fit into this mechanism? A larger population contains more individuals who are not needed for agriculture and can therefore be employed in manufacturing. Population growth enables the development of an export industry, which is the only way to increase the wealth of the State. If wealth is gold, the way to get it is through population growth. This is what lies behind Jean Bodin's exclamation (in *Les Six Livres de la République* [Six Books of the Commonwealth], 1576) that "there is no wealth nor strength except in men." The population policy propounded by mercantilists was therefore first and foremost a pro-marriage, -large families, -immigration (preferably of manufacturing workers), and anti-emigration (except to populate the colonies).

[5] Some of these exports of manufactured goods may be used to cover imports of subsistence goods if the national production does not suffice. But the balance of trade, in gold, must remain positive (Cantillon, 1755).

The leading theorists of mercantilism are Jean Bodin (1530–1596) (*Les Six Livres de la République*, 1576), Barthélemy de Laffemas[6] (1545–1612) and Antoine de Montchrestien (1575–1621) (*Traité de l'économie politique*, 1616). Broadly, they believe the earth to be non-depletable, France underpopulated, and that a larger and better managed population could substantially enrich the king and the kingdom. Their population policy is promarriage and for controlled migration, and their economic policy focused on food self-sufficiency and the export of high added value manufactured goods. There are only benefits to population growth ("one should never fear there being a surfeit of subjects or citizens", says Jean Bodin in *Les Six Livres de la République*, 1576). At worst, should the nation's farmers be unable to supply sufficient foodstuffs, the State can always import food using the added wealth gained from the increased population. Montchrestien adds a significant nuance to classical mercantilist doctrine: the source of wealth is not just the population, but the population at work (*Traicté de l'oéconomie politique*, 1616), from which stemmed his preoccupation with an employment policy and advocacy for the restoration of the tradition of census taking (abandoned in Europe after the fall of the Roman Empire) to identify the shirkers and scroungers (those that cannot be put to work will be removed from the kingdom).

Colbertism is surely the most perfected form of demo-economic policy inspired by mercantilism. Colbert, Louis XIV's principal minister during the first part of his personal rule from 1661 to 1683, went to extraordinary lengths to foster population growth, industrialization, and exports, fearing neither to overturn customs and privileges, nor to oppress the populace. Notable among his many initiatives was the 1666 edict encouraging marriage and reproduction by granting tax exemptions to couples who married before the age of 20 and to families with at least 10 legitimate children. This statute was only rarely and poorly applied, and had to be repealed in 1683. Colbert's bid to drive out religious celibacy by limiting the numbers of priests, monks, and nuns, and raising the minimum age of female noviceship, had little more success. He encouraged immigration and banned emigration, except for settlement in the colonies (Canada). To increase the volume of work, he cut the number of public holidays and set up workshops for sturdy beggars. At the same time as reorganizing the navy and the colonies to increase French influence on the channels of international trade, he laid the costly foundations of a mainly export-oriented luxury goods industry manufacturing tapestries, silk and other costly fabrics, fine glassware, watches and clocks, and more. To maximize the export selling prices of these products, he used the prestige of Louis XIV and his court to constantly launch new fashions in clothing (ribbons, lace, accessories), which Europe's other courts would rush to imitate from suppliers in Lyons, Lille, and Paris. The well-oiled mercantilist machine was running smoothly and population policy was a big part of it.

To Colbert and his interest in population we also owe the regular publication of vital records in Paris from 1670 on, a failed attempt at a national statistical survey, and the testing out in Canada of the systematic collection of population data.

The very end of the 17th century witnessed the emergence of what may be called critical mercantilism, propounded by humanist philosophers (Fénelon (1651–1715),[7] Vauban (1633–1707),[8] and Boisguilbert (1646–1714)[9]) who agreed with the broad theory, but minus the totalitarian excesses that absolutism and Colbertism had produced. They believed that if the king's glory derived from population numbers, the misery of the populace could only tarnish it. Their concern for the population's standard of living led them to a major rethinking of orthodox mercantilism. They sought to restore the status of farming, unfairly neglected in favor of industry, to add morality to and lighten the tax burden, and to liberalize trade and the movement of goods to and from the home market. From there, they implicitly came to a new definition of national wealth: no longer gold reserves, but a well-fed, gainfully employed population.

In England and Germany, mercantilism coexisted with antipopulationist arguments informed by the fear of overpopulation, too varied to form a uniform school of thought. Thomas More (1478–1535) (*Utopia*, 1516) harked right back to Plato's stationarity and authoritarian communism, while the deeply pessimistic Thomas Hobbes (1588–1679) (*Leviathan*, 1651) doubted that emigration to the colonies would suffice to drain off the surplus population.

Few mercantilists accepted the possibility of a limit to the volumes of subsistence that national agriculture or international trade could supply, or the obvious ceiling that would set on population growth. The most highly developed such theory is that of the Italian Giovanni Botero (1540–1617) whose *Delle cause della grandezza e magnificenza delle città* (1593, repub. 1930)

[6] See in particular *Septième traité du commerce* (1601).

[7] See the complete works published by Gallimard in 1983.

[8] See in particular *La dîme royale* (Vauban [1707] 1992).

[9] See in particular *Le détail de la France*, 1695, republished by INED in 1966.

postulated that the life of humankind depends on two factors: the *virtus generativa*, which is humankind's propensity to reproduce, and the *virtus nutritiva*, which is the ability to produce subsistence. Botero's axiom is that the former is naturally more expansive than the latter: humankind's tendency is always to increase to the maximum allowed by the available means of subsistence. Point-in-time food stocks set a maximum level to absolute population size. Once the population rises above this maximum, a range of ills (famine, plague, war) will drain off the surplus. A state that wishes to increase its population must therefore first work to increase its food production; an increase in population will automatically follow.

After the death of Louis XIV, theorists taking stock of his disastrous rule took on a gloomier tone. For some, the mercantilist machine created its own problems, and the glimmering fears of overpopulation, whose harsh control by famine and wars could be more humanely achieved by emigration to the colonies. For others, population growth certainly added to the state's wealth, but its attendant excesses, vice, urbanization, and capitalization were all adverse to reproduction.

2. The Physiocrats

A breakaway theorist from the mercantilist school, Richard Cantillon (1680–1733), set the foundations for a radical new philosophy—physiocracy—which was only to be fully developed during the latter half of the 18th century. Cantillon (*Essai sur la nature du commerce en général*, 1755) reasoned that the real source of wealth was land (through labor). Landowners hold the reins of the economy, labor, and the population. The economic allocation they give to their lands (agriculture, stock farming, or amenity) dictates the volume of subsistence produced, labor demand, and hence maximum potential population growth. Cantillon also points out that, for a given quantity of subsistence, population size is determined by lifestyle: a population accustomed to excess cannot increase as much as a frugal population.

The principle behind all physiocratic theories is that of a fundamentally good and balanced natural order. Governing consists in enforcing that natural order and restoring it when it is upset. The physiocrats are therefore diametrically opposed to the economic interventionism advocated and practiced by the mercantilists. The real wealth of a nation is not its gold reserves or its population, but its land and agriculture. Only the land generates wealth, producing more than goes into it (François Quesnay's [1694–1774] "net product"; the articles "Fermiers" and "Grains" in the *Encyclopédie*, 1756–1757; *Tableau économique*, 1758); all other activities merely adapt the goods extracted from the land, adding nothing to national wealth. Logically, therefore, the nation's happiness lies in a fertile land, well-exploited by a sufficient population respectful of the natural order.

Population growth is no longer the absolute priority that it was for the mercantilists. It is desirable only insofar as the land is underused. And there is never any need for a specific policy to promote population growth, which occurs spontaneously to the full measure permitted by the means of subsistence available: "The only limits to [population] increase are those set by subsistence" (Quesnay, "Fermiers," *L'Encyclopédie*, 1756); "There are always as many people as can be fed" (Mirabeau, *L'Ami des hommes*, 1756). The best-fit criterion between population and subsistence is a decent standard of living. Unlike classical mercantilists, the physiocrats saw no benefits to a large but impoverished population. The aim of government must be not population increase but increased agricultural output, which will generate population growth and raise living standards. Physiocratic policy is therefore an agricultural policy to develop a prosperous farming community and promote scientific agriculture. Victor de Mirabeau[10] (*L'Ami des hommes*, 1756) even professes a belief in limitless agricultural progress that will not only enable but automatically produce population growth with no drawbacks. But he also believed that the population of Louis XV's France was declining, for which he blamed the architects of royal despotism, unfair taxation, luxury, and underexploitation of the land. François Quesnay reasoned far more pessimistically that the population inevitably tends to increase beyond what the available means of subsistence can support, so that the wage will always be minimal and poverty inevitable. The legislature's main concern, therefore, should be the populace's quality of life, not its numbers.

The controversy over depopulation of the known world sparked off around 1720 by Charles de Montesquieu (1689–1755)[11] (*Lettres persanes*, 1721) and Robert Wallace (1694–1771)[12] would impel many European intellectuals to propound theories on the origins of population growth that were more conviction- than science-based. Has the world or European population increased or decreased since ancient times? Is the population of a given country increasing or decreasing? With no population statistics at hand, sub-

[10] Marquis Victor Riquetti de Mirabeau (1715–1789) was the father of the powerful revolutionary orator.

[11] See in particular letters CXII to CXXII.

[12] See in particular his *Dissertation on the numbers of mankind* (Reed, 1969).

stantiation for views was sought from any useful source, and many drew on an implicit axiom that the population increases under sound government and declines under despotism.

Victor de Mirabeau's belief in the limitless possibilities of progress—not just in agricultural science, but science in general—to increase subsistence without restriction or maximum finds echoes in William Godwin (1756–1836) (*An Enquiry Concerning Political Justice*, 1793) and Marie-Jean-Antoine de Condorcet (1743–1794) (*Esquisse d'un tableau historique des progrès de l'esprit humain*, 1795), who were flatly opposed by Thomas Robert Malthus (*An Essay on the Principle of Population*, 1798). More realistically, Jean-Jacques Rousseau (1712–1778) (*Social Contract*, 1762) believed that there was a right relation to be kept between size of territory and population size (Book II, Chap. 10), but remained in favor of some population growth, seeing depopulation as a sure sign of degenerated government (Book III, Chap. 10).

3. Adam Smith

The extraordinary demo-economic rise of North America inspired the father of free-market capitalism, Adam Smith (1723–1790), to evolve a theory that pushes the mechanistic nature of population growth to the limit. Just as the production of goods is demand-driven, population changes are labor-demand-driven, with wages interfacing between supply and demand. A labor shortage that pushes up wages will have a knock-on effect on marriage and fertility rates, "the demand for men [. . .]necessarily regulates the production of men" (*An inquiry into the nature and causes of the wealth of nations*, 1776). Malthus was to adopt much of this approach.

V. THE FIRST DEMOGRAPHERS

Demography dates from 1662. Its founder, John Graunt (1620–1674) (Vilquin, 1976), and its earliest theorists (e.g., William Petty [1623–1687],[13] Edmund Halley [1656–1742],[14] Antoine Deparcieux [1703–1768],[15] Gregory King [1648–1712],[16] Louis Messance [1734–1796],[17] Jean-Baptiste Moheau [1745–1794],[18] Johann-Peter Süssmilch [1707–1767][19]) strove mainly to observe and describe the population, its composition and changes. Their ultimate goal was to uncover laws governing population akin to those being discovered in the natural sciences. And they did indeed highlight many statistical laws (sex ratio at birth, age- and cause-specific mortality, the relative invariability of the crude birth and death rates, etc.), which produced astonishment from within a wide audience. They argued that the scientific results of demography were essential to inform government choices and policy, but seldom ventured further into political or economic theorizing. Many gave the backing of science to the theory of geometric progression: the natural tendency of humankind is to multiply by geometric progression, but obstacles of many kinds act to make it grow much less quickly in reality. Johann Peter Süssmilch (*The Divine Order*, 1741) sees it as not only a natural but a divine law (as the title of his treatise clearly indicates) which it would be blasphemous to oppose; population growth is therefore inherently good since it is "ordained" by Providence; also, adds Süssmilch to still his own objections, there is plenty of room on the Earth.

Over time, the development, systematic production, and publication of population statistics, and advances in analyzing them, meant that caveats had to be added to population theories, which were now at risk of being scientifically disproved.

VI. THOMAS ROBERT MALTHUS

Thomas Robert Malthus had many forebears—he did not originate his theory. But so glaring a light did he shed on the law of population as to blind his ardent followers and fiercest detractors alike, setting off a furious debate that still rolls on today. Population cannot now be addressed without referring to Malthus and taking a stand on his thought. Sadly, Malthus's thinking, clumsily propounded by self-styled Malthusians, has also continually been deformed by rigid theorists more concerned with castigation than dispassionate examination. Emphatic refutations of the Malthusian system by some big names have been accepted as authoritative, and history's belated attempts to right the worst wrongs done to Malthus and his thought have not yet made sufficient inroads.

[13] See in particular, *An essay concerning the multiplication of mankind* (1682).

[14] See *An estimate of the degrees of the mortality of mankind* (1693).

[15] See his *Essai sur les probabilités de la durée de la vie humaine* (1746).

[16] See *The LCC Burns Journal* (1973).

[17] See in particular his *Nouvelles recherches sur la population de la France* (1788).

[18] See *Recherches et considérations sur la population de la France* (1778).

[19] See *Die göttliche Ordnung* (1741).

1. Thomas Robert Malthus (1766–1834), A Noble Intellectual

Born in 1766 into a family steeped in the utopian ideals that were the seeds of the French Revolution, Thomas Robert Malthus became a fellow of Cambridge University and a country parson after a broadly based education inspired by the educational theories of Jean-Jacques Rousseau. He was a man of nobility and rectitude, esteemed and liked. He had a reputation for wisdom without severity. Initially fired by his father's enthusiasm for all the social, reforming, and revolutionary ideas converging from all sides, both in France in the wake of the Revolution and in England then in the throes of turmoil, he nevertheless sought to distance himself from the dangerous optimism of some ideal society theories with the publication in 1798 of a pamphlet which caused a furor—his *Essay on the Principle of Population*. This was a bare-knuckle demonstration of the misguided theories held by Godwin, Condorcet, and other egalitarians or pre-Socialists on the causes of poverty and hence the means of eradicating it. Spurred on by the unexpected success of his pamphlet, he wrote a second edition (1803), which was both more developed and especially more meticulously documented with the collected "evidence" gleaned from his lectures and travels. From 1804 to his death (1834), he was a professor of history and political economy at the East India Company College, publishing, in addition to four new editions of his *Essay*, other significant works that rank him among the founders of the science of economics. He married late in life, at age 38, had three children, and seems to have lived his life by the virtues he preached.

2. Egalitarian Utopias

While almost all the political philosophies of previous centuries had been dominated by a single obsession (how to enrich the State or the Ruler), thinkers in the latter half of the 18th century were increasingly concerned with an issue of a very different order: how to stem the sharp rise in the number of poor. But rising poverty was nothing new. Renaissance England had witnessed an economic change that enriched the few by condemning the many to poverty: the land enclosures, which turned ploughed land to pasture, replaced corn by wool, cereal-growing by sheep-rearing. Sheep-farming is much less labor-intensive than crop-farming, forcing tens of thousands of displaced peasants to flee the land to huddle in poverty-stricken masses on the outskirts of towns. In 1516, Thomas More, in his *Utopia*, accused sheep of devouring men! As the canker continued to spread, Queen Elizabeth I revived a primitive form of social assistance in 1601–the poor laws. This body of legislation made each parish responsible for feeding paupers residing within their boundaries. Most English parishes discharged this obligation by levying a tax on ratepayers to provide bread to the unemployed and vagrants. Two centuries later, the poor laws were generally criticized for failing to prevent the unrelenting spread of poverty, despite the increasing cost burden they represented. The social visibility of poverty had reached unacceptable levels, and the remedies were clearly not working, which made poverty a universal subject of concern to philosophers and economists in the closing quarter of the 18th century.

Theories abounded, but some were keener to lay blame than try to understand why, while others preferred simply to dismiss something that they could not realistically conceive of. It was this latter attitude that shocked Thomas Malthus, and the subtitle of his *Essay* (". . . with remarks on the speculations of Mr. Godwin, M. Condorcet and other writers") is a thinly veiled statement of intent to lay bare the errors of Godwin, Condorcet, and others.

For these noble souls, poverty was all about the distribution of wealth. In an inegalitarian society, the arrogation of wealth for the benefit of the few is institutionally organized (in particular through ownership), plunging the many into abject poverty. Revolution will overturn these unfair institutions, at the same time sweeping away all obstacles to progress. With wealth then being fairly distributed, excess and poverty will be no more; also, humankind's biological and spiritual progress will lead it inexorably to an earthly paradise where the indefinite prolongation of human life will result in the diminishing and disappearance of the sexual instinct.

Poverty, thought Malthus, was too serious and acute a problem to be dealt with through such naive conjectures. And, refuting the theory of the indefinite perfectibility of mankind and society led him to develop his own explanation of poverty.

3. The "Principle of Population"

The centerpiece of his theory of poverty is encapsulated in a universal natural law, which many other writers (since renamed pre-Malthusians) had previously expounded, calling it even then the principle of population: the propensity of living species to multiply (driven by the instinct of reproduction) is at all times apt to exceed the Earth's capacity to produce the subsistence necessary to support that multiplication. The human species, like the animal species (a bold comparison for the time, which Darwin would draw on for support), is subjected by divine will to two forces pulling in opposite directions: the naturally powerful, fierce, and unreasoning instinct of reproduction is at variance with Nature's fundamental inability to produce means of subsistence at the same unchecked pace. The pattern of food production necessarily imposes its pace on population growth. The renowned geometric (of population) and arithmetic (of subsistence) progressions of increase that have elicited so much fruitless discussion are, as Malthus himself would point out, no more than rough depictions of the dizzying nature of the divergence between the two trends. He cites the example of the United States of America, a vast, wealthy, and clearly underpopulated country: its population doubles every 25 years, and even every 15 or 10 years in some regions. So long as neither space nor means of subsistence are lacking, therefore, the population naturally tends to increase in a geometric progression. Sooner or later, however, both space and means of subsistence will inevitably become scarce.

4. Misery, Its Causes and Remedies

The demands of the instinct of reproduction can be kept down to the level of food production possibilities in different ways, all of which will lead to the same necessary and inevitable result: that the Earth never has more than the number of inhabitants that it can feed. Animals—and, sadly, most humans—leave it to Nature to operate this check, and Nature has at its disposal only the grim means of ridding itself of the surplus of live-born individuals it lacks the means to feed.[20] Famines, epidemics, wars,[21] malnutrition, unhealthy living conditions, unemployment, and low wages,[22] all factors of premature death that Malthus refers to together as "positive checks," constitute misery. This is the first aspect of the thesis, the origin of misery. But a wide gulf separates Malthus's analysis from that of his detractors: the cause of misery lies not in human institutions that can be changed by revolution, but in a universal (divine) natural law that no revolution can change.

No living species can evade the principle of population, but misery—Nature's means of applying that law—is an inevitability only for animals. This is because Man's capacities for reason, self-assessment, and forethought give him other resources than to submit to Nature. He can take control of his life and replace the destructiveness of Nature (misery) with a preventive human strategy proper of ensuring that there are no surplus individuals condemned to an early death for want of subsistence; it is enough for them not to be born. Misery can be averted by rendering it redundant, that is, by replacing it with another way of keeping population commensurate with the means of subsistence: voluntary fertility control. All men have the choice of avoiding misery. They may (by default) be driven purely by reproductive instinct to produce more children than they can feed, or they may have only the number of children for whom they can decently provide, which, unless they are wealthy, will certainly lead to a smaller family size than in the former hypothesis. This is the second aspect of Malthusian theory: solving misery through the use of preventive checks.

As a Christian moralist, Malthus dismisses vice and acknowledges contraception as an effective but wrongful means of fertility control. All that remains, therefore, is marriage, but only after having acquired the means to support a family, remaining pure in the celibate state ("moral restraint").

The first way of applying this solution to the problem of misery in practice is to explain to those most patently in need of explanation—the poor—how their situation has come about: the natural mechanism of the principle of population and the false inevitability of the resulting misery; then to describe the solution to them, and urge and help them to choose wisely. For a liberal economist like Malthus, this is purely a policy of education, unsullied by restraint. No social or economic system (which he systematically reviews) offers any way of avoiding the choice between the preventive and destructive obstacles imposed by the principle of population.

[20] In the 1803 edition of his *Essay*, Thomas Malthus expressed this in the celebrated "Nature's feast" passage: "A man who is born into a world already possessed, if he cannot get subsistence from his parents on whom he has a just demand, and if the society do not want his labour, has no claim of right to the smallest portion of food, and, in fact, has no business to be where he is. At nature's mighty feast there is no vacant cover for him. She tells him to be gone, and will quickly execute her own orders, if he does not work upon the compassion of some of her guests. If these guests get up and make room for him, other intruders immediately appear demanding the same favour. The report of a provision for all that come, fills the hall with numerous claimants. The order and harmony of the feast is disturbed, the plenty that before reigned is changed into scarcity; and the happiness of the guests is destroyed by the spectacle of misery and dependence in every part of the hall, and by the clamorous importunity of those, who are justly enraged at not finding the provision which they had been taught to expect. The guests learn too late their error, in counter-acting those strict orders to all intruders, issued by the great mistress of the feast, who, wishing that all guests should have plenty, and knowing she could not provide for unlimited numbers, humanely refused to admit fresh comers when her table was already full." So great was the outcry caused by this passage that Malthus deleted it from later editions.

[21] Surplus consumers contending for too small quantities of space and food.

[22] Surplus demanders contending for an insufficient quantity of available jobs, and the jobs are allocated to those who accept the lowest wages.

5. Challenging False Solutions

If misery results from the principle of population, any putative solution of misery that does not go to the levers of that principle is unrealistic. Malthus systematically examines such various solutions as social revolution, public assistance, and private charity, and concludes that in every instance these are bogus remedies that in no way affect the conditions for application of the principle of population. They neither check the insensate drive of the instinct of reproduction, nor radically increase the productive capacity of the land, which are the only possible avenues to a real solution. A simple redistribution of existing wealth by whatever means changes neither the number of consumers, nor the quantity of subsistence available, nor the conditions in which both are produced. Justice and charity are lofty moral values, certainly, and Malthus in no way disavows them, but they can only alleviate specific misfortunes, solving the social problem of misery is beyond their ken. In the final analysis, palliating misery simply displaces it, replacing one set of poor by a larger number of new poor. Malthus's opposition to the poor laws of England is wholly consistent, therefore. Palliating the distresses of the poor allows them to reproduce without having the wherewithal to provide for their children, who will therefore increase the poor population. This Malthus summed up in a short phrase that set British society in uproar: "the poor laws create the poor which they maintain"! (*Essay*, 1798, p. 83).

6. Liberal Philanthropist or Cynical Reactionary?

The ease with which Malthus dismantled the false solutions outraged many who took their vengeance through smear campaigns, painting him as a cold and cynical reactionary. Although fundamentally an economic liberal, Malthus was no less fundamentally progressive. His foremost concern was to increase the standard of living for the whole population, and he was more than ready to compromise his economic liberalist principles so that the poor did not bear the brunt. He called for sweeping social reforms (democratization of education, medicine, savings, guaranteed minimum wage, etc.) to promote the rise of the poor up the social scale, in the belief that this would dispel the fatalism and resignation that stopped them from taking control of their own future.

Nonetheless, objectively speaking, Malthusianism relieves the governing and privileged classes from all responsibility for misery and its elimination, cutting the ground from under the feet of any revolutionary theorizing;[23] objectively, while ever people's standard of living is low, marriage and families are bourgeois privileges; it is also true that Malthus was an apologist for social inequality, which he saw as morally reprehensible but socially necessary as the only effective engine of human activity, work, and progress. His crusade against the putative right of the poor to be maintained at society's expense is little short of shameful. Much remains still to be said about the alarming repercussions of Malthusian theory.

VII. THE 19TH CENTURY–PRO- OR ANTI-MALTHUS

1. Malthusians and Neo-Malthusians

The very first Malthusians were naturally the bourgeoisie. Living handsomely on the profits generated by an unheeding, unquestioning capitalism, they arrogated wealth and power, right and respectability to themselves everywhere in Europe. Malthus providentially salved their consciences of any unease about the plight of the laboring classes. When French minister François Guizot made his celebrated "Make yourselves wealthy" speech[24] to the Chamber of Deputies (Lower House), the noble parliamentarians' ovation was given with no hidden agenda.

The 19th century saw the flowering of classical economic theories. All the writers in this vast movement were familiar with Malthus's *Essay* and, treating it as established scientific fact, drew arguments or at least inspiration from it on which to base and develop their views on the role of population in the mechanics of socioeconomic processes. Most were doomsayers who believed that the principle of population was an inescapable fact, and even sought to refine the case for it.

David Ricardo was a friend of Malthus, and a Malthusian: the "law of diminishing agricultural returns" (*On the Principles of Political Economy and Taxation*, 1817) reinforces the principle of population; wages act to regulate the population because, naturally settling at subsistence level, they embody the maximum means of subsistence to which the absolute size of population always necessarily adjusts. Many economists make the so-called wages fund theory a central article of faith: when the population increases, wages decrease, and wages can rise only if the population decreases.

[23] As was perfectly well understood by bourgeois governments and socialist purists, the former embracing Malthusianism as a fixed scientific truth, the latter showering insults on Malthus's head.

[24] Given on 1 March 1843.

John Stuart Mill (1806–1873) (*Principles of Political Economy*, 1848) also argued that the working population must restrict its numbers as the only means of achieving full employment with decent wages. From a given population density, any population growth brings the law of diminishing returns into operation and the standard of living falls. For Mill, the ideal tendency of the entire connection between volumes and prices (including population and wages) is toward a stable equilibrium.

Some argued that population was proportioned to capital stock, while others believed that capital accumulation and population growth were in competition. Jean-Baptiste Say (1762–1832) (*Catéchisme d'économie politique*, 1815) recommended "saving rather than procreating."

There was a broad consensus of liberal thought to reinforce the class theory element of Malthusianism. Malthus was cited as authoritative proof for the argument that the misery of the working classes was caused not by free-market institutions, but the ignorance and improvidence of workers, and that fertility regulation by the working classes would result in the elimination of misery without calling the principles of economic liberalism into question.

Jean Charles Léonard Sismonde de Sismondi (1773–1842) (*Nouveaux principes d'économie politique*, 1819) stood apart from this trend as a self-professed Malthusian who was opposed to economic liberalism. He believed that self-interested capitalism bore some responsibility for the emergence and spread of misery, and advocated legislative protection of workers against abuse. He was a somewhat unorthodox Malthusian, arguing that misery made the poor improvident and prompted them to reckless reproduction; foresight therefore had to be inculcated in them (education, saving), which would lead them to regulate their fertility.

This era also saw the development of neo-Malthusianism, chiefly through those who subscribed to Malthus's view that the poor must limit their family size to improve their lot, but without his moral scruples, believing that all or almost all means were justified to reduce fertility. More activists than theorists, they ran information campaigns to persuade the working class of the benefits of contraception. They included luminaries of the non-Marxist far left, feminists, and anarchists calling for a "womb strike" to choke off capitalism by depriving it of the "labour fodder, cannon fodder, and pleasure fodder" (Chapelin, cited by Bureau, 1920) which were its lifeblood! In England and France, these movements encountered draconian suppression by law and justice.

While English and French socialists were on the whole fiercely anti-Malthusian, some German socialists like Adolph Wagner (1835–1917) (*Foundations of Political Economy*, 1876) believed that regulation of working-class fertility was essential. But, believing the education in individualism and providence counseled by Malthus to be ineffective, they argued for State-imposed restriction of fertility, advocating measures, some of which foreshadowed the worst perversities of Hitlerism, like the forced castration of poor male children.

As if to validate Malthus, for whom the United States was the exemplar of a country where the principle of population had not yet taken effect, Malthusianism would not find a foothold in North America before the 20th century.

2. The Anti-Malthusians

On the fringes of liberal capitalism which had embraced Malthus's views as a revelation, some optimistic economists, like Frédéric Bastiat (1801–1850) *Harmonies économiques*, 1850), argued that the opportunities brought by industrialization and technological progress (division of labor, productivity growth) disproved Malthusianism and held out the prospect that subsistence would increase consequent on the rate (exaggerated by Malthus) of population growth.

The Church shared that optimism and overlaid it with theological arguments to refute Malthusianism, which it argued denied the affirmation of Holy Writ that the multiplication of living things, and mankind in particular, was a good thing and God's will (Genesis). Moreover, it was unacceptable for some men to be declared surplus and so eliminated (or prevented from being born). Malthusianism was a self-seeking and blasphemous theory. Society must strive to increase subsistence so that any nascent human being may come into life.

Following William Godwin, vehemently refuted by Malthus, the early socialists argued that misery was caused by the organization and functioning of society. Standing chronologically between Malthus and Marx, they were more concerned with repudiating what they saw as the heinous principle of population than constructing serious foundations for their own explanation and solution for misery.

Very broadly, the pre-Marxian socialists challenged the idea that the working class were to blame for poverty-creating overpopulation, and repudiated a theory that sought to justify the preservation of existing social structures with their glaring injustices.

Inequality of resource allocation, regardless of absolute population size, was what produced misery among some of the populace, and created a false impression of overpopulation. Some went further to argue that, were the working classes to follow Malthus's counsel and limit their family size, then far from improving their standard of living, social inequalities would actually widen, as the wealthy would arrogate to themselves all the benefits of the demographic slowdown.

The most utopian argued simply that an egalitarian, collectivist, or socialist society would spontaneously generate regulating mechanisms which would prevent population growth from outstripping the increase in subsistence (Charles Fourier, 1772–1837) (*Le nouveau monde industriel*, 1829). Pierre-Joseph Proudhon (1809–1865) (*Système des contradictions économiques*, 1846) ("There is only one man too many on Earth: Mr Malthus!") and Louis Blanc (1811–1882) (*L'Organisation du travail*, 1839) turn Malthus's argument on its head: misery causes overpopulation; when their standards of living rise, the working class will become provident and limit their fertility. They cited anecdotal evidence that the wealthy limited their family sizes to maintain their standard of living, while the working class living on subsistence level had nothing to lose or gain from an inegalitarian system.

Karl Marx (*Das Kapital*, 1867, Vol. 1) would draw together and give substance to the disparate arguments of the earliest socialists, without thereby clarifying the role of demographic variables in the socialist economy and society. He refuted the idea of a principle of population discrete from the economic context, arguing for a law of population peculiar to each society, and more specifically, to each mode of production. The misery of the working class was a necessary consequence of the capitalist mode of production.

The capital accumulated by individual capitalists comprises both constant (buildings, machinery, intermediate goods) and variable capital (stock of money exchanged as wages for the labor of workers). One effect of competition between individual capitalists is the invariable tendency for the constant to increase much more than the variable capital. Like capital, production increases, and so the economy should be able to satisfy the needs of an increased population. But the individual capitalist is coldly calculating and always more inclined toward labor-reducing investments that consign to unemployment workers who would otherwise have been employed in accumulating more capital. This is why Marx talks of a "relative surplus-population" (Book 1, Chap. 25, Sect. 3). In this way, an "industrial reserve army" (ibid) is formed, comprising not only workers unable to find work and those thrown out of work by mechanization, but also a whole range of other laborers without stable jobs, farm laborers ready to leave the land, and all society's rejects—(widows, orphans, the sickly, etc.).

This mass of unemployed labor is not only a by-product, but also an essential instrument of the capitalist mode of production; from it, individual capitalists source and return the labor they need, and consign their supernumerary labor according to the fluctuations in demand. The unemployed also press down on the wage demands of the active labor force, ensuring that rising wages do not undermine profit rates.

However, while Marx carries out a searching critique of the law of population of the capitalist mode of production, he is extremely noncommittal on the law of population in a socialist system, simply asserting that as private ownership of the means of production disappears, so too will the mechanism of relative surplus-population. His ostensible argument is that rising standards of living in an egalitarian society will reduce natality, but little more; Marx offers no theory of population. He does not consider the problem of population growth relative to the means of subsistence or even production; he analyzes only one aspect of it—relative surplus-population—but seems unable to carry this partial analysis over into the broader problem.

Ill-equipped with the bases of theory on this by its founder, Marxism can only signal its implacable opposition to Malthusianism by attacking it for objective complicity with bourgeois privilege and professing a dogmatic populationism, largely founded on a blind faith in progress, which, after the revolution, will no longer be held back by the private interests of capitalists and resource owners. Only Friedrich Engels, albeit taking a more optimistic view than Marx (socialist production would outstrip capitalist production), raised the possibility of an absolute surplus-population in a socialist society. But he believed in a socialist society's ability to carry out population no less than production planning, if need be.

Revisionist socialists accepted that contraception could bring immediate relief to the lot of the working population. But orthodox socialists argued that the poor should not yield to Malthus by limiting their family size, since he was merely an apologist for his own class and capitalism. Instead, they took refuge in Marxist theory, deferring the issue to the advent of socialism, when the problem would no longer arise since it was a problem specific to capitalism. Most claimed (but brought no proof) that socialism would admit steady population growth above that permitted by capitalism but far from the geometric progression of Malthus. They also rejected the law of diminishing returns as a Malthusian axiom espoused by classical economists.

Vladimir Ilyich Ulianov, popularly known as Lenin, took the Marxian view, describing neo-Malthusianism (*Pravda*, 16 June 1913) as typically "petty-bourgeois." Contraception and abortion are personal liberties, but strategically against the interests of the proletariat. This line of argument exerted a strong influence on socialist thought right up to the late 20th century.

VIII. MATHEMATICAL, BIOLOGICAL, AND SOCIOLOGICAL THEORIES

1. Logistic Growth

In about 1840, Adolphe Quetelet (1796–1874) who had met and corresponded with Malthus, thought to add the final touch to Malthusianism by making it a mathematical model. He tasked a young associate, the mathematician Pierre-François Verhulst, with formalizing a self-limiting geometric growth model. After a series of attempts, all unsatisfactory to Quetelet, Verhulst (1845, 1847) devised the "logistic function" as a common model of time-dependent population growth. The graph of this function takes the form of a sigmoid curve. After an initial phase of near-exponential growth, the population reaches its "normal" absolute size when all productive land is farmed; growth is continuous in time as long as new land can be put under crop, but slows more rapidly (proportional to the difference between current population and normal population) when it passes a point of inflection and approaches a horizontal asymptote, that is, a maximum absolute population size.

Rediscovered and popularized circa 1920 (Pearl and Reed, 1920), the logistic law has attracted increasing criticism for omitting too many key factors involved with population social and cultural development. Its explanatory capacity is of little value and its forecasting capacity highly unreliable. The few attempts to incorporate economic and psychological factors into logistic theory have been short-lived.

2. Biological Approaches

Herbert Spencer's (1852) biodemographic theory describes the antagonisms between the propensities of species to survive and to reproduce. He argued that development of the nervous system decreases the tendency to multiply in favor of self-maintenance. Fertility is inversely proportional to development of the organism. Accordingly, the least developed individuals will multiply, but being exposed to the principle of population, only the fittest will survive.

Francis Galton (1869) believed that differential fertility was present between races and social classes, and recommended that the process should be directed toward the development of a superior race.

Corrado Gini's (1934) cyclical theory assumes that the rising standard of living (among the higher social classes) diminishes reproductive energies. After World War II, Josué de Castro (1952) gave a similar theory a deceptive scientific veneer, which lasted only 20 years or so. He claimed that increasing protein intake in the diet led to sterility, while certain dietary deficiencies, by contrast, increased fecundability ("the poor are under-fed, but misery makes a fertile bed"), but no reliable scientific evidence was produced to give credence to these assertions.

3. Sociological Theories

Many sociologists have stressed the importance of population growth, especially increased population density, in the process of civilization, intellectual, economic, and social progress. For Emile Durkheim (1893), *inter alia*, population density is an essential driver of progress because it enables the division of labor. For Eugène Dupréel (1928), increasing density develops individuals' activity in all fields through the development of linkages between them; he also argues that a population that perceives itself as growing is imbued with optimism.

Another focus of concern for sociologists was to find an explanation for fertility decline in industrialized countries. Their many and varied theories involve growing individualism resulting from rising standards of living, women's economic activity, urbanization, rationalism, and so forth.

Arsène Dumont's (1890) theory of "social capillarity" argues that the desire for upward mobility leads to restriction of family size, since fertility is a financial burden that competes with other possible investments (see Chapter 92); the desire to secure higher social status for their children, moreover, will often impel parents to invest more in their education than in their number.

IX. OPTIMUM POPULATION

The basic idea underpinning the theory of the optimum is very simple: if a population can be considered as too small or too large, there must be an intermediate stage between underpopulation and overpopulation that is more balanced than these two extremes. Adolphe Landry (1934) saw the ultimate aim of demography as being to determine the theoretical conditions of this optimum and put them into practical terms for specific populations. The idea of an optimum is at least implicit in several older theories: the golden mean of China, Plato's static equilibrium, the optimum combination between population and subsistence or welfare theorized by Richard Cantillon and many others. But the concept of optimum population came under methodical study only from the 1920s onward.

Even the definition of an optimum raises controversial issues, for in fact it involves determining the main demographic characteristics (mainly absolute size) of a population that will maximize its happiness, the assumption being that happiness is not maximized where population is too high or too low. The crucial issue is, if not to define happiness, at least to select a reliable and preferably measurable indicator of it: longevity, individual income level, for example. Any change in one constituent of the definition of happiness will produce a shift in the optimum.

One source of this theory lies in the assertion of some classical economists that an increase in the absolute size of a population produces first an increase, then a decrease, in the return on its economic activities (Cannan, 1861–1935) (*Elementary Political Economy*, 1888). There is, therefore, an absolute size of population that will yield a maximum return on its activity. One objection to this line of argument is that there is probably a different optimum for the different activities of a population.

In economic terms, conjecturing a certain quantity of wealth produced by a certain quantity of population, for distribution among that population, the optimum may lead to conclusions with which its theorists themselves may be uncomfortable. By applying the laws of increasing and diminishing returns and the (average and marginal) productivity trends, it can be concluded that maximizing individual welfare depends on reducing the number of individuals who must share in the national wealth (where the wealth generated decreases more slowly than the population). But a new question arises (previously raised by Victor de Mirabeau, Jean Charles Sismondi, and others): is it better for a few to enjoy a very high level of welfare, or for a slightly lower level of welfare to be shared by a greater number? The economic optimum offers no answer to this question.

Optimum theory therefore leads to a dilemma: either a satisfactory definition of happiness is achieved through mostly nonmeasurable elements, for which no quantitative relation can be made with absolute population numbers and so no optimum absolute size can be calculated, or happiness is reduced to a quantifiable indicator, and that simplification unduly constrains the scope of the results. Some proponents of this theory have therefore accepted that between underpopulation and overpopulation—which can be defined—there lies a broad optimum area in which the choice of an optimum point is discretionary and depends on population- and time-specific, especially noneconomic, criteria to define that population's happiness.

Attempts to further refine the theory have been made with the concept of a dynamic optimum. Because the optimum is never achieved and constantly shifts with the continual changes in the context in which a population evolves, rather than matching absolute population size to a putatively fixed context, population development should be matched with the evolution in its environment. Although conceptually more pleasing, this new theory of the optimum resolves none of the technical contradictions of its predecessor: its key constituents are and remain unquantifiable; the optimum rate of change can no more be determined than the optimum absolute size.

Alfred Sauvy (1963) put an interesting gloss on the definition of the optimum: it is the absolute size of population that best permits achievement of the self-appointed objectives of a given society, from which a wide range of optima can be defined, linked to the various possible societal objectives: geopolitical supremacy (optimum power), population living standards (optimum welfare), economic efficiency, elimination of poverty, economic self-sufficiency, attainment of a philosophical ideal, and so on. Alfred Sauvy also sought to bring into the theory previously neglected elements like the linkages between the growth and changes in the structure of a population (his marked preference was for population growth from what he regarded as the all-around beneficial effects of the resulting rejuvenation of the population).

X. DEMOGRAPHIC TRANSITION

Population thought in the latter half of the 20th century was dominated by the theory of demographic transition. This theory, whose raw makings predate World War II (Landry, 1934; Thompson, 1929), grew out of the observation that all industrialized countries had, over approximately the last two centuries, followed the same demo-economic development from a starting stage (agrarian society) to an end stage (industrial society). As this trend seemed to be independent of the physical, economic, and political characteristics of the populations concerned, it was hard to resist the beguiling temptation to develop it into a universally applicable theoretical model, particularly so for the comforting answers it appeared to bring to the huge and distressing question mark raised by the demographic future of the Third World, and hence the planet.

Adolphe Landry (1934) divided time-dependent population change into three phases or "demographic regimes." In the primitive regime, the pressure of subsistence (or, more generally, the economy) on the population affected only mortality, not fertility. Nature was left to regulate the population by taking off the excess numbers. Through this process of "natural" adjustment, the population always tends toward the maximum supportable by available subsistence. Both fertility and mortality are very high. The absolute size of this population experiences very moderate long-term growth that is frequently and suddenly reduced by crises of excess mortality (epidemics, wars, etc.).

In the intermediate regime, scientific progress gradually improves mortality, while economic progress brings the population a standard of living that it aspires to maintain or further improve. This calculation leads to a measure of fertility regulation (initially through postponed marriage). The third stage, or contemporary regime, brings a sharp drop in natality, not just as a result of individual economic motives; this, said Landry, disproved the idea of a universal law.

Various later authors produced classifications somewhat different from that of Adolphe Landry, but the broad pattern remains unchanged, focusing chiefly on time-dependent sequences of fertility and mortality trends.

There are two aspects to the transition phenomenon. First, it is a transition from equally high mortality and natality to equally low mortality and natality. Between these two, spectacular economic progress brings major changes in the population whereby mortality and fertility are reduced by different and independent mechanisms. Second, there is a time lag between these two downturns, and it is the scale of that lag that determines the pace of transition, since it dictates the trend in the population growth rate. In most of the historical situations for which the theory tries to account, mortality decline has preceded fertility decline, such that transition has commenced with a relatively spectacular population growth and ended with population stationarity or even decline.

Following so many others, this theory was seized on as offering a universal model, and used in particular to predict the demographic future of the Third World (having recognized that its "demographic revolution" was the effect of mortality decline, while fertility remained very high). Its proponents therefore naturally looked closely for the signs of the onset of fertility decline. In some cases, feeling it to be too long delayed, they forced it in order to give credence to a theory that they deemed not only scientifically right but also politically expedient.

At the same time, the theory was attacked by demographers and historians as an over-simplification of the real and often quite different population processes of different industrialized countries. Their conclusion was that its lack of general value in describing the past applied equally to forecasting the future. They also stressed that the onset of transition in the Third World today is taking place in a radically different context from that of the onset of Europe's own transition at the end of the 18th century.

One major weakness of this theory is that it merely describes the course of a phenomenon but not its mechanisms, the factors involved, or the cause–effect relationships. The search for a theory to explain fertility decline is probably key to a renewed impetus for transition theory.

XI. MAJOR CONTEMPORARY DEBATES

At this close remove, a detailed review of the ferment of 20th century thought on population issues is clearly not possible. Interest in population issues will obviously be increasingly acute and widespread at a time in history of utterly unprecedented population events and situations: wars that wrought the greatest carnage ever, the largest scale genocide, the Third World population boom, the most coercive population policies, the inexplicable inconsistencies and collapse of fertility in wealthy countries, and more. The final quarter of the 20th century saw the average annual growth rate of the world population reach and exceed its historical peak. Scientists and politicians fear unbridled growth in the total numbers of humankind during most of the 21st century, which will be very unequally distributed between rival worlds. Population awareness is more than ever global, and population thought arguably more discrete than ever. The scale and pace of the demo-geopolitical upheavals in progress are fertile ground for the resurgence of theories of every kind.

Demographic load remains very much part of the world political agenda, keeping nationalist populationism very much alive.[25] The arguments propounded

[25] Michel Debré, French Prime Minister from 1959 to 1962, is said to have dreamed of a "France of 100 million people."

for it are varied: renewed fertility efforts, possibly combined with selective immigration, to prevent white Western Christian civilization from becoming a numerical minority, to ensure the survival of a language or a culture, to balance and outweigh others. The recurrent theme of improving fertility often cloaks fear of immigration and its social and cultural effects. In some cases, the exclusionary and racist overtones are clear: increasing the number of births so that society's needs do not have to be met from immigration, selecting only prospective immigrants that match the host country profile, driving out foreigners that cleave to their cultural identity, cleansing the nation.

One of the most hotly discussed topics of recent years among population theories is that of population and development (Chapter 86). The interaction between population growth and economic development is widely acknowledged as self-evident, but theories as to the meaning and nature of this relationship are irreconcilable and all the more strenuously argued (in particular at the big United Nations conferences: Bucharest, 1974; Mexico City, 1984; Cairo, 1994) where the stakes run into the billions of dollars! The wealthy who see population growth consuming the bulk of the crumbs they spare for the economic development of the Third World take comfort in the belief that funding family planning programs is much more effective and profitable than lending assistance for economic development; while the poor, who see themselves trapped in inescapable economic inferiority, feel that only an economic takeoff will help check population growth.

The industrialized world is engaged in a different, internal debate on population aging. The question is, in order to avert the collapse of pension and social security systems in the meantime, should fertility be encouraged to increase the contributor base for when the number of beneficiaries increases, or should the way the systems work be adjusted to changes in the shape of population pyramids?

It is also in wealthy societies that an almost total schism has silently come about between the sexual behaviors of couples and individuals and the codes of sexual morality propounded by the great religions, mainly Catholicism. The Roman Catholic Church, with what is often described as obdurate consistency, has regularly renewed its prohibitions on contraception (with the exception of "natural" forms of birth control), abortion, remarriage after divorce, and euthanasia (see, *inter alia*, Pope Paul VI's encyclical *Humanae Vitae*, published in 1968). But no more in the 20th century than in the 1st has it expressly linked its moral injunctions to demographic issues. In the debate on population and development, it argues that the Malthusian view is no more than a self-seeking blind refusal to help build a society of fairness and fellowship.

Malthusianism received a new lease of life after World War II with the emergence of environmentalism. The Club of Rome's (Meadows *et al*., 1972) work on the linkages between world population trends, forms of economic development, and the future of the planet's energy resources, was the starting point for a major extension of the Malthusian argument. The imperatives of prudence and providence apply both to human fertility (of rich and poor alike) and management of the natural heritage; uncontrolled demographic change foredooms the long-term, sound management of the earth. The most alarmist of these arguments gave rise to the radically new concept of the right of successor generations, but while it has become a bandwagon issue, its content has not really been spelled out.

What will inform population thought in the coming century? All the old lines of reasoning and argument will be more alive than ever. But, probably more than ever, the issues will be those of space and food, ethics and realism, urgency and indecision, differences and injustices.

Acknowledgment

This chapter was translated from the French version by Glenn D. Robertson.

References

Sacred Works

New American Standard Bible. Study Bible, 1999. Grand Rapids, Michigan, Zondervan Publishing Company, 2055 p.

The Koran, 2003. Translated by J. M. Rodwell. London, Phoenix Press, 509 p.

Other References

Aristotle, 1998. *Politics*. Translated by Ernest Baker. Oxford and New York, Oxford University Press, 423 p.

Bastiat Frédéric, 1850. *Harmonies économiques*. Paris, Guillaumin, 463 p.

Blanc Louis, 1839. *L'organisation du travail*, 5th ed. Paris, Bureau de la Société de l'industrie fraternelle, 284 p.

Bodin Jean, 1576. *Les six livres de la République*, 6 vol. Paris, Fayard, 340 + 123 + 209 + 216 + 229 + 550 p.

Boisguilbert (de) Pierre, [1695] 1966. Le détail de la France. La cause de la diminution de ses biens et la facilité du remède, en fournissant en un mois tout l'argent dont le Roi a besoin et enrichissant tout le monde, *in:* Alfred Sauvy and Jacqueline Hecht (coord.), *Pierre de Boisguilbert ou la naissance de l'économie politique*, p. 581–662. Paris, INED, 2 vol. 1032 p.

Botero Giovanni, [1593] 1930. *Delle cause della grandezza e magnificenza delle città*. Turin, Istituto Giuridico, 84 p.

BUREAU Paul, 1920. *L'indiscipline des mœurs.* Paris, Bloud et Gay, 608 p.

CANNAN Edwin, 1888. *Elementary political economy.* London, Oxford University Press, 152 p.

CANTILLON Richard, [1755] 1952. *Essai sur la nature du commerce en général.* Paris, INED, 196 p.

CHAPELIN Émile, 1920. *Ayez peu d'enfants ! Pourquoi ? Comment ?* quoted in P. Bureau, *L'indiscipline des mœurs*, Paris, Bloud et Gay, 608 p.

CONDORCET Marie Jean Antoine Nicolas Caritat (Marquis de), 1795. *Esquisse d'un tableau historique des progrès de l'esprit humain.* Paris, Agasse, 389 p.

CONFUCIUS, 1981. *Entretiens de Confucius.* Translated by Lunyu, with introduction, notes, maps, and chronology by Anne Cheng. Paris, éditions du Seuil, 180 p.

DE CASTRO Josué, 1952. *Géopolitique de la faim.* Paris, Les éditions ouvrières.

DEPARCIEUX Antoine, 1746. *Essai sur les probabilités de la durée de la vie humaine.* Paris, H. L. Guérin and L. F. Delatour, 208 p.

DUMONT Arsène, 1890. *Dépopulation et civilisation. étude démographique.* Paris, éditions Lecrosnier et Babé, 520 p.

DUPRÉEL Eugène (ed.), 1928. *Deux essais sur le progrès.* Bruxelles, Maurice Lamertin, 272 p.

DURKHEIM Émile, 1893. *De la division du travail social.* Paris, F. Alcan, 471 p.

FÉNELON (de Salignac de La Mothe) François, 1983. *Fénelon, œuvres*, 2 vol. Edited by Jacques Le Brun. Paris, Gallimard, 1638 p., 1830 p.

FOURIER Charles, 1829. *Le nouveau monde industriel et sociétaire, ou Invention du procédé d'industrie attrayante et naturelle distribuée en séries passionnées.* Paris, Bossange père, 576 p.

GALTON Francis Sir, [1869] 1892. *Hereditary genius: an inquiry into its laws and consequences*, 2nd ed. rev. London, Macmillan, 390 p.

GAY-STERBOUL Sylvie, 1974. Confucius, ses disciples et la population, *Population*, vol. 19(4–5), p. 771–794.

GINI Corrado, [1930] 1934. *Nascita, evoluzione e morte delle nazioni: teoria ciclica della popolazione e i vari sistemi di politica demografica.* Rome, Libreria del littorio, 137 p.

GODWIN William, 1793. *An enquiry concerning political justice and its influence on general virtue and happiness*, 2 vol. London, G. G. J. and J. Robinson.

GRAUNT John, 1662. *Natural and political observations made upon the bills of mortality.* London, J. Martin, J. Allestry and T. Dicas, 150 p.

HALLEY Edmund, 1693. An estimate of the degrees of the mortality of mankind, *Philosophical Transactions*, vol. 17(196), p. 579.

HOBBES Thomas, 1651. *Leviathan, or the matter, forme and power of a common-wealth ecclesiasticall and civill.* London, Andrew Crooke, 396 p.

IBN KHALDOUN 1968. *Discours sur l'histoire universelle (Al-Muqaddima).* Translated and edited by Vincent Monteil. Beyrouth, Commission internationale pour la traduction des chefs-d'œuvre, 1434 p.

KING Gregory, 1973. The L. C. C. Burns journal: a manuscript notebook containing workings for several projected works (composed c. 1695–1700), *in:* Peter Laslett (ed.), *The earliest classics John Graunt [. . . and] Gregory King*, p. 1–292. Weastmead, Farnborough, Hants, England, Gregg International Publishers Limited, 457 p.

LAFFEMAS (de) Barthélemy, 1601. *VIIe traicté du commerce, de la vie du loyal marchand, avec la Commission du roy et bien qu'il faict aux peuples et royaumes.* Paris, Léon Cavellat, 15 p.

LANDRY Adolphe, 1934. *La révolution démographique. Étude et essai sur les problèmes de la population.* Paris, Sirey, 231 p.

LENIN Vladimir Ilyich Ulianov, 1913. Paper in *Pravda* no. 137, June 16, 1913.

LENIN, 1987. *Essential Works of Lenin.* Mineola, NY: Dover Publications, 372 p.

MACHIAVEL Nicholas, 1952. Histoires florentines (Istorie fiorentine), *in:* Edmond Barincou (ed.), *Machiavel: œuvres complètes.* Paris, Gallimard, 1639 p. (La Pléiade).

MALTHUS Thomas R., 1798. *An essay on the principle of population.* London, McMilan/J. Johnson, 396 p.

MARX Karl, 1867. *Das Kapital, Kritik der politischen Ökonomie. 1. Band.* Hambourg, O. Meissner and New York, L. W. Schmidt, 784 p.

MEADOWS Donella H., MEADOWS Dennis L., RANDERS Jorgens and BEHRENS III William W., 1972. *The limits to growth.* New York, Universe Books, 310 p.

MESSANCE Louis, 1766. *Recherches sur la population des généralités d'Auvergne, de Lyon, de Rouen, et de quelques provinces et villes du Royaume, avec des réflexions sur la valeur du blé tant en France qu'en Angleterre, depuis 1674 jusqu'en 1764.* Paris, Durand, 330 p.

MESSANCE Louis, 1788. *Nouvelles recherches sur la population de la France, avec des remarques importantes sur divers objets d'administration par Monsieur Messance, receveur particulier des finances de Saint-Étienne en Forez.* Lyon, Les Frères Perisse, 131 p.

MILL John Stuart, 1848. *Principles of political economy, with some of their applications to social philosophy*, 2 vol. London and New York, J. W. Parker and C. C. Little and J. Brown.

MIRABEAU (Riquetti Marquis de) Victor, 1756. *L'ami des hommes, ou Traité de la population*, 3 vol. Avignon, 431 + 578 + 577 p.

MOHEAU, 1778. *Recherches et considérations sur la population de la France*, new ed. 1994, Paris, INED, 652 p.

MONTCHRESTIEN (de) Antoine, 1616. *Traicté de l'oeconomie politique.* Geneva, Droz, 452 p. (Les classiques de la pensée politique).

MONTESQUIEU de Secondat Charles, [1721] 1998. *Lettres persanes.* Paris, Pocket, 390 p.

MORE Thomas, [1516] 1997. *Utopia.* Mineola, NY: Dover Thrift Editions, 96 p.

PAUL VI, 1968. *Humanae vitae.* Paris, Centurion, 63 p. (Lettre encyclique de Sa Sainteté le Pape Paul VI sur le mariage et la régulation des naissances).

PEARL Raymond and REED Lowell J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation, *Proceedings of the National Academy of Sciences*, vol. 6(6), p. 275–288.

PETTY William, [1682] 1686. *An essay concerning the multiplication of mankind*, 2nd ed. London, Printed for M. Pardoe, 50 p.

PLATO, 1997. *Complete works*, Edited by John M. Cooper and D. S. Hutchinson. Indianapolis (Indiana) and Cambridge, Hackett Publ., 1808 p.

PROUDHON Pierre-Joseph, [1846] 1872. *Système des contradictions économiques ou Philosophie de la misère.* Paris, Librairie Internationale.

QUESNAY François, 1756. Fermier, *in: Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers, par une société de gens de lettres*, vol. 6, p. 528–540. Paris, Briasson.

QUESNAY François, 1756. Grain, *in: Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers, par une société de gens de lettres*, vol. 7, p. 812–831. Paris, Briasson.

QUESNAY François, [1758] 1958. Tableau économique, *in:* INED (ed.), *François Quesnay et la physiocratie. II Textes annotés*, p. 675–682. Paris, INED, 1005 p.

RICARDO David, 1817. *On the principles of political economy and taxation.* London, J. Murray, 589 p. (Republished in *Ricardo. Complete works*, vol. 1. Cambridge, 1951).

Rousseau Jean-Jacques, 1762. *Du Contrat social*, Edited by Pierre Burgelin, col. GF. Paris, Flammarion, 187 p.

Sauvy Alfred, 1963. *Malthus et les deux Marx: le problème de la faim et de la guerre dans le monde.* Paris, Denoël, 353 p.

Say Jean-Baptiste, 1815. *Catéchisme d'économie politique, ou Instruction familière qui montre de quelle façon les richesses sont produites, distribuées et consommées dans la société*. Paris, Impr. de Crapelet, 160 p.

Sismondi Jean Charles, 1819. *Nouveaux principes d'économie politique ou de la richesse des nations dans ses rapports avec la population*, 2 vol. Paris, Delaunay, 379 p.

Smith Adam, 1776. *An inquiry into the nature and causes of the wealth of nations.* London, George Routledge, 780 p.

Spencer Herbert, 1852. A theory of population, deduced from the general law of animal fertility, *The Westminster Review*, vol. 57, p. 468–501.

Spengler Joseph J., 1936. French population theory since 1800, *Journal of Economic History*, vol. 44(1), p. 577–610.

Spengler Joseph J., 1936. French population theory since 1800, *Journal of Economic History*, vol. 44(2), p. 743–766.

Spengler Joseph J., 1942. *French predecessors of Malthus, a study in eighteenth-century wage and population theory.* Durham, (North Carolina), Duke University Press, 398 p.

Spengler Joseph J., 1945. Malthus' total population theory: a restatement and reappraisal, *Canadian Journal of Economics and Political Science*, vol. 11(1), p. 83–110.

Spengler Joseph J., 1945. Malthus' total population theory: a restatement and reappraisal, *Canadian Journal of Economics and Political Science*, vol. 11(2), p. 234–264.

Stangeland Charles E., 1904. *Pre-Malthusian doctrines of population: a study in the history of economic theory.* New York, Columbia University Press and The Macmillan Company, 356 p.

Stassart Joseph, 1957. *Malthus et la population.* Liège, Université de Liège, Faculté de Droit, 342 p.

Süssmilch Johann Peter, 1741. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen.* Berlin, Kultur-Verlag, 360 p.

Thompson Warren S., 1929. Population, *American Journal of Sociology*, vol. 34(6), p. 959–975.

Vauban (de) Sébastien, [1707] 1992. *La Dîme royale.* Paris, Imprimerie Nationale, 293 p.

Verhulst Pierre-François, 1845. *Recherches mathématiques sur la loi d'accroissement de la population.* Bruxelles, Académie Royale des Sciences et Belles Lettres, 38 p. (Nouveaux Mémoires de l'Académie Royale des Sciences et Belles Lettres de Belgique, vol. 18(1)).

Verhulst Pierre-François, 1847. *Deuxième mémoire sur la loi d'accroissement de la population.* Bruxelles, Académie royale des sciences et belles lettres, 32 p. (Nouveaux mémoires de l'Académie royale des sciences et belles lettres de Belgique, vol. 20(3)).

Vilquin Éric, 1976. La naissance de la démographie, *Population et Famille*, vol. 3(39), p. 145–164.

Vilquin Éric, 1982. La doctrine démographique de Platon, *European Demographic Information Bulletin*, vol. 13(1), p. 1–18.

Wallace Robert, 1969. *A dissertation on the numbers of mankind in ancient and modern times*, 2nd ed. New York, A. M. Kelley, 338 p.

CHAPTER

98

History of Ideas in Population since 1940

PAUL DEMENY

Population Council, New York, United States

INTRODUCTION

Labor in industrial societies is highly specialized. A large proportion of gainful activities are directed—through exchange in economic and political markets—to the production of goods and services aimed at satisfying the needs of the day and the foreseeable future. Within the social division of labor, however, in modern economies a growing proportion of human effort is allocated to science: to activity aimed at a better understanding of natural and social phenomena and processes through observation, description, and the derivation of regularities. Some of these activities may be best seen as ornaments that an affluent society desires and can afford. What once may have been amateur pastimes only, now find systematic support within the walls of academia and in specialized research institutions. Thus, studies are devoted to uncover the architecture of ancient Pompeii, to chart the evolution of medieval music, or to fathom the origins of the solar system. Utilitarian arguments, narrowly construed, poorly justify such work: its aim is satisfaction of curiosity and aesthetic interest now elevated to the level of a social need.

The analysis of the state and dynamics of human populations, the core subject matter of the science of demography, may provide its practitioners intellectual stimulation and, when a job is well done, a sense of accomplishment similar to what those students of Etruscan vases or of distant galaxies might experience. But it is important to remember that what society demands of the study of population is far more directly tied to considerations of practical use. If the goal in describing demographic facts were simply satisfaction of curiosity—contemplation of intriguing numerical oddities and regularities in a mirror, as it were, held up to a population—demography no doubt would still flourish in any affluent society. It would be, however, a much more limited enterprise than what it is today. Society through its institutions—governments, foundations, international organizations and the like—seeks to have access to a polished and multifaceted mirror, providing an accurate demographic image, because it suspects, indeed knows, that what appears in the mirror is likely to be less than fully in harmony with the collective interest, or with the interests of various collective constituencies. Similarly, predictions concerning population change are sought because of the anticipation that demographic trends, evaluated by various value criteria, are less than optimal.

Optimality does not mean a state of bliss. The notion of demographically "less than optimal" merely implies that through appropriate collective rearrangements, improvements could be achieved. Optimality would be reached when this is no longer possible: that is, when, under the existing rules of the political system, proposed further rearrangements could not command support that is sufficiently strong for their adoption. Demographic knowledge is needed, and population research is socially supported, first and foremost because it is expected, directly or indirectly, to yield useful information to the political process searching for opportunities for social improvement. Awareness of this social need suggests a simple fourfold classification of the functions of demography and its practitioners. Demographers ask four types of questions that start, respectively, with the words: What? Why? With what effect? and, finally, What (if anything) can and should be done about it? In other words demography, and demographers, in their special domain focused on the state of population and its dynamics, provide *description*, investigate *causes*, study *consequences*, and engage in studies of *policy*. The massive volumes of this treatise address in great detail the topics delineated by this fourfold division. But while the classification just outlined is a useful device in thinking about and in understanding what demographers do, particular products of demographic research do not necessarily fall in just one of the four boxes identified by the classification. Description, explanation of causes, discussion of consequences, and even discussion of policy matters may be bundled rather than neatly separated. Still, one of those ingredients typically tends to dominate, suggesting the right pigeonhole. But as the present volume amply demonstrates, investigating the substantive issues as delineated above requires appropriate methods—instruments of research. Such instruments are sometimes imported ready-made from neighboring social science disciplines, but more often must be laboriously home-grown. The importance of such instruments in demography, and the appreciable share of research effort demographers devote to developing and polishing them, warrants that separate consideration be given to a fifth topic: *tools*. Most, although by no means all, demographic tools serve the goal of better description. Thus, discussion of research instruments may be thought of as subordinated to the topic of description.

I. LANDMARKS OF POPULATION THOUGHT SINCE 1940: AN OVERVIEW AND A LITERATURE GUIDE

At least two weighty reasons support the seemingly arbitrary period–demarcation line of 1940 indicated by the title of the present paper. First, the steady accumulation of demographic work, traceable back to the 18th century has markedly accelerated after 1940. A number of notable developments signaled this shift. These include the launching of a series of ambitious demographic studies at Princeton's Office of Population Research in the early 1940s under the sponsorship of the by-then moribund League of Nations; the extensive scientific activities of the Royal Commission on Population in Britain, set up in 1944; the establishment of the Institut national d'études démographiques (INED) in France; and the launching of the first two major specialized scientific journals in demography: *Population*, in 1945, under the auspices of INED, and *Population Studies*, in 1947, under the sponsorship of Britain's Population Investigation Committee. Other novel initiatives were also important. The United Nations established its Population Commission in 1946 to "provide advice and assistance on matters affecting or affected by population changes" (Trygve Lie, 1995). At the same time, to carry out population studies and to serve as a source of expert analysis and advice in population matters, the Population Division was created within the UN Secretariat. In 1947 the International Union for the Scientific Study of Population (IUSSP), originally founded in 1928, was reconstituted as an international association for individuals interested in population studies. The establishment of the Population Council in 1952 and the subsequent entry of private U.S. foundations, notably of the Ford Foundation and the Rockefeller Foundation, into the field of population gave impetus to the establishment of vigorous university-based demographic research programs in the United States and in a number of developing countries. The 1954 World Population Conference, held in Rome, marked the beginning of an influential series of international scientific congresses on population that were sponsored, respectively, by the United Nations and by IUSSP. Within individual countries, the immediate postwar years witnessed a greatly expanded role claimed by governments in economic and social matters that generated a strong demand for more timely and more extensive demographic data. National statistical organization supplying these data, in turn, set up research units for the analysis of demographic information. Thus, post-1940, within the social sciences demography became an established independent discipline.

A second consideration that justifies 1940 as a demarcation line in population studies is the strong shift in the subject of demographers' interest that can be best linked with that particular date. In the 1920s and 1930s there were two major preoccupations in the "cutting edge" of demographic research: interest in eugenics—(in changes, actual or potential, in population quality) and the related concern with low fertility in industrial societies, pointing toward eventual population decline. The trauma of World War II brought discredit to eugenics, resulting in the virtually total dissociation of demographers from work in that field.

And the unexpected upturn of Western fertility, that culminated in the postwar baby boom but was already evident in the early 1940s, made the concern with below-replacement fertility at least temporarily moot. At the same time, the accelerating tempo of mortality decline in the economically less developed world and the resulting dramatic acceleration of the rate of population growth, coinciding with the process of decolonization, shifted many demographers' interest to the study of population dynamics in the countries of Asia, Africa, and Latin America—a topic only marginally present in pre-1940 demographic research. The perceived negative economic consequences of rapid population growth greatly amplified demand for research on the demography of the developing world and also strengthened demographers' interest in consideration of the policy implications of their findings. This, in turn, generated demand for research on the factors influencing population growth and its proximate and underlying demographic and socioeconomic determinants, and stimulated demographers' involvement both in research on the nondemographic—economic and social—consequences of demographic change and in policy studies based on findings from such research.

Thus, ironically, at the same time that demographic studies gathered strength and demography became a recognized independent academic field, its disciplinary boundaries also became fuzzier and less well-defined. In descriptive demographic work, and in fashioning tools for that work, demographers are in possession of their own uncontested proprietary terrain. That is still true when demographers seek to explain a change in a particular demographic variable in terms of changes in some other demographic variable. But in searching for deeper explanations for population processes, demography must become an "interdiscipline": demographers need to acquire skills and cooperate with practitioners from other fields of study—from history, sociology, anthropology, economics, political science, to name only disciplines most evidently capable of contributing to a better understanding of demographic phenomena. Making use of a broad array of skills is even more of an obvious necessity in research on the consequences of population change and in the domain of policy studies; there, demographers are often only junior partners in interdisciplinary cooperation.

Not surprisingly, there is a tendency among demographers to resent this dilution of their professional identity and a preference for drawing narrower, hence clearer, boundary lines around their discipline. Doing so assures easy communication with colleagues and provides the familiar academic and intellectual comforts of conducting discourse only with the like-minded. Such discourse is important for advancing knowledge in the core areas of demography, and wise institutional design should assure that it is sufficiently fostered and protected. But the temptation to make it exclusive should be resisted; wise institutional design should also seek to stimulate an opening up of the cloistered world of demography to wider intellectual influences through interaction with other specializations. As is, few demographers would be brave enough to claim that existing arrangements in the ways demographers are trained, research is conducted, research results are scrutinized, published, and disseminated have found satisfactory solutions. Such solutions require maintaining a delicate and shifting balance between disciplinary exclusivity (with the attendant risk of marginalization in terms of the social significance of the resulting research products), on the one hand, and interdisciplinary cooperation (with the attendant risks of losing professional identity and intellectual overextension), on the other.

The ideas of demography are hatched in demographers' heads. If the ideas emerge from that locus—and often they do not—they may take the form of words, spoken in a classroom, or in professional debates, or in public forums. Each of these sources can be influential and important, but each of them is elusive and limited in outreach and permanence. Ideas in the scholarly world can be communicated fully satisfactorily only if they are committed to paper, pass critical peer review (admittedly a fault-ridden process), and get published in journals and books, accessible to anyone interested and open to sustained critical scrutiny. Thus, in a first approximation, an account of ideas produced by an academic discipline is a review of the published literature in that field. The references to the chapters in this treatise provide a comprehensive introduction to the important contributions to modern population thought and its grappling with the four questions noted above.

The limitations of such an introduction, however well intended, should be, nevertheless, evident. The first is the limitation of the present reviewer, but in most instances also of the other contributors, in terms of access to languages and in geographic outreach. The references cited here, and in most other contributions to the enterprise, are in English and French only. This is a bias only partially excusable by the fact that by sheer numerical bulk and by ease of access to demographers, irrespective of their geographic location, the terrain covered constitutes a disproportionately large chunk of the population literature.

The second bias is one of selectivity, that reflects, once again, limitations of the reviewer's, and also of other commentators', outlook and judgment, but is also imposed by the sheer size of the literature that carries demographers' ideas. One bibliographic research tool familiar to demographers, *Population Index*, now (unfortunately) defunct, even if professing selective coverage, carried many hundreds of new citations every year, and has done so during the entire period of the six decades under consideration here. The *Review of Population Reviews*, published by the Paris-based organization CICRED lists the contents of some 80 journals in the field of demography. *POPIN*, the population information service maintained by the United Nations Population Fund, provides extensive databases on population publications. It also offers a *Worldwide Directory of Population Institutions* which lists, for example, more than 100 population research and training institutions both in China and in India.

Clearly, the short overview offered below cannot claim to be representative. The aim, instead, is to illustrate main strands of population thinking that appear to have been especially important in population policy debates during the last six decades or during parts of that period. As has long been the case in the natural sciences, contemporary social science work tends largely to ignore publications that are not of recent vintage. The overview below seeks to avoid that practice by paying attention to significant work during the period covered regardless of date of publication.

Where should the reader look for enlightenment about the flow of ideas in the field of demography that is far more detailed than what follows below and that also permits forming a judgment uninfluenced by a guide? Those with ample time would be well advised to settle in the reading room in a well-equipped library and to leaf through the proceedings volumes of IUSSP's General Conferences. These are held every four years (and special conferences sometimes in-between) and the raw (unedited) papers presented at each meeting fill either a single massive tome, or more often two or three, and on occasion up to six, hefty volumes, and, more recently, a CD-Rom, which is deceptively slim but offers equally rich fare. Their content is obviously too sprawling to be sampled and characterized here, but their inspection and selective perusal would well reward the critical reader (who may also imagine hopping from conference to conference in places like Rome, Vienna, Stockholm, New York, Ottawa, Belgrade, London, Mexico City, Liège, Helsinki, Manila, Florence, Delhi, Montreal, Beijing, Salvador de Bahia, or Tours) The papers in most cases are volunteered by IUSSP's members, subject only to the restriction that they fit a fairly accommodating overall topical frame. Thus, the content of these volumes gives a good representative picture of what demographers, or the institutions that support their research, have been interested in, hence considered important, during the post–World War II decades.

A similarly simple recipe for an exploratory excursion in pursuit of modern ideas on population would be perusal of past volumes of specialized journals. The reader would have the advantage of being exposed only to material that had been filtered through a competitive process for limited space and had been subjected to prior editorial screening and revision, albeit of varying severity. Those with ample yet not unlimited time at their disposal could apply themselves to a subset of what is available in this genre and go through only some of the journals, such as (a) INED's *Population* (completing in 2005 its 60th annual volume), preferably with a side excursion to INED's ambitious monograph series appearing under the general title *Travaux et documents* (and, more recently, titled *Les Cahiers de l'INED*: some 150 volumes since the inception of the series); (b) *Population Studies* (in its 59th year); (c) *Demography*, the journal of the Population Association of America (in its 42nd year); (d) *Population and Development Review* (in its 31st year) and its eleven book-length supplements on special topics; or (e) *Genus*, the population journal with the greatest longevity. Reading the book review sections of the journals just cited (unfortunately, not all of them are engaged in providing this service so essential for the good health of a discipline) also would be a rewarding guide into the population literature that takes the form of books and monographs.

Turning to books that provide a comprehensive overview of the field and experienced guidance combined with original scholarship can be an efficient way of avoiding the hazards of a random walk across the periodical literature. Eminent examples of this genre include Alfred Sauvy's *Théorie générale de la population* (Sauvy, 1956, 1969), and classic French and American population textbooks in population, such as Adolphe Landry (1949), Philip M. Hauser and Otis Dudley Duncan (1959), Donal Joseph Bogue (1969), and William Petersen (1975). A less ambitious book is a "reader," offering selections of special merit from the periodical literature, with sometimes extensive, sometimes perfunctory commentary that imposes some coherence on the works presented. The main merit of such books is that they provide easy access to items, typically mostly journal articles, that readers otherwise may find difficult to locate—a problem that is now being solved by access to specialized literature through the Internet, such as through the indispensable journal storage *JStore*. Some recent examples include, each with a focus or special emphasis on economic aspects of population, Nancy Folbre (1996), Paul Demeny and Geoffrey McNicoll (1998), T. Paul Schultz (1998), and Julian L. Simon (1997).

As a rigorously edited early guide to key issues that occupied demographers in the past, the UN-sponsored study *The Determinants and Consequences of Population Trends* (United Nations, 1953) is still indispensable. The follow-up volumes to that project (United Nations, 1973) are also valuable, as are the volumes commissioned as back-up materials for the UN's first large intergovernmental conference on population, held in Bucharest in 1974 (United Nations, 1975). In general, the United Nations, primarily through work in its Population Division, its Statistical Office, and through its specialized agencies (more recently especially the UN Population Fund), is a steady producer of population reports, analyses, and statistics that find wide use in the field of demography. The United Nations, often through its regional economic and social commissions, also is a frequent convener of specialized population meetings, typically resulting in valuable technical publications with contents either produced within the organization itself or contributed by invited academic researchers.

In contrast to the United Nations' work, with its resolutely internationalist outlook, country-specific examinations of population issues sponsored or initiated by national governments reflect the special interests of domestic constituencies. Reports of this type, often published annually—such as part of, or a supplement to, compendia of up-to-date national demographic statistics—have long been a steady presence in the population literature in most industrialized countries.

More recently, their equivalents on a regional level have also increasingly made an appearance, as organizations with access to strong professional staff resources, such as the Organization for Economic Cooperation and Development (OECD), the Council of Europe, or the European Commission, publish general surveys of the population of their member countries or report on special population–related issues of their region. Despite their national or regional frame of reference, the best examples of these products should be of substantial interest to students of population everywhere.

The same remains true, despite a degree of inevitable obsolescence, of two especially ambitious national programs aimed at examining population issues at a particular point in time. The first of these, the work of Britain's Royal Commission on Population, which stretched over a period of four years, was already mentioned above (Great Britain, 1949). Its papers, rich also in technical–analytic material, constituted six volumes. The other, carried out in the United States, was the Commission on Population Growth and the American Future (1972). Its report was based on deliberation that took nearly as much time as did its British predecessor's, and was backed up by eight large volumes of research papers, covering demographic, social, economic, and spatial aspects of population growth, and discussing the relationship of demographic growth to such issues as resources, environment, and governance. The considerable scientific merits of the work of the two commissions notwithstanding, both projects turned out to be somewhat ill-timed, hence less influential on policy matters than could have been expected. The perceived problems that prompted these enterprises were polar opposites: in the first instance it was low fertility and the prospect of population decline; in the second, the problem of sustained high fertility and the resulting rapid population growth and the prospect of an increasingly negative influence of these phenomena on the quality of life. Both expectations proved to be misguided. As a member of the Royal Commission is said to have observed (Grebenik, 1997), the Commission was established when there was lots of money and few babies, but issued its report when there were lots of babies but little money. The American effort encountered similar interference from hard-to-predict demographic trends. When the enterprise was conceived, the U.S. baby boom—which at its peak brought total period fertility rates above 3.5—was still a reality. By the time the Commission's report was published, the baby boom was history and the total fertility rate was below replacement.

Ideas about population include ideas by practitioners in the field of population studies about how well, or how unsatisfactorily, the discipline is performing its main functions: producing and disseminating important and sound research findings. Such stock-takings, assessments, criticisms, and forecasts of future developments by demographers are often influential in protecting the field from unhealthy tendencies, provide guideposts for new directions, or at least enhance demographers' self-awareness about problematic aspects of their work. Even more optimistically, they prompt academic decison makers and administrators in allocating funds for population research more wisely and with deference to unpopular ideas.

The genre typically takes the form of an essay; sober analysis is often combined with expressions of controversial personal views and occasionally filled with polemics. Some examples are Dennis Hodgson (1983, 1988), who, using the example of the theory of demographic transition, documents a shift in demographic research from a social–scientific to a policy-oriented stance; Paul Demeny (1988), who assesses and criticizes the institutional arrangements that tend to relegate policy-oriented research to performing ancillary functions to existing population programs in the Third World; Samuel H. Preston (1989), who identifies deficiencies and conflicts in the diagnosis of population problems; Dirk J. van de Kaa (1991) who discusses emerging issues for demographic research agendas in Europe and the ambiguities of the role of the demographer as a scientist and as an actor concerned with socially useful results; McNicoll (1992), who complains that increasing sophistication in the analysis of population processes has been accompanied by a lessening of interest by demographers in the larger, related questions of social and behavioral change; Simon Szreter (1993), who argues that further intellectual progress in fertility studies, a key area of demographic research, requires emancipation from the idea of transition and its associated obsolescent methodology; and Greenhalgh (1996), who finds that demography's sophisticated methods and noted achievements are undercut by outmoded thinking rooted in modernization theory and analyzes the tension between demography's use for applied purposes and its claims to being a science.

Round-numbered institutional anniversaries occasioned much self-examination in a similarly critical vein. *Demography*'s 30th birthday was marked by several such essays, notably by Keyfitz (1993), who examines the influence of population journals and comments on the tension between researchers wishing to focus on conclusions about substantive matters and those pursuing technical questions of methodology; and by Preston (1993), who discusses the scope of demography and the various research approaches that can legitimately claim that label, and identifies promising research areas in which demographers can be expected to be called on to address issues of national and international concern. From INED's 50th anniversary celebration sprang a volume (Chasteland and Roussel, 1997) containing some 20 highly informative articles assessing national experiences in demographic research and discussing the future of demography. Particularly notable among these are two contributions by the editors of the volume, and essays by Ronald Lee, Samuel Preston, Alain Blum, and Graziella Caselli. Marking the first half century of the publication of *Population Studies*, the closing issue of volume 50 of that journal published an equally important set of articles examining the achievements, status, and prospects of various subfields of demography. The lead essay, by John Caldwell (1996), questions the extent to which demography is a social science (and gives a less than conclusive answer) and examines how changes in the outside world affect what demographers do and what and where they publish.

A recent comprehensive appraisal of the field of population studies that surveys contemporary thought on demographic processes in their multifarious connections to social phenomena at large is offered by *The Encyclopedia of Population* (Demeny and McNicoll, 2003). Population issues are often defined fairly narrowly, focused on description of the core demographic processes or, in a policy context that was dominant in the second part of the 20th century, on rapid population growth and measures to curtail it.

An aim of the *Encyclopedia* is to make a compelling demonstration of the need for a more universal interpretation of the subject matter. Even a partial listing of the topics covered by the nearly 300 contributors of this 2-volume survey suggests the breadth of contemporary population thought. The topics include the entrenchment of very low fertility and the growing problems of old-age support; the retreat from marriage and the diversification of family forms; the new medical technologies affecting reproduction and longevity; the resurgence of a variety of infectious diseases; increased South-North migration and refugee movements; the press for women's equality and fuller reproductive rights; a widened array of environmental effects of demographic dynamics, notably climate change; and global shifts in the relative population size of countries.

The *Encyclopedia* also gives attention to the evolutionary bases of human development and behavior, with insights drawn from genomic research and from comparisons with other species. Many of these topics have ethical debates associated with them—long-standing, like abortion and asylum-seeking, or newfound, like genetic engineering and animal rights—debates that are unlikely to be satisfactorily advanced without expert input from demographers. And the traditional interest in population history and prehistory also brought new findings and interpretations during recent decades, settling some controversies and raising others. All this territory is staked out in the *Encyclopedia*.

In a somewhat different genre, the present series, *Demography: Analysis and Synthesis*, also offers, as was noted above, a broad overview of the population field. There, the coverage is less encompassing, but with the advantage that the narrower focus permits greater in-depth treatment of the various topics. These two recent works demonstrate that any attempt in a single chapter that would aspire to offer a comprehensive survey of population covering the last 60-odd years could necessarily excel only in superficiality. To avoid that pitfall, my assignment must be interpreted as one calling for drastic selectivity. The ambition of the present chapter is not to duplicate specialized discussions of population thought but to focus selectively on some major issues that the in-depth but piecemeal approach may have missed. The prime opportunity for that is offered by examining the bearing of population dynamics on public policy as perceived by those studying demographic issues during the last 65 years. In doing so, I seek to distill some of the major themes in population thought from the literature cited above.

II. THE GLOBAL POPULATION PROBLEM

The study of demographic dynamics as a scholarly pursuit has its own considerable intellectual and aesthetic rewards to both researcher and consumer of demographic research. But such study acquires social importance only to the extent that particular patterns of population change are proven to affect, or are perceived to affect, social welfare differently: some patterns are, or are thought to be, better than others. Such differences raise the possibility of deliberate, purposeful modification of the demographic status quo in a mutually agreed direction that enhances the common welfare. If the cost of such an engineered change is less than the improvements obtained, ameliorative collective action—crafting beneficial population policies—becomes a potential possibility. Population thought in the second half of the 20th century has been preoccupied by exploration of the opportunities modified demographic trends might offer for enhancing human welfare.

Population policies are often discussed with reference to particular countries. This is the case also in the present treatise. Although the choice is in many ways natural—states are the locus of many important political decisions affecting demographic behavior—a country-by-country study may miss some important policy issues and leave their common underlying logic unexplored.

In the early 1940s, students of population anticipated the emergence of two kinds of population problems in the post–World War II years. One was the continuation of below-replacement fertility observed in almost all countries of the Western world during the 1930s, implying negative rates of population growth in the coming future. To the extent that this was seen a "problem," it clearly required responses to be determined by the affected countries themselves.

A second type of problem was seen as a global or at least a broad regional one: emerging as a result of exceptionally rapid rates of population growth in countries where fertility was still traditional (i.e., "high" in comparison to Western levels) but mortality was declining. Economists as well as demographers expressed doubt about the ability of many low-income countries to cope with a rapid population surge. Some concluded that the economic and social consequences of mortality improvements will not be sustained or will even be reversed unless rapid fertility decline also ensues.

By the late 1940s, the first type of the anticipated postwar problem seemed to evaporate quickly as birthrates rose in most low-fertility countries. Although the magnitude of this "baby boom" varied from country to country, vigorous pronatalist policies no longer seemed justified. Indeed, at the height of the baby boom, in some instances, notably in the United States where period total fertility rate reached 3.7 in 1957, serious consideration was given to policies that would actively seek moderation of population growth.

Predictions about accelerating population growth in the less developed world were, however, on a firmer footing. What fueled this growth was an often-precipitous drop in the level of mortality—a welcome development. But a doubling or even tripling of population size by the end of the 20th century was likely to set back already low living standards or at least greatly retard their improvement. To a significant extent, mortality decline was sparked and sustained through the adoption of new methods of controlling infectious disease, greatly aided by international aid programs. This suggested that international programs could and should also be designed to bring birthrates down faster than could be expected in the absence of deliberate intervention. An acceleration in demographic transition, it was argued, would speed up economic development.

This was not a universally shared view. In an influential article published in 1949, Alfred Sauvy argued that the global population problem is a false problem; indeed the notion of world population is largely devoid of meaning. The politically relevant units are the sovereign countries; only countries can assess their own problems and find their own appropriate solutions.

Yet many of the newly independent developing countries of Asia and Africa were not well prepared to pay sufficient attention to potentially deleterious demographic changes or to mount effective programs that could modify these changes for the better. An elaborate set of economic–demographic projections prepared by researchers at INED (Balandier, 1956) showed the magnitude of the economic difficulties that accommodation of rapid population growth would involve, and the potential alleviation of that burden if fertility decline were to follow more closely the falling rates of mortality. For example, the INED projections showed that the population of India and Pakistan (within the borders prior to the independence of Bangladesh) with constant fertility would grow from 460 million in 1955 to 1.3 billion by the year 2000. Early and rapid fertility decline would result in a population of 866 million in the year 2000: still a rapid growth by any historical standard, but presumably more conducive to economic development. (The actual 2000 population of these countries is estimated by the United Nations as 1.288 billion: a result that reflects both a lower mortality trajectory than was assumed in the INED projection and an appreciable decline of fertility in the last two decades of the 20th century.)

Concerns with the harmful effects of rapid population growth manifest in academic studies and eventually also in developed countries' public opinion and in official policy formulations by economic development agencies, national and international, were not only humanitarian. Economic instability and impoverishment in the less developed countries were deemed as having potentially ominous consequences in the context of Cold War rivalries. Quite apart from the latter consideration, sharpening income inequalities among countries were seen as a source of international instability. And, in the long term, the growing imbalance between population sizes in the low-growth regions as compared to regions experiencing rapid demographic growth had unwelcome geopolitical implications.

The search for policy instruments suitable for accelerating the fertility transition in less developed countries led to extended and sometimes acrimonious debates. Much of this discussion among countries, with the conspicuous exception of China, which found its own unique solution for the population problem, was conducted in academic forums in the developed world, primarily in America. The developed world of course had relatively low fertility and much historical experience on why birthrates fell. Social science analysis was virtually unanimous in interpreting this experience. The explanation centered on the role of changing structural conditions of the economy, conditions to which microlevel units of the population readily responded, in demographic as well as in other matters. Demand for smaller families was seen as the primary force determining birthrates; the means by which couples regulated their fertility was not unimportant, but was seen as a distinctly secondary factor. If the demand were strong enough, fertility would be low, even if birth control technology were primitive. A transition to low fertility presupposed changing preferences, and such preferences were responses to market signals. If policy were to have an explicit role, it would be through reinforcing those signals through appropriate institutional arrangements. The specifics of the needed socioeconomic arrangements were known and well elaborated. They were development oriented in a broad sense: the demographic side effects were essentially by-products of the overall design.

But assembling the policy instruments so identified into a coherent development strategy of institutional-structural reforms, reforms justified also by demographic objectives, remained a task inconsistent with the spirit of post–World War II development philosophy. In promoting development, governments came to see their roles not in supporting institutions harnessing the market, but in replacing the market in key developmental tasks through specific goal-oriented categorical programs. In the matter of population policy, the rapid postwar progress in the technology of birth control provided an appealing, apparent shortcut for achieving fertility decline through programmatic means. Markets, it was held, could not be relied on to bring that technology to those wishing to practice birth control. Governments could, instead, organize free delivery of birth control information and provide effective means for preventing births to all those who wished to plan their families. Surveys indicated that there existed a substantial latent demand for such services. Satisfied customers, in turn, would serve as role models, bringing new clients to the program.

The effectiveness of family planning programs in reducing fertility remains a matter of controversy. According to international guidelines, programs recruit their clients on a strictly voluntary basis. By accepting the service voluntarily, clients demonstrate that they value that service. But some of the more successful programs, notably in Asia, tended to increase acceptance by often heavy-handed methods of persuasion, and, in the especially important case of China, by coercion backed by legal sanctions. Where fertility fell in less developed countries with active family planning programs, it is typically found that program-provided free services account for a large percentage of those practicing contraception. This result is then often taken as an indicator of success in reducing aggregate fertility. But what would have happened in the absence of the program is conjectural, hence routinely ignored. Some less developed countries that lacked government programs also experienced major falls of fertility: Brazil is a conspicuous example. Similarly, if programs have seemingly only minor success in reducing fertility, this can be taken as evidence that the program is inadequately financed, organized, and managed: greater efforts would have led to better results.

Family planning programs as they were commonly conceived had a strong resemblance to health programs. But given the special priority accorded to family planning services in foreign assistance, typically they were organized as a separate "vertical" program, or kept administratively distinct within the broad health program. The justification for such treatment was that while clients of family planning services are recruited because the program satisfies their individual needs, the program also serves a national developmental need by helping to reduce aggregate population growth and hence deserves priority. Once a family planning program is organized, its managers and professionals form a natural advocacy group strongly interested in the program's sustenance. Invoking the public interest in lowered fertility, as distinct from simply serving the needs of the program's clients, has long served as a key supporting argument in that endeavor.

Over time, this developmental prop has eroded. This was in part a result of criticisms of the intrinsic scientific merit of the argument but most of all a reflection of the extensive decline of fertility that has occurred, a decline often attributed to the success of the family programs themselves.

Accordingly, the development rationale of family planning programs was gradually dropped and was replaced by the argument that the programs satisfy important health needs and help people exercise a fundamental human right. The Cairo conference formalized this shift: even though the name of the conference for the first time included a reference to development, scant attention was paid to that concept. Family planning programs were redefined, instead, as reproductive health programs, responding to a broader range of women's health needs, such as prevention of unsafe abortions and sexually transmitted diseases, including HIV/AIDS. But beyond this, new emphasis was put on some requirements that would contribute to women's empowerment: reduction of infant and maternal mortality and improvement in girls' education and women's opportunities for employment and political participation. Although the connection was not highlighted, these are conditions that are likely to help reduce the birthrate through stimulating the demand for smaller family size. The Cairo conference, in effect, reverted to some key elements of earlier population thought, advocating a demand strategy for reducing birthrates.

The future of family planning service programs—the centerpiece of the dominant population thought since the late 1960s—is thus left in a somewhat tenuous state. Without invoking a collective interest in a wider practice of birth control, it is not clear what level of priority should be accorded to such programs as just one part of publicly financed health programs, or indeed relative to any other social welfare programs that also serve demonstrable human needs. Not surprisingly, there are increasing efforts in national programs to rely on the market to facilitate access to contraceptives and to provide program services on a fee-for-service basis.

III. THE PROBLEM OF BELOW-REPLACEMENT FERTILITY

During the second half of the 20th century, debates about both population policy and consequent programmatic action were centered on the issue of rapid population growth in the less developed world. Toward the end of this period, however, a quite different demographic phenomenon began to attract increasing attention: aggregate fertility levels that are inadequate for the long-run maintenance of the population. Analytically, the potential population policy issue raised by low fertility is identical to the problem inherent in rapid population growth: it is caused by the disjunction between the sum total of individual reproductive decisions and the collective interest in a long-run demographic equilibrium. But this time individual aspirations generate a deficit rather than an excess in population growth. The syndrome, is not entirely novel: it was detectable in fertility trends in the West, especially in Europe, in the 1920s and 1930s, and in some instances, notably in France, even earlier. But in the decades immediately following World War II, the baby boom seemed to make the issue of low fertility moot. Indeed, by any historical standard, population growth was rapid during the second half of the 20th century, even in the developed world. Europe's population, for example, grew during that period from 550 million to about 730 million.

The baby boom was, however, a temporary interruption of the secular downward trend in fertility. By the 1970s, the net reproduction rate was at or below unity in most countries in Europe and also in the United States. In the U.S., fertility stabilized at or very close to that rate, but in Europe, fertility continued to decline. By the beginning of the 21st century, the average total fertility rate was 1.4. Such a level, if maintained indefinitely, would result in a population loss of one-third from generation to generation, that is, roughly, over each period of some 30 years. In some countries, notably in Southern, Central, and Eastern Europe, period fertility rates were at low levels without historical precedent for large populations. If continued, in the absence of large compensatory immigration this would not only lead to rapid population decline but also result in very high proportions of the population at old ages. It might be expected that in the affected countries such prospects would generate not only concern, but also vigorous remedial policy action.

By and large, however, this response has not been evident. Most governments, as well as the general public, tend to view below-replacement fertility with an equanimity quite unlike the alarmed reaction that the same phenomenon elicited when it first emerged between the two World Wars. And explicit pronatalist policies, common in the 1930s, are conspicuous by their absence. There are a variety of reasons explaining this indifference.

First, the preeminent population issue confronting policymakers in the post–World War II period was rapid global population growth. Programs aimed at moderating fertility in the developing world received assistance or at least encouragement from the rich, low-fertility countries. Although the rationale was modified over time, such assistance and encouragement has continued, as indeed substantial further population increase in the less developed countries is still anticipated in the early decades of the 21st century. Even though population issues tend to be sui generis, reflecting differences in demographic behavior country-by-country, there was, and remains, a perceived dissonance between fertility-lowering assistance to other countries and engaging in action at home serving the opposite aims. Faulty logic notwithstanding, the international terrain has not been favorable for domestic pronatalism.Second, the natural rate of increase—the difference between the number of births and the number of deaths—is still positive in many of the countries with fertility well below replacement. This is the result of age distributions that reflect past fertility and mortality, and notably the effects of the postwar baby boom, that still favor population growth. While this momentum effect is temporary, the long-term implications for population decline and population aging are only dimly perceived by the general public and provide an excuse for inaction on the part of policymakers.

Third, when those long-term demographic effects are understood, a calmer attitude still prevails. There is an inclination, reinforced by increasing concern with the quality of the natural environment, to regard a degree of demographic "decompression" as a not necessarily unwelcome prospect, especially in countries with already dense populations. And it is assumed that the economic and social disadvantages that might be imposed by a declining population can be effectively dealt with through institutional adjustments and social policy measures other than measures aiming for a higher birthrate. A demographic policy often regarded as potentially helpful in this regard is encouragement of immigration. That willing immigrants are available to compensate for low birthrates is taken for granted—a realistic assumption in high-income countries.

Fourth, there is a vague expectation that the population decline, impending or already begun, will in due course trigger corrective homeostatic mechanisms, leading to a spontaneous rebound in the level of fertility. Another baby boom may not be in the offing, but fertility may rise sufficiently to once again reach, or at least approximate, replacement level. Governments, it is assumed, would be ill-advised to interfere with this natural process by trying to increase birthrates and then seek to fine-tune them at the desirable steady-state level. According to this view, a laissez-faire fertility policy is justified since, apart from broad upper and lower limits, governments are not competent to determine what constitutes an optimal fertility rate, or growth rate, or population size in any given year, decade, or even longer time interval.

Finally, even if the will were there, there is a paucity of effective pronatalist policy instruments. Exhortation from governments is not promising, and in any case unlikely to be tried in a democratic polity. Restrictions imposed on access to modern contraceptive technology are not politically acceptable and would be certain to fail. This leaves the traditional levers of social policy: dispensing material incentives and disincentives so as to increase the willingness of couples to have children. Such incentives can be engineered by the government through fiscal measures, such as differential taxation, provision of services in kind, or both. This approach was tried in the interwar years, but with limited success. After World War II many similar measures continued to be applied; in fact, with the steadily expanding welfare state, they were often upgraded and their scope, too, was extended. However, they were no longer considered "pronatalist" but were absorbed within the more encompassing frames of family and general welfare policy. The new label partly reflected a political–ideological preference, but in part also the fact that some distinctive features of pronatalism—such as differential rewards that favored large families, and non-means-tested or even regressive allocation of family and child benefits—were generally no longer acceptable.

Although the redistributive policies of the contemporary welfare state are biased in favor of the elderly and the poor, government-organized transfers to parents of children, or to children directly (such as through publicly financed day-care services and free or subsidized education often beyond the secondary school level, which lessens the cost of children to parents), are substantial in all low-fertility countries.

Indeed, it is typically assumed that existing family and welfare policies sustain fertility above a level that would ensue in their absence. Accordingly, making these policies more generous—socializing an even larger share of child costs—is often seen as a means toward increasing fertility, whether as an outright policy objective or, more in the prevailing spirit of the time, as an unintended but welcome by-product. Such extension, however, is difficult, given the fiscal constraints of already overcommitted welfare states. And more to the point, the net effect of family-friendly redistribution of incomes and provision of services is uncertain. It is notable that in the United States, where such schemes are distinctly less well funded than, for example, in Western Europe, fertility is, nevertheless, relatively high.

In recent decades, in modern industrial economies, the participation of women in the formal labor force has expanded rapidly. This tendency, reflecting market forces, but also encouraged by government policy (partly as an antidote for deteriorating dependency ratios as the population becomes older), is likely to continue. Among the factors explaining the low level of fertility despite general material affluence, many observers point to the double burden on women of both raising children and working outside the home. To the extent that higher birthrates are seen to be socially desirable, the derived policy prescription is to adopt measures that make motherhood and women's labor force participation more compatible. The higher fertility in countries (notably in Scandinavia) where such measures are strongly applied, compared to countries (especially those in Southern Europe) where they are largely absent, suggests that enhanced compatibility (through day-care services, flexible work hours, liberal sick-leave allowances, and the like) is an effective pronatalist policy even if motivated by other considerations. But it is far from clear whether the fertility differential so generated is high enough to bring the total fertility rate back to replacement level. Steady labor force participation of women during the child-bearing years can certainly be made compatible with having one child or even two. It is likely to be far less compatible with sustaining, or even increasing, the proportion of women who have more than two children. Many career-oriented women voluntarily remain childless; many others prefer a single child. It follows that, to achieve average replacement-level fertility, the proportions of such women need to be counterbalanced by high enough proportions of women who have chosen to have three, four, or more children. There is little indication at present that policies directed at enhanced compatibility achieve that result.

When fertility is high, as it still is in most developing countries, it is a safe prediction that with economic development it will eventually decline, at least to replacement level. In the long term, apart from out-migration, the only alternative is higher mortality. The record of the high-income countries indicates, however, that replacement-level fertility is not a necessary resting point. Once fertility is lower than that, predictions become highly hazardous. European and East Asian experience suggests that fertility has a tendency to settle below an average of two children per woman, hence tending toward sustained population decline. The question, to which no good answers exist at the dawn of the 21st century, is "how far below?" "Scandinavian"-style family policies may stabilize fertility only modestly below replacement—around a total fertility rate of 1.8. That would imply a fairly moderate relative shortfall of births compared to deaths, and population stability in rich countries with such vital rates could be fully or almost fully compensated with a modest level of controlled immigration. Population aging would be then kept within relatively narrow limits, which postindustrial economies could readily adjust to. The demographic weight of such countries within the global total in the foreseeable future would continue to shrink, raising possible problems of a shifting geopolitical balance. Still, such demographic configurations would be likely to push the day of demographic reckoning beyond the policy horizons that governments feel an obligation to be actively concerned with. Pronatalist interventions would find at best a marginal place on governments' policy agendas.

On the other hand, fertility levels in the lowest-fertility countries—countries with a total fertility rate of 1.3 or below around the turn of the 20th century—might stabilize at that level or even shrink further, reflecting the decentralized and uncoordinated decisions of individuals and individual couples. Such an outcome might also foreshadow future reproductive behavior in countries in which fertility is still fairly close to replacement level. This would create a qualitatively different demographic situation for which there are few precedents in modern history. It would represent a clear threat to the continuing viability of the countries affected. Compensatory immigration flows would have to be so large as to be inconsistent with any reasonable degree of cultural and ethnic continuity. Alternatively, population aging in the absence of immigration would create virtually unsolvable challenges, and likely there would be a drastic loss of relative geopolitical status. Spontaneous homeostatic mechanisms may not come into play to save the day, or may do so too sluggishly to matter. A radical rethinking of fertility policy would then become a necessity for social—and national—survival.

References

Bogue Donald Joseph, 1969. *Principles of demography.* New York, Wiley, 917 p.

Balandier Georges (ed.), 1956. *Le "Tiers-Monde." Sous-développement et développement.* Paris, INED and PUF, 392 p. (Travaux et documents, Cahier no. 27).

Caldwell John C., 1996. Demography and social sciences, *Population Studies*, vol. 50, no. 2, p. 305–333.

Chasteland Jean-Claude and Roussel Louis, 1997. *Les contours de la démographie au seuil du XXIe siècle.* Paris, INED and PUF, 434 p.

Commission on Population Growth and the American Future 1972. *Population and the American future : the report of the Commission on Population Growth and the American Future.* Washington DC, U.S. Government Printing Office, 182 p.

Demeny Paul, 1988. Social science and population policy, *Population and Development Review*, vol. 14(3), p. 451–479.

Demeny Paul and McNicoll Geoffrey (eds.), 1998. *The Earthscan reader in population and development.* London, Earthscan Publications Ltd., 363 p.

Demeny Paul and McNicoll Geoffrey (eds.), 1998. *Encyclopedia of population*, 2 vol. New York, Macmillan Reference, 1040 p.

Folbre Nancy, 1996. *The economics of the family.* Cheltenham (United Kingdom) and Brookfield (United States), E. Edgar, 687 p.

Great Britain, 1949. *Royal Commission on Population: Report*, London, His Majesty's Stationery Office, 259 p.

Greenhalgh Susan, 1996. The social construction of population science: an intellectual, institutional, and political history of twentieth-century demography. *Comparative Studies in Society and History*, vol. 38, no. 1, p. 26–66.

Grebenik Eugene, 1997. On Population Studies: its origins and outlook, *Population and Development Review*, vol. 23(3), p. 575–577.

Hauser Philip M. and Duncan Otis Dudley (eds.), 1959. *The study of population: an inventory and appraisal.* Chicago, University of Chicago Press, 864 p.

Hodgson Dennis, 1983. Demography as social science and policy science, *Population and Development Review*, vol. 9(1), p. 1–34.

Hodgson Dennis, 1988. Orthodoxy and revisionism in American demography, *Population and Development Review*, vol. 14(4), p. 541–570.

Keyfitz Nathan, 1993. Thirty years of demography and *Demography*, *Demography*, vol. 30(4), p. 533–549.

Landry Adolphe, 1949. *Traité de démographie*, 2nd ed. Paris, Payot, 658 p.

McNicoll Geoffrey, 1992. The agenda of population studies: a commentary and complaint, *Population and Development Review*, vol. 18(3), p. 399–420.

Petersen William, 1975. *Population*, 3rd ed. New York, Macmillan, 784 p.

Preston Samuel H., 1989. The social sciences and the population problem, *in*: J. Mayone Stycos (ed.), *Demography as an interdiscipline*, p. 1–26. New Brunswick, Transaction Publishers, 214 p.

Preston Samuel H., 1993. The contours of demography: estimates and projections, *Demography*, vol. 30(4), p. 593–606.

Sauvy Alfred, 1949. "Le faux problème" de la population mondiale, *Population*, vol. 4(3). (English translation in *Population and Development Review*, vol. 16(4), p. 759–774.)

Sauvy Alfred, 1956. *Théorie génerale de la population*, 2 vol. Paris, Presses Universitaires de France.

Sauvy Alfred, 1969. *General theory of population.* New York, Basic Books, 550 p.

Schultz T. Paul (ed.), 1998. *Economic demography*, 2 vol. Northampton, MA, Edward Elgar, 562 and 621 p.

Simon Julian L., 1997. *The economics of population: key modern writings*, 2 vol. Cheltenham, UK, Edward Elgar, 533 and 555 p.

Szreter Simon, 1993. The idea of demographic transition and the study of fertility change: a critical intellectual history, *Population and Development Review*, vol. 19(4), p. 659–701.

Trygve Lie, 1995. Trygve Lie on Population in the United Nations' Agenda, *Population and Development Review*, vol. 21(4), p. 867–873.

United Nations, 1953. *The determinants and consequences of population trends*, (Population Studies, No. 17). New York, United Nations, 404 p.

United Nations, 1973. *The determinants and consequences of population trends: new summary of findings on interaction of demographic, economic and social factors*, 2 vol. (Population Studies, No. 50). New York, United Nations, 622 + 152 p.

United Nations, 1975. *The population debate: dimensions and perspectives*, 2 vols., New York, United Nations, 676 and 726 p.

Van de Kaa D. J., 1991. Emerging issues in demographic research for contemporary Europe, *in*: Michael Murphy and John Hobcraft (eds.), *Population Research in Britain*, Supplement to vol. 45 of *Population Studies*. London, Population Investigation Committee, p. 189–206.

CHAPTER

99

Ethics and Demography or Macrodemus and Microdemus in the Country of Ethicists

FRANÇOIS HÉRAN
Institut national d'études démographiques (INED), Paris, France

Macrodemus and Microdemus were long-standing members of the same research center. They were both demographers, but it was hard to believe they were both in the same profession. Macrodemus practiced macrodemography, Microdemus microdemography.[1]

Macrodemus spent most of his time processing the data series on births, marriages, and deaths from around the world, which he received via the Internet from national statistical offices, official health organizations, and the United Nations Population Division. He was an expert in formal demography, as his English-speaking colleagues put it, and also a fan of demographic projections. Reputed for his rigorous analyses, he had learned demography from Nicodemus and Mathodemus, and was eternally grateful to his masters. Macrodemus had never observed or interviewed anyone in his life. He saw the world from high above and far away. He believed that one could learn about human subjects all the more easily by studying them in the aggregate and at a distance.

Microdemus was 10 years younger than Macrodemus. Trained by Epidemus and Geodemus, he liked to explore the field on foot, observe behaviors at length, and converse with men and women of all ages and ranks. He was fond of stating that a researcher's first task was, using the 18th-century phrase, to "know society." This meant possessing not only the art of conversation but also the knack of getting on with all social groups, the ability to discover the exotic in everyday life, to decipher the words of all tribes, to construct and deconstruct institutions, and to reveal the hidden side of things. He felt a special attraction to fringes and margins, where, in his view, the fate of society was determined. While not disdaining quantitative analysis, he was convinced that qualitative analysis was the key to human behavior.

Macrodemus and Microdemus were not the Center's best-known researchers but, to use the language of sociologists, they embodied the two "ideal types" at the opposite poles of its activity. Other researchers filled the space in between.[2] Some covered

[1] These demographers should be of either sex. In the French text, I have given them masculine-sounding names and have used male gender agreements, in the hope that female readers will be happy not to be identified with such overly typical male characters. Apart from that, I borrow the following guideline from the Quebecois textbook by André Duhamel and Noureddine Mouelhi (2001, p. vi): "The masculine form is used here to refer to both sexes, without discriminating between men and women, and for the sole purpose of streamlining the text."

[2] They often came in pairs: Epidemus and Biodemus, Ecodemus and Geodemus, Agrodemus and Cosmodemus, Paleodemus and Neodemus, Calodemus and Mesodemus, Erodemus and Gynodemus, Iatrodemus and Necrodemus, Oligodemus and Polemodemus, etc.

a very broad spectrum by switching effortlessly from macrodemography to microdemography and vice versa: they could analyze data covering the entire planet in the morning and follow an interviewer testing long questionnaires on fertility practices in a distant suburb in the afternoon. But there were few of these hybrid, semi-micro/semi-macro people around. Most often, the distribution of labor at the Research Center was polarized between the two models. The two worlds coexisted without really mingling. Macrodemus and Microdemus were, so to speak, caricatures, but they merely embodied extreme forms of widespread practices.

The inclination of the Center's researchers toward macro- or microdemography was not determined by their original disciplines. None had been initially trained in demography. All had discovered it while working in another discipline: sociology, economics, history, political science, geography, statistics, epidemiology, biology, or more seldom, social psychology and anthropology. But some had hastened to forget their initial training in order to enter into a jealous marriage with demography, while others were repeatedly unfaithful to it by cultivating a strong interest in social science or the life sciences: they thought it natural to put demographic analysis in a broader framework.

This divergence in practices and cultures had a deep structural influence on the Center. Both groups invariably focused their research on human subjects, but their relationships to the subjects were radically different. The approach to ethical issues was substantially altered in consequence—so much so that one might ask whether demography had a specific ethic or whether, instead, each type of demography had a separate ethic.

The philosophical tale told in this chapter seeks to answer the issue I have just defined. It shows that the discovery of ethics by demographers is a laborious process, marked by advances and setbacks. To provide a better grasp of these difficulties, we shall deliberately begin our reconnaissance trip in the company of a hardline "macrodemographer": Macrodemus. He seems to be the antithesis of the ethically concerned, because he sincerely believes that his working style poses no particular ethical issues. We shall follow him in his unlikely and involuntary visit to Ethics Island. This initiation voyage will give him the opportunity to make many discoveries, until the day when, like Robinson encountering Friday, Macrodemus finds a companion in the guise of Microdemus. He eventually understands that they will need to join forces to face the ethical challenges of tomorrow.

I. FIRST EXPLORATION OF THE ETHICISTS' ISLAND

When our story begins, at the start of the Third Millennium, Macrodemus has not yet met Microdemus. Through circumstances beyond his control, which it would be superfluous to recount here, Macrodemus has landed unexpectedly on an unknown island with his laptop computer—and no hope of return. He has taken refuge at the top of a promontory, where he has gradually built a rugged shelter. Thanks to wireless Internet, he continues to receive and process his beloved data. Through the same channel, he fulfills contracts for national and international organizations and receives his salary and research funds. Macrodemus is a modern Robinson. He has made the unsurprising discovery that the island is inhabited. He has promptly introduced himself to the locals, but without arousing any feelings among them except polite attention. He accepts this indifference and rarely leaves home.

Macrodemus has not taken long to discover that the island is called Ethics Island and is dominated by an elite of formidable experts: the Ethicists. They are subsidized by the island's governor for the task of addressing all the moral dilemmas arising on the archipelago, and even beyond. The Ethicists describe the dilemmas, exacerbate them (a good dilemma is a life-or-death dilemma), and sometimes try to resolve them. They produce a prolific literature, in which several debates rage. The most prominent have seats in houses for the initiated, called "Ethics Committees," which the governor consults periodically. The Ethicists practice the cult of ancestors but do not elevate them to the same degree of divinity as sociologists do. They venerate founding fathers from the late 18th and early 19th centuries: Immanuel Kant, founder of the ethics of duty; and Jeremy Bentham and John Stuart Mill, founders of utilitarianism. But the cult of Aristotle—for whom virtue is the quality of virtuous men and must be nurtured through education—still has its followers and is even enjoying a certain revival.

At the start of each academic year, the Ethics Island bookstores display stacks of textbooks and collections of readings that ethics professors recommend to their students. Almost all of these works are in English (Mackie 1977; Singer, 1993; LaFollette, 2000; Thomson, 2000). [CB: order cites late to early year]A few manuals in French, imported from Quebec, have slipped into the list (Duhamel and Mouelhi, 2001).

Macrodemus is not a big reader and, to tell the truth, his intellectual curiosity before landing on the island was very limited. But one day, in a moment of forced idleness (the macrodata he was expecting had not been delivered on time), he fought off his boredom by taking refuge in a bookstore. There, he struck up a conversation with a very learned and highly persuasive young salesgirl, Philarethes, who invited him to browse the works of the Ethicists.

This was a discovery for Macrodemus. The issues addressed were the life-and-death issues recurrently raised by events in the news. The Ethicists pondered the legitimacy of certain actions such as preventive war, capital punishment, suicide, euthanasia, abortion, tissue and organ removal for grafting, recovery of stem cells from excess human embryos, gay marriage, and gay parenting. Macrodemus was surprised to discover that some authors discussed these sensitive issues in an intelligible, lively manner. Philarethes advised him to purchase a remarkable encyclopedia.[3] At home, he started reading it in small installments, and this prompted him to buy more books in the months that followed.

From these readings, Macrodemus derived some simple ideas:

- Ethics is a critical questioning of the principles that guide our choices and actions. It has a practical goal: the point is not only to wave principles around but also to act in accordance with them.
- Ethics is meaningful only for players who are in a position to make responsible choices: totally constrained actions are not choices and cannot be subjected to ethical assessment.
- Ethics questions the moral legitimacy of our actions, whether private or public. The issues raised are at two levels: meta-ethics and applied ethics.

Applied ethics covers the practical dilemmas already listed. Meta-ethical issues touch on the principles for distinguishing good actions from bad actions. In particular:

- Does the value of actions lie in those actions or in their consequences? If we ought to judge them by their consequences, what time frame should we take into account: the near future, the next generation, or future generations? And on what portion of humanity should we assess the effects of the action: myself, my immediate circle, my partners, strangers, neighbors, or all of humanity?
- Can we fail in our duties (e.g., abiding by contracts or respecting truth) to save a person's life?
- How should we define the welfare that we ought to promote, given the number of persons and the diversity of their conditions? Is it the growth in aggregate welfare, the growth in welfare per capita, the alleviation of suffering for the less fortunate, the definition of a basic quality of life, economic justice, the universal respect for the individual, the fulfillment of personal aspirations in all their diversity—or something else?
- Are there absolute or sacred values that apply to all persons and are, strictly speaking, unchallengeable, that is, impossible to discuss?
- What is the moral value of an action that is good for me but unacceptable to others?
- Can the same society harbor several moral systems?
- Is every individual free to define the good to which he or she aspires—in which case the State should remain neutral and impartial toward individual preferences—or should the State rank priorities and steer behaviors in accordance with a vision of collective welfare?
- Should we entrust to experts or specialized institutions the task of examining or even resolving our moral dilemmas?

Macrodemus understood that the distinction between applied ethics and meta-ethics was classic but not crystal clear, owing to a necessary alternation between the two registers: the dilemmas of applied ethics can be settled only by relying on a thoughtful, critical vision of principles, while principles are revised or deepened with the occurrence of practical cases that raise new issues.

The list of questions raised by ethics is not improvised. As the young bookseller confirmed, they had taken shape in the course of long discussions that had begun in antiquity and had intensified in the late 18th century. As religious references faded, two major families of theories had emerged (Table 99–1). Because of their relationship, marked by sharp mutual criticism, they would polarize the field of ethical reflection until our time. At one end was the Kantian ethic of duty (or "deontologism"). It does not define the nature of Good but subjects the rules of action to the universalization test: a rule is moral only if it can be universal, which means not instrumentalizing persons but respecting their dignity in all circumstances (Table 9–2). At the other end stood "consequentialism,"

[3] *Dictionnaire d'éthique et de philosophie morale* edited by Monique Canto-Sperber (1996). Updated in 2001, reprinted in 2004 in paperback (two volumes of 1,000 pages each), it has 330 entries, averaging eight pages, with a detailed bibliography for each. Two hundred sixty authors contributed, including many from outside France; the indexes are very helpful. There is a glaring gap between the dictionary's international quality and the weakness of ethics teaching in the French education system, in both mainstream and vocational programs. The situation is entirely different in English- and Spanish-speaking countries.

TABLE 99–1 The Leading Theorists of Ethics, from Antiquity to the Present

Author		Main work and date of publication	
Aristotle	(384–322 BC)	*Nicomachean Ethics*	320 BC
Pyrrho of Elis	(360–272 BC)	No writings (views known via Sextus Empiricus)	—
Marcus Aurelius	(121–180)	*Thoughts*	180
Thomas Aquinas	(ca. 1225–1274)	*Summa theologiae*	1273
Baruch Spinoza	(1632–1677)	*Ethica, ordine geometrico demonstrate*	1677
David Hume	(1711–1776)	*Enquiry concerning the principles of morals*	1751
Immanuel Kant	(1724–1804)	*Kritik der praktischen Vernunft* (Critique of practical reason)	1788
Jeremy Bentham	(1748–1832)	*An introduction to the principles of morals and legislation*	1789
Immanuel Kant	(1724–1804)	*Grundlegungen zu einer Metaphysik der Sitten* (Foundations of the metaphysics of morals)	1797
Jeremy Bentham	(1748–1832)	*Deontology*	1834
John Stuart Mill	(1806–1873)	*Utilitarianism*	1861
Henry Sidgwick	(1838–1900)	*Methods of Ethics*	1874
Francis Herbert Bradley	(1846–1924)	*Ethical studies*	1876
George Edward Moore	(1873–1958)	*Principia ethica*	1903
Émile Durkheim	(1856–1917)	*L'Éducation morale*, lectures given in 1902	1925
Richard Mervyn Hare	(1919–2002)	*The Language of Morals*	1952
Elizabeth Anscombe	(1919–2001)	*Intention*	1957
Richard Mervyn Hare	(1919–2002)	*Freedom and Reason*	1963
John Rawls	(1921–2003)	*A Theory of Justice*	1972

TABLE 99–2 The Three Versions of Immanuel Kant's Categorical Imperative

Original German	English translation
Handle so, daß die Maxime deines Willens jederzeit zugleich als Prinzip einer allgemeinen Gesetzgebung gelten könnte.	Act only on that maxim whereby you can, at the same time, will that it should become a principle of universal law.
Handle so, als ob die Maxime deiner Handlung durch deinen Willen zum allgemeinen Naturgesetze werden sollte.	Act as if the maxim of your action was to become, through your will, a universal law of nature.
Handle so, daß du die Menschheit sowohl in deiner Person als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst.	Act in such a way that you always treat humanity, whether in your person or in the person of any other, never simply as a means, but at the same time as an end.

which judges actions (or the rules of action) not on their intrinsic value but by measuring their consequences as a function of a specific good to be promoted (welfare, pleasure, happiness, utility, etc.). The main variant of consequentialism was the utilitarianism of Jeremy Bentham and John Stuart Mill, which aimed at the "greatest possible happiness for the greatest number of people," asserting that each person should "count as one and only one."

Examining the list of major ethical questions, Macrodemus had the feeling that some might directly involve his discipline or that he could have a say in the manner of addressing them. In particular, the number of humans seemed to be a key element in utilitarian ethics. However, a search of the bookstore's database with Philarethes failed to yield any reference work that presented a balance sheet of the relationship between ethics and demography. The young bookseller admitted there was a gap here.

Macrodemus's attention was drawn to a singular fact: contrary to a widespread belief, ethics was not principally concerned with issues of sexual morality. Most Ethicists were even convinced that these issues were secondary. Some leading religions stubbornly continued to place them at the center of their doctrine, but the entire history of moral philosophy was a long, patient effort to emancipate ethics from this religious dominance. The divine commandments could not constitute an autonomous source of ethics. For either the divine commandments were moral because they were already consistent with preexisting ethics, or they arbitrarily gave God the exclusive privilege of defining good and evil, which made human ethics superfluous. A divine commandment is an absolute truth, literally undebatable, in the sense that it is impossible to discuss, being sheltered forever from public debate, under the sole control of authorized exegetes; it therefore cannot have the slightest ethical value. There was no absolutist ethic but only absolutism *tout court*. The same reasoning applied to the references to "natural law" or "anthropological invariants," which in fact were merely variants of the argument from authority, often applied to sexual relationships or family configurations. True ethics could not rest on divine or natural foundations.

Macrodemus had never thought about these issues too hard. In conversing with Philarethes, he had the feeling that the Ethicists' approach, unlike the naturalist or religious approaches, maintained a relationship to truth that was very close to the scientific spirit: truth is not given, but is built up through a succession of proofs and counter-proofs. It must be subject to debate. It was not relativism. Some truths were now established, such as the Earth's rotation around the Sun, or the primacy of the respect due to the human person. But the only reason why they had been established is that they had been conquered. They would always need to take account of the novel issues that the latest events would keep raising.

But Macrodemus's curiosity had its limits. His interest in Ethicist literature mainly satisfied his concerns as voter or consumer, as patient or parent. He confessed to Philarethes that, as a demographer, he felt little concern for the Ethicists' dilemmas. The young bookseller conceded that demographic issues occupied only a tiny space in this vast literature. Macrodemus thought he understood why. The demographer is never in a position to resolve the life-or-death dilemmas explored by the Ethicists. He always arrives after the battle, after the action. He can, of course, estimate the losses due to war or genocide, identify the share of suicides in death statistics, count the number of abortions, measure the effects of medically assisted procreation on fecundity and fertility, and so on. More generally, he is bound to take an ever-greater interest in new reproductive methods, in new forms of union, in lives extended by medicine or cut short by violent deaths, and he does so chiefly in order to understand these phenomena and assess their impact on population dynamics. But when he steps in, the moral dilemmas have already been settled—or discarded. Unlike the physician, psychologist, or social engineer, the demographer does not operate on bodies or souls. He decided nobody's fate. To borrow the term used in ethical analysis, demography is not an "interventionist" science. It is only exceptionally associated with clinical or therapeutic trials. It even disregards the in vivo experiments of social psychology. We can say of the demographer what Proust said of the statistician: he takes a "bird's eye view," because his goal is not to find out about an individual life—even less, to influence its course—but to grasp the aggregate of all the existences that compose a population.[4] From his refuge on his promontory, Macrodemus appreciated this commanding position and was not about to abandon it.

One exception, however, caught his attention. While the Ethicists were very discreet about demographic issues, some chimed in with the ecologists in stressing the responsibility of present generations toward future generations. This topic took up some room in textbooks on environmental ethics (Pojman 2005). Monique Canto-Sperber's dictionary had an interesting entry on environmental ethics (Birnbacher [2001] 2004). But Macrodemus felt the issue had been treated rather superficially: the philosophers' plea for sustainable development still referred to a "population optimum," an old notion long since abandoned by demographers. Indeed, where should one locate this alleged population optimum, if not somewhere in the huge space between an underpopulated territory and an overpopulated territory? These, in turn, were extremely fuzzy concepts soon made obsolete by the passing of time. Demographers were not uninterested in the interactions between resource renewal, population size, and human welfare, but most were aware that the issue largely spilled over the field of demography. Just look at the Millennium Goals defined by the United Nations, in which demography took up only a limited share. The leading international researchers in those early 2000s (such as Amartya Sen [1987] and Joseph Stiglitz) were primarily economists with a thirst for justice; they made very limited use of demography, which mainly served as a backdrop for them. True, the ethical issues they raised did find some resonance among specialists of population studies in the developing countries, but without producing a definition of a robust ethical system for demographers.

Philarethes had listened to this speech intently. She did not seem to approve it entirely. But Macrodemus

[4] At the end of *Time Regained*, in the aftermath of World War I, the narrator reviews the social trajectories of the various characters: "Of course, circumstances and individual character have always a share in the matter and it was in quite different ways that Legrandin [. . .] had in his turn penetrated this *milieu*, that Odette's daughter had become related to it, that Swann and finally I myself, had entered it. [. . .] But, from the bird's eye view of a statistician who ignores reasons of sentiment and the avoidable imprudences which lead to the death of an individual and only counts the number of people who die in a year, one could observe that many people starting from the same environment as that with which the beginning of this narrative has been concerned reach another quite different and it is likely that, just as in every year there are an average number of marriages, any other well-to-do and refined bourgeois *milieu* would have furnished about the same proportion of people like Swann, like Legrandin, like myself and like Bloch [. . .]" (Proust, [1927] 1967, p. 346; English trans. 1931, with one amendment).

did not want to linger. He took his leave, promising to resume the discussion again some day.

II. THE DREAM OF MACRODEMUS

One fine day, Macrodemus received a visit from an Ethicist. People had found out about his activities. He was handling computer files produced by the processing of individual data. He should urgently declare them to the ethics committee with authority in such matters, under pain of being found in breach of the law. Macrodemus had no choice but to undertake the necessary formalities, which, all in all, were rather modest.

But he was in for a shock. He had thought he was above all suspicion. The issues pertaining to the black box of individual-data collection were of little interest to him: he had never administered a questionnaire in the field (unless in an earlier life, or as occasional supervisor in an exotic district). He obtained his material from statistical offices, on the principle that they had one single task: that of supplying him in a timely manner with the impeccable files that his sovereign science required. The files had to contain the complete list of unions and their children and, more generally, the exact calendar of all events (so that he could fill out all the half-lozenges of the Lexis diagrams). He had trouble understanding why it was so hard to obtain the exact chronology of migrations: why wasn't it routinely included in the files of all the statistical offices, with the variables clearly describing the origins of each flow? He could not imagine the efforts and transactions involved in producing demographic information. He readily believed that it was enough to remind every player (citizens, foreign residents, civil registrars, doctors, police, etc.) of what the public interest required. Overwhelmed by the force of his arguments, the authorities would finally decide to introduce the right questions into the right schedules, in order to collect impeccable responses and deliver them in the proper format to the demographer. One day, exasperated by the inertia of officials and the failures of statisticians, Macrodemus had not hesitated to publish an open letter in the press solemnly enjoining the authorities to follow his recommendations for producing better statistics.

Recommendations was one of his watchwords. Macrodemus had a very simple way of characterizing the social exchanges needed to produce data. It was a balanced theory of rights and duties. A single duty: that of the authorities to provide the relevant data. A single right: that of the demographer to receive them. To cope with the gaps, biases, and disparities in data from country to country, Macrodemus believed a two-step approach would suffice. First, you analyzed the deficiencies by identifying the data that should have been available and those that were not. Then, you promulgated the national or international recommendations that made it possible to remedy them—by insisting, for example, that a particular country add a specific question to a given register or identity document. The rest was a matter of willingness and enforcement. If the recommendations were not followed, it is because the authorities were incompetent or craven.

Occasionally, Macrodemus had a dream, that of an ideal data collection. It was an unmediated operation without the participation of a statistical office, survey department, or interviewers—a flow of data springing directly from the population to feed his computer. No ethics committee, no authority protecting individual data. And this dream extended to all of humanity: he saw on his screen a unified global system of vital statistics—or, better yet, a planetary population register—that would provide him in real time with all the data and indicators he needed. A large title appeared on the screen: "Universal predetermined harmony of data," written in a new language that he instinctively understood: demospeak. But his euphoria was short-lived. No sooner had Macrodemus started to query the database than the screen went fuzzy and displayed blinking failure messages: "Data concealed!" (by the populations surveyed) and "False data!" (owing to the incompetence of statisticians). Oddly, men and women refused to be counted with the same docility as livestock. They didn't want to be corralled, branded, or tracked. They demanded explanations. Strange words streaked across the screen: "consent," "non-response," "private life," "information technology and civil liberties," "borders," "informal," "irregular." Unfilled questionnaires fluttered in the skies, eventually cluttering up the screen. The awakening was rude.

Macrodemus would have liked to shake off this nightmare, but the vision returned, relentlessly. He lacked a field culture, a personal experience of the organization, and, ultimately, a genuine statistical culture to understand that data come from people, that they are burdened by transaction costs, and that the demographer's art is to be able to work with human data—in other words, imperfect data.

III. WHEN MACRODEMUS FANCIED THAT DEMOGRAPHY COULD MAKE A MOCKERY OF ETHICS

Each year, Macrodemus declared his data files to the Ethicists, irritated at having to comply with what he saw as a superfluous obligation. An idea occurred to him: wasn't there a radical means to expedite the issue of the links between ethics and demography by asserting outright that true science couldn't care less about ethics? After all, the goal of science—and this applies to demography as to any other discipline—is not to maximize human happiness but to augment the quantity of knowledge, irrespective of the use that humans make of it. Science is unconcerned by good and evil. It must only untangle truth from falsehood, and this quest for truth presupposes total freedom of research. Let's be Nietzschean, for heaven's sake! Let's assert the gay science beyond good and evil!

Energized by this revelation, Macrodemus donned Zarathustra's clothes and started traveling across Ethics Island preaching the theory of Superman—the Superman of science, naturally.

But the Ethicists were not impressed. They were experienced and subtle folks. It was even said that they had taken power on the island after the fall of a local autocrat, the "Super-scientist," who had compiled a database of all the islanders without telling them. The Ethicists promptly reminded Macrodemus that his claim to autonomy ignored some crucial points:

- To engage in an activity with no concern for its consequences on one's fellow-humans is, in itself, to take an ethical position, however negative.
- Declaring the free search for truth as an unconditional value is to defend its universal character, which is already an ethical position (we know, since Kant, that a rule of conduct is moral if it can be made universal, that is, if it can be adopted by all without harming anyone).
- The unconditional search for truth cannot prevent us from entering into conflict with other values, given the multiplicity of reference systems in our societies. But a researcher who ignores value conflicts would leave the task of settling them to others—moral philosophers, jurists, legislators, journalists. He would therefore lose the autonomy to which he claims title.
- In social science, data do not spring naturally from the soil. The source of sources is always the persons whose declarations are collected, whose administrative data are gathered, or who need to be observed for relatively long periods. This raises issues of consent, anonymity, secrecy, and information access.
- As people are not things but actors, they voice opinions, implement strategies, and interpret their own history: all these viewpoints, as well, belong to social reality, and deserve to be heard. Objectivating the researcher's viewpoint with respect to the actors' viewpoints is a scientific requirement that also has an ethical impact.
- Scientific activity is both individual and collective. On the one hand, the researcher legitimately seeks to establish his or her reputation in a competitive world. On the other, many agents and institutions are involved in the data production and processing chain. This raises issues of intellectual property and recognition, which are ethical issues as well.
- Knowledge production has a cost, if only to pay for professionals with the time needed to conduct research and methodical investigations. But the relationship with employers, orderers, and financers (very often governmental or international agencies) creates a tension between responsibility and freedom that has an ethical dimension.
- Science does not boil down to an intransigent quest for truth. It is also a path to fame and honors—resting on a pedestal that may be national or international, duly controlled by the profession or broadened to the extreme by the media, at the risk of upsetting the internal ranking of scientific values. Should one counteract distortions of this kind or, instead, exploit them? Does the publication of unverifiable but stimulating tests have its place in the advancement of research? All these issues touch on professional ethics.
- More generally, we cannot skirt around the relationship between science and society, which inevitably has an ethical dimension. Despite many doubts, we expect that scientific research, if anything, will help to increase collective welfare rather than decrease it, to strengthen personal rights rather than erode them. We could, of course, imagine a Nietzschean position that consists in denying the universality of personal rights and privileging a knowledge elite situated beyond good and evil. But this pretension to a superior amorality is itself an ethical position that, as such, must be open to discussion and refutation.

It was decidedly impossible to elude the Ethicists' grasp. Demography was admittedly a temperamental, freedom-loving science, but it could not waive its ethical duties. There had to be an alternative. Our macrodemographer did not give up. The Nietzschean model being of no help, he resolutely embraced a new strategy: outwitting the Ethicists at their own game. They wanted ethics? They would have it!

An opportunity arose: the authorities invited him to give a lecture in a series on "Ethics and society" at the Island's Grand Theater. Macrodemus embarked on intensive preparations. As he did occasionally in times of great need, he started to read some books, but this time without assistance from Philarethes; the young bookseller had defended her doctoral thesis and, apparently, taken another job. Macrodemus spent a whole month polishing his arguments. On the day of the lecture, he was ready.

IV. WHEN MACRODEMUS IMAGINED THAT DEMOGRAPHY WAS ALREADY ETHICS IN ITSELF

Before a large audience of invited guests, Macrodemus began his lecture. Ethics and science were inseparable. But some disciplines meet this requirement better than others, starting with demography. It was, for Macrodemus, the most ethical of the social sciences. It described the general dynamics of the population in a fully neutral, unbiased manner. The distance it spontaneously took from persons made it perfectly harmless. *Primum non nocere* (First, do no harm) proclaimed the Hippocratic oath in its Latin version. Well, Macrodemus argued, the observance of this great principle was enough to make demography ethically self-sufficient. No need to call on the Ethicists: demography was ethics through and through.

Macrodemus was convinced of the radically innocuous character of his activity. The demographer in a research center was neither an administrator nor a judge, neither a healthcare professional nor a social worker. He never handled personal files for the purpose of deciding people's fates but worked only on derived statistical files, which were always anonymized. His statistical activity could therefore have no effect on people, either negative or positive.

Indeed, Macrodemus noted, this neutrality was not specific to demographic analysts working on aggregate data. It also applied to researchers mining individual data, even when the data were of administrative origin. The main factor was not the nature of the source, which could be administrative (e.g., the justice ministry, the interior ministry, or a social-protection agency), but the purpose of the new file was that it was compiled for knowledge purposes, ruling out all intervention concerning people and, from this standpoint, totally disinterested. The demographer practiced better than anyone the art of detachment, far from human agitation.

And Macrodemus challenged his listeners: do you think that the subject of demography is life, love, death, and exile? If that were true, it would be terribly intrusive. In reality, demography talks of other things: births, unions, deaths, and migrations. The demographer is an alchemist who converts the milestones of human life into a prosaic series of dates. Hence, like the ancient sage who rises above to contemplate the restless world and seizes the order of things under the chaos of passions, the demographer remains impassible. He can study infant mortality without shedding a tear on the fate of children. If he counts abortions, it is without taking sides. If he tries to enumerate migrants, it is not to proclaim that there are too many or too few of them, but to measure the impact of migration flows on the demographic or social dynamics of the countries of origin and destination. If he tries to individually identify the old people who break longevity records, it is not out of love for the very elderly but to improve the quality of the data that will allow a modeling of the final phase of life tables. In short, demography is not sentimental.

Indeed, Macrodemus noted wryly, any listeners who fancied going up to his office one day would soon realize this. There's nothing more disappointing that a demographer's office: a computer, statistical compendia, files, a few memos. No cell holding a few guinea-pig inhabitants awaiting interrogation. The demographer lived outside the world. By spending his life handling aggregate series, like a hermit reciting his prayers, Macrodemus had no need to make contact with the anonymous individuals peopling his tables of figures. So what if he was told that such remoteness from the field was unthinkable in most other social sciences—anthropology, social psychology, sociological observation? He accepted this singularity and saw it as a guarantee of objectivity. Macrodemus was not unaware that a population consists of men and women in flesh and bones. But he had no reason to take an interest in people in themselves. Their sole value was as units of account or as underlying elements of variables allowing the construction of anonymous statistical relationships. The data-collection process was so heavily delegated and decentralized that the demographer, working at the end of the chain, could ignore its existence, although he might recall it when the quality of the data fell short of his standards.

It would soon be time for a break. The seminar organizers opened the floor to an initial discussion. A listener thought this conception of the demographer's profession was terribly narrow and reductionist. Macrodemus replied that an office researcher could make an excellent demographer provided he did not move out of his field of competence and remained clearly aware of his power and its limits. Demography had to be reductionist in order to be productive: depersonalization was the price to pay to construct an overall view of the population that remained beyond the reach of individual scrutiny and ensured the researcher's ethical neutrality.

Another attendee declared that, quite to the contrary, Macrodemus's vision did not lack grandeur. It was reminiscent of the contemplative reduction exercises practiced in antiquity, which marked the beginnings of the scientific approach (Box 99–1). Macrodemus accepted the comparison as a good omen. He cited the work of his friend Bibliodemus on the genesis of German demography in the 17th

BOX 99–1 The origin of the demographic viewpoint: the scholar's detachment, according to ancient wisdom

Greek and Roman philosophers, particularly the Stoics, practiced spiritual exercises that reflected their determination to remove themselves from the world by contemplating it from above and from afar. Imagine, they wrote, that you rise above the Earth until it is reduced to a sphere. Seen from on high, persons are merely particles, individual lives merge into a population, and you will see the hustle and bustle of humans give way to regular movements. Scipio's dream, described by Cicero in Book VI of the *Republic*, is the best-known example of this thought experiment. Pierre Hadot has shown that it fit into a long tradition of spiritual exercises of ancient philosophy, ranging far beyond Stoicism.

Emperor Marcus Aurelius, for example, quotes this "fine saying," which he attributes to Plato:

> That he who is discoursing about men should look also at earthly things as if he viewed them from some higher place; should look at them in their assemblies, armies, agricultural labors, marriages, treaties, births, deaths, noise of the courts of justice, desert places, various nations of barbarians, feasts, lamentations, markets, a mixture of all things and an orderly combination of contraries. (*Thoughts*, VII, xlviii)

The lofty Stoic perspective converges with the Emperor's viewpoint here. Significantly, the list includes the decisive events that were later to interest demographers: births, unions, and deaths. In the same period, the Romans had raised census and registration methods to a level of perfection that was not to be reproduced until the Renaissance. In particular, they shifted from an enumeration of noble families to the exhaustive count of individuals, issuing birth certificates to enable the Empire's subjects to prove their age (Nicolet, 1988).

More generally, the Ancients asserted the prerogatives of "theoretical life" or "contemplative life," which allows a handful of free men to observe the battles of interests among the multitude without taking part in them. The contemplative detachment of the wise man, as cultivated and celebrated by the Greeks at a very early stage, played a crucial role in the history of Western science by asserting the social figure of the scientist as disinterested researcher.

century. Stoic thought, which best embodied this distancing, had significantly influenced Leibniz and his disciple, the Protestant minister Johann-Peter Süssmilch, a keen student of demography. For the latter, the regularities of demographic phenomena could be contemplated and understood only from the standpoint of God, who is located at the "geometral point of all perspectives." It is from there that we perceive the "divine order,"that is, the optimal order of the world, which combines—or so Süssmilch believed—the best birth and death rates imaginable (Rohrbasser, 2001). Provided it was duly secularized, this point of view remained valid.

It was objected to Macrodemus that the notion of modeling scientific detachment on divine detachment reflected a very elitist approach. But Macrodemus countered with the example of John Graunt, the man who inaugurated demography by tabulating deaths in the city of London. In his seminal essay of 1662, Graunt expressed his satisfaction that demographic computations enabled a mere citizen of London, a merchant by profession, to have access to a comprehensive science and hence be elected to the Royal Society. Very early on, the pioneers of demography realized that tabulation and computation gave everyone access to a totalization capacity hitherto reserved for kings or theologians. The divine gaze on the demographic order of things was now within reach of the lowest commoner, provided that he mastered the appropriate techniques.

To a sociologist who asked him what he thought of Émile Durkheim's approach, Macrodemus explained that demography realized almost effortlessly the ideal proclaimed by Durkheim in 1894 in *Les Règles de la méthode sociologique* (The Rules of the Sociological Method): it spontaneously treats social facts "as things," it does not let itself be thrown off course by players' motives, and it steers clear of preconceived ideas and value judgments.[5]

[5] Émile Durkheim's *Le Suicide* (1893) is the textbook example of a statistical analysis that may be conducted in the absence of the interested parties or their close circle (note, however, that Durkheim mentions in *Le Suicide* an important collection of letters left by persons who committed suicide and that he returns to it at the end of the work, taking care to check that these testimonies do not contradict his typology of suicides). Throughout his career as sociologist and anthropologist, Durkheim remained a desk-bound scholar, remote from the field. When he agreed to run the Commission for welcoming Jewish refugees from Russia in World War I, he performed his duty with dedication, but confided to his close friends that he suffered from this change of register: "To decide people's fates on the basis of written reports without seeing the interested parties leaves the door so wide open to error that it makes me shudder. We're working in the live flesh, a job that's not for me" (Durkheim, 1998, letter to Marcel Mauss, March 12, 1916).

V. UNIVERSAL, IMPARTIAL, AND NEUTRAL?

After the break, Macrodemus resumed his presentation. He explained that the demographer treated people in a universal, impartial, and neutral mode. Universal, because in good demography everyone counts. Impartial, because results are not influenced by the observer's position and interests. Neutral, because the goal is not to describe what must be but only what is. And Macrodemus set out to elaborate these three points.

He began by quoting a remark by Jacques Vallin at the beginning of an introduction to demography: "We cannot count individuals or compile statistics about them unless we treat them as if they were all the same" (Vallin 1992). This statement contained a dual concept: individual enumeration and a fertile reductionist approach. The notion of enumerating individuals may seem simple but it is not self-evident. For a long time, authorities counted intermediate management units (houses, lineages, hearths, concessions, etc.) or pivotal individuals (freemen, men fit to bear arms, taxpayers, producers, the faithful, *de jure* citizens, natives, etc.). The notion of counting identically people of either sex and of all ages—without discrimination by status or origin, and irrespective of the price that can be assigned to them in the market for economic or social values—is a late development. It first emerged in Rome in around the second century AD (Nicolet, 1988). It did not reappear until the 18th century.

To realize this, we need only look at the lifelines in the Lexis diagrams: demographic analysis rests on the individual, not the family; it is not intrinsically family centered but individualist.[6] Family and household statistics are a useful but fortuitous extension of the demographer's work. Reproduction indicators, for example, do not start by separating couples that reproduce from those that do not; they aggregate individuals of the same birth cohort and try to see if this set collectively produces enough children to replace the cohort. The fact that some individuals remain childless and others have an above-average number of children does not influence the construction of the reproduction rate, although it is a useful element to take into account when explaining behaviors. In sum, the demographers' reproduction indicators can operate just as well in a Platonic society that collectivizes children as in a liberal society that individualizes them. What matters is the end result of reproduction at the cohort level. This is because demographic analysis bypasses ready-made social groups. It is content with placing individuals on a time axis. The cohort is only a time grouping that ensures a formal equivalence between persons. The equivalence principle likewise amalgamates persons born locally and those who have arrived from elsewhere, that is, natural increase and net migration, interpreted as two growth factors of a population. By the following year, demographic accounting incorporates the two additions into the reference population, because it naturally assumes that one existence is worth another, whatever its origin.

In its very principle, Macrodemus stressed, demography treats individuals as units of account, without weighting them in advance by any factor whatsoever. Such a notion presupposes an equivalence principle defining the membership of all individuals in a common humanity (Desrosières, 1993; Boltanski and Thévenot, 1991). Now this natural presupposition of demographers (shared with many statisticians) did not appear until late, at the same time as the emergence of the individual as an autonomous subject of rights and duties in moral philosophy. Macrodemus underscored the strange affinity here between the demographic calculus and the utilitarian calculus championed by Bentham and Mill: both procedures implement a principle of accounting universality that attests to an equal regard for all members of humanity.[7]

In the second part of his presentation, Macrodemus set out to show that he had adopted some of the Ethicists' basic ideas. The first was that all moral theories begin by defining a reference population whose members possess equal dignity. For Aristotle, the group consisted of free citizens, excluding slaves and half-breeds. For the Founding Fathers of the American Constitution, Blacks and slaves could not be counted on a par with Whites and freemen. The Australians did not start counting Aborigines until 30 years ago. And one could give further examples. By contrast, with utilitarian philosophy, the definition is as broad as possible. It potentially extends to all of humanity and even beyond. Macrodemus did not share the prejudice, still widespread in France, that saw utilitarianism as an economic ideology centered on selfishness and the pursuit of personal interest. He was sufficiently conversant with the founding texts of utilitarianism to understand that it was, in fact, a resolutely altruistic moral philosophy. In the phrase borrowed by Bentham from Francis Hutcheson, ethics sought "the greatest happiness of the greatest possible number" (Bentham [1781]). The balance of ills and goods should be

[6] Contrary to the argument defended by Rémi Lenoir (2003).

[7] Later, as we know, economic theory dealt a severe blow to the civic and community dimension of utilitarian calculations by reducing "utility" to a purely economic function (Pareto's "ophelimity") and treating every individual as an entrepreneur, including when he decides to form a union and have children. But that's another story.

assessed for the aggregate social body and ultimately for all of humanity, as each individual had an equal claim to welfare, regardless of his or her condition. Bentham defined the principle of the computation of benefits and negatives ("pleasures" and "sorrows") in a famous phrase—"everybody to count for one, nobody to count for more" (Bentham [1781])—later endorsed by John Stuart Mill (1861).

All things considered, it was a subversive idea, the product of a long history, of which Macrodemus could provide only an outline. It is potentially embodied in Stoic cosmopolitanism and, in an altogether different manner, in the equivalence established by Christianity between the children of God, all of whom were identically marked by original sin and the possibility of redemption. The notion that each person counts is expounded in the American and French declarations of the rights of man. It was fulfilled in the implementation of universal suffrage ("one man one vote"), which was accepted only at a late stage, in the face of stubborn resistance (Rosanvallon, 2002). For a long time, people thought it impossible to reduce the heterogeneity created by dependency relationships (master/slave, white/black, settler/aborigine, master/servant, man/woman, native/immigrant, etc.). The right to vote was denied for this reason to the homeless, the poor, servants, women, former slaves, indigenous inhabitants, and so forth, and most countries still deny it to immigrants—an archaism that will eventually disappear like all the others. The paradox is that universal suffrage without a means test or weighting was a path to emancipation whose access was denied to the subordinate categories, on the grounds that they were not sufficiently emancipated. They were judged too dependent on dominant groups or heads of families (just as, today, immigrants are barred from the electoral process on the grounds of their presumed dependence on foreign interests). By deciding to count everyone without discrimination, irrespective of their legal status, demographic statistics embedded in the core of its practices a universalist principle. It is also the principle of democracy on the march, which it sometimes takes one step further—even if this means having to face the old condemnation (dating back to Edmund Burke) of the "atomization" of the social bond triggered by the "rise of individualism."[8] Each person counts as one and only one: an unexpected consequence of this utilitarian formulation is that the reference individual cannot place his own interests above those of others when seeking to assess the consequences of his action. The obligation to grant the same weight to everyone, including to oneself, takes the impartiality requirement to the extreme. The utilitarian calculus must logically produce the same collective results, whatever the actors' positions.[9]

The negative part of Bentham's precept ("nobody to count for more") establishes a leveling principle: for the demographer as for the utilitarian philosopher, enumerated human beings are neither big nor small, neither powerful nor destitute. In a further stage, of course, demography, expanded into social statistics, takes heterogeneity into account by introducing all sorts of variables that differentiate situations and behaviors, but it does so a posteriori, by attaching these values to bases initially regarded as equivalent. It is precisely because it uses an equal, neutral background that it can measure differences, detect inequalities and, accordingly, objectify the variations in the probabilities of access (such as differential mortality). This initial leveling is anything but trivial, as attested by the resistance from people of quality, who are not fond of being quantified.

Macrodemus saw another proof of the impartiality of demographic calculus in his refusal to weight the planet's inhabitants according to their distance. Seen at a remove, the world no longer offers a privileged vantage point separating nearby peoples from faraway peoples. Again, this was a break from our habits. We have a natural tendency to weight the existence of human beings according to the distance between them and us. Adam Smith had already posed the question: if an earthquake destroyed China and its millions of inhabitants, would the English be affected? The famous apologist for the Chinese mandarin expresses the same notion: if I can make a fortune by simply pressing a button that remote-destroys a Chinese mandarin whom I've never met, would I hesitate for one second? And in its present version, the apologist for the bomber or cruise missile says the same thing: it is easier to obey an order to kill anonymous populations at a distance than to execute nearby or visible individuals point-blank. A modern variant: we feel more interest and compassion for victims that television brings close to us than for invisible victims.

[8] Edmund Burke, in *Reflections on the French Revolution*, castigates "the metaphysical and alchemistical legislators" who "have attempted to confound all sorts of citizens, as well they could, into one homogeneous mass [. . .]. They reduce men to loose counters, merely for the sake of simple telling, and not to figures whose power is to arise from their place in the table" (Burke [1790] 1980, p. 398).

[9] This impartiality constraint, a sort of "veil of ignorance" before its time, is an equivalent of Kant's universalization test: a point of view that is not interchangeable with another person's point of view cannot have moral value. That is why the present defenders of the utilitarian theory staunchly deny that it allows the sacrifice of the fate of a random individual to the benefit of an impersonal collectivity, as frequently objected by Ethicists of Kantian obedience.

The issue is all the more acute today with the disappearance of physical obstacles to information gathering and transmission in our global village. There remains a pure distance, the psychological or cultural distance, which measures our real degree of interest in others. It is taken into account by the media, who select images from information agencies depending on what they expect to be the level of public interest. The giant tsunami that devastated the South Asian coasts on December 26, 2004, causing almost 300,000 deaths (a figure whose size was not known until several weeks later), captured the attention of Western media for a sustained period. The event triggered an unprecedented solidarity movement, coordinated by the United Nations, which had castigated the selfishness of the leading countries. Was it the number of victims that succeeded in offsetting the effect of distance? Would we have shown equal concern if the victims had not included many Western tourists? How would we have reacted if the catastrophe had mainly struck a poorer region less attractive to tourists, such as East Africa? The principle of "everybody to count for one" is at variance with our general propensity to measure the value of human life in terms of its physical or social distance. In this convergence with the Benthamite imperative, the demographic approach shows its solid ethical foundation.

A recent episode had particularly shocked Macrodemus. The intrinsic impartiality of the demographic approach had been undermined during the preparation of the 2000 U.S. census, at the end of the Clinton presidency (Anderson and Fienberg, 2001). The Census Bureau had suggested improving the census coverage by performing in-depth surveys in the most disadvantaged urban areas. Republican members of Congress had objected, fearing that this improvement would increase the electoral weight of the Democratic Party, which has a stronger presence in these districts than the Republican Party. The census results determine electoral redistricting in a country that takes for granted that minorities should be locally represented by elected officials of the same ethnic origin. Even as the partial defense of political interests was intruding into data collection, it was the Census Bureau that was accused of political partiality: the contradiction was such that the Bureau's director preferred to resign. The ethics of demographic and statistical neutrality had been openly violated, since the electoral weight and partisan weight of individuals had been taken into account to determine whether they were worthy of entering demographic statistics or not. For Macrodemus, this example simultaneously illustrated the dangers of too close a connection between the collection of demographic data and political organization, of which the United States offered an extreme example.

Macrodemus proposed strengthening the impartiality principle by extending it from spatial distance to temporal distance. This extension was a logical consequence of demographic analysis. A generation that wants to make an unbiased assessment of the impact of its behaviors should not place its own interests above those of future generations. The only way to achieve true impartiality would be to neutralize temporal proximity in the same way as physical proximity. The Ethicists who had pondered the issue were wondering about how to make present generations sensitive to the interests of future generations, which, by definition, are composed of distant, anonymous, and unknown beings. For Macrodemus, demography could play a decisive role in instilling such awareness. The calculation of reproduction rates, combined with long-term demographic projections, is indeed consistent with this impartial perspective that puts the interests of all generations on the same footing: it places inter-generational links in a dynamic framework that, by construction, grants equal weight to the players and output of successive generations, whereas economists do just the opposite: taking our preference for the present for granted, they apply discount rates to long-term investments, devaluing the distant future to the benefit of the near present.

Macrodemus refrained from giving other examples. He concluded his lecture with a call for mutual recognition of demographers and ethicists. Our objectives are the same, he assured.

Macrodemus was warmly applauded. The Ethicists had listened to his presentation in silence. They announced that a young Ph.D., recently promoted spokeswoman for the Higher Council of Ethicists, would respond. Macrodemus was surprised to recognize Philarethes, the young lady who had welcomed him a few months earlier in the Island's general bookstore.

VI. THE LIMITS OF THE ETHICAL CONTENT OF MACRODEMOGRAPHY: THE ETHICISTS' RESPONSE

Our guest, Philarethes began, has just made an interesting contribution to research ethics. His thesis is simple: the demographic approach is universal, detached, impartial, indifferent to personal cases, and disinterested. Do these qualities suffice to proclaim the intrinsically ethical value of demography? I don't think so, and here's why.

First, your ethics, dear Mr. Macrodemus, are essentially negative. You quoted the abstention phrase at the start of the Hippocratic oath: *Primum non nocere*. But that's just the beginning of the ethical approach, a basic reminder of what we Ethicists call the Golden Rule: "Do not do unto others what you would not have done unto yourself." Does doing no harm suffice to meet all one's obligations? Surely not.

Second, you tell us that the demographer practices ethics spontaneously, without having to give it a thought. Now an ethics worthy of the name cannot remain tacit. It must be practiced cognizantly, and rest on principles that can be stated and discussed publicly. "Everybody to count for one" may be a mere advertising slogan concocted by communication agencies in charge of launching a census. We now need to give it meaning.

Third, the invocation of ethical principles cannot be confined to self-justification. Many demographers and statisticians are convinced they are serving the public interest. The notion that one could suspect their attachment to the ethics of neutrality seems radically inappropriate and unfair to them. Convinced that their practice can harm no one, whether by its goals or by its methods, they reckon that their only problem is to get the public to share this belief, and that a large share of so-called ethical issues are, in fact, just communication problems. In sum, their practices are good; all that remains is to advertise the fact. This is to ignore that the ethical approach is relational. It's built up through interaction with colleagues, partners, the public, and limited partners—not proclaimed on a solitary basis. Where are the documents that constitute your ethics code and regulate your dealings with your interlocutors?

Philarethes took the example of the impersonal, anonymous character of data. She quoted one of the few French statisticians to have explored these issues: "The need to know individual data and the need not to take an interest in them are the basis of statistical confidentiality" (Padieu, 1991, p. 38). The problem is that the second need does not have the same status as the first. It's a form of abstention obtained automatically, without special efforts. In practice, the statistician working on thousands of individual recordings has no merit in obeying this rule. If he happens to handle individual files that have not yet been anonymized (e.g., when updating a longitudinal survey), we can easily imagine that his attention may be drawn here and there to a person displaying a rare characteristic (a woman with 10 children, a household with exceptionally high income, an exotic place of birth, a TV show host, etc.), without this ephemeral attention leading to public disclosure. We may find this situation fairly reassuring. If respect for confidentiality had to rely solely on the virtue of each person whose work involved handling the data, we would doubtless have reason to worry (with 500 people processing 5,000 individual data files each, the risk that at least one worker will commit an indiscretion is high). The fact that the manager of a statistical operation has no interest in divulging an individual datum offers the best possible protection. But this is more a consequence of data-protection arrangements stipulated by law than of a personal ethical disposition. Let's not overestimate the moral merits of the statistician who merely respects statistical confidentiality: this secrecy depends very little on his moral qualities.

Philarethes expressed her skepticism about the demographer's intrinsic impartiality. Far from being cosmopolitan, demography had often lent itself to nationalist manipulations. To say the least, demography did not escape the tendency for each country to focus on its own problems and to reduce enumerations to the "accounts of power." In many European countries, international migration boiled down to the immigration question, that is, the place of foreigners "on our turf,"—a real problem, admittedly, but one that was far from exhausting the issue. The United Nations' sustained use of demographic data, thanks to the compilation and harmonization undertaken by its Population Division, clearly showed that demography spoke the same language on all continents and practiced the same methods everywhere, but the UN's work was not itself without incidence on the fate of nations. And Philarethes gave an example inspired by one of her past conversations with Macrodemus.

The latest UN 50-year projections revealed that Yemen might overtake Russia in population size, making it one of the world's demographic giants. Paul Demeny, demographer at the Population Council and editor of the journal *Population and Development*, relied on these projections in a doomsday article that used this scenario—Russia overtaken by Yemen—as the symbol of the demographic threat from the Muslim world to an old Europe resting on privilege (Demeny, 2003). The article had a public impact. And the Yemeni authorities were able to exploit this unexpected publicity to obtain a mega-loan from international monetary organizations. The problem is that some demographers specializing in the Arab world challenged the quality of the data used: it turns out that Yemen's fertility rate is not seven to eight children per woman but rather five, which means that its demographic transition is already well under way and that its future growth would be less robust than what the UN projections announced (Courbage, 1999).

I don't have the means to settle this dispute among demographers, Philarethes hastened to add, but I'm struck by the fact that the international image of an entire country and, consequently, its economic fate can be so dependent on the uncertainties surrounding the demographer's calculations. I thought, she added, that the statistician could frame his measurements in a "confidence interval." Isn't there a statistical ethic that the demographer should have followed in this case? In any event, the notion that the remote handling of demographic macrodata protects local populations from all intervention (benevolent or malevolent, it makes little difference) is clearly refuted by such an example. Generally speaking, the main justification for demographic coverage of the planet since World War II has been that the data gathered should allow an expansion of action programs in developing countries—for reducing fertility, combating HIV/AIDS, or intervening in migration flows. This is not reprehensible in itself, and is perhaps even positive, but we must conclude that the practice of demography is not without consequence to the fate of the countries concerned and that the demographer would do well to reflect on it.

Philarethes paused. We need to go further, she continued. You stressed the demographer's impartiality, his indifference to social weightings, his lack of interest in individuals. But impersonality, when linked to bureaucratic indifference, is a double-edged quality. It can, depending on circumstances, provide people with the most efficient protection possible or cover up the worst excesses. The demographic study of populations largely owes its neutrality toward individuals to the fact that it retrieves the standardization work already performed by society: the main life events have been ritualized, then recorded, by the Church (series of baptisms, marriages, and burials), before being recovered by civil registration officials (births, marriages, deaths). Today, the information is sent by postal mail or computer networks to statistical institute delegations, then centralized and put through the statistical number-cruncher. Survey data follow similar paths. The whole process unfolds, as Alfred Schütz put it, "like a letter in the mail," through channels that are not those of interknowledge or personal exchange but require a new form of trust, which has become crucial in our societies: "systemic trust" (Luhmann, 1989), that is, a reasonable confidence by individuals in the proper functioning of an anonymous bureaucracy.

You are not unaware, she went on, that one of the most mind-boggling phenomena of contemporary history is the ease with which ordinary people ensured the success of the final solution perpetrated by the Germans against the Jews, as if it was a mundane business matter. An army of functionaries and employees contributed to it by applying measures of which many were—in Raul Hilberg's concise phrase—"of a bureaucratic nature, anchored in routine, habit, and tradition" (Hilberg, 1985, p. 219). This required not only the contribution of railroad employees and police officers but also that of ordinary clerical workers "attached to the respect of form" and for whom the Jews "were only a word in a subheading of their files." A careful examination of the evidence reveals that "normal procedures were applied to abnormal situations, as if the decisions were in no way extreme." One had to hate the Jews in order to decide to persecute them, but (contrary to what Daniel J. Goldhagen imagines) it was not even necessary to hate them in order to contribute to the smooth functioning of the persecution machine. Human problems were kept at a distance, largely neutralized and de-realized: they became technical problems. Correcting Hannah Arendt's famous phrase on Eichmann, Raul Hilberg concludes that the "banality of evil" operated at the ground level rather than among top decision makers: they were fully aware of what they were doing, and knowingly relied on the inexhaustible resources of bureaucratic and "systemic trust" (Hilberg, 1985). In sum, a core of hate surrounded by a gigantic halo of indifference proved capable of fueling the enterprise of extermination.

Macrodemus wanted to interrupt her. He couldn't stand this penchant of the Ethicists for continually referring to History's worst misdeeds, as if statisticians or demographers were potential murderers.

But the Ethicists' spokeswoman had anticipated the objection: I know what you feel—you feel miles removed from such horrors, and you think this historical reference merely shrouds the issue we're discussing in useless pathos. You're wrong: this old history, which neither you nor I have known, will never be over. It gives us a permanent lesson, namely, that indifference to people as guaranteed by automated procedures has no ethical value in itself. It may protect the people concerned as it may protect their persecutors. Ordinary bureaucrats, Hilberg writes, acted "with deep indifference toward the nature and consequences of the acts performed" (Hilberg, 1985, p. 233). Let us apply this vigilance to demography (Selzer and Anderson, 2001; Selzer, 1998; Rémond, 1996). In ordinary times, the demographer's distanced, neutral treatment of persons has a negative ethical virtue: it does not harm people. In times of crisis, the reverse is true: neutrality can become reprehensible. It is neither beneficial nor detrimental, neither good nor bad.

And, to support her statement, Philarethes went back to census files and population registers. The notion of anonymity is more complex than you think, Mr. Macrodemus. Anonymized data may remain indirectly personalized if they include uncommon local or individual identifiers. If your file shows a doctor in a given age group with a given number of children in a particular locality, it's easy to track him down through a medical directory. Small groups localized with too great a precision are vulnerable to identification attempts. Also, the demographer, as you well know, does not always work on anonymous files. Increasingly, he uses longitudinal files, which, by definition, require the long-term archiving of individual profiles. Population registers (in some countries), residence-permit registers, social-insurance databases, and so forth are also long-term databases accessible to demographers and requiring special protection. There's nothing reprehensible about working on such data, provided that proper confidentiality precautions are taken. But all this requires special protocols. Plain routine won't do.

Philarethes offered a final argument. The demography practiced by Macrodemus did not embrace all of demography. The Ethicists were receiving ever more requests for consultation from demographers who combined macro-analyses and micro-observations. Witness this recent example: French statisticians are thinking of setting up, for research purposes, a vast cohort of children, similar to the Millennium cohort in Britain. The project would track 10,000 to 15,000 children from pregnancy to adulthood by combining direct surveys of parents, collection of biomedical data from children, information retrieved from social-protection agencies and schools, and environmental information gathered by means of sensors installed in homes. An undertaking as multifaceted and as long-lasting in its purpose takes us far, very far, from the national or planetary macrodata dear to the demographers of yore. It requires special ethical arrangements. The researcher no longer keeps human subjects at a distance but follows them closely throughout their lives: this raises serious issues of consent and confidentiality.

It's an extreme case, you will say, but we could mention simpler operations that consist in deepening macrodata by means of a microsurvey. One example is the international program for checking, individual by individual, the civil-registration records of all residents of a country who state their age as 110 or older. The demographer of mortality embarks on this type of operation because he wants to make a detailed description of mortality distribution at the oldest ages; he wants to see how it operates at the farthest tip of the population pyramid, where even the largest populations are inevitably reduced to a handful of individuals. Don't tell me, Mr. Macrodemus, that there's an unbridgeable gap between macrodemography and microdemography. On the evidence, they are tied by a continuum that is being explored in ever-greater depth. Moreover, who can fail to notice that the macrodemographer's field is narrowing? Marriages, which he stubbornly keeps tabulating as he did 30 years ago, have ever less significance, owing to the increase in cohabitation without marriage. If he wants to capture the informal chronology of union formations, he must go and ask the interested parties directly by means of a questionnaire. The chronology of migrations poses the same problem. The macrodemographer must ally himself with the microdemographer to update our knowledge base on the evolution of demographic behaviors. The ethics of macrodemographers and microdemographers will increasingly converge.

Philarethes offered her conclusion in the form of a question: dear Macrodemus, you're a demographer, but have you ever asked yourself why? There are a host of reasons for wanting to count the men and women who inhabit our territories. We may want to count because the interested parties want to be taken into account. We may want to count because of a desire to understand. We can count to assess needs, to measure what separates real-life situations from our ideals of justice. We can count because we want to speculate on the probability of survival and estimate the yield on annuity investments. A government may count to render accounts of its management. It may also count to control, and sometimes to exclude. In short, one may have the best reasons and the worst reasons to count. One thing is certain: nobody counts for the sheer pleasure of counting. Falling back on a technicist position does not dispense you from asking yourself about the purpose of your work and, ultimately, about its impact on the population and minorities.

The lecture was about to end. Macrodemus had neither the strength nor the time to reply to Philarethes. That would come later, no doubt. He returned to his outpost, firmly resolved to limit his contacts with the Ethicists to a bare minimum. However, the future would decide otherwise.

One fine morning, when Macrodemus was seasonally adjusting a Maltese or Cypriot series, there was a knock on the door of his workshop. He recognized an old colleague, whom he'd lost sight of, but whose name he still remembered. It was Microdemus.

Microdemus, too, had washed up on the Ethicists' island and had heard about the hermit on the promontory. He explained to Macrodemus what he was doing: a longitudinal survey on migrations between the islands of the archipelago and the mainland, based on a long questionnaire combining sociodemographic and economic questions, plus a series of qualitative interviews. He was looking for framework data, which Macrodemus perhaps could supply. That was the beginning of their relationship.

Macrodemus and Microdemus had very little in common. But the fact that they both lived abroad brought them close. They discovered that they were complementary. Microdemus told Macrodemus he had settled in the plain, near a village, where he was living in a house open to all visitors. He had trekked up and down the island and held long conversations with the Ethicists—including Philarethes, whom he knew well. He urged Macrodemus to get out of his workshop and explore the island at length. Soon, their relationship became more informal. They called each other by their abbreviated names, Mic and Mac. Each initiated the other in his science. The couple they formed was sometimes reminiscent of Robinson and Friday, sometimes of Bouvard and Pécuchet. As nothing in their appearance told them apart, the children nicknamed them Tweedledee and Tweedledum. In fact, they had become inseparable.

VII. DIALOGUE BETWEEN MACRODEMUS AND MICRODEMUS ON CHANGES IN DEMOGRAPHIC SURVEY METHODS

I wonder where demography's heading, Macrodemus confided to Microdemus one day. At times, I feel like a dinosaur, the last of the Mohicans. My master, Mathodemus, who still drops by to see me from time to time, is desperate; he feels his lifetime's teachings are lost, and that no one will ever be trained in demographic analysis anymore.

I don't share his pessimism, answered Microdemus. They're still training analysts, but of a new kind, who put demography in the broader context of social statistics. They can no longer merely adjust for structural effects by constructing "typical populations" free of disruptive events. They try to analyze files of individual data and to subject them to multivariate analyses that take durations into account and go beyond mere description, toward a form of statistics that is ever more explanatory.

But is it still demography? Macrodemus retorted.

It's still demography if it tries to shed light on the genesis and evolution of demographic behaviors: births, unions, deaths. But the novelty is that these behaviors are now viewed against a backdrop comprising a wide range of social, economic, educational, and institutional factors. The analysis of individual files requires new analytical techniques. So Mathodemus lamented the decline of demography? One could counter his complaint with a symmetrical one: is it conceivable that a demographer trained 30 years ago by Mathodemus, who had developed his expertise 15 years earlier, might never have subsequently received the slightest additional training to supplement his technical capabilities? Can we accept the notion of a researcher not enriching or upgrading his methods in 30 years? Tell me frankly, Mac, what new methods have you learned since your initial training? Have you, for example, familiarized yourself with duration models, which our young doctoral students and new recruits effortlessly master today?

Macrodemus carefully avoided answering. The idea of incorporating a training plan into his career plan had never occurred to him. He simply replied:

Why, then, change methods of demographic analysis that work? Sex and age will remain sex and age, fertility and mortality distributions will always need to be modeled, and Lotka's equations are here to stay.

It's the relationship to people, to human subjects, that's radically changed, Microdemus continued. The researcher must now keep close track of behaviors and question respondents in depth. This means asking for their cooperation and, ever more often, their informed consent. As you well know, Macrodemus, the new demographic behaviors elude official recording. Marriages no longer comprise all unions; families are reconstituted; new kinship ties emerge; people adopt children in lieu of biological children or, on the contrary, to add to the number of siblings; couples now form without cohabiting, calendars are getting complicated, complex ties are developing between demographic events (unions, separations, births, recovery of children) and the various paths followed by individuals (educational, residential, migratory, occupational, health-related). Length of life, quality of life, and number of persons are interacting variables. How can you possibly grasp all this richness by the sole grace of official aggregate data? Ever more often, demographic information must be obtained directly from the interested parties. To know what's going on, you have to go and ask people—it's that simple. You have to ask them questions about informal or unofficial behaviors. You need to jog their memories to reconstruct past events that have generated present situations. In short, you need to establish a new kind of relationship with people.

That's precisely what worries me, Macrodemus reacted. How do you make sure that people aren't making things up?

It's very hard to make things up. Responses given at random in a questionnaire would immediately be spotted as aberrant and removed from the data file. If they're simply errors, we can screen most of them through consistency checks upstream, inserted into the computer-aided personal interviewing (CAPI) programs activated during the interview; or we can correct them downstream through consistency checks that restrict the field of compatible responses. The problem doesn't lie there. Rather, it's in the respondents' tendency to standardize, minimize, and simplify (male bluff is not a prejudice; it explains many gaps between male and female responses). We accordingly need to estimate response bias, for example by benchmarking to external framework data or, more subtly, by combining different questions.

OK, but there are other sources of bias. I imagine there are a lot of things that people don't dare admit to you?

Only avowable practices are avowed, Microdemus conceded, but that's true of any survey or administrative record: income surveys never reveal tax evasion or embezzlement. If you want to study deviant or clandestine practices, you need other investigation methods. In fact, we draw a distinction between the private and the intimate. Respondents quickly understand that a question about private life (date of first-union formation, for example, or ex-spouse's occupation) is not prompted by an unhealthy personal curiosity but reflects a scientific inquiry of general import. If the questionnaire is well designed, if the survey managers are clearly identified as working in the public interest, then the operation will come across as professional and legitimate, and the ethical issue for us will be, instead, not to abuse the trust that respondents readily tend to place in us.

But, Macrodemus insisted, some surveys go further, like surveys on drug addiction, sexual behavior, and abortion. How do you manage?

I was coming to that. Such surveys are always tricky, of course. They need to follow precise protocols, guaranteeing confidentiality and anonymity (extreme cases may involve erasing individual contact information at the start of the telephone conversation). They must be conducted for a public-interest purpose: AIDS prevention, fight against discrimination, identification of target beneficiaries of a social program, and so on. That purpose must be stated fairly: for example, one must not announce a benefit that could only be, at most, an indirect consequence of the study. Above all, the surveys must win the informed consent of the persons concerned, when the topic is sensitive, that is, when it may reveal membership in a vulnerable or stigmatized minority, or offend the sensitivity of interviewees by the very fact that the question is asked. "Informed consent" means written and duly verifiable consent. If all these conditions are met, it becomes possible to ask intimate questions concerning—as it were—private life within private life. That said, each question must be proportionate to the questionnaire's overall purpose. This means, by the way, that the purpose must be clearly stated to respondents: the questionnaire plan must not throw them off balance in the process.

That seems obvious, Macrodemus.

Not all that obvious, said Microdemus. You'll find questionnaire theorists who, on the contrary, recommend jumping from one subject to another during the interview, on the assumption that the answers will be more spontaneous and more authentic.

Indeed, Macrodemus recalled, I had a visit one day from an interviewer working for a private firm, who questioned me in the manner you mentioned about my tastes and attitudes in a wide variety of areas. It was heavy going.

Actually, it's an entirely different conception of surveys, consistent with the psychologists' vision. The point is not to go and see people to gather the sociodemographic information they have, but to design the survey situation as a revelation experiment in itself: the interviewer isn't trying to collect your information but to trigger and analyze your reactions. The ethics of the questionnaire as we practice it in sociodemography generally excludes this type of experimental relationship. We don't see the respondent as the subject of an experiment: he's cooperating with us to produce missing information. He's almost a coproducer of data, so, even if it's technically difficult, the least we can do is to give him a summary of the results at the end of the survey.

It's an approach that suits me fine, Macrodemus concurred. I have a better understanding, now, of all the preparation and pedagogy that this type of survey requires. On the other hand, I'm more skeptical about the subjective questions that are increasingly common in these questionnaires.

You mustn't confuse them with the general opinion questions scattered throughout polling-company questionnaires such as "What do you think of School in general? of the Justice system? of the Police?" and so on. What respondents can describe accurately are their reactions to an event or an institution that we've just discussed with them (a recent divorce, their kid's school, the healthcare facility they've used). It's interesting to compare properly contextualized judgments and assessments with the corresponding objective practices. On this score, sociologists have learned a lot from American ethnomethodologists. They now believe respondents are not "cultural dopes," but are capable of describing what they do, and the survey designers must make intelligent use of this accountability.

What do you think of surveys on values, taken as explanatory factors of demographic behavior?

It's problematic, Microdemus admitted. Typically, these are batteries of fairly standardized questions. Some are even patented and cannot be reproduced without paying royalties to their inventor. Their purpose is to reveal people's deep-rooted attitudes, arranged in classic antagonistic pairs such as progressive/reactionary, tolerant/closed, individualist/collective, hedonist/ascetic, idealist/materialist, and so forth. In a way, it's a return to the system we were discussing before, that is, the transformation of questions into small experimental devices for gathering subjects' live reactions. I find the explanatory power of these typologies limited or relatively tautological: Why do some people expect morals to change? Because they're more open to moral change. Why do young couples want fewer children? Because they have a mindset that drives them to have fewer children. The notions of materialism, individualism, or hedonism seem too general and very moralistic to me.

Gérard Calot, a top expert in macrodemography if ever there was one, freely resorted to this type of explanation to account for the evolution of demographic behaviors.

True, said Microdemus, but he was aware of the weak scientific status of this type of explanation. When he had to write a scientific article about the decline in fertility, he simply made a cautious allusion to it. In fact, when you jump off the platform of demographic technique with your two feet and you have no sociological baggage whatever, you tend to use such spontaneous explanations as a parachute. If you believe in the "decline of fertility," you're sure to explain it by the decline in moral values, if not of civilization. Adolphe Landry was already complaining about this in *La Révolution démographique* back in 1934. And he painted a very grim picture of moral decay. Pornography, he assured, had already invaded the smallest French villages (he was targeting the movie industry and the image industry). What would he say today?

Are you telling me, Macrodemus asked, that you're against putting questions on values in a questionnaire on demographic behaviors?

No, Microdemus replied. If I had to choose, I'd prefer introducing the issue in the questionnaire itself, so I could check its relevance scientifically, rather than introducing it surreptitiously into the commentary on the data, as an uncontrolled heuristic resource. Like any sociologist, I'm reasonably Weberian. I think the special problem in social science is the need to study actors' values without indulging in value judgments, to take their moral standards seriously without lecturing them. Values do indeed exist, in more or less crystallized forms. They can function as intermediate drivers of action—as vectors or obstacles. I just think you should always inquire into their origins and, most important, never turn them into a deus ex machina that explains everything without ever being explainable. One of the ethical requirements of social science, in my view, is to abstain from judging actors' behavior by one's own moral values. There's nothing more insufferable than a researcher posing as an altruistic ascetic to disqualify someone else's hedonism and selfishness. That's how old folks have been judging the young since earliest antiquity, and how the rich judge the poor. An adult demography should shun this type of bias altogether, which would make it fall back on the glibness of common sense.

I don't feel targeted by such diatribes, Macrodemus mumbled. Personally, I prefer to explain the demographic by the demographic, as Durkheim explained the sociological by the sociological. All the rest is fluff.

You're wrong, countered Microdemus. It's not because the explanation of behaviors is subject to pressure from spontaneous and uncontrollable psychological theories that you should fall back on sex and age and never step outside the frame. I wholeheartedly applaud your indignation at a researcher who neglects structural effects, selection bias, or calendar effects. Identifying them is a vital prerequisite. But it's never more than a prerequisite. The notion that you need to "explain the demographic by the demographic" proceeded from a strategy to autonomize the discipline in the 1960s and 1970s. A similar phenomenon occurred in other social sciences: Durkheimian sociology, of course, but also pure Walrasian economics and structural anthropology. It belonged to a necessary moment in the development of demography. Today, the notion of pure demography has not vanished but it fits into a broader vision of social phenomena, where the groundwork for demographic analysis gives way to a wider social analysis. Concretely, this means that at both ends of the sequence running from questionnaire design to statistical processing, demographic behaviors are now incorporated into a complex of socioeconomic and sociocultural factors that put individuals back into their context. Demographic analysis is no longer separated from statistical modeling in general. It does make its specific contribution: a privileged attention to the calendar of events, to interference between demographic events, and to the interaction between age effects, period effects, and cohort effects. However, it does not cut itself off from statistical models with a wider descriptive and explanatory intent, such as duration models, multilevel models, and ceteris paribus regressions.

And, Microdemus insisted, that's where the Ethicists call us to account. With each successive survey project, they clearly see that most demographers now work on individual data, produced from surveys, where events and durations are studied in the light of social, educational, economic, health or cultural information—and, ever more often, combined with information on knowledge skills, preferences, intentions, motives, values, and habits. It's demography seen from below that now complements demography seen from above.

Microdemus drew an analogy with the work of historians. In their recent study of the immense historiographical output on World War I, Antoine Prost and Jay Winter (2004) observed that the history of the war long gave precedence to the viewpoint of diplomats and officers, while soldiers and civilians remained invisible. Historians studied "war without warriors." It was only in 1959, 40 years after the conflict, that the perspective began to change: under pressure from amateur historians, who were veterans themselves, professional historians have finally trained their spotlights on the men who had fought the war and the populations that had been subjected to it. Generals and diplomats have given way to soldiers and the populace, "the war seen from above' has been replaced by "the war seen from below" (Prost and Winter, 2004, p. 27–33). *Mutatis mutandis*, demography has followed a similar path—and so, incidentally, has a large sector of sociology. The changes in behavior have forced researchers to modify their way of looking at things and their perspective.

To support his claim, Microdemus cited the results of the Demography 2000 international online survey, conducted between October 1999 and 2000 among 637 demographers and population-studies researchers (Chasteland *et al.*, 2004, p. 142). Respondents were asked to name the main factors of change in demography from a long list provided. Their top three picks accounted for 48% of total responses: "development of information technology," "use of new categories from other disciplines," and "techniques borrowed from other disciplines." This was a clear acknowledgment of the force of external pressures on demography. More than other social sciences, population studies now combine very diverse techniques: collection of statistical data, collection of individual data, and the event-history or longitudinal approach. These methods were long separate. Social scientists who handled aggregate statistical data seldom practiced the intensive methods dear to sociologists of the family or anthropologists, and vice versa. On the one hand, you had large numbers, the standard questionnaire, the cross-section, and anonymity; on the other, interknowledge within a community, the qualitative interview, long-term tracking, and the personalized relationship based on trust. The two worlds are now joined. Individual data can now be mass-processed; lifelines can now be tracked for years, even for an entire lifetime (through long-term cohorts or administrative registers).

An ever-richer content, ever more intensive observations, ever longer tracking: these changes inevitably have ethical implications. Until now, demographers focused on just a few traces of our existence: a snapshot in the population censuses, and a few milestones in civil-registration documents. They can now capture entire lives in their webs. At the scientific level, this move toward multiple competencies and comprehensiveness must be encouraged as long as it produces knowledge, but it clearly adds to the researcher's moral obligations: the more information he or she extracts from respondents, the more explanations and guarantees he or she must give them.

In fact, Microdemus continued, the researcher's interest coincides here with the individual's interest. As you realize, the most grandiose observation systems remain at the mercy of popular consent. If they are targeted by a media campaign, they will be discredited.

I know something about that, Macrodemus acquiesced. There have been no census data for the Netherlands since 1971 or for Germany since 1987. I also learned that these antienumeration campaigns were aimed not so much against the census itself as against the plan to match it with population registers.[10]

It's not just the census, Microdemus added. In France, the population census is well accepted, despite the increasing difficulty for enumerators to enter buildings. The systematic recording of social-

[10] We cannot cite the suspension of the German and Dutch censuses to justify the expansion of population registers (argument sometimes used in France), for it is precisely the potential enhancement of the registers through the census that has caused problems. In the Netherlands, the first significant protests were triggered by the 1971 census, in university towns; they pointed to the danger of collating the census and population registers. In the preparatory tests for the 1981 census, the authorities observed such high rejection rates that they decided to call off operations. In West Germany, the 1983 census faced identical opposition for the same reasons: the protests were sparked by the automatic transmission to town halls of a page from the census schedule for the purpose of updating the population registers. The Constitutional Court ordered a halt in census preparation until passage of a law to protect personal data, which took place in 1985. The government authorized a new census in 1987, accompanied by a large-scale advertising campaign, but the Green Party called for a boycott, widely followed in the "alternative" neighborhoods of Berlin and university towns. Since then, the Länder and the federal government have failed to reach agreement to finance a new census.

insurance contributors (the notorious "Safari" plan) had to be halted in 1976 after a press campaign. Do you remember? They set up the Tricot commission, after which Parliament passed the Information Technology and Civil Liberties Act in 1978, establishing the like-named Commission. What strikes me, Macrodemus concluded, is the vulnerability of the databases used by demographers. Whether it's the census or administrative files, these data are as solid yet as vulnerable as a deposit bank. If you build a trusting relationship with the customer, they grow bigger and more beautiful. But if the public withdraws its trust, you get instant panic: the system collapses, and for a long time. Hence the usefulness of High Authorities for protecting the data: they remind us that censuses, administrative registers, and continuous surveys are arrangements requiring specific review and protective action. Personally, I've always argued that these watchdog bodies in each country should rely on the cooperation of statisticians and demographers, and should help, in exchange, to protect their work from frivolous claims.

But how do these high authorities, Macrodemus asked, follow the changes in morals and questionnaires? Do they easily accept the new survey style now prevailing in demography, which seems to be driving the researcher's curiosity ever further?

I won't conceal from you, Microdemus responded, that they have trouble following. The example of the problems encountered by the Family Survey associated with the 1999 French census is illuminating (Héran, 2005). The survey nearly went under: the High Authority that reviewed the questionnaires in 1994 didn't understand why a demographer would be interested in cohabiting unions of less than six months (these were necessarily "affairs"), in "living apart together" couples (they were not real couples), in the complete calendar of unions and separations (that stirred family secrets), and in adopted children (more secrets to protect). The survey managers had to negotiate some wordings and search the jurisprudence themselves to demonstrate the legitimacy of these questions. Now some far more formidable issues are being raised by the new generations of surveys.

Such as?

I'll only mention a few. How do you obtain the informed consent of "statisticized" people when part of the information is gathered unbeknown to them by cross-tabulation with administrative files? How do you guarantee anonymity if the longitudinal survey requires you to store personal data for long periods—up to an entire lifetime? Is it enough to deliver general information when the source files are opened, or do you need specific information for each new series of connections? How do you obey the neutrality imperative if the extension of the observation period tends to modify the behavior observed? Can you avoid intruding in private lives if the questions end up covering all aspects of life? Should respondents be told in advance how long the operations are expected to last? As you see, a lot remains to be done to settle these issues without jeopardizing the interests of science or the interests of individuals. One approach under consideration (which Benoît Riandey is promoting in France) is to develop encryption systems that segment the identification numbers. They allow citizens filling out administrative forms or survey respondents to enter sensitive information in one section of the questionnaire; that information will be inaccessible to the administrative body but can be decrypted by a properly authorized third party.

VIII. RESEARCH ETHICS: AN EPIDEMIOLOGICAL MODEL ILL-SUITED TO DEMOGRAPHY

In the hope of finding out more, Macrodemus and Microdemus surfed the Web looking for documentation that would give them a more practical answer to their questions. In their search, they discovered a variety of ethics more targeted than the one practiced by the Ethicists—which, in principle, should have been more relevant to them. Its name was "research ethics," a booming specialty in English-speaking colleges and universities. Did it offer guidance on the ethics of demographic research?

Macrodemus and Microdemus soon realized that research ethics was strongly anchored in the life sciences. It was only just beginning to spread to the behavioral sciences, in particular the disciplines that conducted experiments on humans, or that subjected them to intensive, long-term monitoring, such experimental psychology, cognitive psychology, and anthropology. The sciences that resorted to demographic and social statistics were virtually absent from this group, because they did not conduct real-life experiments but only thought experiments.

Microdemus paid a visit to the Ethicists' main bookstore. He found the research ethics guide written by the director of bioethics programs at the University of Montreal (Doucet, 2002). It was a remarkable summary of the regulations and legal arrangements successively adopted in Europe, the United States, and Canada. The guide pointed out in passing that the founding texts suffered from several imprecisions or terminological confusions, which ultimately revolved around the legitimate definition of science:

- Often, in international ethics texts, the term *research*' is used for "experimental research" and the expression *research subject* is simply a shortcut for "subject hired for experimental research purposes." This is misleading, as experimentation does not designate all research but only one of a number of research methods. The social sciences claim the right to use other methods. Macrodemus was happy to note that the Montreal specialist began by citing statistical modeling for descriptive and explanatory ends, and regarded it as a full-fledged scientific approach (practiced even by hard sciences such as astronomy). He also mentioned qualitative observation methods, of which Macrodemus was not very fond: these sought to reconstruct and understand the logic of behaviors using various techniques such as detached observation, participant observation, directive interviews, comprehensive interviews, focus groups, and the analysis of documents produced by actors and institutions. From one method to another, the researcher's relationships with human subjects are extremely variable. However, if the relationship is properly defined and controlled, the qualitative approach need not be experimental in order to qualify as scientific.
- Some ethical regulations do not merely reduce research in the general sense to "experimental research" but go further by equating the latter with "intervention-based clinical research," which covers pharmaceutical trials (as well as clinical trials of medical devices and surgical trials). The purpose of such tests is to measure the pharmacological and clinical effects of a product and establish at what level of effectiveness and safety it is absorbed by the subjects in the experiment. This marks the maximum degree of intervention by research on the human body. But there are experimental research studies that, so to speak, act not on the body but on the mind. This is true of experimental psychology, which conducts experiments in laboratories and natural environments. It is also true of studies involving the administration of tests or exams—a standard practice in education science and sociolinguistics. More broadly, any administration of a questionnaire to a sample of subjects may be qualified as experimental, since the same person is invited to respond to a series of questions (or give multiple-choice answers) and, reciprocally, a series of persons is invited to respond to the same question.
- Conversely, intervention on the human body is not necessarily synonymous with scientific intervention. There are forms of innovative care that use untested methods, whose main purpose is not the advancement of science but a therapeutic procedure of last resort (often, compassionate care). Such procedures are referred to as "experimental treatment," insofar as it involves a "procedure that is performed for the direct and immediate benefit of the recipient, but that, in scientific terms, has not yet fully proved its worth" (Law Reform Commission of Canada, 1989, quoted by Doucet, 2002, p. 148). Now it is not always easy to distinguish intervention-based clinical research from such innovative care not validated by science. The confusion is all the greater as clinical research is itself generally conducted on patients undergoing treatment and has not yet obtained the scientific validation that normally follows the experiment. To draw the sharpest possible distinction between the two types of intervention, clinical trials are no longer qualified as therapeutic research in North America because the expression may lead patients to believe that the research will have an immediate therapeutic effect on them. The term *clinical research* is now used instead of *therapeutic research*: the scientific connotation therefore outweighs the medical connotation, but the degree of intervention on the body may be of equal scope in both cases.

Microdemus had to read the texts twice to grasp the meaning of all these distinctions, which he tried to explain to Macrodemus. He understood the basic point: research, scientific research, experimental research, and intervention-based research are realities that communicate but do not blend. Research ethics must be careful to clear up these terminological confusions, which mirror confusions in the areas covered. The ethical principles that apply in one case are not necessarily valid in the other.

Likewise, Microdemus noticed that ethical guidelines were structured according to dichotomies ill-suited to demographic research: research was described as quantitative or qualitative, with human subjects or without human subjects, clinical or explanatory, with or without therapeutic (or diagnostic or medical) purposes, and so on. These distinctions were combined with a distribution by field, which, for example, allowed Hubert Doucet to devote the chapters of his textbook to four categories of research: intervention-based clinical, epidemiological, genetic, and qualitative (Doucet, 2002). But, Macrodemus wondered, where does demography fit into these typologies?

Into non-medical research, of course, but (to use the official Canadian phrase) was demographic research conducted with human subjects or without human subjects? Macrodemus had never really thought about it. He had the impression he was working on human subjects and at the same time without them. Admittedly, at a certain point, the subjects had had to participate in order to supply the basic information; the government had had to ask them to fill out census schedules, population registers, civil registration forms, or survey questionnaires, unless the declarations had been filed by friends or relatives, registrars under oath, or doctors. Operating at the end of the chain, Macrodemus also realized that individual data sometimes had gaps that needed to be filled by editing or imputation methods. But he preferred to leave this tinkering to the specialists at the national statistical offices. All that a macrodemographer asks of a statistical institute is the timely delivery of a clean file that meets the requirements of accounting consistency. Little does it matter to him to know the efforts and compromises needed to extract the data from the population.

On occasion, when he had to end a set speech with a lyrical flourish or answer a radio-show caller who insisted on describing his personal case, Macrodemus held in reserve a phrase that he would utter in an earnest tone: "Let's never forget that behind the indicators, you have people." This dash of humanism was always welcome. It was an inexpensive way to counter the unfortunate impression given by the demographer's impersonal treatment of the data, with decimals of children and mean life expectancies. But Macrodemus no longer dared use his fine phrase after Microdemus politely noted that it had no connection to any professional practice. For the true precept informing Macrodemus's practice was none other than Aristotle's old adage: "There is no science except of general things." A private individual could never be more than an anonymous purveyor of data for him. To study the individual as such was to run the risk of sinking into a bottomless pit, as we can clearly see from Jean-Paul Sartre's unfinished and desperate attempt in *L'Idiot de la famille* (The Family Idiot): the singular human subject (Gustave Flaubert, in this case) is infinite, and thousands of pages are not enough to describe him (Sartre, 1970–1971). To turn the human subject into an object of science, one must reduce him, and, ultimately, reduce him to insignificance. Microdemus found Macrodemus better inspired when he forthrightly defended the advantages of reduction instead of indulging in humanism or personalism.

In fact, while demographic research always needs human help, even when working on highly aggregated data, we must admit that this reliance is more systemic than individual. It is channeled through largely institutional and depersonalized mechanisms. Initially, individuals know that the registration of their existence (or death) by the authorities incorporates them into a broader enumeration. By the end of the process, when the demographer steps in, there is no personal dimension left. We can see this clearly in the borderline example of demographic analyses of world population: each of the world's 6.4 billion inhabitants is included in the tabulation and counts as one, but there is no way they can give the researcher their consent whenever they are "statisticized." Where should we classify this human yet impersonal science, which, on the one hand, can take all human beings as its object, through a vast compilation of individual data, and, on the other, never intervenes on the body and never performs experiments on human subjects? Where should we put a discipline that works both with human subjects and without them? The textbooks on research ethics offered no answer.

Macrodemus and Microdemus tried to determine in which cell research ethics placed social-science research. They realized that it was wrongly classified under qualitative research, on the pretext that it was not biomedical research of an experimental nature. "Qualitative" was a euphemism for "nonquantitative" and "nonexperimental"—in other words, "nonscientific." In countries such as Canada and the United States, the ethics committees set up in each university hospital no longer simply examined biomedical projects; they had ended up examining social-science research projects. This extensionist approach had been implemented in Canada by the Tri-Council Policy Statement (1998) devoted to "Ethical Conduct for Research Involving Humans," the three councils being specialized in medical research, natural sciences and engineering, and social sciences and humanities respectively (Doucet, 2002). To cover the three research fields, the Canadian authorities prescribed the "proportionate approach to ethics assessment," that is, "the more invasive the research, the greater should be the care in assessing the research." If the risk for individuals was minimal, one could authorize expedited review (called "simplified" review in France).

At first, it was thought that projects in sociology and anthropology, owing to their very limited potential harmfulness, would benefit from these streamlined procedures. But they were often viewed with great suspicion by committees of experts trained exclusively in the experimental sciences. As Doucet notes, these experts' incomprehension of social science is not so much at the ethical level as at the scientific level. They are worried not about the immorality or danger of social-science research projects but about their methodological weakness. They fail to see how causality links can be proven without using the experimental method. They fear that, in these conditions, research will give free rein to an ideological or subjective analysis, especially since social science jargon can be worryingly obscure whereas it is supposed to describe realities known to all. Because of their scant familiarity with social science, experts in experimental science tend to assess sociodemographic projects in terms of their common sense or prejudices.

Microdemus was deeply shocked by the notion that an approach was irretrievably subjective if it did not implement the experimental method. In his view, demography provided proof of the contrary: a social science, it was neither qualitative nor experimental, yet a science all the same, at least as much as economics and social statistics. To define the proper ethical approach for it, shouldn't demography be placed somewhere between experimental research and qualitative research? But Microdemus noticed that the research-ethics manuals had installed another discipline in this intermediate space: epidemiology. Did this mean that the ethical approach for epidemiology could be extended to demography?

As noted earlier, Microdemus had received some training from Epidemus. He had a rough idea of the relationship between demography and epidemiology. They do have in common the fact of being focused more on populations than on the individual. But the resemblance stops there. Epidemiological studies can treat human subjects using methods as interventionist as those of clinical trials. In fact, epidemiology is not consistent on this count: it can perform observation studies as well as experimental studies:

- The observation studies conducted in epidemiology may have a purely descriptive intent (measuring the distribution and propagation of health statuses to steer the allocation of healthcare expenditures) or, more ambitiously, they may seek to explain (by identifying behaviors or risk factors that play a decisive role in the etiology of illnesses). In ethical terms, these observation studies are interventionist or invasive when they involve taking biological samples from human subjects. The ethical problem becomes particularly sensitive if the samples consist of genetic material: the information gathered concerns not only the individual's identity and origins but also those of his blood relatives, who become, de facto, complementary subjects of the study.
- As for the epidemiologists' experimental studies, they are interventionist in the full sense of the term: they rely on the same principle as clinical trials. The difference is that causality is measured by comparing groups of subjects. The variations in health status induced by new products are not validated separately for each individual but established on a probabilistic basis at the group level. They must be statistically significant. For this purpose, the experimenter chooses the subjects on whom the medical procedure is performed and compares the changes in their status to those of the group that has not undergone the procedure. A crucial aspect of the method is that statistical causality must be measured by identifying insofar as possible the effects attributable to other factors at work.

Microdemus was tempted to place demography on one of these two branches. It could not be on the experimental studies branch. Should he conclude that demographic methods had more in common with those used in the epidemiologists' observation studies? On closer inspection, he saw that the latter could apply three types of protocol: cross-sectional studies, cohort studies, and case-control studies.

The only protocols found in identical form in demography were cross-sectional studies: they consisted in gathering the sociodemographic characteristics of a given population from all its members at the same time (these characteristics are either explanatory factors or variables to be checked), as well as information on health status, individual behaviors, and environmental data (interpretable as risk factors). By contrast, demographers did not conduct comparative cohort studies. Epidemiologists use them to detect changes in health status (or even deaths) associated with a risk factor. For this purpose, they conduct a long-term observation of two groups of subjects, one exposed to the risk factor, the other not. These studies are generally performed on small samples of patients treated in a hospital ward. The demographer preferred nationally representative samples, with as large a set of potential explanatory variables as possible. In this respect, he was close to the sociologist or social statistician.

This plunge into research ethics had greatly excited Microdemus, who had tried to interest Macrodemus in it—but in vain. Macrodemus had not examined a science other than his own in a long time. He let Microdemus conclude. No, the demographer's methods were not congruent with those of the epidemiologist. They were not interventionist, nor did they give precedence to groups of patients or subpopulations exposed to risk factors. They could not be subjected to the same ethical requirements, except for cross-sectional surveys.

But what were those ethical requirements? They were particularly strong for intervention-based clinical research and, from there, had been extended to other epidemiological studies. The Nuremberg tribunal, which had judged Nazi criminal doctors after World War II, had defined the main guidelines (known as the

Nuremberg Code), later broadened by the Helsinki declaration and various international resolutions (Box 99–2). These principles were stated as follows. No experiment on human subjects can be performed without their informed consent. They must not be coerced; in particular, they should be free to leave the scientific experiment at any time. Vulnerable individuals—children, the elderly, the mentally ill, the handicapped, the homeless, immigrants, members of disadvantaged categories—needed special protection: they could not be asked to participate in a medical experiment unless they were the intended privileged beneficiaries and unless the experiment could not be conducted first on other categories of individuals. The three principles were, therefore, securing informed consent, respecting the subjects' autonomy, and protecting vulnerable minorities.

The ethics committees' extensionist approach consists in applying these clinical-trial guidelines to epidemiological studies in general, including the least intervention-based. The main adjustment involves taking into account the collective or "populational" dimension of epidemiological studies. The informed consent of groups (via their legitimate and recognized representatives) is obtained in the same way as individual consent; the autonomy of groups (or communities, to use the standard North American term) is respected in the same way as personal autonomy. The groups likely to benefit most from the study are made to carry the bulk of the survey burden; reciprocally, steps are taken to ensure that the vulnerable groups are not burdened with studies that do not specifically concern them and in which they would only serve as guinea pigs.

This extensionist approach may seem exaggerated. Its explanation lies in the fact that some epidemiological studies are effectively performed on the bodies of subjects or patients, with procedures resembling those

BOX 99–2 Morals, ethics, deontology: semantic drift

The word *ethics* traces its roots to late antiquity; it derives from the Greek è*thos* (morals). In linguistic terms, *ethical* is the Greek counterpart to the word of Latin origin *moral*—its "didactic and literary equivalent," says the French dictionary *Le Robert historique de la langue française* in its discussion of the French forms *éthique* and *morale*. Ethics, therefore, is nothing other than morals, and professional ethics is the moral code of a profession. Academic English has given precedence to the term *ethics* and its derivatives, later imported as such into France (where this Anglicism revived the outmoded usage of *éthique*). In France, there is a perceptible difference in register between "ethics" (*éthique*) and "morals" (*morale*): the first has a more formal connotation, referring to the set of principles that enables us to decide between good and bad conduct; by contrast, *morale* tends to be confined to more informal usage, the assumption being that morality is a routine discourse or a discourse from authority enunciated in the first degree, without a clear knowledge of the underlying principles. Hence the expression *faire la morale*—to give someone a moral lesson.

To denote the corpus of moral rules that a profession defines for its internal use, English-speaking countries use the phrase *professional ethics*; Romance-language countries tend to use *déontologie* (deontology) or its cognates. The vocabulary varies according to whether the national language is Romance or Germanic: *code de déontologie* in French, *codice di deontologia* in Italian, *código de deontología* in Spanish. Deontology, in this sense, is simply the systematic exposition (*logos*) of duties (*deonta*): a compound that did not exist in classical Greek, since it was coined by Jeremy Bentham in ca. 1820. Bentham already mentioned it as a known term but used in a special meaning that persists to this day among English-speaking moral philosophers: they apply the qualifier *deontological* to theories that, like Kant's philosophy, make the unconditional fulfillment of duty the supreme criterion of good action; this is in contrast to consequentialist theories, which judge the moral value of action by its consequences on general welfare (the best-known consequentialist theory being utilitarianism).

In the English-speaking countries, the equivalent of "deontological codes" or "charters" is designated by a variety of roughly synonymous expressions: code of ethics, code of professional ethics, code of ethical practice, code of practice, code of conduct, code of professional conduct, code of good professional conduct. French often replaces code (*code*) with charter (*charte*), to denote that this is not a set of laws but a document freely adopted by an institution or profession. More recently, the expression *livre de style* has made its way into the Francophone press. This is a literal translation of the United States style book, initially a compendium of typographic rules, in which newspapers now gather not only their rules of writing and conduct, but all their ethical rules.

of clinical trials. But this is not always so, far from it. One must therefore attenuate the extensionist assessment approach by requiring ethics committees to abide by the principle of "proportionate approach to ethics assessment" set out in the Canadian Tri-Council Policy Statement.

But could one really curb the Ethicists' propensity to extend their powers of review to all sorts of research? Microdemus was openly skeptical. The Ethicists were not known to give up easily. What he heard from Canadian and U.S. epidemiologists seemed to indicate, on the contrary, that the tendency to generalize a legal or ethical procedure beyond its initial scope of application was irrepressible. If abuses of power had already been committed in extending ethical rules from the clinical to the epidemiological fields, there was reason to fear the effects of an additional extensionist thrust, which would result in demographic studies being placed under the jurisdiction of ethics committees specializing in epidemiology.

Clearly, Macrodemus and Microdemus concluded, demography did not have a place in the divisions drawn by the research Ethicists. The unsuitability of epidemiological standards was blatant as regards the procedures for obtaining informed consent. If a control model imitating clinical trials were applied to sociodemographic research, the obtaining of consent would become the default standard, and a research project would not escape the requirement except by special waiver, if it proved its harmlessness for human subjects. In demographic studies, the opposite prevails: the subject's consent is taken for granted, and is sought only in exceptional cases (when the data gathered are sensitive or too localized, or when a longitudinal survey makes it necessary to store names and addresses). It is generally assumed that questionnaire respondents have agreed to reply. It is also assumed that by declaring their personal data to civil registration officers, individuals accept once and for all that their lives are amalgamated along with those of thousands of others into anonymous overall statistics (in some countries, the civil registration forms advise people of this statistical use of their data). It would be absurd to ask demographers to obtain the individual consent of all the subjects taken into account whenever they calculate birth and death rates. Contrary to epidemiological studies, obtaining explicit consent is therefore not the rule but the exception in standard demographic research.

Macrodemus and Microdemus had no trouble concluding. Because it is centered on biomedical studies, research ethics has serious difficulties in specifically addressing the issues raised by demographic studies. It tends to draw them either toward the camp of qualitative methods, as practiced by social anthropology, or toward the camp of epidemiology, which in turn is drawn toward clinical trials. In both cases, research ethics misses its target. Another approach was therefore needed. But Macrodemus had no time for this: he had just received superb macrodata that required immediate processing.

IX. STRENGTHS AND WEAKNESSES OF ETHICS CODES

Time was going by. Like Saint Simeon on his pillar, Macrodemus was riveted to his suspended workshop. Too busy with his survey, Microdemus had spaced his visits but their e-mail exchanges had continued.

One day, while searching the Web for a reference on data quality, Macrodemus ended up by mistake on the site of an international statistical association, which published a code of professional ethics. The document referred to other guidelines and, by following the thread, one could peruse most of the codes of professional ethics promulgated in the past half-century by national and international associations of statisticians. The latter had been followed by associations of sociologists and anthropologists.

At first glance, this internal regulation literature had seemed more concrete to him than the Ethicists' speeches. He alerted Microdemus, who voiced a great interest in the discovery. Microdemus broached the topic with the Ethicists he knew, starting with Philarethes. What did she think of the ethics literature? Oddly enough, she didn't think much of it, despite her passing mention of it in the lecture/debate at the Grand Theater. The Ethicists barely cited it in their works. Professional ethics was just a narrow segment of applied ethics. Rarely did it discuss the deadly dilemmas that the Ethicists had made into one of their specialties. Moreover, the rise of professional ethics had begun long ago, as far back as the Hippocratic oath. There had been a rebirth in late 19th-century America—and, in between, a prophetic remark by Chancellor Francis Bacon in the early 17th century.[11]

[11] As early as 1605, in his astonishing treatise on the organization of science, Bacon explained that "duty," defined as the good rendered to society, comprises two parts: duties toward the State and the duties of professions and vocations. The theory of these duties should include a section on the "frauds, cautels, impostures, and vices of every profession." Instead of leaving the subject to satirists, we should make a serious and realistic study of it, for we cannot defend virtue without a good knowledge of evil: "we are much beholden to Machiavel and others, that write what men do, and not what they ought to do" (Bacon [1605] 2000, p. 216). This realistic program of professional ethics was promising, but Bacon never pursued it.

The phenomenon had not awaited the recent revival of moral and political philosophy, or the proliferation of ethics chairs in universities. For some Ethicists, this self-regulation enabled the professions enjoying exorbitant power—such as doctors, journalists, police, and insurers—to do two things: first, to protect themselves against customers increasingly worried about their abuse of power; second, to finally conduct an external communication campaign, albeit of little ethical value.

For Macrodemus and Microdemus, the Ethicists' disdain, if anything, spoke well for professional codes of ethics. Our two demographers didn't like radical speeches on good and evil. The modesty of the professional-ethics enterprise was a good sign for them. They were interested to discover a lecture by Jean-Marie Delarue, a jurist who had extensively frequented statisticians[12] (Delarue, 1998). It put professional ethics in perspective, as one of the many forms of regulation, alongside law, contract obligations, and private regulations. If I break the law, I expose myself to punishment by public authority. If I breach a contract unilaterally, I will be sanctioned on the initiative of the other party. If I ignore the regulations that surround me (internal regulations in businesses, schools, hospitals, condominiums, nonprofit groups, supermarkets, concert halls, sports facilities, Internet forums, etc.), I will be called to order, or even excluded, by the people—generally private individuals—tasked with preserving order. But if I infringe the code of ethics of the profession, I expose myself primarily to censure by my peers. This can range from a simple reprimand to disciplinary proceedings, and my peers can apply sanctions on their own initiative or in response to a user's complaint. The code of ethics is therefore a new source of normativity whose special feature is that it is basically internal to an organized group; more accurately, it provides internal prevention of possible external challenges. René Padieu notes, however, that the different sources of normativity communicate: the ethics codes prepared by professionals may precede and inspire laws.

Macrodemus and Microdemus started to read the codes of statistical ethics available (Box 99–3). They fell into two categories.

Some (such as the code for *administrateurs* [senior statisticians] of the French National Institute of Statistics) covered the entire process of a statistical operation, stating the guidelines to be followed at each step: how to define the initial framework, construct the data, ensure representativeness, administer proof, verify that the results are significant and of general import, prevent interpretation errors, inform readers of the implications and limits of the published figures, recognize one's mistakes and, lastly, make it possible to conduct a scientific critique.

[12] A member of the Council of State, Delarue was the former chairman of France's National Council for Statistical Information, the body in charge of vetting all questionnaires in official statistics.

But the codes of international associations followed another plan: they went through the long list of the statistician's partners—subjects interviewed, colleagues, employers, sponsors, and the general public—and stated the duties to be fulfilled toward each. Here is a summary of the prescriptions that appear in most of these codes:

Obligations toward the discipline:

- promote advancement of research and dissemination of knowledge;
- defend freedom of research;
- refuse contractual provisions restricting freedom of research;
- promote research-friendly conditions;
- welcome young generations of researchers.

Obligations toward colleagues:

- fully acknowledge their contribution, cite them;
- read them;
- avoid all plagiarism;
- do not exploit other researchers (particularly young ones waiting to fill positions);
- provide training;
- avoid over-personalizing relationships, which could create dependency situations;
- avoid all discrimination (by sex, origin, age, sexual orientation, or religious, labor-union, or political affiliation, etc.);
- foster equal opportunity in hiring and promotion.

Obligations toward persons participating in the research (survey respondents):

- ensure that the persons' physical, social, and psychological welfare is not adversely affected by the survey work, or, if so, to the least possible extent;
- particularly for persons of precarious or vulnerable status due to age (children, the elderly), health (illness, handicaps), social marginality (prisoners, homeless, illegal migrants, etc.), hardship caused by a special event (separation, death of a close relative, abortion, domestic violence, etc.);
- work to minimize the disruptions caused by the survey operations; special attention should be given to the possibility of appropriate compensation (its relevance, type, and amount, bearing in mind that national traditions diverge on the issue);

BOX 99–3 Specialized organizations and codes of professional ethics

1. Selected organizations and committees specializing in professional ethics in the social sciences

American Statistical Association (ASA), Committee on Professional Ethics

Social Research Association (SRA), United Kingdom

2. Ethical guidelines and professional ethics in statistics and epidemiology

International Statistical Institute (ISI), 1985. *Declaration on professional ethics.* http://www.cbs.nl/isi/ethics.htm

United Nations, Statistics Commission, 1994. *Fundamental principles of official statistics*

International Council of Scientific Medical Organizations (CIOMS) and World Health Organization (WHO): *International guidelines for ethical review of epidemiological studies*, Geneva, 1991 (under revision, 2004).http://www.cioms.ch/frame_1991_texts_of_guidelines.htm

3. Ethical guidelines and professional ethics in social science

Social Research Association (SRA), 2003. *Ethical Guidelines*

American Statistical Association's Committee on Professional Ethics

UNESCO, 1994. *Ethical guidelines for international comparative social science research in the framework of M.O.S.T.* (Management of social transformation). Paris.

The Research Council of Norway, 1994. *Guidelines for research ethics in the social sciences, law and the humanities.* Oslo.

Institute for Employment Studies, 2004. —*Respect Project: An EU Code of ethics for socio-economic research.* Project funded by the European Commission's Information Society Programme (IST). Brighton. http://www.respectproject.org

4. International bioethical guidelines (chronological order)

Nuremberg Code, 1947, judgment by U.S. military tribunal against Nazi doctors (text regarded as starting point of ethical guidelines on experimentation).

English version: http://ohsr.od.nih.gov/guidelines/nuremberg.html

World Medical Association

Declaration of Helsinki, Recommendations guiding physicians in biomedical research involving human subjects, adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964 (text regularly amended in later Assemblies, latest ed. 2004).

Website: http://www.wma.net/e/policy/b3.htm

International Bioethics Committee (IBC)

Universal Declaration on the Human Genome and Human Rights, Paris, UNESCO, 1997. http://www.unesco.org/ethics.

Council of Europe

Convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine (Oviedo, 4 March 1997), European Treaty Series, no. 164, 1997.

Council for International Organizations of Medical Sciences (CIOMS) and World Health Organization (WHO): *Manila Declaration*, 1981. *Proposed international guidelines for biomedical research involving human subjects*, Geneva, 1982. Revised edition: *International ethical guidelines for biomedical research involving human subjects*, Geneva, 1993. Website: http://www.cioms.ch/draftguidelines

5. National bioethical guidelines

Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Washington, D.C. (U.S.):

Research involving prisoners: Report and recommendations, and Appendix, 1976.

Research involving children: Report and recommendations, and Appendix, 1977.

Research involving those institutionalized as mentally infirm: Report and recommendations, and Appendix, 1978.

The Belmont Report: Ethical principles and guidelines for research involving human subjects, 1978, Washington DC: U.S. Government Printing Office. http://ohsr.od.nih.gov/guidelines/belmont.html

Medical Research Council (MRC), Natural Sciences and Engineering Research Council (NSERC), Social Sciences and Humanities Research Council (SSHRC), Ottawa (Canada)

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 1998 (with 2000 and 2002 updates).

Website: http://www.pre.ethics.gc.ca/english/policystatement/policystatement.cfm

Comité Consultatif National d'éthique pour les sciences de la vie et de la santé (CCNE), Paris (France).

Publishers *Cahiers du CCNE*.

Website: http://www.ccne-ethique.org

- alleviate the survey burden (duration, mental burden, psychological burden, guilt); obtain the informed consent of persons, including in the exceptional cases where the survey work is compulsory for reasons of public interest. These should be defined by an independent, legal body, democratically mandated to decide whether the survey should be compulsory: problem of surveys conducted in prison at the request of the prison director; problem of surveys conducted in non-democratic countries with the visible support of the authorities, and so on;
- inform persons of the degree of anonymity and confidentiality, which must be as high as possible. In exceptional cases (longitudinal surveys requiring long-term storage of personal details), obtain their informed consent;
- inform persons of the purpose of the study;
- do not give them illusory hopes of positive, immediate effects;
- make sure that the use of an informant in the household is relevant, is approved by other household members (wife describing her husband's characteristics, or vice versa), and concerns objective facts only—not opinions or judgments;
- make sure that the use of a proxy (informant) has a legal basis (for example, guardian, parent, or licensed caregiver for a handicapped person) if the interested party cannot give his/her consent;
- provide them with every guarantee that the information gathered will not be supplied to subsequent respondents, to the authorities, or to third parties in general;
- take special precautions when questioning children, by obtaining both parental consent and permission to question them in private, without parental pressure;
- do not offend persons' ethical and religious sensitivities; if the survey arrangements conflict with special rules (e.g., men's refusal to have their daughters or spouses questioned individually by men), adjust the survey arrangements in consequence, without losing sight of the survey objectives and without condoning potential discriminatory behavior;
- abandon the survey if it proves impossible to conduct without infringing personal rights.

Obligations toward sponsors and employers

- clarify the contractual arrangements for research by specifying the nature of the relationship with the customer, reciprocal expectations, type of services to be provided, timetable of operations, financial arrangements, authorized use of results by sponsor, right of researchers to publish them freely and re-use them later;
- inform sponsors beforehand of inherent limitations of methods used, as well as ethical restrictions on data access or use;
- do not accept orders that would impose methods or even results consistent with the sponsors' wishes, or enable them to block or delay discoveries they did not like;
- do not seek to undercut competitors by accepting an order whose budget precluded compliance with basic scientific standards;
- mention support from orderers and sponsors in the research publications.

Obligations concerning data quality

- refuse participation in research contracts without prior determination that you will be able to use or produce data of sufficient quality to answer the questions raised (this implies bringing together all the required competencies in the research team);
- give full information on data-collection method, so other teams can criticize it later;
- be attentive to data selection and construction biases, particularly when the data have been compiled by organizations for nonresearch purposes;
- verify data quality even if the results seem plausible at first sight;
- do not engage in international comparisons without assessing the variations due to the diversity of definitions, collection methods, and processing methods among countries.

Reading the good advice contained in the codes of statistical ethics, Macrodemus thought that demography, as a branch of social statistics, could readily adopt some of them. His highly technicist view of the profession led him to prefer the first expository method, the one that tracked every step of a statistical operation. As noted earlier, the notion that the data had to be built and negotiated through a complex network of social relationships was rather alien to him. But, in his defense, we must admit that the good advice dispensed by the statisticians' codes of professional ethics were curiously diverse. Some were extremely general, while others went too far into the details.

The general precepts could elicit unanimous approval. Do not harm the people whose data you handle, combat fraud and deceit of every kind, be honest and modest in your attempts at interpretation, recognize your mistakes if need be, respect other people's work, give each his due, quote your sources, colleagues, and sponsors fairly—who could object? The problem was how to put all these precepts into practice. Should oversight procedures be set up? If so, at what step in the statistical process? Early on, at the research-project review stage? In a later phase, when editorial committees vet planned publications? Or periodically, when researchers come up for evaluation? Should ethical vigilance be distributed among the different players involved in data production, or should it be concentrated in the hands of a committee acting on behalf of the entire institution? Should violations be sanctioned?

The detailed recommendations set out in the codes posed the opposite problem. They described the "how," but starting from principles that were not unanimously accepted. Macrodemus was not convinced, for example, that a code of ethics should go into details about the methods for administering proof or measuring significance. These methods not only sparked disagreements between statisticians, epidemiologists, and demographers, but also were hotly debated within each of the three tribes. Adapting such codes to the needs of demographers required much fine-tuning.

Macrodemus and Microdemus continued their exchanges. By dint of seeing each other, they had evolved. They began to understand that you cannot examine the links between ethics and demography without scrutinizing the nature of the demographic discipline itself. The professional ethics of demography had to vary depending on whether demography was practiced in a macro or micro approach. But that did not justify splitting the discipline in two and assigning separate ethics to each. Ever more commonly, demography was practiced in a macro perspective enhanced with micro approaches. Long-term cohorts offered a good example. This shift entailed an increasingly delicate alternation between statistical individuals and real persons. To cope with this difficulty, demographers concerned about professional ethics had to build their toolboxes by combining elements of different origin. They had to respect the ethics of other disciplines whenever they took part in interdisciplinary projects, but that was not enough. Macrodemus and Microdemus set themselves the goal of developing an original synthesis that would address the entire spectrum of ethical issues raised by demographic research, whatever the degree of proximity to human subjects.

They rolled up their sleeves and went to work.

Acknowledgments

This chapter was translated from the French by Jonathan Mandelbaum.

The author would like to thank Benoît Riandey for his extremely valuable help.

References

Anderson Margo J. and Fienberg Stephen E., 2001. *Who counts? The politics of census-taking in contemporary America*, 2nd ed., updated. New York, Russell Sage Foundation, 357 p.

Audard Catherine, 1999. *Anthologie historique et critique de l'utilitarisme*, 3 vol. Paris, Presses universitaires de France, 1,024 p.

Bacon Francis, [1605] 2000. *Of the proficience and advancement of learning divine and humane*. London, Library Binding.

Bentham Jeremy, [1781] 1996. *An introduction to the principles of morals and legislation*. Edited by J. H. Burns and H. L. A. Hart. Oxford, Clarendon Press, 456 p.

Birnbacher Dieter, 1988. *Verantwortung für zukünftige Generationen*. Stuttgart, Reclam, 298 p.

Birnbacher Dieter, 1996. Éthique de la population et des générations futures, *in*: Monique Canto-Sperber (ed.), *Dictionnaire d'éthique et de philosophie morale*, p. 1169–1176. Paris, Presses universitaires de France.

Boltanski Luc and Thévenot Laurent, 1991. *De la justification. Leséconomies de la grandeur.* Paris, Gallimard (NRF essais series), 496 p.

Bonnet Doris (ed.), 2003. L'éthique médicale dans les pays en développement, *Autrepart*, no. 28, special issue, Éditions de l'Aube/Institut de recherche pour le développement, 192 p.

Browning Christopher R., 1994. *Ordinary men: Reserve Police Battalion 101 and the final solution in Poland.* New York, Harper Academic, 304 p.

Burke Edmund, [1790] 1980. *Reflections on the revolution in France.* London, J. Dodsley, 356 p.

Banto-Sperber Monique, ed., [1996] 2004. *Dictionnaire d'éthique et de philosophie morale*, 2 vol. Paris, Presses universitaires de France (Quadrige series), 2,199 p.

Chasteland Jean-Claude, Loriaux Michel, and Roussel Louis (eds.), 2004. *Démographie 2000: une enquête internationale par Internet auprès des démographes.* Louvain-la-Neuve, Belgium, Academia Bruylant, 368 p.

Chippaux Jean-Philippe, 2004. *Pratique des essais cliniques en Afrique.* Paris, IRD Éditions (Didactiques series), 318 p.

Courbage Youssef, 1999. *Nouveaux horizons démographiques en Méditerranée.* Paris, INED and Presses universitaires de France (Travaux et documents series, *Cahier* no. 142), 200 p.

Delarue Jean-Marie, 1998. *Pour une statistique éthique.* Paris, INSEE, 80 p. (Paper given at the conference "Statistiques sans conscience n'est que ruine . . . ," organized by the Institut National de la Statistique et des Études Économiques chapter of the CGT labor union, Paris, November 4, 1998. http://cgtinsee.free.fr/Kolok/kollok2

Demeny Paul, 2003. Population policy dilemmas in Europe at the dawn of the twenty-first century, *Population and development review*, vol. 29(1), p. 1–28.

Demeny Paul and McNicoll Geoffrey (eds.), 1998. *The Earthscan reader in population and development.* London, Earthscan Publications, 393 p.

Cench Sally, Iphofen Ron, and Huws Ursula, 2004. *An EU Code of ethics for socio-economic research.* Brighton (UK), Institute for Employment Studies, 96 p.

Desrosières Alain, 1993. *La politique des grands nombres: histoire de la raison statistique.* Paris, La Découverte, 441 p.

Desrosières Alain, 1998. *The politics of large numbers: a history of statistical reasoning.* Translated from the French by Camille Naish. Cambridge (Massachusetts), Harvard University Press, 384 p.

Doucet Hubert, 2002. *L'éthique de la recherche: guide pour le chercheur en sciences de la santé.* Montreal, Les Presses de l'Université de Montréal (Paramètres series), 267 p.

Duhamel André and Mouelhi Noureddine, 2001.*Éthique: histoire, politique, application*. Boucherville (Quebec), Gaëtan Morin, 306 p.

Durkheim Émile, 1992. *Professional ethics and civil morals*. New preface by Bryan S. Turner. London, Routledge Sociology Classics, 304 p.

Durkheim Émile, 1998. *Lettres à Marcel Mauss* [1893–1917]. Edited by Philippe Besnard and Marcel Fournier. Paris, Presses universitaires de France, 593 p.

Glover Jonathan, 1977. *Causing death and saving lives*. London, Penguin Books, 328 p.

Goldhagen Daniel J., 1996. *Hitler's willing executioners: ordinary Germans and the Holocaust*. New York, Alfred Knopf, 622 p.

Greenhalgh Susan, 2003. Science, modernity, and the making of China's one-child policy, *Population and Development Review*, vol. 29(2), June, p. 163–196.

Halévy Élie, [1901–1904] 1995. *La formation du radicalisme philosophique*, 3 vols. Paris, Presses Universitaires de France (Philosophie morale series), 1,168 p.

Héran François, 1986. La seconde nature de l'habitus: tradition philosophique et sens commun dans le langage sociologique, *Revue française de sociologie*, vol. 28(2), p. 385–416.

Héran François, 2000. Qu'est-ce que la démographie? Voyage historique et critique au pied des pyramides, *in*: Yves Michaud (ed.), *Université de tous les savoirs*, vol. 2: *Qu'est-ce que l'humain?* Paris, Odile Jacob, p. 171–187. (Reprinted *in*: *Université de tous les savoirs*, vol. 1: *La géographie et la démographie*, Paris, Poches Odile Jacob, p. 9–30).

Héran François, 2005. Du public, du privé et de l'intime dans l'enquête famille 1999: témoignage d'un concepteur sur une enquête qui revient de loin, *in*: Cécile Lefèvre and Alexandra Filhon (eds.), *Histoires de familles: l'enquête famille 1999*, p. xi–xxxviii. Paris, Institut National d'Études Démographiques (Les Cahiers de l'INED series, no. 642).

Hilberg Raul, 1985. La bureaucratie de la Solution finale, *in*: *L'Allemagne nazie et le génocide juif. Colloque de l'École des hautes études en sciences sociales* p. 219–235. Paris, Gallimard/Le Seuil (Hautes études series).

Isambert François-André, 1992. *De la religion à l'éthique*. Paris, Cerf, 432 p.

Jonas Jans, [1979] 1984. *Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation*, 2nd ed. Frankfurt am Main, Suhrkamp Taschenbuch, 425 p.

Kant Immanuel, [1785]. *Grundlegung zur Metaphysik der Sitten* (Foundations of the metaphysics of morals). Riga, Hartnoch.

Kant Immanuel, [1788] 1971. *Critique de la raison pratique*. Translated by François Picavet. Paris, Presses universitaires de France, 192 p.

La Follette Hugh, 2000. *The Blackwell guide to ethical theory*. Oxford, Blackwell Publishing, 446 p.

Landry Adolphe, [1934] 1982. *La Révolution démographique: études et essais sur les problèmes de la population*. Paris, Librairie du recueil Sirey, 227 p. (Reprint 1982, Paris, Institut National d'Études Démographiques, 231 p., Classiques de l'économie et de la population series).

Lenoir Rémi, 2003. *Généalogie de la morale familiale*. Paris, Seuil (Liber series), 587 p.

Lesthaeghe Ron (ed.), 2002. *Meaning and choice: value orientations and life course decisions*. The Hague/Brussels, NIDI CBGS Publications, 335 p.

Luhmann Niklas, 1989. *Vertrauen: ein Mechanismus der Reduktion sozialer Komplexität*, 3rd ed., rev. and expanded. Stuttgart, Enke, 119 p.

Mackie John L., [1977] 1990. *Ethics: inventing right and wrong*. London, Penguin Books, 249 p.

Marcus-Aurèlius, [ca. 170–180] 1964. *Pensées pour moi-même*. Suivies du *Manuel* d'Épictète. Translated, prefaced, and annotated by Mario Meunier). Paris, Garnier-Flammarion, 248 p.

Mill John Stuart, [1861] 1991. Utilitarianism, *in*: *On liberty, and other essays*, p. 129–201. Introduced and annotated by John Gray. Oxford, Oxford University Press (Oxford World's Classics).

Moore George Edward, [1903] 1959. *Principia ethica*. Cambridge, Cambridge University Press, 232 p.

Nicolet Claude, 1988. *L'Inventaire du monde: géographie et politique aux origines de l'Empire romain*. Paris, Fayard, 345 p.

Padieu René, 1991. La déontologie des statisticiens, *Sociétés Contemporaines*, vol. 7(Sept.), p. 35–62.

Padieu René, 1999. La déontologie statistique, *Journal de la Société Française de Statistique*, vol. 140(1), p. 5–21.

Pettit Philip, 1993. Consequentialism, *in*: Peter Singer (ed.), *A companion to ethics*, Malden (Massachusetts)/Oxford, Blackwell Publishing, p. 230–240.

Pojman Louis P., [1990] 2005. *Environmental ethics: readings in theory and application*, 4th ed. Belmont (California), Thomson-Wadsworth, 678 p.

Prost Antoine and Winter Jay, 2004. *Penser la Grande Guerre: un essai d'historiographie*. Paris, Seuil (Points—L'Histoire en débats series), 345 p.

Proust Marcel, 1931. *Remembrance of things past. Part 8: Time regained*. Translated from the French by Stephen Hudson. London, Chatto & Windus.

Proust Marcel, [1927] 1967. *À la recherche du temps perdu. Vol. VII: Le temps retrouvé*. Paris, Gallimard/Le Livre de Poche.

Rémond René, 1996. *Le 'Fichier Juif': rapport de la commission présidée par René Rémond au Premier ministre*. Paris, Plon, 233 p.

Rohrbasser Jean-Marc, 2001. *Dieu, l'ordre et le nombre: théologie physique et dénombrement au XVIII^e siècle*. Paris, Presses universitaires de France (Philosophies series, no. 144), 127 p.

Rosanvallon Pierre, 1992. *Le sacre du citoyen: histoire du suffrage universel en France*. Paris, Gallimard (Bibliothèque des histoires series), 504 p.

Rottmann Verena and Strohm Holger, 1986. *Was Sie gegen Mikrozensus und Volkszählung tun können* (What you can do against the microcensus and the population census), 2nd ed., rev. Frankfurt-am-Main, Zweitausendeins, 340 p.

Sartre Jean-Paul, 1981–1994. *The family idiot*, 5 vol. Translated from the French by Carol Cosman. Chicago and London, University of Chicago Press, 638 + 444 + 652 + 366 + 632 p.

Schutz Alfred, 1962. *Collected papers, vol. 1: The problem of social reality*. Edited by Maurice Natanson and H. L. van Breda. Dordrecht, Martinus Nijhoff, 320 p.

Seltzer William, 1998. Population statistics, the Holocaust, and the Nuremberg trials, *Population and Development Review*, vol. 24(3), Sept., p. 511–552.

Seltzer William and Anderson Margo, 2001. The dark side of numbers: the role of population data systems in human rights abuses, *Social Research*, vol. 68(2), p. 481–513.

Sen Amartya, 1987. *On ethics and economics*. Oxford, Basil Blackwell, 131 p.

Singer Peter (ed.), 1993. *A companion to ethics*. Malden (Massachusetts)/Oxford, Blackwell Publishing, 565 p.

Singer Peter, 2004. *The president of good and evil: the ethics of George W. Bush*. New York, Dutton, Penguin Group, 280 p.

Slote Michael, 2000. Virtue ethics, *in*: Hugh LaFollette (ed.), *The Blackwell guide to ethical theory*, p. 325–347. Oxford, Blackwell Publishing.

Süssmilch Johann Peter, [1741]. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen*. Berlin, Kultur-Verlag, 360 p.

SÜSSMILCH Johann Peter, [1741] 1998. *L'ordre divin dans les changements de l'espèce humaine, démontrépar la naissance, la mort et la propagation de celle-ci.* Translated and annotated by Jean-Marc Rohrbasser, Paris, Institut National d'Études Démographiques, 358 p.

SUMNER L. W., 2000. Rights, *in*: Hugh LaFollette (ed.), *The Blackwell guide to ethical theory*, p. 288–305. Oxford, Blackwell Publishing.

TEITELBAUM Michael and WINTER Jay, 1985. *The fear of population decline.* New York/Orlando, Academic Press, 201 p.

THOMPSON Mel, 2000. *Ethics.* London, Hodder Headline (Teach Yourself series), 201 p.

VALLIN Jacques, 1992. *La démographie*. Paris, La Découverte (Repères series, no. 105), 128 p.

C H A P T E R

100

History of Population Policies up to 1940

MICHAEL S. TEITELBAUM
Alfred P. Sloan Foundation, New York, United States

It is possible to interpret the title of this chapter in such a broad sense as to be utterly unmanageable. "History of population policies up to 1940" may be interpreted, especially in English, to mean all the ways in which, over the millennia from antiquity to the onset of World War II, human societies have debated and/or influenced the demographic forces that are changing them. Such a breathtaking scope could encompass essentially the whole of the history of present-day human societies. Each society in some sense contemplates whether it will survive, and embodies at the very core of its culture a set of institutional arrangements and belief systems that have a powerful impact on demographic phenomena, including religion(s), family structures, economic systems, political systems, and so forth. At this level of generality, a history of population policies becomes a global history of humanity.

In French usage, the term "politiques de population" normally has a more restricted meaning, limited to actions taken by the state (for discussion, see Prost, 2002, Chapter 13). For the purposes of this chapter, it therefore seems wise to limit the frame of our attention to those actions taken by states that are deliberate, that is, in which conscious decisions have been taken to create or modify societal arrangements with the expressed intention of affecting the rates of fertility, mortality, or migration, and thereby the trends of demographic growth, decline, or composition. The key ingredient then is volition and intent, or at least an awareness of the demographic impacts of political, economic, and cultural arrangements that are proposed or implemented. The requirement that such actions be taken by states means that population policies in this sense are to be found mostly since 1800, once the modern state had emerged and began to consider demographic characteristics as elements of military power or as sources of poverty or prosperity.

Even so constrained, the topic remains a sweeping one. Throughout human history, governments have taken actions to affect demographic change in ways that serve their interests. As Éric Vilquin noted in Chapter 97 of this volume, the political implications of this subject have been a topic of discussion by political philosophers, historians, and others during many periods of recorded human history, going back to the empires and city-states of antiquity. Moreover it has constantly engaged the attention of some of the intellectual giants of human history, from the Classical Greeks and Old Testament scribes of the Mediterranean to the classical economists of 19th-century Europe.

A common theme has been that larger populations imply greater power for the State, whether ruled by monarch, prince, emperor, chief, leader, president, or premier. Such perspectives are surely understandable, given that much of human history has been one in which societies struggled to survive by producing enough children to compensate for high mortality rates, and in which the number of soldiers or warriors commanded by the state had much to do with collective security and survival. Such a theme tended to dominate discussions of the subject in the ancient world, but also during the period of three centuries (16th–18th centuries) when mercantilist ideas dominated European perspectives. The 19th century saw sharp challenges to these assumptions that a bigger population is better, with Malthus and others arguing that unrestrained population growth spelled poverty and instability. But by the late 19th century into the 1960s, there was a return to the idea that population growth implied strength and prosperity. These ideas waned again due to concerns about the population explosion of the last half of the 20th century and the rise of environmentalist sensibilities, only to reappear in the 1990s as fertility rates fell to very low levels in much of the industrialized world.

While discussions about population policies have been part of much of history, the high point of such debates can be found during the two centuries from roughly 1800 to the present. Although there were strong ebbs and flows of concern during this period, both about whether population growth was too low or too high, these debates were not closely tied to the deep ideological rifts about political and economic policy that dominated this period. The 1930s and 1940s, for example, were decades dominated by fundamental (and ultimately catastrophic) conflicts among the leaders of fascist Germany and Italy, liberal democracies such as France, the United Kingdom, and the United States, and the authoritarian communism of Stalin's USSR. Yet the rhetoric and policy frameworks on population questions among this diverse array of leaders and countries tended to parallel one another in very significant ways. Moreover, within liberal democracies themselves, there was a surprising level of consensus among otherwise disputatious groups of political radicals, progressives, socialists, conservatives, and reactionaries.

I. ORIGINS: FROM THE ANCIENT WORLD TO MALTHUS AND MARX

1. Ancient Roots

Given space limitations, I can touch only lightly on discussions of population issues before the 19th century. The earliest recorded discussions on the subject, appearing in writings from ancient Greece, the Bible, and India, tend toward emphasizing the advantages of larger populations for military and economic power. The Old Testament includes several relevant passages, including the oft-quoted "Be fruitful" blessing in Genesis,[1] still cited today by religious pronatalists, and a precursor to mercantilist ideas appearing in Proverbs.[2]

Yet Genesis also points to the necessity of limits on numbers that can be supported by a given set of resources. Indeed, the historic separation of Lot and Abraham is attributed to such resource constraints.[3]

In ancient Greece, the historian Herodotus wrote as early as 440 BC of a conscious policy of pronatalism among the kings of Persia,[4] and Aristotle reported that the Spartan city-state under Lycurgus embraced explicit pronatalist laws such that ". . . the father of three sons shall be exempt from military service, and he who has four from all burdens of the state" (Aristotle, *Politics*, Book II, Chap. 9). Others noted that in classical Athens there were laws against celibacy and only fathers of legitimate children could appear as orators before the people or hold certain military commands (cited in Hutchinson, 1967, p. 9).

Yet Plato and other Greek philosophers saw the ideal state as consisting of a constant number of citizens numbering 5,040. The precise number was selected perhaps for mathematical reasons, but the general idea of a constant number of citizens on this small scale was likely seen as both sufficient for self-defense while small enough to maintain efficient governance and equal division of limited property. Aristotle argued that political leaders must

> . . . take care that the increase of the people should not exceed a certain number; and in determining that, to take into consideration those children who will die, also those women who will be barren; and to neglect this, as is done in several cities, is to bring certain poverty on the citizen, and poverty is the cause of sedition and evil. (Aristotle, *Politics*, Book II, Chap. 6)

In his comments on the pronatalist policies of Lycurgus in Sparta, Aristotle noted that it was "evident . . . that if the people increased there must many of them be very poor" (Aristotle, *Politics*, Book II, Chap. 9). In

[1] "God blessed them and said to them, 'Be fruitful and increase in number; fill the earth and subdue it. Rule over the fish of the sea and the birds of the air and over every living creature that moves on the ground'." Genesis 1:28 (New International Version)

[2] "A large population is a king's glory, but without subjects a prince is ruined." Proverbs 14:28 (New International Version)

[3] "Now Lot, who was moving about with Abraham, also had flocks and herds and tents. But the land could not support them while they stayed together, for their possessions were so great that they were not able to stay together. And quarreling arose between Abraham's herdsmen and the herdsmen of Lot. . . . So Abraham said to Lot, 'Let's not have any quarreling between you and me, or between your herdsmen and mine, for we are brothers. Is not the whole land before you? Let's part company. If you go to the left, I'll go to the right; if you go to the right, I'll go to the left.' " Genesis 13:5–8 (New International Version)

[4] "Next to prowess in arms, it is regarded as the greatest proof of manly excellence to be the father of many sons. Every year the king sends rich gifts to the man who can show the largest number: for they hold that number is strength" (Herodotus, *The History*, Book I, Chap. 136, cited in Hutchinson, 1967, p. 8).

this juxtaposition of population size and poverty, Aristotle anticipated many of the debates of the 19th century.

Perspectives on such matters in Rome were quite different. Instead of limiting its boundaries to those of a small city-state, and dealing with any population increase through trade and emigration, the Roman conception was of a growing population that would expand Roman territories via military power. Subsequent maintenance of the resulting empire required continued demographic increase. The goal of increases in the numbers of potential soldiers to serve imperial interests was reflected in Roman legislation (Hutchinson, 1967, p. 13–14).

2. Medieval Pauses

The classical tradition of thinking and practice regarding population matters diminished in Europe during the Middle Ages. Medieval thinkers generally believed that population growth was a sign of God's benevolence, and considered large families as one of his blessings (Spengler, 1942, p. 4–5, cited in Hutchinson, 1967, p. 15). The classical tradition was, however, maintained to some degree in the Arab world; in the 14th century, the eminent Arab historian Ibn Khaldun devoted an entire chapter of his *Prologomena* to a discussion of demographic change (Hutchinson, p. 15–16). He argued that the economic well-being of a society would be promoted by population growth, and described how states can promote the fertility of their subjects through wise and lenient government.

With the Renaissance, demographic discussion reemerged in the West as well. Already, during the 15th century, Francesco Patrizzi of Siena argued that while a large population might be desirable, this was so only with sufficient resources to support its numbers without mass poverty. In the 16th century, Niccolo Machiavelli was characterizing population increase as desirable for the monarch (Machiavelli, *The Prince*, Chap. 10). but also noting that demographic increases in excess of available resources would be constrained by poverty, disease, floods, plagues, or famines (Machiavelli, *Discourses*, Book I, Chap. 1 and Book II, Chap. 5, cited in Hutchinson, p. 16–17). Six decades later, the French political theorist Jean Bodin argued further for the virtues of larger populations as a source of political stability for the state and monarchy.[5] In 1588, the Jesuit thinker Giovanni Botero produced a treatise on population that both advocated a large and growing population as a source of strength and security for the state, while anticipating by two centuries the arguments of Malthus that the power of human reproduction exceeds the potential of agricultural subsistence.[6]

[5] Bodin was not only a theorist, but also counselor to the Duke of Anjou and a member of the Parliament of Paris.

3. From Mercantilism to Marx

European economic and political thought was dominated, for much of the three centuries from 1500 to 1800, by mercantilism. In this tradition, the population controlled by a monarch determined the power of his or her monarchy. For Frederick the Great, "the number of [the] people makes the wealth of states." People represented the monarchy's source of labor and soldiers, and hence the monarch was right to intervene to maximize the rate of population growth and to limit outmigration.

By the 18th century, utopian thinking had embraced mercantilists' view that population size was a principal determinant of wealth and prosperity. But they rejected the mercantilists' support for absolute monarchy, arguing instead that mankind could be perfected such that there would be no need for coercive social institutions, land could be held in common, and the resources produced by the application of the population's labor to the land distributed on the basis of need. Thus in their view, any level of population size and growth would be consistent with continuing progress. Among the leading utopians were the Marquis de Condorcet, William Godwin, and Daniel Malthus, the father of Thomas Robert Malthus.

A very different stream of thinking evolved first among physiocrats such as François Quesnay. Like the utopians, they opposed the authoritarian interventions favored by the mercantilists, arguing instead for what they termed "laissez-faire." Land rather than labor produced economic surplus, and hence the utopians erred in arguing that excessive population growth would lead to poverty. The physiocrats were the first to seek out and use the then-emerging data on demographic patterns produced by John Graunt, Edmund Halley, William Petty, and Gregory King.

The arguments of the physiocrats were influential with the classical economists who came to dominate the economic thinking of British political leaders by the early 19th century. For Adam Smith, David Ricardo, and Thomas Robert Malthus, there were three factors of production—land, labor, and capital—but of these only land was of a fixed supply. Hence increased

[6] Discussed in some detail in Hutchinson, *op cit*., p. 18–20. Botero, like Bodin, did not limit his role to theoretical writings. He served as Secretary to Federico Cardinal Borromeo, longtime Archbishop of Milan, and as an aide to the ruler of Savoy.

inputs of labor or capital applied to the land would lead to diminishing returns, with resulting limits on agricultural and industrial production.

It was Malthus who focused most deeply on the demographic dimensions. His first *Essay on the Principle of Population*, written in 1798, was a polemic deficient in empirical evidence, and directed principally against the utopian writings of William Godwin, Marie-Jean Antoine de Condorcet, and Malthus's own father Daniel. The perfected human society of the utopians was not possible, he argued, because without conscious constraints the basic biological propensity toward sexual behavior and unlimited reproduction would overwhelm the limited increases that could be extracted from the land.

Criticisms of his first Essay prompted Malthus to undertake more serious research and analysis, incorporated into subsequent revised versions. He estimated that a country with an excess of land relative to its population might double in size every 25 years. However, in more typical situations in which land was not in surplus, population growth would be limited by what he termed "positive checks" such as "sickly season, epidemics, pestilence, and plague." The alternative to such terrible positive checks was logically "prudential restraints." By this he meant both "moral restraint" such as late marriage accompanied by celibacy outside marriage and "immoral vices" such as promiscuity, prostitution, homosexuality, birth control, and abortion. Of these, only moral restraint was acceptable to Malthus, a former Anglican curate (Teitelbaum, p. 179–180). No humane society could ignore the need for prudential restraints, as the alternatives would necessarily be increased poverty, starvation, and suffering as the positive checks came into force.

While Malthus opposed both contraception and abortion on moral grounds, those who decades after his death called themselves "neo-Malthusians" embraced his arguments about the links between excessive population growth and poverty, but promoted use of voluntary means of fertility control. Their embrace of Malthusian arguments led to sharp rhetorical clashes between them and those on the left, who argued that the cause of poverty was not excessive population growth, but rather an economic system that failed to equitably distribute the output of economic activity.

Both Malthus and the neo-Malthusians, in turn, came under concerted fire from Karl Marx and his associates. For them, poverty was caused by capitalism, which allowed owners of capital to siphon off part of the surplus product produced by labor. Under socialism, these workers' products would be provided to the workers, thereby eliminating poverty. Hence any number of people could be supported, as people were at once producers and consumers; this was a view shared with the utopians, although Marx was otherwise quite critical of the utopians' views.

II. FROM THE 19TH TO THE 20TH CENTURY

This all-too-brief summary of more than two millennia brings me to the two centuries since 1800, which I will examine in more detail. The dramatic declines in both mortality and fertility that characterized this period were the topic of much debate and concern. The theory of the demographic transition (discussed in Chapters 68 and 69 of Vol. III), which came to prominence during the 1930s, explained these trends as consequences of forces of modernization such as industrialization and urbanization. (For detailed descriptions of the theory and empirical tests thereof, see Coale and Watkins, 1986; Teitelbaum, 1984.) Yet, contrary to the timing patterns implicit in the theory, it was the agrarian and rural France of the 18th century that experienced the earliest recorded sustained declines of fertility (van de Walle, 1974)—at least a century earlier than that of Great Britain, the pioneer and crucible of the Industrial Revolution, and even further in advance of rapidly industrializing Germany (see Chapter 69, Vol. III).

Evidently there were no obvious systematic patterns in the timing of such fertility declines, but all industrializing societies ultimately did experience fertility declines, some in the 19th and others in the 20th century. Moreover, despite the large differences among these societies in their histories, religions, languages, politics, and cultures, a few common population issues, which were of key concern during the period from 1870 to World War II, can be identified. It may be useful to categorize these political themes as follows:

Strategic demography, nationalisms, and the "Volk"
Concerns about poverty
The embrace of science: the rise and fall of eugenics and Social Darwinism
Taxes, medals, family allowances

1. Strategic Demography, Nationalisms, and the "Volk"

Perhaps the most common theme related to demographic change within Europe after 1870 concerned its implications for nationalism, national power, and related imperial interests. This has been described previously by the shorthand term "strategic demography" (Teitelbaum and Winter, 1985), a perspective that was most vigorously expressed in France.

The early decline of French fertility beginning in the middle of the 18th century coincided with a period in which the relative power and influence of France initially rose and then plummeted. By the late 19th century, France had declined in relative terms from perhaps the dominant state in Europe to one struggling to compete with the surging power of Britain and Germany. Of course the same century saw deeply destabilizing developments in France,[7] but it is perhaps unsurprising that many leading 19th-century French intellectuals focused instead on the early French fertility declines, and the resulting diminishment of the size and dynamism of the French population, as a central cause of the decline a nd "decadence" of France that they perceived to be underway.

While such concerns were present during the first half of the 19th century, they came into full cry following the French military collapse before the armies of Prussia in 1870–1871, which culminated in the creation of both the German Empire of Bismarck and the French Third Republic. In the ensuing final decades of the 19th century, the French Assembly considered numerous legislative proposals with pronatalist intent, clearly population policies in the restricted meaning discussed earlier. These included tax exemptions for heads of large families, tax surcharges and state appropriation of inheritances for unmarried adults and the childless, and restrictions on voting rights of unmarried men aged 26 to 40. While many such policies were proposed, the unstable coalitions that characterized the Third Republic meant that few of these provisions were adopted, and some of those that were passed were later rescinded.

In 1896, a coalition of prominent French intellectuals and politicians founded the National Alliance for the Increase of the French Population (*Alliance nationale pour l'accroissement de la population française*). This quickly emerged as an energetic and effective lobby group that framed numerous legislative proposals and pressed them on members of the government and the National Assembly.

By 1900, a resolution to create a commission on depopulation gained support from over 130 members of the French Senate. This commission, which began its work in January 1902, took only six months to conclude that the "prosperity and grandeur of France" required financial incentives to encourage French couples to increase their fertility (Beale, 1911, p. 171–267, as cited in Quine, 1996). Within a decade, this view had become the norm in France, as evidenced by the emergence of a broad-based Parliamentary Group for the Protection of Natality that by the election of 1914 had attracted more than half of the deputies in the National Assembly (Quine, 1996, p. 66).

Germany and Great Britain experienced their fertility declines much later than in France, and both elite and mass attention to low fertility also appeared much later. While in the 1870s the political class in Paris was actively debating the causes and consequences of French fertility decline, the British equivalent was more focused on the perceived dangers of too-high fertility in increasing poverty levels, with active debate especially about the merits or otherwise of the neo-Malthusian arguments and activism then underway in London.

In Italy, where fertility declines also occurred far later than in France, there was little visible support for promoting births until the 1920s, with the rise to power of Benito Mussolini's fascists.[8] Indeed, Mussolini himself had embraced neo-Malthusian ideas only 9 years earlier, just before the onset of World War I (Wanrooij, 1990, p. 79, as cited in Quine, 1996, p. 25). Yet in the ensuing slaughter, Italy lost over 680,000 soldiers (vs. 765,000 British and nearly 1.4 million French), and notwithstanding Mussolini's 1913 embrace of neo-Malthusianism, one of Mussolini's central goals after taking power in 1922 was to increase the population of Italy.

In a famous speech in 1927 and an essay published in 1928 titled "Il numero come forza" ("Numbers as Force"), Mussolini articulated a rejection of neo-Malthusian principles in favor of expanding the Italian population by at least 25% or 10 million (Mussolini [1928]). The goals were almost textbook examples of strategic demography: to increase Italian economic power by counteracting underconsumption in the economy; to secure Italian national security in a European neighborhood populated by 90 million Germans, 200 million Slavs, and much larger populations controlled by the empires of Great Britain and France; and to provide military manpower for the dispatch of colonial settlement abroad that was intrinsic to the heirs of

[7] The French Revolution of 1789, the Terror of the mid-1790s, the rise to power of Napoleon Bonaparte (1799–1815) and the depredations of the Napoleonic wars, the Bourbon restorations of 1814–1815, the revolution of 1848, and the defeats at the hands of Prussia during of the Franco-Prussian War of 1870–1871.

[8] Indeed, Mussolini became *Il Duce* (Leader) of Italy a year before Hitler's first attempt to seize power in Germany in the failed Beer Hall Putsch. By the time Hitler, another survivor of World War I, became the *Führer* (Leader) of Germany, Mussolini had ruled Italy for more than a decade.

the Roman Empire. The Italian fascists would lead to the rebirth of a "young and fertile race," the "national resurrection" of a powerful New Italy.

2. Concerns about Poverty

As noted earlier, there were long-standing concerns about poverty resulting from high rates of fertility and population growth, concerns that became more visible during the first half of the 19th century in response to the writings of Malthus, his followers, and his critics. Such concerns were most visible in Great Britain, where Malthusian ideas originated and then generated sustained interest and debate for many decades.

Indeed, one of the ironic twists that has heavily affected population debates over the past two centuries is the fact that Marx resided in London during the 1870s when the neo-Malthusians launched a highly public and energetic campaign, of which Marx was presumably aware. The radical leaders of the movement deliberately sought to test British laws prohibiting publication of information about birth control by openly reprinting a half-century-old text by an American physician, Charles Knowlton, *The Fruits of Philosophy*. This was one of the first test cases, in which Charles Bradlaugh and Annie Besant openly announced the book's re-publication, were arrested, and then in 1877 tried in a highly visible trial that was amply covered by the British press. This very public campaign by the neo-Malthusians may have something to do with Marx's exaggerated denunciation of Malthus and his theories, a denunciation which, as subsequently interpreted by Marxians, produced a century of controversy that to some degree still continues.

Meanwhile, in Germany, one of the leading Marxists of the period, Karl Kautsky, published a book arguing that in reality there was no conflict between Marxist and Malthusian theory. "The two," he asserted, "are not in principle incompatible"(Kautsky 1880, p. 166–192). For this, Kautsky was bitterly excoriated by none other than Friedrich Engels, close confederate and coauthor of Marx, and by then the defender of Marxian orthodoxy. In 1910, Kautsky recanted his apostasy, thereby solidifying the notion on the left that Malthusian theory was fundamentally reactionary (Kautsky, 1910, Chap. 15–16).

Support for Malthusian and neo-Malthusian ideas was more muted in France, without doubt due to its early fertility declines. Yet even in France, a neo-Malthusian society *(Ligue de la régénération humaine)* was founded in 1896, along with other efforts to promote the practice of birth control. Indeed these organizational efforts occurred during the same year as the founding of the National Alliance, and their near-simultaneous appearance may not have been unrelated.

Italy, unified only in 1871, was a latecomer to industrialization, modernization, urbanization, and the demographic transition. Its population was largely agrarian and poor; indeed even by the late 1930s, nearly half of the economically active population worked in agriculture. Its fertility and mortality rates had not shown signs of decline until close to the turn of the 20th century, and remained high throughout World War I and into the 1920s. The average rates disguised large regional disparities: the North experienced rapid industrialization and relatively early declines in both fertility and mortality, while the South was dominated by low-productivity agriculture performed by peasant farmers who sustained long-standing patterns of high fertility and high mortality.

By the turn of the 20th century, these conditions of deep poverty in the South led to increasing interest among Italian intellectuals and reformers in Malthusian ideas, and eventually to the development of a small but active movement supporting the adoption of contraception. In 1912, two leaders of this movement (Luigi Berta and Secondo Giorno) were brought to trial for publication of an allegedly obscene manual titled *L'arte di non far figli: neomalthusianismo pratico* (The Art of Not Having Children: Practical Neo-Malthusianism), of which some 27,000 copies had been sold. Like the London trial of Charles Bradlaugh and Annie Besant 35 years earlier, this Milan trial provided much free publicity to the cause, and in the end resulted in acquittal. Berta and Giorno promptly founded a new neo-Malthusian League and, until the onset of World War I, published a periodical with the evocative title *Educazione sessuale: rivista di neomalthusianismo e di eugenica* (Sexual Education: The Journal of Neo-Malthusianism and Eugenics).

3. The Embrace of Science: The Rise and Fall of Eugenics and Social Darwinism

The 19th century was in many ways the century in which science was first embraced enthusiastically by both the elite and the mass public, in part attributable to the evolutionary ideas of Charles Darwin. Many of the natural and social scientists who rose to prominence in this period played a powerful role in stimulating interest and concern about low fertility and depopulation.

In France, the national harbinger of concern about low fertility, early alarms came from prominent social scientists such as Frederic Le Play, a sociologist, engineer, and founder (in 1856) of the Society of Social Economy, and Paul Leroy-Beaulieu, professor at the College de France and founder and editor of *L'Économiste française*. While they differed in their prescriptions (with Le Play urging a domestic pronatalist strategy while Leroy-Beaulieu advocated French colonial expansion and the annexation of Tunisia), they agreed that the decline of France relative to Germany would continue unless the growth of the French population could be augmented.

More biologistic explanations were popular across the English Channel, most notably in the rise of eugenics. The word itself was a neologism created in 1883 by the English polymath Sir Francis Galton (1822–1911) from the Greek "noble in heredity." Galton, a cousin of Charles Darwin, was himself one of the most prominent scientists of his day: he was a founder of modern psychology, the creator of now-standard statistical correlation and regression analyses, inventor of identification via fingerprints, as well as a geographer, explorer, and meteorologist.

Galton and his followers took many steps beyond where Darwin had trod. Their core argument was that modern science for the first time made it possible to improve the human condition by cultivating "desirable" inherited traits and eliminating "unsuitable" ones. Eugenics and its supporters formed part of a larger intellectual current called Social Darwinism, in which the evolutionary insights of Darwin were to be applied to human societies, which came to be seen as subject to the same Darwinian natural laws of "struggle for existence" and "natural selection of the fittest."

Contrary to current-day perceptions of that period, in these early days there was nothing intrinsically reactionary or right-wing about such ideas. For decades surrounding the turn of the 20th century, the lure of nominally scientific, eugenic, and Social Darwinist ideas attracted fascists, socialists, liberals, racists, and feminists alike. Ultimately, however, both eugenics and Social Darwinism were captured by the forces of reaction, racism, and fascism, and both were fatally discredited by the horrors perpetrated by Nazi Germany. (For a comprehensive treatment of the intellectual and political history of the eugenics movement, see Kevles, 1995.)

Although French intellectuals did not embrace eugenics as enthusiastically as did their British colleagues, those in France who were supportive of eugenics sought to closely align themselves with the pronatalist advocates led by the National Alliance. The founder (in 1896) and president of the alliance, Jacques Bertillon led a large French delegation to the First International Eugenics Conference at the University of London in 1912, and an alliance cofounder Charles Richet served as vice president of the French Eugenics Society, also founded that year (Quine, p. 67).

In Italy too, the pronatalist enthusiasms of Mussolini and the Italian Fascist Party drew on a substantial body of arguments and claims emanating from Italian intellectual and scientific circles. In particular, the fascists embraced the ideas of Italy's most prominent anthropologist, Giuseppe Sergi, who long before the founding of the Italian Fascist Party had embraced the then-burgeoning notions of "scientific racism." In 1900, Sergi had put forward the argument that the Italian "race" had inherited from its Roman forebears an innate drive toward empire, a disposition toward conquering and colonizing lesser "races." These "imperial Italian race" notions converged with related intellectual currents supporting the rising influence of eugenics and Social Darwinism.

In 1913 (just after the Milan trial of the Italian neo-Malthusians Berta and Giorni), eugenics had converged intellectually with pronatalism and Social Darwinism in the founding of *Comitato Italiano per gli Studi di Eugenica* (Italian Committee for the study of eugenics), which in 1919 became the *Società Italiana di genetica e Euginica* (*Italian Eugenics Society*). The intellectual leaders of this group were Guiseppi Sergi and the demographer Corrado Gini. In 1912, Gini (1884–1965) had published his *I fattori demografici dell'evoluzione delle nazioni* (Demographic Factors in the Evolution of Nations), in which he argued for a comprehensive set of pronatalist policies to enhance the economic prosperity and military security of the Italian nation. These included positive incentives such as welfare benefits and tax reductions to those who bore many children, negative incentives imposed on the unmarried and married but childless, and in general encouragement toward early marriage and large families.

The fascist government under Mussolini embraced many of these ideas, including both Sergi's notions of the "imperial Italian race" and Gini's prescriptions as to how its innate imperial proclivities might be facilitated by increased fertility.

4. Taxes, Medals, Family Allowances

Many of these ideas, promoted effectively by advocacy groups (e.g., the National Alliance in France), were translated into formal government policies during the first decades of the 20th century. In certain cases, these ideas resulted in the emergence of population policies during the chaotic political-economic period of the interwar years and during World War II. The three most prominent instances of this phenomenon occurred in France, Italy, and Germany.

a. France

In France in 1913, the French parliament enacted a new family allowance act advocated by pronatalist groups, which provided substantial payments to fathers for each child between the ages of 3 and 13. In 1914, French income tax rates were modified to reduce taxes imposed on larger families. However, many of these provisions were delayed by World War I, and the family allowances adopted in 1913 were allowed to diminish in real terms with inflation.

After the chaos and mass slaughter of the war, it is ironic to note that in 1919, in the French Senate debates on the Treaty of Versailles, the French Premier and Minister of War Georges Clemenceau saw fit to address himself to the strategic threats posed by low French fertility:

> The treaty does not say that France must undertake to have children, but it is the first thing which ought to have been put in it. For if France turns her back on large families, one can put all the clauses one wants in a treaty, one can take all the guns of Germany, one can do whatever one likes, France will be lost because there will be no more Frenchmen. (*Journal Officiel: Débats du Senat*, 11 October 1919, p. 1625–1626; as cited in Tomlinson, 1983)

Following its election in 1919, the center-right government formed by the *Bloc National* created a *Conseil superieur de natalité* to develop new pronatalist policies. It also overwhelmingly passed legislation to prohibit "antinatalist propaganda," and adopted criminal penalties against providers of abortion. In 1920, the *Conseil superieur de natalité* organized the first national celebrations of *Journée nationale des mères de familles nombreuses* (National Day for Mothers of Large Families), which became Mother's Day in 1928 and Mother's Festival in 1941, during which some 30,000 mothers with five or more children received bronze, silver, or gold medals recognizing their service to the French nation. These celebrations were emulated two decades later by the Vichy Government. Later that same year, legislation was adopted that imposed a 25% surcharge on unmarried males and females over age 30, along with a 10% surcharge on childless marriages of 10 years or more in duration. In 1921, special train fares were legislated for large families, and in 1922 large families received preference to rent apartments in subsidized public housing (Talmy, 1962, vol. 1, p. 162–165, and Tomlinson, 1983, cited by Quine, 1996, p. 74.).

During the 1920s, French pronatalism developed from elite activism to a mass movement. There was a coalescence among national political movements ranging across much of the political spectrum, religious defenders of both the family and traditional values, and widely dispersed local and regional groups promoting aid to families. Many of the latter had their roots in clerical circles, yet reflected the goals of the lay Catholic establishment. Their memberships reached some 300,000 persons by 1930, far outnumbering the 40,000 supporters of the secular National Alliance for the Increase of the French Population. Though they differed in particular ways, they shared a common commitment to increase financial support for French families so as to facilitate increased childbearing.

In 1932, this loose coalition achieved a major success with passage of the Family Allowance Act. The act made compulsory the payment of family allowances *Allocations* that had been introduced voluntarily by many industrial employers. Most of the financing continued to come from employers, but by 1938 nearly two-thirds of the 9 million industrial workers in France received such supplementary payments. These payments could be substantial relative to earnings, especially at the lower end of the social scale, although the nominal amounts varied during the economic crisis of the 1930s and during World War II, and their real value often declined when inflation accelerated.[9] Still, following World War II, an unskilled French worker with three children received a family allowance equal to half his pay, while the allocation to a factory worker with five children exceeded the average wage for such occupations (Prost, 1989, p. 155).

Finally, in 1941 the Vichy government organized festivals of family culture that included a Mother's Day celebration in Reims under Maréchal Philippe Petain's patronage. Some 21 years after the 1920 celebrations organized by the *Conseil superieur de natalité*, the collaborationist Vichy government reflected continuity with the practices of earlier nationalist governments by awarding medals to women with large families: bronze for women with 5 or 6 children, silver for those with 7 to 9, and gold for those with 10 or more (Teitelbaum and Winter, 1985, p. 38–39).

b. Italy

In Italy, during the mid-1920s, Mussolini's government declared *una battaglia delle nascite* (battle for births) to be conducted on many fronts. In 1926, a celibacy tax was imposed on unmarried men. It was portrayed as an evolution of a tax imposed by the Roman Emperor Augustus, but may well have been based more immediately on the model of the 1920

[9] Even so, following World War II an unskilled French worker with three children received a family allowance equal to half his pay, while the allocation to a factory worker with five children exceeded the average wages for such occupations (Prost, 1989, p. 155).

French income tax surcharge imposed on men and women who were unmarried and on married couples with no children. The Italian celibacy tax was based on both income and age, and for unmarried men between 35 and 50 represented an income tax surcharge of fully 25% plus an additional flat-rate tax. The fascist government openly discussed additional measures, including a tax on married couples with few children, annulment of marriages that were childless after 5 years, exclusion of unmarried and low-fertility persons from attractive public sector employment, and the appropriation of wealth left by persons with no direct heirs or inherited by unmarried heirs (Quine, 1996, p. 41).

In his 1927 speech, Mussolini had forecast that under the influence of fascist pronatalist policies, the Italian population by 1950 would have increased from about 40 million to 60 million. Yet it was not to be, and by 1950 the population of Italy was still only 47 million (United Nations, 2003). The war could not be blamed for this failure as in 1938, in the build-up to World War II, the regime acknowledged that its battle for births had failed, and indeed that Italian fertility and marriage rates had continued to decline over the preceding decade.

c. Germany

But it was in Germany that the most aggressive, and finally, the most disastrous political measures were taken by the Nazis. As Paul Weindling comments (1988, p. 111), these measures were derived partly from the comparison made in 1921 by Oswald Spengler of low fertility to a "racial senescence." This idea which was then adopted by the German doctor, statistician, and demographer Richard Korherr in his study of the German fertility decline, the 1928 edition of which was honored with a preface by Benito Mussolini himself, while the preface of the 1935 edition was written by Heinrich Himmler (moreover, Richard Korherr later became head of the SS Department of Statistics).

It is not possible to disassociate the Nazi regime's population policy from its racial policy, which was a combination of three principal beliefs: the superiority of the "Aryan" race, the primacy of the state's interests over those of the individual, and women's main function being motherhood. These beliefs led to the three most ferocious political campaigns of the Nazi government: the sterilization and extermination of millions of "non-Aryans," the promotion of population growth among "racially pure" Aryans, and settlement of the resulting increasing number of Aryans in lands conquered by the German army or liberated through ethnic cleansing.

It is undoubtedly the second of these three campaigns, the promotion of growth among the racially pure Aryan population, which should be discussed here, even though the other two beliefs aimed at, and produced, direct demographic consequences. From the beginning of the Nazi regime, in 1933, interest-free loans were given to young married couples on the condition that they met the necessary racial conditions (being Aryan) according to the criteria established. These loans were also partly viewed as a means of creating employment for the millions of young unemployed males, as another condition was that the women showed proof of having been employed for at least 6 months and that they undertook to stop working once married. These loans were financed by the creation of a tax on unmarried men and women who earned a certain income, a "tax on single people," obviously copied from the Italian fascist legislation of 1926. The loans were provided for 8 years and the repayment capital was reduced by 25% on the birth of each child (McIntosh, 1983, p. 63).

The Nazi regime did not prohibit contraception or induced abortion before 1941, possibly to enable the eugenic use of these methods. However, from the very beginning, the government considerably limited distribution and access to birth control methods by closing all the family planning clinics and by prosecuting abortionists. The Nazis also harshly repressed people who were guilty of *Rassenschande* (racial shame), that is, sexual relations between Aryans and non-Aryans. Even more severe punishment was reserved for homosexuals, including among the Nazis themselves.

At the same time, a vigorous ideological campaign was undertaken in order to convince German women of their duty to produce many German children. The entire panoply of the formidable Nazi propaganda machine was mobilized to convince German women that motherhood was their sacred duty to the German state and its people. Oddly enough, given its rhetoric in support of the German family, the regime even encouraged Aryan Germans to have illegitimate children, provided that it was in the context of racially pure sexual relations (Bleuel, 1973, p.148–179, quoted by McIntosh, 1983, p. 62).

Did this policy have a major and durable influence on German fertility? Measured by the total fertility rate, fertility in Germany had, by the end of the 19th century, declined far less than in neighboring countries in Western Europe. Indeed, in France, the first country to experience fertility decline, nationalists used the high fertility rate in Germany as an explanatory factor for the military disaster in 1870 against the Prussians (Teitelbaum and Winter, 1985, p. 18–30). During the

1920s, after the defeat of Germany in World War I, German fertility fell to a level comparable with that observed in France and elsewhere in Europe. The total fertility rate (admittedly prone to significant variations due to postponing or bringing forward the fulfilment of desired fertility) fell in Germany to a level below the replacement of cohorts during the mid-1920s, before dropping to the historical low of 1.6 children per woman in the same year as Adolf Hitler came to power in 1933. Under the reign of the Nazis, the total fertility rate in Germany increased over a few years to 2.4 in 1939 at the beginning of World War II, but fell to less than 1.6 by the end of the war.

SUMMARY

Population policies, in the restricted sense of conscious actions by states to affect demographic trends, occupy an important place in the history of the 19th and 20th centuries. As the modern state evolved during this period, two foci of population policy emerged, relating first to military power of the state relative to its neighbors or competitors, and second to demographic trends internal to the state, relevant to perceived problems of poverty and population "quality."

Demographic components entered strongly into 19th-century debates about the sources of poverty and prosperity, especially in the passionate disputes among Malthusians, neo-Malthusians, and Marxians. Beginning around the 1870s, concerns began to emerge (especially in France) about "strategic demography," that is, threats to the state resulting from fertility decline. By the turn of the 20th century, parallel concerns about perceived decline in population quality appeared. Such concerns, perhaps exacerbated by the mass slaughter of the "lost generation" during World War I, attracted growing and broad-based enthusiasm for the new "science" of genetics and race, among intellectuals and politicians alike. During the 1920s, such concerns came to be co-opted by political movements of the right, including Social Darwinism. During the 1930s, such ideas flowered in some countries, ultimately morphing into the murderous depredations of the Nazi movement and its supporters elsewhere. Meanwhile, during the same decade more benign alarms spread about the by-then very low fertility rates experienced during the Great Depression.

By the end of World War II, ideas about differential "biological quality" on which Social Darwinism, eugenics, and scientific racism were based had been discredited by the Nazi genocides. Yet concerns persisted about the implications of very low fertility, with many governments adopting policies designed to increase fertility, as well as supporting the creation of demographic research efforts such as Institut national d'études démographiques (INED) in France and the Royal Commission on Population in the United Kingdom.

References

Aristotle, 350 BCE. *Politics* (Book Two, Chapter IX). Translated by Benjamin Jowett. http://classics.mit.edu/Aristotle/politics.html.

Beale Octavius Charles, 1911. *Racial decay: a compilation of evidence from world sources.* London and Sydney, A. C. Fifield and Angus & Robertson Ltd., 439 p.

Bleuel Hans P., 1973. *Strength through joy: sex and society in Nazi Germany.* Translated from German by Maxwell Brownjohn. London, Seker & Warburg, 272 p.

Coale Ansley J. and Watkins Susan Cotts, 1986. *The decline of fertility in Europe. The revised proceedings of a conference on the Princeton European fertility project.* Princeton (New Jersey), Princeton University Press, 484 p.

Gini Corrado, 1912. *I fattori demografici dell'evoluzione delle nazioni.* Turin, Fratelli Bocca, 143 p.

Holy Bible, 1990. *New international version.* Grand Rapids (Michigan), Zondervan Bible Publishers, 1970 p.

Hutchinson Edward Prince, 1967. *The population debate.* New York, Houghton Mifflin, 466 p.

Kautsky Karl, 1880. *Der Einfluss der Volksvermehrung auf den Fortschritt der Gesellschaft* [The influence of population increase on social progress]. Vienne, Bloch and Hasbach, 195 p.

Kautsky Karl, 1910. *Vermehrung und Entwicklung in Natur und Gesellschaft.* Stuttgart, J. H. W. Dietz Nachf, 268 p.

Kevles Daniel J., 1995. *In the name of eugenics: genetics and the uses of human heredity.* Cambridge (Massachussetts), Harvard University Press, 426 p.

Korherr Richard, 1928. *Geburtenrückgang: Mahnruf an das deutsche Volk* [Fertility decline: a warning to the German people]. Republished in 1935 with a preface by RSS Heinrich Himmler. Munich, Deutsche Monastshefte G. m. b. H., 48 p.

Machiavelli, Nicolo, 1515. *The Prince.* Translated by W. K. Marriott, 2004. http://etext.library.adelaide.edu.au/m/machavelli/niccolo/m149p.

McIntosh C. Alison, 1983. Population policy in Western Europe: responses to low fertility in France, Sweden, and West Germany. Armonk (New York), M. E. Sharpe, 286 p.

Mussolini Benito, [1928] 1937. Il numero come forza, *in*: Richard Korherr, *Regresso delle nascite, morte dei popoli*, Preface. Rome, Libreria del littorio, 179 p.

Prost Antoine, 1989. Catholic conservatives, population, and the family in twentieth-century France, *in*: Michael S. Teitelbaum and Jay M. Winter (eds.), *Population and resources in Western intellectual traditions*, p. 147–164. Cambridge and New York, Cambridge University Press, 310 p.

Prost Antoine, 2002. *Republican identities in war and peace: representations of France in the nineteenth and twentieth centuries.* Translated by Jay Winter with Helen McPhail. New York and Oxford, Berg, 339 p.

Quine Maria Sophia, 1996. Population politics in twentieth century Europe: fascist dictatorships and liberal democracies. London and New York, Routledge, 152 p.

Spengler Joseph J., 1942. *French predecessors of Malthus, a study in eighteenth-century wage and population theory.* Durham (North Carolina), Duke University Press, 398 p. (Republished: New York, Octagon Books, 1965).

Talmy Robert, 1962. *Histoire du mouvement familial en France (1896–1939)*, 2 vol. Paris and Aubenas, Union nationale des caisses d'allocations familiales, 317 + 285 p.

Teitelbaum Michael S., 1984. *The British fertility decline. Demographic transition in the crucible of the Industrial Revolution.* Princeton (New Jersey), Princeton University Press, 269 p.

Teitelbaum Michael S., 1988. Demographic change through the lenses of science and politics, *Proceedings of the American Philosophical Society*, vol. 132(2), p. 173–184.

Teitelbaum Michael S. and Winter Jay M., 1985. *The fear of population decline.* Orlando and New York, Academic Press, 201 p.

Tomlinson Richard, 1983. *The politics of Dénatalité during the French Third Republic, 1890–1940*, Ph.D. thesis. Cambridge, University of Cambridge, 382 p.

United Nations, 2003. *World population prospects. The 2002 revision*, 2 vol. New York, United Nations, Department for Economic and Social Affairs, Population Division, 781 + 939 p. and CD-Rom. (ST/ESA/Ser.A/225).

van de Walle Étienne, 1974. *The female population of France in the 19th century: a reconstruction of 82 départements*. Princeton (New Jersey), Princeton University Press, 483 p.

Wanrooij Bruno P. F., 1990. *Storia del pudore: la questione sessuale in Italia, 1860–1940.* Venise, Marsilio, 234 p.

Weindling Paul, 1988. Fascism and population in comparative European perspective, *Population and Development Review*, vol. 14, special issue. "Population and resources in Western intellectual traditions," edited by Michael S. Teitelbaum and Jay M. Winter, p. 102–121.

CHAPTER

101

An Introduction to Population Policies

STEFANO BALDI AND RAIMONDO CAGIANO DE AZEVEDO

Permanent Mission of Italy to the United Nations, United Nations Plaza, New York, United States

Facoltà di Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy

As expanded on in the preceding chapter, the history of our societies shows that population policies, which tend to influence the size, structure, or dynamics of the population, even though they have not always been explicitly presented or perceived as such, are deeply rooted in the past. While being developed they are based, consciously or not, on ideological, military, or social grounds. Formerly, although it still happens nowadays, such policies were based on hypotheses or positions of superiority of one group vis-à-vis another one. In any case, it must be emphasized that they are developed somewhat independently from the actual circumstances that prevail at the time when they are adopted, and consequently from the possibility of implementing them, an aspect which is beyond the scope of this chapter.

Religion is an important source of inspiration for population policies. Persecution has been perpetrated in its name, sometimes leading to massive emigration of minorities, the creation of new human settlements in certain regions in the world, or even to the extermination of whole groups of the population. Religion has also been an important source of legislation in terms of private life regulation—marriage, family building, and reproduction.

Military motivations, particularly in the past, have been ground for nationalistic and imperialistic policies. The various motivations behind those two policies, that is, defense of the national territory for the former and expansion of zones of influence for the latter, have had the same effect: they tend to entail high pressure on the population. In the past they were often used as central themes in governmental policies and action in the developed countries, and even nowadays, they have provided motivation for the governments of certain developing countries.

Moreover, ideological factors have influenced the implementation of population policies—and still do. The major migration flows related to ethnic and racial discrimination provide recent examples of such a close relationship between ideologies and populations. Moreover, history is full of instances when abuse of political power was used to alleviate inequality and linguistic differences. The effects on population policies resulting from different schools of thought are subtler: for example, nowadays, the idea of a stationary population or of zero growth has had significant effect, both implicit and explicit, over population policies.

Most prominent among matters related to population policies, are those relating to fertility and reproduction, particularly two specific models, family policy and family planning. Family policy refers to a set of measures, rules, and behaviors that regulate the relations between governments and families, usually so that families can expand. Family planning, though not in opposition to family policy, is based first and foremost on a view of quality of life in which health and social conditions in human reproduction tend to be emphasized; when it is of a stronger type, family planning may include demographic objectives clearly expressed in quantitative terms.

The real difficulty, in defining a population policy, comes from the fact that most political actions eventually have direct or indirect effects over the population. Moreover, population policies cut across several sectors, affecting as they do many different areas, or interacting with them.

In fact, the following statement is quite right: "Population policy involves a good deal more than making family planning centers available to the rural poor. It also involves more than a recognition that poverty is a root cause of high fertility rates" (Dasgupta, 1995).

Actually, every nation will take political measures that affect the population and the size of families. However, it is sometimes useful to speak of population policies only in those cases when governments make explicit statements about those being their objectives. In many countries, population or family policies are only implicit within social policies and are not dependent on the demographic situation.

> Most countries have never had explicit and coherent policies on population, although all aspects of demographic change are affected by many government actions, usually taken for other reasons. The range of policies affecting population characteristics is so large that full consideration would include virtually all aspects of social and economic policy. (Pressat, 1985, p. 177)

I. DEFINITIONS, OBJECTIVES, AND COMPONENTS OF POPULATION POLICIES

Population policies may have various, multiple objectives—population size or more frequently, growth pace, age structure, geographic distribution, and so forth.

A policy can be considered an instrument for identifying the main objectives, combined with a specific set of means to reach them. The success of a policy is closely dependent on the capacity to relate objectives to the means available. Population policies are sets of various measures taken by a government with a view to influencing the size, growth, structure, and spatial distribution of the population.

Among other possible approaches to population policies, two definitions seem of particular relevance:

1. Strictly speaking, a population policy is a deliberate, purposeful policy aimed at influencing the number of the population in a given country. For example, a pronatalist policy will aim at increasing the size of the population, whereas a Malthusian, antinatalist policy would aim at reducing the size of it, or at least at bringing down the fertility rate and consequently the growth pace.
2. Broadly speaking, a population policy is a set of laws and other measures that directly or indirectly, deliberately or unconsciously affect the population, both in qualitative and quantitative terms.

According to point 3.5 in the Programme of Action of the International Conference on Population and Development, which was held in Cairo in 1994:

> At the international, regional, national and local levels, population issues should be integrated into the formulation, implementation, monitoring and evaluation of all policies and programmes relating to sustainable development. Development strategies must realistically reflect the short-, medium- and long-term implications of, and consequences for, population dynamics as well as patterns of production and consumption. (United Nations, 1995a, p. 14)

The characteristics of a population may be affected by policies that are implemented in a wide variety of sectors (health, housing, employment); consequently it is important to draw a distinction between those policies that have an indirect (or implicit) demographic effect and explicit ones that directly aim at (and were designed with a view to) affecting the population. Population policies have widely differing origins and substantially different effects, depending on their targets (housing policies, health policies, employment policies). A very good example is provided by policies directly aimed at reducing child mortality and in the area, significant success has been achieved within the social and health policies adopted in the latter part of the 20th century.

Mortality, fertility, and migration are the main components of population policies and represent the three facts that determine the quantitative evolution of the population. Whereas mortality and fertility rates may change in a gradual way, migration, both internal and international, may cause sudden variations in the structure of the population. Migration can be considered the most flexible of the three factors in implementing population policies, though it may also have negative effects over social stability when it is exceedingly massive.

The main methods that can be used to change the level of fertility rates include (a) measures relating to women's status; (b) systems on the protection of children, youth, and high-risk groups; (c) education and information; (d) strategies for an improvement of the distribution of the population; and (d) family planning and reproductive health services.

It is only recently that governments have started explicitly considering population issues as a stake in public policies, and in some countries there is still a tendencyto view population policies solely in terms of growth control.

Most population policies are adopted at the national level, although ongoing international debates on the subject clearly have strong influence over the attitude of national governments, particularly through the major international conferences (we shall return to this point at the end of the chapter, but the reader is also referred to Chapter 120 of this volume for further details on that point).

"An important first step in formulating population policy is to identify population problems and policy measures on the basis of population studies and research" (United Nations, 1993, p. 204). As with many other policies, implementing a population policy often implies hazards and requires occasional adjustments. A major factor is that population problems, cutting as they do across national borders, from a local scale take on an international one, which makes institutional responses still more complex.

II. THE RELATIONSHIP BETWEEN POPULATION POLICIES AND POPULATION THEORIES

1. Different Types of Population Policies

Roughly speaking, there are two major groups of population policies: the policies devised with a view to limiting population growth, and those aimed at encouraging growth or slowing down decrease. The former are usually called "Malthusian."

Population theories provide the basis for population policies to be adopted. According to Christopher Wilson and Roland Pressat (1985), "a population policy requires a clear, explicit statement of aims and a plan, backed by a theoretical framework, of how to achieve these aims" (p. 177). Concern is often expressed about governments interfering, through population policies, in this type of sector which is so intimately related to human private behavior. The debate on how justified any intervention in that area can be is a long-standing one. Among the wide variety of answers from various authors, the following, by Alfred Sauvy, seems to offer the most appropriate conclusion: "Interference exists anyway" ("L'intervention existe en tout état de cause") (Sauvy, 1966, Vol. 2, p. 371).

2. Policies Aimed at Limiting Population Growth

Those population policies that aim at addressing issues related to size of the population deal with growth rates rather than with the size itself.

In some cases, governments can implement antinatalist policies by adopting measures to encourage contraception. Basic measures taken to control population growth usually include supplying free contraceptives and campaigning for contraception. However, other initiatives, which vary from one cultural and religious context to another, may also affect the growth of the population. The most significant ones include rules on legal minimum age at first marriage, and laws concerning divorce or remarriage.

Governments' attitude to migration may also affect the growth of the number of the population and even its composition. If the overall legal framework is in favor of immigration, it may become a factor in accelerating population growth. In many cases, as in the United States or Canada, it may also delay the aging process, as a result of the specific contribution of young immigrants. On the contrary, those countries that allow large numbers of people to emigrate may consider such strategies as a means to alleviate demographic pressure. Permanent migration, which is now usually due to economic factors, still has significant demographic consequences. We return to this subject in Chapter 110, which is on migration policies.

Women's education has significant influence over fertility, in that it entails such effects as postponing first marriage, bringing the demand for children to lower levels and increasing contraceptive use (United Nations, 1995e). Any measure that encourages women's employment also has an indirect, though significant effect over decline in fertility rates.

3. Policies Aimed at Increasing the Size of the Population (or Slowing Down Its Decline)

Policies aimed at increasing population size are usually adopted by governments that are concerned with the negative evolution of the population growth rate (or even an actual decrease in the size of the population). Several methods and mechanisms (financial, fiscal, social ones) can be used to implement such policies, while taking into consideration the various factors that influence them (see Appendix 1).

Many different financial and fiscal means can play a role in encouraging population growth—low interest rate loans, benefits for young people and large families; benefits and assistance to working mothers (e.g., by allowing mothers to have paid leave), benefits for families based on the number of their children; cutting income taxes (in proportion with family size); allowances for the birth of a child (with amounts increasing proportionately to the size of the family); subsidizing day care and nursery schools.

Many social measures (concerning housing, schools, health care, and employment) can also be taken with a view to encouraging population growth—for example, the creation of special conditions for young couples renting or buying an apartment or house, access to information on modern contraceptive methods, or measures encouraging part-time employment.

Most of those measures or provisions are part of the so-called welfare state, which is currently being reappraised, and consequently should be reexamined in many developed countries, particularly in Europe.

Letting immigrants in is another possible option aimed at increasing the size of the population, though the issue is now an extremely controversial one, particularly among developed countries.

The psychological environment may also play a part in the success of population policies, particularly when they are aimed at increasing the population. Moreover, where raising children is concerned, the parents' expectations play an important part. In a society in which employment is so uncertain and the social cost of a child so high (in terms of education), parents are psychologically deterred from increasing the size of their family, even though their living standards may be quite acceptable.

In trying to stimulate birthrates, pronatalist policies sometimes aim at countering certain negative effects of population decline, particularly in relation to labor supply, human capital, and cultural heritage.

The Soviet Union under Stalin (in the 1930s) provides one example of implementing a staunchly pronatalist policy. Pursuant to the Marxist theory—which in fact does not differ much from mercantilist views, at least on this point—Stalin's pronatalist policy was based on the idea that population growth could be viewed as a stimulus to economic progress. In recent years, pronatalist policies have been implemented, to varying degrees, not only in such countries as Taiwan, South Korea, and Singapore (as will be expanded on in Chapter 103), but also in some extremely poor African countries.

4. The Adoption of Population Policies and Their Efficiency

An increasing number of countries are concerned with their population growth rates, and adopting policies aimed at solving the problem; and in international fora more and more developing countries have been expressing their preoccupation with the consequences of population growth (United Nations, 1996).

Though the proportion of countries that wish to effect major changes about the pace of their population growth has not much changed (Table 101–1), that of those countries who deem their growth rate exceedingly high has much declined: only 25% of the countries thought so at the beginning of this century, as against 83% only five years before!

The perception of fertility levels is approximately the same. The small proportion of countries that find their level of fertility satisfactory has significantly increased, rising from 11% to 18% in 25 years (Table 101–2). However, the proportion of countries who deem their fertility rates too high has also sharply risen to as much as 45% in 1996 (those countries represent 70% of the world population) (United Nations, 1997–2002). The evolution is clearly in close relation to the fact that fertility is declining in developing countries just as a growing number of governments are enforcing fertility regulation policies.

Health policies affect mortality and morbidity, and in the last few decades a favorable evolution of those two demographic variables has been observed. Life expectancy has been increasing everywhere in the world since World War II. However, few countries consider that the mortality level is wholly satisfactory. A survey by the United Nations on governments' atti-

TABLE 101–1 Governments' Views on Their Population Growth Rate (1976–2001): The Proportion of Countries (%)

Year	Satisfactory	Minor change desired	Major change desired	Total	Number of countries
1976	23	45	32	100	150
1986	16	45	39	100	164
1996	15	44	41	100	193
2001	15	44	41	100	193

Source: United Nations, 2002.

TABLE 101–2 Governments' Views on Fertility Levels (1976–2001): The Proportion of Countries (%)

Year	Satisfactory	Minor change desired	Major change desired	Total	Number of countries
1976	11	53	37	100	150
1986	13	46	41	100	164
1996	15	40	45	100	193
2001	18	38	44	100	193

Source: United Nations, 2002.

tudes toward mortality rates (United Nations, 1998) has shown that in 1992, 67.4% of countries considered their level of life expectancy as unacceptable. Most of those countries were developing ones, and their life expectancy levels are still quite far below those prevailing in developed countries.

In the mid-1970s the vast majority of governments (89%) considered that the spatial distribution of the population was inadequate (Table 101–3); the policies conducted in the last 25 years, however, have reduced such dissatisfaction, as the proportion of countries that wish to encourage further evolution in the area has gone down to 66%. Moreover, the majority of the problems that governments have had to face resulted from difficulties in dealing with urbanization, and the policies aimed at slowing down the growth of large cities have generally proved ineffective. Above all, it has been admitted that "the major urbanization problems of developing countries have little to do with the absolute size of the individual cities, but result from the failure to manage rapid urban population growth" (United Nations, 1993, p. 38).

The significance of censuses and surveys in formulating coherent population policies will depend on how often they are conducted (10-year intervals between censuses cause serious problems), and how comparable and analyzable they are. Another crucial element, particularly for developing countries, is the governments' level of commitment to population policy, as reflected in the way they develop national plans.

Two important aspects should be briefly mentioned: the relationship between population and development policies on the one hand, and population policies and environment on the other hand.

Development and environment are taking an increasingly important part in developing and formulating population policies. The relationship between population policies and environment has also been included in point 3.27 in the Programme of Action adopted at the International Conference on Population and Development in Cairo in 1994: "Implementation of effective population policies in the context of sustainable development, including reproductive health and family planning programs, require new forms of participation by various actors at all levels in the policy-making process" (United Nations, 1995a, p. 19–20). Unfortunately, the 2002 Johannesburg Conference failed to mention population issues in any explicit way.

TABLE 101–3 Governments' View on the Adequacy of the Geographic Distribution of the Population (1976–2001): The Proportions of Countries (%)

Year	Satisfactory	Minor change desired	Major change desired	Total	Number of countries
1976	11	37	52	100	150
1986	11	43	46	100	164
1996	29	30	42	100	192
2001	34	27	39	100	193

Source: United Nations, 2002.

The relationships between population and development are ambiguous. They depend on the social system, which may motivate individual, family, and social group strategies in various ways. At national level, development may have little influence over population growth in the short term. At lower levels, even over short periods of time, development plans that affect job creation may have strong influence on the evolution of the population through migration (United Nations, 1993, p. 6).

Another aspect of population policies currently under discussion is how they relate to local authorities. Many decisions relating to population issues depend on the local context (at the village, town, or region level) and, due to subsidiarity, efficiency requires a degree of decentralization in implementing them. The parties involved, the methods, and the areas of intervention of population policies clearly vary according to which levels are concerned. At the world and major regional levels, the United Nations and the cooperating institutions of the major zones (based on common geopolitical interests) are concerned with the large international migratory movements as well as the impact of conflicts and major shortages. At intranational levels, local authorities (from communes to federated States) and certain private institutions (e.g., large corporations) take an interest in those decisions that may have repercussions on the future of their populations (or workers, or customers). However, this chapter is focused on the national level, and we consequently deal mostly with governments' population policies, though it should be kept in mind that the most appropriate solutions to the problems and conflicts of interests caused by demographic evolution are not always to be found at the State level only, particularly where migration is concerned. The success of population policies implies political commitment, at local, national, and international levels.

5. Examples from Some Developing Countries

During the past two decades, the Governments of many developing countries have recognized the importance of for-

> mulating and implementing comprehensive national population policies. Many countries, however, do not have effective national population policies, while others have formulated policies that are not broad-based, comprehensive, integrated, multisectoral and adaptable to their socio-economic needs. Furthermore, although some countries have officially enunciated a population policy, it is not strongly supported by policy makers, planners, government officials and/or the general public. Other policies have not succeeded because they are not backed by proper legislation, plans of action, institutional arrangements and effective programmes and projects. (United Nations, 1993, p. 203)

In developing countries that are experiencing economic difficulties serious enough to affect the development process, it is particularly difficult to formulate any development program. For example, the implementation of authoritarian family planning methods in India only yielded rather uncertain results from a demographic point of view (see Chapter 113, which is on India). Likewise, in neighboring China, imposing family planning as a basic national policy required strenuous efforts, before the "one couple, one child" principle was eventually enforced in 1979 (see Chapter 112).

Despite the ethical and political reservations that necessarily come up concerning some of the measures that were taken, it must be admitted that China's population policy has been more efficient than India's, at least once it is acknowledged that the mortality and fertility transition in the former, which occurred much faster than in the latter, was a result of its population policy. However, such success perhaps was due to the economic, social, and cultural environment, rather than the population policy, strictly speaking. The example of India (particularly as far as certain States, such as Kerala, are concerned), as well as that of China, shows that the effects of population policies are all the greater as they are backed with such balanced economic growth as equitably benefits the different social classes. If the evidence of economic and social development is to be accepted as a prerequisite for an efficient population policy, then anti-Malthusian positions are eventually challenged. The evolution has been very clear in many countries, and still more so at the level of international relations (as is discussed in Chapter 103 of this volume).

Actually, except for China, very few governments have adopted coercive fertility policies, whether for ethical reasons or for the sake of efficiency. Most have chosen the opposite method, based on incitement and persuasion. In the latter part of this section are important illustrations of this, through various country case studies (Chapters 112–119).

6. Concerning Developed Countries

Two different concerns have characterized population policies in developed countries in the last few decades: foreign immigration and fertility decline, both of which have direct consequences over population aging. Consequently, most population policies explicitly adopted in those countries focus on those two aspects.

The case of France (which is discussed at length in Chapter 119) provides a good example of a country that implemented a pronatalist policy. To various degrees, other countries, such as Belgium, Luxemburg, Germany, and the Netherlands, also have chosen to enforce the same policy. Roland Pressat (1991) has very aptly summarized the situation of France:

> The French family policy is still currently broadly inspired from the principles contained in the Family Code, though a very large number of amendments and additions have been made to the earliest provisions established during the war. At the heart of the system lies the principle of various benefits paid in cash—prenatal or maternity benefits, family allowances from the birth of the second child on. . . . (p. 34)[1]

In the past the debate on population policies often took an ideological turn, between Malthusian and anti-Malthusian proponents. Only in recent times have the United Nations conferences on population (in 1984 and 1994) eventually achieved international consensus on the fact that population policies (or social policies that affect population dynamics) form a decisive part of development policies.

III. THE ROLE OF INTERNATIONAL ORGANIZATIONS AND AID IN PROMOTING POPULATION POLICIES

Though we shall not anticipate Chapter 120, in which the action of the international community is expanded on at length, this introductory chapter on population policies just cannot be concluded without mentioning, if only briefly, the part played by international organizations in that area.

The most comprehensive analysis of population policies at the world level was conducted by the United Nations (1996, 2002), which tried to summarize

[1] La politique familiale française s'inspire encore actuellement dans les grandes lignes des principes contenus dans le code de la famille et cela après les très nombreuses retouches et compléments apportés aux premières dispositions prises pendant la guerre. Le principe de l'octroi de diverses prestations en espèces est au coeur du dispositif: allocations prénatales, de maternité, allocations familiales à partir du 2e enfant. . . .

every government's perceptions of and every country's approaches to population policies. Every general policy is described in its institutional framework. Furthermore, demographic evolution, the size and age structure of the population, mortality and morbidity, family and fertility, international migration and spatial distribution, urbanization, and women's status are briefly described. That comprehensive survey, published in three volumes, makes it possible to compare the direct measures adopted by the different countries, and through which they implement their population policies.

Considering the nature and contents of population policies, as could be expected, many international organizations (particularly those of the United Nations system) are stakeholders in those issues. Most prominent is the UN Secretariat, through the Population Division of the Department for Economic and Social Information and Policy Analysis (DESIPA) in New York, while the United Nations Population Fund (UNFPA), whose headquarters are also in New York, is the most substantial source of aid funding for population-related matters. The United Nations Development Programme (UNDP) in New York offers a large number of aid programs to developing countries. Among the specialized agencies, the World Health Organization (WHO), the International Labour Organization (ILO) and the International Organization for Migration (IOM), all three of which are based in Geneva, and the Food and Agriculture Organization (FAO), whose headquarters are in Rome, have paid particular attention to the issue of population policies. The World Bank also plays an important financial role in population and development policies. "An increasing number of countries actively seek to work with the Bank on population policies—and borrow from it as well" (World Bank, 1994). At the regional level, the Council of Europe, based in Strasbourg, France, must be mentioned, as well as Eurostat for the production and dissemination of demographic data.

Many non-governmental organizations (NGOs) have played a vital role in population and development policies. They were pioneering in that area at a time when governments, restrained as they were by political, ideological, or bureaucratic considerations, were not in a position, or were unwilling, to take action (Sadik, 1991, p. 245). They played a crucial role in many countries in the implementation of aid programs. Organizations such as the International Planned Parenthood Federation (IPPF), the Population Council, the Pathfinder Fund, and the Rockefeller Foundation have been using extensive networks to disseminate their ideas and programs; on the other hand, religious organizations have had significant influence on population policies (or decisions) in many countries, and not only those where religion is a State institution. Religious organizations, to various degrees, are actively in favor of population growth.

The specific international conferences of the United Nations that have been held every 10 years since 1974 have also drawn the attention of governments to population policy issues; in fact, it was only after the 1970s that these issues were first recognized as essential in the development process. During the first conference, which was held in Bucharest in 1974, the first Population World Action Plan was adopted, and a relationship between population and development was thus officially recognized for the first time. During the Mexico City Conference in 1984, the Action Plan was amended so that the demographic evolution that had occurred since 1974 could be taken into account. The most important point included in the recommendations was on the development of population policy objectives. The 1994 Cairo Conference was titled "International Conference on Population and Development" (whereas for the first two conferences the word *development* was not mentioned in the official title, which included only the word *population*). For the first time the issue of population was viewed as part of the larger development issue (Najam, 1996), and the Conference adopted an Action Plan covering the next 20 years. The idea of one more conference on population in 2004 was rejected by the relevant authorities in the United Nations system.

Acknowledgment

This chapter was translated from the French by Mireille Rabenoro.

References

Cagiano De azevedo Raimondo, 1991. Osservazioni sulle politiche di popolazione, *Affari Sociali Internazionali*, vol. 19(2), p. 5–30. (Franco Angeli, Milano).

Dasgupta Partha, 1995. The population problem: theory and evidence, *Journal of Economic Literature*, vol. 33(4), p. 1879–1902.

Najam Adil, 1996. A developing countries' perspective on population, environment and development, *Population Research and Policy Review*, vol. 15(1), p. 1–19.

Pressat Roland, 1979. *Dictionnaire de démographie*. Paris, PUF, 295 p.

Pressat Roland, 1985. *The dictionary of demography.* Translated and edited by Christopher Wilson. Oxford and New York, Blackwell Reference, 243 p.

Pressat Roland, 1991. La population de la France jusqu'à la deuxième guerre mondiale, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European Population. I. Country analysis*, p. 19–39. Paris and Montrouge, INED, John Libbey Eurotext Ltd., 450 p.

Sadik Nafis (ed.), 1991. *Population policies and programmes—Lessons learned from two decades of experience.* New York and London, New York University Press, 464 p.

Sauvy Alfred, 1966. *Théorie générale de la population*, 2 vol., 3rd ed. Paris, PUF, 373 p. and 403 p.

United Nations, 1986–1993. *Population growth and policies in mega cities*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, 12 booklets (A set of studies on Cairo, Calcutta, Manila, Bombay, New Delhi, Dhaka, Bangkok, Madras, Karachi, Jakarta, Mexico City and Sao Paulo).

United Nations, 1987, 1989, 1990. *World population policies*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 3 vol. *Vol. I: From Afghanistan to France*, 247 p., *Vol. II: From Gabon to Norway*, 254 p., *Vol. III: From Oman to Zimbabwe*, 264 p. (Population Studies n° 102, ST/ESA/SER.A/102).

United Nations, 1993. *Population policies and programmes. Proceedings of the United Nations Expert Group Meeting on Population Policies and Programmes. Cairo, Egypt, 12–16 April 1992*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 268 p.

United Nations, 1994. *Programme of action of the conference, a note from the Secretariat.* New York, United Nations, 85 p. (Draft programme of action of the International Conference on Population and Development).

United Nations, 1994. *Programme of action of the United Nations International Conference on Population and Development.* New York, United Nations, 100 p. (Cairo).

United Nations, 1995a. *Population and development. Vol.1, Programme of action adopted at the International Conference on Population and Development, Cairo, 5–13 September 1994*. New York, United Nations, Department for Economic and Social Information Policy Analysis, Population Division, 100 p.

United Nations, 1995b. *Population consensus at Cairo, Mexico City and Bucharest—An analytical comparison*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 166 p.

United Nations, 1995c. *Report of the World Summit for Social Development, Copenhagen, 6–12 March 1995*. New York, United Nations, 154 p. (Document A/Conf.166/9).

United Nations, 1995d. *Report on the International Conference on Population and Development, Cairo, 5–13 September 1994.* New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 193 p. (Document A/CONF.171/13/Rev.1).

United Nations, 1995e. *Women's education and fertility behaviour: recent evidence from the Demographic and Health Surveys*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 113 p. (Document ST/ESA/SER.R/137).

United Nations, 1996. *Global population policy, Database 1995.* New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 214 p.

United Nations, 1997–2002. *World population monitoring*. New York, Commission on Population and Development and Population Division.

United Nations, 1998. *World population monitoring, 1997*. New York, United Nations, Population Division, 204 p.

United Nations, 2002. *National population policies 2001.* New York, United Nations, Population Division, 427 p.

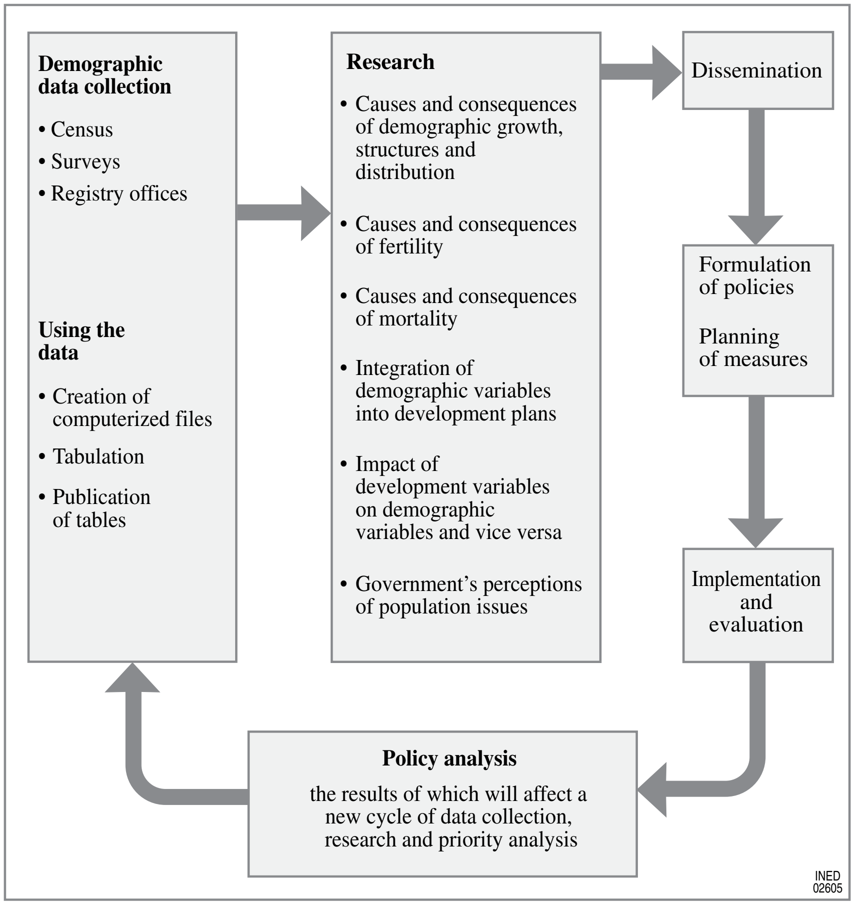
World Bank, 1994. *Population and development: implications for the World Bank*. Washington, World Bank, 134 p. (Development in Practice).

APPENDIX 101–1 Table of Factors, by Sector, that Influence Population Policies

Agriculture	Soil preservation and land tenure
National defense	Recruitment and distribution of troops
Economic policy	Natural resources conservation, full employment, investment adequacy, land development and management
Education	School age limits, reduction of obstacles to girls' education
Finance	Fiscal systems (in particular, tax deductions for families), direct and indirect taxes, local taxes, access to credit and income opportunities, particularly for women
Industry and commerce	Concentration of companies, decentralization of industrial zones
Home affairs	Population census, registry offices, policies concerning abortion
Justice	Legislation on marriage and divorce
Housing and town planning	Building low-rent blocks of flats, distribution of flats, rent regulations and town planning management
Public health and population	Health policies (particularly directed at mothers and children), immigration, naturalization, abortion, contraception, population policy coordination
Employment	Full employment, welfare system, women's working conditions, pension systems, immigration, training programmes
Public works and telecommunications	Fares for transport and communications public services (trains, postal services), infrastructure
Civil service	Staff recruitment and distribution

Source: Adapted from an original table by Alfred Sauvy (1966).

APPENDIX 101–2 The Sequence of Activities in Population Policy Planning



Source: Nafis Sadik (1991, p. 4).

P A R T

II

FROM FERTILITY CONTROL TO FERTILITY SUPPORT

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Paul Demeny pointed out in Chapter 98 that two big developments converged to focus population policy thinking on population growth issues in the aftermath of World War II. In the northern sphere, the baby boom temporarily masked fundamental questions already being asked in some countries, especially France, about what happens after demographic transition. But, in the wider world more especially, the onset of very rapid transition in Third World countries propelled the problem of rapid population growth to the top of the agenda, as spreading health improvements in these countries led to a faster rate of mortality decline than ever observed in any developed country, and a subsequent population explosion. The focus of concern then turned to excess fertility, to the point where northern public opinion was apt to accuse the South of endangering planetary equilibriums through unrestrained reproduction. This was in clear disregard of the fact that developing countries, like northern ones historically, needed time to add the new concept of fertility control to their very recent health improvements. Even so, this context focused population policy on fertility control, if not basic family planning.

This volume on population thought and policies plainly had to break away from this narrow and skewed approach; hence the historical chapter (Chapter 100) reviewing developments in population policies before 1940, placed at the end of Part I. More especially, however, in the belief that fertility is anything but the one issue of a population policy, Parts III and IV also separately address the other spheres that population policies cut across: health, migration, and age structures. But, because their dominance of the agenda in recent decades has tended to overshadow the fact that demography is not synonymous with fertility, this part (Part II) first addresses policies that deal with fertility and the family.

Until recently, the issue has been essentially very different in the developing and developed countries. Over the past half-century, the (presumptively excess) fertility of developing countries is what has exercised the international community, whereas low developed country fertility has, in most cases, only become a much more recent government concern. In fact, there are two different issues here, which are addressed separately: the developing countries are addressed in Chapters 102 and 103, and the developed countries in Chapters 104 and 105.

In Chapter 102, Thérèse Locoh and Céline Vandermersch give a wide-ranging review of fertility control policies implemented in Third World countries. It is an issue that has long been the focus of more ideological than scientific debate, but has ultimately been universally acknowledged as essential to economic and social development, notwithstanding that, at the same time, birth-control policies proper have had less real impact where not closely connected with social and cultural improvements in society in particular. But with or without policies, the second phase of demographic transition—fertility decline—is now everywhere under way, and well advanced in many countries. In some, it has even exceeded policy aims of replacement-level fertility. Some countries that were still only recently grappling with too-rapid population growth are now concerned at the prospect of population loss, and some have already made policy U-turns. This is the issue addressed briefly by Magali Barbieri in Chapter 103.

It is an issue that has long bedeviled many European countries, some of whom have long-standing fertility-support measures in place. Two questions remain open to debate, however: whether to let matters take their course, or try and curb the fall, or even reverse the trend, to restore a replacement capacity that has been largely eroded by current developments? Gustavo de Santis charts the history of fertility-support policies in the industrialized countries in Chapter 104. The other question is whether it is possible to pursue really effective policies that also respect fundamental human rights? A meticulous demographic analysis by the late, lamented Gérard Calot of the results of pronatalist policies in some countries where they were most openly pursued addresses this issue in Chapter 105.

C H A P T E R

102

Fertility Control in Third World Countries

THÉRÈSE LOCOH AND CÉLINE VANDERMEERSCH

Institut national d'études démographiques (INED), Paris, France

The fertility decline that has occurred in the majority of developing countries is indisputably one of the major events of the last 50 years (see Chapter 69, Vol. III). Concern regarding the rapid growth of world population began at the end of World War II. The developed countries were especially worried by the deceleration in their growth and had, up to that point, shown little interest in the developing countries which, at that time, constituted two-thirds of the world population and were not yet known as the Third World. From the 1960s, population forecasts established by the United Nations' Population Division have shown the growth potential of all the underprivileged countries of the planet, and over the last half-century world population has increased at the unprecedented rate of nearly 2% per year (see Chap. 77, Vol. III). Nevertheless, the increased control of fertility currently witnessed everywhere is occurring at different rates, and present situations are extremely diverse (see Chap. 69 and 71, Vol. III). At the beginning of the 21st century, the annual population growth of Third World countries is still, on average, 1.5% and as high as 3% per year in some of them. On a map of the world, Africa clearly appears as the last continent with high fertility (see Figure 69–9 of Chap. 69, Vol. III). Overall, Asia and Latin America have fertility levels that are significantly lower. There is no longer a uniform Third World with high fertility that can be compared to a developed world that has completed its demographic transition; we now have Third Worlds. Some countries have fertility levels that are comparable with those of the old European and North American countries; others, mainly in Africa, continue to value large families and still have relatively stable fertility levels.

This evolution, which we will discuss briefly in the first section of this chapter, is largely a result of general development policies that caused a decline in mortality, mainly in child mortality; an increase in education levels, particularly those of women; and an improvement in health and social infrastructures; but it is also the result of policies and programs created to reorient the fertility levels that we will discuss in this chapter.

In the second section, we will trace the evolution of fertility policies and the central role played by the United Nations which, during the four large conferences on world population, made an essential contribution in raising awareness of the population dynamic of developing countries, in developing theoretical knowledge on the factors that determined it, and in implementing national and international birth control policies and programs. We will then demonstrate how governmental opinions regarding policies on fertility control have developed. In the third section, we will discuss the means by which numerous countries implemented fertility control policies, in the forefront of which is the spread of contraceptive methods within family planning programs. In the fourth section, we will analyze the evolution of contraceptive practices in the three continents of the Third World, using a few countries as examples. In conclusion, we will mention some of the emerging issues regarding fertility control policies, for the governments of Third

World countries as well as for the whole international community.

I. SINCE 1950 FERTILITY HAS DECREASED BY HALF IN THIRD WORLD COUNTRIES

In 1950, the total fertility rate (TFR) was a world average of 5 children per woman. In 2002, it was only 2.8. The average completed fertility of women has therefore more or less been halved in 50 years, mainly due to the decline that has occurred in Third World countries (Table 102–1) where the total fertility rate has decreased from 6.2 to 3.1 (United Nations, 2002c). This is an extraordinary change that presupposes considerable changes in family dynamics and individual behavior in these societies. Figure 102–1 clearly illustrates the scale of the trend in the 1950–2000 period. Admittedly, sub-Saharan Africa and some Middle Eastern countries have so far resisted the general decline, but their fertility transition also began a few years ago (Locoh and Hertrich, 1994). East Asia and Latin America are the two developing regions for which the changes in fertility have occurred most rapidly.

Figure 102–1 summarizes the 50 years of change (see also Appendix Table 102–1). There are considerable differences between continents, and between subregions within the continents. But what is particularly striking is the similarity of the levels at the beginning of the period, between the different regions of the developing world (all with a TFR of between 5.5 and 6.5 children per woman), and the divergence in their current situations which have led to a TFR of 1.8 in China to 5.6 on average for the sub-Saharan African countries as a whole.

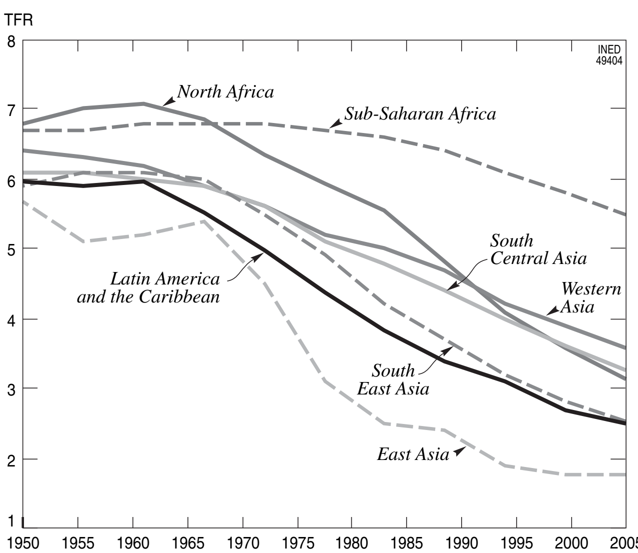


FIGURE 102–1 Total fertility rate, regions of the developing countries, 1950–2005. (Source: United Nations, 2003a.)

In Latin America, temperate South America (Argentina, Chile, Uruguay) has been characterized by relatively low fertility (3.5 children per woman) since the 1960s. But Central America and tropical South America at that time still had a rate of over 6 children per woman. Mexico only began its rapid fertility transition in 1970. The Caribbean began its transition with 5 children per woman. Fifty years later, in a rapid movement, the differences had diminished and in 2002 the total fertility rate remained at 3 children per woman only in Central America (United Nations, 2003a). Thus, significant gaps remain between the countries of the Latin American subcontinent. Guatemala, Honduras, Haiti, Bolivia, and Paraguay have rates of over 4 children, while Cuba, Trinidad and Tobago, and other small states of the West Indies have rates below 1.8 children per woman, which places them with the countries that are at the end of their fertility transition (United Nations, 2003b).

The changes were even more rapid in Asia, where the average rate was of 2.6 children per woman in 2002. The country that first springs to mind is the immense country of China which, following an initial unsuccessful attempt to introduce family planning (during the 1960s, see below), implemented a draconian program of birth limitation in 1970. This country, or rather this subcontinent, has succeeded in decreasing its fertility from more than 6 children per woman to fewer than two 2 between 1965 and 2000 (1.7 children in 2002). Other countries in East Asia also have experienced similarly spectacular changes. This is the case of Taiwan and South Korea (TFR of 1.3 according to the most recent estimate). Fertility levels had already started to decelerate in Hong Kong in the 1950s. The current fertility level is the outcome of a spectacular decline (1.2 in 2000).

TABLE 102–1 Trend in Total Fertility Rate (Number of Children per Woman)

Region	1950–1955	1960–1965	1970–1975	1980–1985	1990–1995	1995–2000
World	5.0	5.0	4.5	3.6	3.0	2.8
Developed countries	2.8	2.7	2.1	1.9	1.7	1.6
Developing countries	6.2	6.0	5.4	4.1	3.4	3.1

Source: United Nations, 2003a.

Generally, Southeast Asia is also following the same process if we disregard disaster-stricken countries such as Cambodia (TFR of 4 according to 2002 estimates), or very isolated countries such as Laos (TFR of 4.9 in 2002). Vietnam (TFR of 2.3 in 2002) is very advanced on the path to fertility control, as the example of its great Chinese neighbor was quickly followed in North Vietnam with the adoption of public health and family planning programs, extended since 1973 to South Vietnam.

In South Central Asia, that is, the Indian subcontinent, the fertility decline has been slower and there are numerous disparities, including between the Indian States. This country, which like China is a subcontinent in itself, is a mosaic of people and cultures, and the differences in fertility levels reflect this. The northern Muslim states still have fertility levels of 4 children per woman, while the state of Kerala, where the income level is more or less comparable, has a total fertility rate below 2 children per woman (IIPS and ORC Macro, 2000). In the region, Sri Lanka has been a frontrunner, with a sustained practice of family planning since the 1970s. Fertility, which has decreased regularly, was only 2 children per woman in 2000 (United Nations, 2003b). Another spectacular example is that of Bangladesh, a country that the media periodically projects as one of the poorest countries in the world; there the total fertility level declined from 6.4 children per woman in 1970 to 3.6 according to the most recent estimates (United Nations, 2003a). This is an example that contradicts the idea that only an improvement in economic conditions can lead to a deceleration in fertility (Adnan, 2002).

In Western Asia, changes in fertility have been slower, decreasing nevertheless from 6.4 children in 1950–1955 to 3.8 children according to a 2002 estimate made by the United Nations. Several countries (Cyprus, Israel, Turkey, Lebanon) now have fertility levels below 3 children per woman, while Yemen, Iraq, and the Palestinian territories remain, for diverse historical reasons, countries with high fertility (5.5 children and more in 2002). The former Soviet republics that belong to this geographic zone (Armenia, Azerbaijan, Georgia) have particularly low levels of fertility, under 2 children per woman. Iran, with a very rapid decline (from 6.5 children in 1980–1985 to 2.5 in 2002) contradicts another generally accepted idea according to which the Islamic religion is a systematic obstacle to fertility reduction. (Tunisia and the other North African countries also contradict it.)

Africa remains the continent with the highest fertility levels. Since the 1980s the first fertility declines have been observed in some sub-Saharan African countries and in all the North African countries. The continent is no longer as homogeneous as it was in the past, in terms of fertility levels. West Africa (the Sahel, the countries on the coast of the Gulf of Benin, including Nigeria, the most populous African country with its 115 to 135 million inhabitants) and East Africa (the Great Lakes, Tanzania, Zambia) are the two regions where fertility remains high (on average around 5.5 children per woman) despite the first signs of decline which have been observed, particularly in Kenya (see below). Southern Africa (South Africa and neighboring countries) is resolutely committed to fertility decline. It has decreased from 5.7 children on average in 1950–1960 to 3.3 children in 2000. Central Africa is a paradoxical case, where fertility has traditionally been lower than in the other African regions. It experienced a slight increase in the 1950–1990 period. This increase has been attributed to the improvement in health systems, which has resulted in a decrease in the incidence of sterilizing diseases that were widespread in some provinces of these countries. Finally, North Africa, which was still more fertile than the rest of Africa around 1950 (more than 7 children per woman on average), now has fertility levels close to 2 children per woman. Fertility in Tunisia has experienced a sustained decline since 1966 when it opened the door to modernity; Tunisia's success has been ensured by social development programs associated with one of the Third World's first family planning programs (around 1965). The other two North African countries, in very different political situations, are in the process of catching up with Tunisia.

The fact that nearly all of the African countries are currently involved in this process of fertility decline, even though some of them have made much more progress than others, confirms the universality of the second phase of the demographic transition, that is, fertility decline, which generally responds to the mortality decline that occurred during the first phase (see Chap. 68 and 69, Vol. III).

This decline in fertility in the regions of the world where, up until the beginning of the 1950s, large families were the rule, is revealing of the social and economic progress that has occurred since World War II. It is no surprise that, with a few exceptions,[1] the poorest countries are the last in this process.

Societies have slowly accumulated fertility control knowledge and practices. But the effective techniques that are now available, even in the poorest countries, were not at our disposal before the second half of the

[1] The oil-producing countries of the Persian Gulf (Saudi Arabia, Iraq, Kuwait, Oman), which are among the countries with the highest gross national product (GNP), are also countries with high fertility.

20th century. And since the 1950s the international community, expressing itself through large conferences organized by the United Nations and its agencies, has seized the issue of fertility control and has progressively included it in its programs and policies.

II. THE ELABORATION OF FERTILITY CONTROL POLICIES

After a brief reminder of the emergence of policies for controlling fertility, we will discuss the key role played by international organizations in their development, the arguments in their favor, and the evolution in governmental attitudes.

1. The Emergence of Fertility Control Policies

a. From Ancient Practices . . .

While the populations of developing countries have always been concerned with controlling their fertility, up until the 20th century it was most frequently in order to guarantee their survival and if possible their growth, in very unfavorable conditions of high mortality and all kinds of hazards (e.g., climatic events, epidemics, raids, or wars) that affected them. On the whole, high fertility was preferred to ensure the renewal of clans, lineages, and family groups that shared common objectives. As formerly in the countries of the North, before the initial progress made on mortality, families and couples were generally concerned more with having a sufficiently large family than with limiting it.

Historians of ancient populations have testified to concerns regarding birth limitation in some societies, well before modern times. Infanticide, the exposure of newborns, and induced abortion have existed in many societies of the North and the South as a means of avoiding imbalances between the production of food and the number of mouths to feed. In China, for example, studies carried out on fertility within marriage tend to prove that fertility was, during the pretransitional period, lower than it was in Europe, and we have proof of the existence of birth limitation practices, induced abortions, and the use of various substances for avoiding pregnancies (Zhao, 2002; Campbell, 2002).[2]

These studies support the hypothesis that birth limitation measures were present, at least in some circles, in pretransitional populations. Withdrawal was also a method used within all societies and was practiced by couples well before any family planning program. The same is true for induced abortion, mostly illegal. In many societies couples practiced and still practice in some countries, particularly in sub-Saharan Africa (Page and Lesthaeghe, 1981), postpartum abstinence of varying duration, intended to ensure a greater likelihood of survival for the child already born by arranging a sufficiently long interval between two births. This was a clear expression of the desire to control fertility, well before modern birth control methods became accessible.

But in many Third World countries, as was the case in developed countries, the main lever for intervening in reproduction was without contest, until the 1960s, social control on the formation of first union by young adults. In some countries, for example in North Africa, the recent decline in first union formation is a major factor in the fertility decline (Ouadah-Bedidi and Vallin, 2000), as the early marriage of women in previous decades was an essential element in the very large families observed there.

b. . . . to the First Fertility Reduction Programs

The conjunction of the discovery of effective contraceptive methods and the special attention given by the international community from the 1960s onward to the speed of population growth, rapidly encouraged some Third World countries to create or simply support programs for encouraging fertility reduction[3] and family planning. From this period onward, some large countries announced objectives for fertility reduction. From the beginning of the 1950s, India, with the assistance of international experts, adopted the first large-scale family planning program in the Third World (in 1952). Fertility reduction was its main objective (see below, Chap. 113). China, following an initial inconclusive family planning program (1956–1962) committed itself at the beginning of the 1970s to a very coercive program for controlling sexuality and fertility, under the rule of Zhou en Lai, to limit population growth as quickly as possible (Attané, 2002). Political control and police control of the Chinese population was consequently used to achieve this national objective (see Chap. 112).

[2] "China has a long history of using medical substances and other measures to induce abortion and to prevent pregnancy. Potions were used to cause abortion 2000 years ago. Medical substances and prescriptions that were believed to prevent or terminate pregnancy or cause sterilization were detailed in many medical writings published during the last two millennia" (Zhao, 2002, p. 756).

[3] At least initially because recently, certain countries of the South have been showing concern at the excessive decline in fertility and have opted for policies to increase fertility (see Table 102–2). The question is reviewed in more detail in Chapter 103.

In Latin America, where several countries had already started their fertility decline in the 1970s, the most common objective was fertility reduction but also a more equal distribution of the benefits of development (Guzman, 1994). This was the case of Mexico, for example, where fertility control programs were based on ambitious health policies (see Chap. 114).

In North Africa, Tunisia is another example of early focus on fertility control policies and Habib Bourguiba, after some equivocation, took a clear position in 1968 supporting a slowing of population growth and the establishment of family planning services (Vallin and Locoh, 2001b; Gueddana, 2001). The same occurred in Kenya, which implemented a national family planning program in 1967 and in Ghana in 1969 where the first African document defining a population policy based on quantified objectives for reducing population growth appeared. However, in sub-Saharan Africa, with the exception of these two countries, family planning was barely discussed until the 1970s. Some private organizations, one of the main ones being the International Planned Parenthood Federation (IPPF), and denominational organizations (notably Protestant churches, by the means of health centers that they controlled) were the first to offer contraceptive services, and not without opposition. Faced with governments that were mainly reticent, the first programs, both private and public, announced only the objectives of well-being for families and couples and the possibility of having the desired number of children, rather than demographic objectives for fertility reduction. In the Francophone countries of West Africa, which were for a long time the most reticent toward any ideas of birth limitation, only "programs for family well-being" were spoken of until the 1980s, and in the same period in Zaire (now called the Democratic Republic of the Congo) only the "program of desirable births" was mentioned.

2. Toward Fertility Control Policies: The Role of the United Nations (1950–2000)

The idea of adopting policies with the aim of slowing population growth is a very recent one. It was initially introduced into international circles by the richest countries, at a time when, having largely finished their transition and significantly slowed their natural growth, they became concerned with the acceleration in the population growth of the poorest countries. The population policies in the Third World countries were, from the beginning, designed within the context of the complex relationships between rich countries and poor countries, which initially did not facilitate their adoption by the governments concerned.

Given the influence of the United Nations' large conferences (United Nations, 2003b) in the evolution of governmental positions, it is legitimate to trace the evolution of ideas regarding fertility control policies in five stages, separated by the large international meetings. They set the rhythm for the last half-century and provided reports on the debates that animated both the international community and the decisions adopted by population specialists and then by political leaders.

a. The 1950–1970 Period and the Awareness of the Speed of Population Growth in the Developing World

During the first years of the United Nations' existence there were still many gaps in demographic data, notably in a certain number of developing countries that did not traditionally record data on population. The possible link between population growth and economic development was not yet a widespread idea. However, the first requests for assistance in data collection came from Brazil, India, Indonesia, and Thailand (United Nations, 2003b, p. 11).

The first World Conference on Population took place in Rome in 1954. It was co-organized by the United Nations and the International Union for the Scientific Study of Population and brought together demographers, rather than governmental representatives, as was later the case (Lebrun, 1985). During this meeting, an ideological debate was initiated between the Western countries and representatives of the USSR, with the latter supporting the opinion that rapid population growth was only a problem in the capitalist countries and that the real issue was the unequal distribution of wealth between the wealthy and the exploited. Indeed, during the years that followed, the incentive to investigate the effects of rapid population growth in the Third World countries came from the United States. In a famous publication, Ansley Coale and Edgar Hoover (1958) estimated that in India, reducing fertility by 50% would, in one generation, increase income per capita by 40% more than in the absence of this fertility reduction.

In the early 1960s, several developing countries began paying more attention to population growth and adopting family planning programs. This was the case of India, first of all, whose government in 1959 gave its support to the adoption of all family planning methods, including sterilization. The Egyptian government implemented a national family planning program in 1962. Other programs were initiated by private organizations in Pakistan and Singapore. These different initiatives took place at the time when the United Nations Population Commission, in an important declaration on population policies, proclaimed that each country was responsible for establishing its own population policy and programs. It also declared that the United Nations should provide, to any government that requested it, assistance in managing population problems. In 1962, the United Nations General Assembly called for a special effort to be made in the research into the linkages between population growth and economic development.

This was what was done in the Belgrade conference in 1965 where, as in Rome, the participants represented themselves rather than governments. The same ideological separations were expressed, with some placing doubt on the legitimacy of the United Nations' involvement in population policies and particularly in the implementation of family planning programs. However, approximately 20 national family planning programs were presented and discussed during this conference (Macura, 1986). A few years later, the International Conference on Human Rights, held in Tehran in 1968, explicitly included access to family planning as a human right. The conference asserted that parents have the right to freely decide the number of children they would have and the spacing of their births. This declaration encouraged movements in favor of the spread of family planning, but it also gave governments formal notice to provide couples with the means of planning their families.

Parallel to these debates in the international arena, important progress was being made in the area of contraception. The contraceptive pill and the intrauterine device (IUD or coil) provided women with new possibilities for controlling their fertility, which would significantly increase and diversify the provision of family planning services.

b. 1970–1980: A Decade Marked by the Creation of the UNFPA and the Bucharest Conference (1974)

The 1970s began with the creation of the United Nations Fund for Population Activities (UNFPA) (recently renamed the United Nations Population Fund). Its mission was not only to develop research and knowledge as did the Population Division, but also to support activities likely to have a direct influence on demographic variables, in particular fertility. In 1972, more than 50 countries committed themselves to contributing to UNFPA's funding. This new decade was marked by a renewed approach to population issues within international circles, issues that have become a major stake in international development strategies.

The International Conference on Population held in Bucharest in 1974 was prepared from within this perspective. This time, it was not a meeting of experts only,[4] but of individuals commissioned by their countries to represent their governments' positions. An ambitious World Plan of Action for Population was conceived by the United Nations Population Division. Notably, it proposed quantitative objectives for the deceleration of population growth. This plan and the position of the Population Division caused impassioned debates between supporters of vigorous actions in favor of the adoption of family planning programs and those, mainly from Third World countries (in particular, a group led by representatives from Algeria and Senegal), who were strongly opposed to it. They were defending a firm ideological position which affirmed that it was development that would lead to a slowing of population growth, and not the reverse. The slogan "Development is the best pill" summarized their position. Furthermore, the issue of the sovereignty of the decisions of each state regarding population was a very sensitive one for these newly independent countries that did not want restrictions on their freedom to act, a freedom for which they had paid so dearly, by allowing quantitative objectives for the control of their population growth to be dictated to them.

The conference therefore ended without retaining the quantitative objectives proposed by the experts for slowing growth. The final declaration recalled that the ultimate objective of the World Plan of Action was development and not the control of population growth. Nevertheless, the idea of reacting to this growth acquired a new legitimacy during this conference. At the same time, an increasing number of developing countries started resorting to international population assistance, some for programs to integrate the population variable into development and others for programs to increase the provision of family planning.

c. 1980–1990: A Decade Less Marked by Ideological Positions, the Mexico City Conference (1984) Draws Lessons from Ongoing Programs

At the beginning of the 1980s, fertility showed signs of easing in numerous developing countries (other than in sub-Saharan Africa), although in the majority of cases, contrary to the slogan of Bucharest, this phenomenon could not be attributed to an improvement in the economic situation. The international commu-

[4] In parallel, they had their own forum and their debates would play an important role in the decisions of the political representatives present at the conference.

nity, having taken note of 10 years of experience of introducing family planning programs into various social contexts, began to discuss the association between population growth and development with more moderate positions than it had in 1974.

The Mexico City conference in 1984 was firmly prepared by four regional meetings. That of Arusha, for the African continent, was a very important stage, and the Kilimanjaro Declaration (CEA, 1984) was adopted there. It recalled the rights of States and couples to freely decide their options regarding fertility but also the legitimacy of family planning programs to fulfill the new aspirations of individuals regarding the size of their families. This position was adopted by the conference during discussions in Mexico City. The reference to the right of individuals to have access to contraception was greatly stressed and resulted in a consensus, including that of developing countries which, in Bucharest, had been reticent to the idea of supporting family planning programs.

Compared to the Bucharest debates, some surprising reversals of positions of some governments were observed during this conference. The African states, sensitized by the Arusha debates, became much more favorable to the idea of making family planning programs available. The same applied to some Latin American states which, up until then, had remained rather conservative on the issue, due to the influence of the Catholic Church. On the other hand, the United States, supporters in 1974 of the adoption of strong fertility control in developing countries, had a different position and declared that they would not support any family planning programs that included the practice of induced abortion. This was due to the United States government's concern with satisfying the conservative right-wing, which had been returned to power.

d. 1990–2000: From Family Planning to Reproductive Health: The Cairo Conference (1994)

The International Conference on Population and Development (ICPD), the last conference of the 20th century was held in Cairo in 1994 (United Nations, 1995). Armed with the experience of previous conferences, the United Nations Population Division actively prepared this meeting, preparing national reports on population, convening meetings of experts on priority issues and, via the United Nations regional representations, organizing meetings on each continent to encourage regional consensuses (United Nations, 1994). New partners including charitable nongovernmental organizations (NGOs) and feminist movements played a key role in this conference. The former came in force to present their action and to lobby government delegates. The latter used the preparatory meetings in a masterly manner to integrate their legitimate claims for women's rights and demands for equality into the new Cairo Plan of Action. Up to that point, in the official documents relating to world population, the improvement of women's status was included mainly as a method of achieving demographic objectives and, more particularly, family planning objectives. The Cairo document adopted a new tone that was firmer on the principle of male and female equality, more open with regard to areas in which the conference should intervene in favor of women, and more precise in the actions that were recommended. The concept of women's empowerment was the cornerstone of the argument. The concept of reproductive rights was proposed, and it was clearly declared that equal rights for men and women was an objective in itself (United Nations, 1995, Chap. IV) and not only one of the methods of ensuring fertility reduction.[5] Thanks to this new approach, the document bravely tackled a certain number of crucial problems: infanticide, rape, trafficking of women, sexual harassment, and more. In fact, this was the first population conference where sexuality was openly discussed.

Furthermore, reproductive health, the new concept developed in the Cairo program, which includes family planning, maternal and infant health, sexual health, and particularly the prevention of AIDS, has, since the Cairo conference, irrigated every program carried out under the aegis of UNFPA, which remains the United Nations' main organizer of population programs. Fertility programs are increasingly justified by arguments for health and their integration into the health system of developing countries is reinforced, notably by the synergy between AIDS prevention programs and family planning programs.

3. Fertility Control Policies: Why?

Over the last half-century, the United Nations Conferences have included debates based on figures but which were often also impassioned, bringing together researchers and decision makers, scientific objectives, and geopolitical aims. The fertility control of Third World countries has made great progress, but it still remains one of the objectives of the international community at the beginning of the 21st century. Family planning and fertility control programs draw their legitimacy from two main arguments: that of the

[5] The new directions in favor of equality between men and women which were adopted by government representatives in Cairo found increased support in the following year during the Beijing conference (1995).

control of population growth and that of health and well-being, which is popularized by the expression "reproductive health."

a. Reducing Population Growth

Inspired by the trend of Malthusian ideas, numerous demographers and economists since the 1950s have estimated that the planet's population growth, mainly that of developing countries, was untenable and should be controlled for fear of a serious imbalance between population and resources. The analysis of Ansley Coale and Edgar Hoover (1958) on India's development clearly illustrated this point of view. During the 1970s (with the Bucharest conference, 1974) supporters of increased fertility control based on the adoption of quantitative objectives for growth, contained in a world plan of action for population, put forward these demographic arguments, insisting on future imbalances between population and resources. Developing countries based their counterattack on this (see Point B-1, below). The theme of the growing tension between population and resources remains one of the major arguments for the legitimization of population programs and policies. It was recently enriched (notably following the Rio de Janeiro conference on the environment in 1992) by the concept of protecting the environment and the urgency of preserving global public goods, such as water and air, which although they are not measured in terms of commercial value, are no less essential to the survival of the planet. It is therefore no longer only the prolific populations of the South, which contravene the balance between population and available resources, but also the rich countries that are predators of the environment due to their consumption methods. This is a novel view on the balance between population and resources that the rich countries are unlikely to acknowledge.

b. Promoting Well-being and Reproductive Health

Since the 1990s and the Cairo conference, population specialists, under the aegis of the UNFPA, have taken reproductive health as the key concept. It integrates family planning in the more general perspective of the respect of the rights of individuals with regards to reproduction and health promotion of mothers and children, but also that of the population as a whole, through combating sexually transmitted diseases, one of the main ones being AIDS. This new direction for the legitimization of fertility control makes it possible to refocus the objectives of demographic control in developing countries within a global strategy of social well-being (Seltzer, 2002). In the context of the AIDS, epidemic which is devastating developing countries, and with the reinforcement of numerous studies that expose the links between the survival of children, women's health, and better control of fertility, this new perspective has garnered a large consensus. It has also made it possible for feminist movements to develop arguments based on the respect of human rights and the reminder that sexual and fertility choices are one of the principal expressions of these rights and that promoting reproductive health can only help in the struggle for equality between men and women (Locoh, 1997; United Nations, 2002c).

The justification for the adoption of fertility control programs is based on research that has exposed the beneficial effects of these programs in terms of the survival of women and children. The reduction of infant mortality, maternal mortality, and adolescent fertility, which should result from these programs, are also essential arguments for their spread.

4. The Evolution of Governmental Attitudes Regarding Fertility Control

Fertility statistics are, at the beginning of this new millennium, available in the majority of developing countries. In countries where the national structures for collecting data are deficient, the demographic and health surveys (DHSs) have nevertheless made it possible to obtain measurements of trends over the last 15 years. This is notably the case for sub-Saharan Africa. These data make it possible to measure the road traveled (Figure 102–1), both by the population that has changed its behavior and by the national organizations for bilateral and multilateral cooperation that incited the change. The debates aroused by the world conferences made it possible to compare experiences and to benefit from successes and failures. These debates played a major role in the evolution of programs for fertility control. Nevertheless, very different tendencies continued to inflect population policies in developed countries. Since 1976, the United Nations have been carrying out a period survey on governments' assessment of the demographic situation of their countries and their positions regarding fertility, mortality, and migration (United Nations, 2003b). In 1976, 47% of Third World countries estimated that their fertility was too high, while 45% judged it to be satisfactory. In 2001, 58% (60% in 1996) considered their fertility to be too high, and 35% considered it to be satisfactory (Table 102–2).

For 25 years, the responses of developing countries have been evolving toward an awareness of the responsibility imposed by their fertility dynamics (Table 102–3 and Figure 102–2). In the first survey, the highest proportion of countries that considered their fertility as being too high was in Latin America (Latin

America 59%; Asia 46%; Africa 38%). In 2001, numerous Latin American and Asian countries had reached a level that they considered satisfactory. In Asia, where spectacular fertility declines had occurred, according to the 2001 survey 17% of countries considered that their fertility was too low (compared to 5% in 1976). The proportion of countries that consider their fertility too high has increased most in Africa (from 38% in 1976 to 77% in 2001).

While in 1976, 67% of African governments, 56% of Latin American governments, and 32% of Asian governments declared that they did not intervene in fertility levels, in 2001 only 23%, 36%, and 24% of these governments, respectively, still had the same attitude (Table 102–4). Government intervention to reduce fertility has become common practice over the last 25 years. Of the 145 developing countries, 85 (59%)

TABLE 102–2 Third World Governments' Opinions (%) of Their Fertility Levels, 1976–2001

Opinion	1976	1986	1996	2001
Fertility too low	8	10	6	7
Satisfactory level	45	38	34	35
Fertility too high	47	52	60	58
Total	100	100	100	100
Number of countries	116	130	145	145

Source: United Nations, 2003b.

TABLE 102–3 Governments' Opinion (%) of Their Fertility Levels by Large Third World Regions (1976 and 2001)

Opinion	Africa		Asia		Latin America	
	1976	2001	1976	2001	1976	2001
Fertility too low	10	2	5	17	7	3
Satisfactory level	52	21	49	41.5	33	45
Fertility too high	38	77	46	41.5	59	52
Total	100	100	100	100	100	100
Number of countries	48	53	37	46	27	33

Source: United Nations, 2003b.

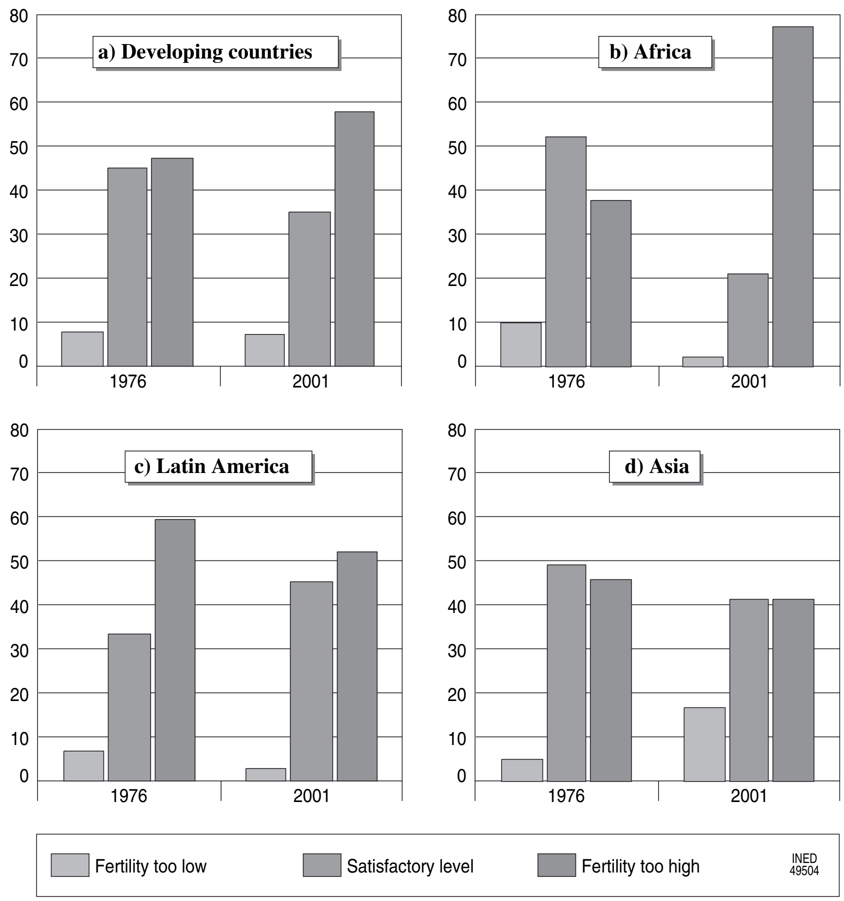


FIGURE 102–2 Governments' opinion (%) of their fertility levels, all the Third World and by large regions, 1976 and 2001. (Source: United Nations, 2003b.)

TABLE 102–4 Governments (%) That Do Not Intervene in Fertility Levels, by Large Third World Regions (1976–2001)

Year	Africa		Latin America		Asia	
	%	*N*	%	*N*	%	*N*
1976	67	*48*	56	*27*	32	*46*
1986	47	*51*	55	*33*	29	*37*
1996	23	*53*	39	*33*	24	*41*
2001	21	*53*	36	*33*	24	*41*

Source: United Nations, 2003b.

TABLE 102–5 Objectives of Fertility Intervention Programs, Declarations of Third World Governments (%), 1976–2001

Governments declaring a wish to:	1976	1986	1996	2001
Increase fertility	5	8	8	8
Maintain fertility	10	8	11	10
Decrease fertility	34.5	42	56.5	58.5
Not intervene	50.5	42	24.5	23.5
Total	100	100	100	100
Number of countries	*116*	*130*	*145*	*145*

Source: United Nations, 2003b.

declare having programs aimed at reducing fertility. There were only 40 (34%) of them in 1976.

It is a sign of the times and of the success of previous programs that the proportion of countries that declare having programs for encouraging fertility has slightly increased since 1976 (from 5% to 8% in 2001) (see Chapter 103) and they are all Asian countries (Table 102–5). However, this does not mean that every polemic regarding these programs has disappeared, especially concerning their demographic argument. But generally, family planning programs are at the very least included within the framework of the improvement of women's health and the respect of individual choice. According to the most recent survey by the United Nations (United Nations, 2002b), 84% of developing countries gave direct support to the provision of contraceptive products and 10% gave indirect support (via NGOs).

The opinions expressed by the governments during the United Nations surveys should be considered with some caution, as they may not correspond to solid commitments, but the trend is clear and the fertility decline of the majority of countries concerned bears witness to the behavioral changes of individuals, generally accompanied by policies related to the issue, sometimes in a coercive manner and sometimes simply as an incentive.

III. POLICIES FOR FERTILITY CONTROL. WHICH METHODS?

Since the leading article by Kingsley Davis and Judith Blake (1956), the analysis of fertility determinants largely has been based on different biological, social, economic, and cultural forces, which can be used to weigh a population's fertility level. These analyses form the substratum of the fertility control policies that were developed from the 1960s onward. The work of John Bongaarts (1978) on the proximate determinants of fertility has had a great influence on the research and implementation of programs. This author identified four proximate determinants of fertility: average age at first union, birth spacing practices (postpartum abstinence and breast-feeding), contraceptive use, and recourse to induced abortion. These determinants are themselves influenced by economic and sociocultural factors (see Chapters 37, 38, Vol. I). This model has made it easier to identify the different measures likely to affect fertility levels. The range of methods used, the direction of the discourse aimed at popularizing them, and the stakeholders responsible for their implementation have resulted in very diverse policies, which are not, as is sometimes believed, limited only to the creation of family planning centers. The use of contraception has indeed been the spearhead of fertility reduction in the Third World, but it has been most effectively adopted in the areas where an arsenal of other accompanying measures have also been applied.

After having stressed the importance of development policies (A) to the success of fertility control policies, in this section we will successively examine (B) the adoption of statutory measures, (C) information and communication campaigns, (D) the choice between incentive measures and restrictive measures, (E) the choice of family planning services, and (F) the problems of funding the programs.

1. Development Policies: An Essential Role

To encourage fertility reduction, which is the objective of the majority of population programs in developing countries, the policies implemented are generally made up of a combination of direct action (information campaigns, family planning programs, various incentive and restrictive measures) and more general measures, arising from development policies. Without these indirect measures, it would not have been possible to implement the direct policies, or they would have had little effect.

Public health policies (dealt with in Chap. 106, 107, and 108) first spring to mind. They resulted in significant progress in survival, and above all, in the survival of children. In the Third World countries as a whole, the decline in infant mortality has significantly contributed to changes in the behavior of couples, and in contraceptive use. The higher the infant mortality, the more difficult it was to accept changes in the ideals that valued high fertility. To date, in sub-Saharan Africa, the high level of infant mortality has been one of the main obstacles to the adoption of fertility reducing behavior. The general improvement in the effectiveness of health systems has also been a centerpiece of the establishment of family planning programs, which are increasingly frequently integrated into health structures. In Mexico, for example, fertility control programs are based on the health services established at a very decentralized level and which had long ago won the confidence of the population (see section E, below). This is also an essential element in the success of the Bangladeshi program (Fauveau, 1994) and that of China (Attané, 2002).

Progress in education—above all, female education—has been another essential factor that has had a demonstrable effect on fertility. In all countries where education has flourished, fertility has begun to decline (United Nations, 2003c) and in proportion to the level of education achieved. The effects of education are both direct and indirect. Educated women and men better understand the available fertility control methods and experience fewer contraceptive failures. It is also easier to recruit trained personnel for contraceptive services. Moreover, educated people marry later, either because they are continuing their education or because they are able to more effectively resist any pressure from their family and wish to reach adulthood before embarking on married life. The increase in women's marrying age is perceptible in numerous developing countries (Westoff, 2003) and plays a significant role in fertility decline. This is a direct effect of the increase in education. Furthermore, educated people have different aspirations for their children and this changes the expectations that they have of them. Men and women create their families by seeking quality rather than quantity of children, are better equipped to protect their health, and are prepared to spend significant amounts on their education; all of these factors indirectly support fertility reduction.

The improvement in female schooling forms part of the battery of measures that aim to promote equal status between the sexes. Development policies in the wider sense are also part of it. Women's access to employment, legislation that protects them from economic and sexual exploitation, the ban on violence toward them, and the adoption of equitable (or better, egalitarian) laws regarding inheritance, marriage, and possession of land and goods are also development policies that are favorable to controlling fertility. But despite the efforts of activist groups and governments' declarations of intent, the problems remain acute in this area. Family planning programs, which can be one of the best expressions of female empowerment and of their ability to freely make decisions about their reproductive life, can also flout women's rights (Locoh, 1997). One only needs to remember the improper sterilizations in Latin America and the ban on having a second child in China, which results in little girls being abandoned or in selective induced abortions; both of these practices are also common in India (Arnold *et al.*, 2002).

In a more general way, in order for family planning programs to be effective, there must be a combination of conditions at the level of both social and economic development: the existence of a legal authority to apply the laws adopted and priority given to infrastructural investment, road systems and trade, as well as contraceptive supply. These conditions are most frequently found in towns than in rural areas, where progress in infrastructure development and in the area of population programs is slower. This results in less frequent contraceptive use in rural areas in a transitional period where urban inhabitants are already involved in the process of fertility deceleration. Urbanization is both the cause and the consequence of development. In cities, marriage occurs later due to a higher level of education, there is more information available and more opportunities to call traditional family systems into question, and there are more problems involved in maintaining a large family, which encourages city dwellers and migrants to change their traditional family ideals. Children are considered less as a resource (extra hands for working on the land) and more as an expense (expenditure on schooling, health). The very rapid urbanization of developing countries has therefore been one of the causes of fertility decline. In some countries, the active policies of regional planning have led to a reduction in the disparities between urban and rural areas. These policies have indirectly supported the implementation of family planning programs.

According to the culture, experience, and proposed objectives (health of mother and child, well-being of families, limiting population growth), or according to the pressure applied within partnerships with funding countries, the decision makers of developing countries, often assisted by international experts,[6] have implemented population and family planning programs with a combination of various ingredients: statutory measures, incentive and restrictive measures, information or persuasion campaigns, the establishment of infrastructure providing contraceptive methods, as well as global development policies which have affected, to varying degrees, each of the proximate determinants of fertility.

2. The Adoption of Statutory Measures

a. Legislation on Age at Marriage

In any fertility control program initiated by a government, the legislative measures must be adjusted to create conditions that are favorable to the achievement of fertility control objectives. A first area is that of age at marriage, which is generally regulated by the law. First union formation, which is not always a formal marriage, is an important factor in fertility. In many developing countries, early marriage of young girls is valued. Thus, the average age at first union is 18 years in West Africa and around 20 years in East Africa and it was earlier still 40 years ago. South Asia (India, Bangladesh, Pakistan) is a region where girls still marry early (see Chap. 28, Vol. I). The increase in age at marriage is generally one of the most reliable predictors of a future decline in fertility. It also indicates an increase in the education level of women, their greater participation in paid employment where it is possible, and generally, a good indication of progress in female autonomy and in a greater freedom in matrimonial choices.

By delaying the legal marrying age, fertility can be affected particularly in countries where the culture or the religion are firmly opposed to births out of wedlock, which is the case in all Muslim countries and in many Asian countries. This was the case in Tunisia, for example, where the increase in the legal marrying age was an effective means of birth limitation (Ben Salem and Locoh, 2001). It led to a decrease in early marriages and also consequently a decrease in births among very young mothers. But the adoption of legislation delaying the legal marrying age is generally an expression behavior changes that have already occurred. This is the case in Morocco where the amendment to the Moudawana, proposed by King Mohammed VI at the end of 2003, increased the legal marrying age from 15 to 18 years for females, whereas the average age at first marriage was already far above this new legal minimum. The increase in the legal age at marriage was one of the efforts used by the Chinese government beginning in 1973 to control fertility. In 1980, a new law on marriage increased the age to 20 years for females and 22 years for males. Moreover, this has been associated with measures for separating married couples, for varying lengths of time, by assigning them to distant production units (Banister, 1987). In sub-Saharan Africa, on the other hand, where cultural traditions tend to be tolerant of births out of wedlock, legislative measures on marrying age have had little effect.

But legislation on the minimum legal marrying age, while it can induce a change at the beginning of a society's modernization process, is far from being the only factor in the increase in age at marriage, which is an essential driving factor, with the adoption of contraception, in fertility reduction. This increase is one of the main expressions of the sociocultural evolution of a society and sometimes of the increasing difficulty of young people in establishing their economic autonomy in family making, as is the case in North Africa (Ouadah-Bedidi and Vallin, 2000) and even in sub-Saharan Africa since the end of the 1990s (see Chap. 27, Vol. I). In the majority of developing countries where fertility decline has occurred, the proportion of young girls married before the age of 20 is low and decreased during the period of fertility decline.

b. Control of Access to Family Planning Programs

The establishment of programs for fertility control[7] should also result in the adoption of appropriate legislation. Some countries are very reticent to open the doors of family planning centers to unmarried adolescent girls. This is notably the case in Africa, where some countries also require the husband's permission to provide contraceptive services to married women. However, under pressure from international bodies these restrictions have been officially lifted in many countries, even though there are still occasions when the staff refuse to provide contraception to young unmarried women.

At the level of statutory measures, all the laws, decrees, and decisions made by health ministries concerning the importation and sale of contraceptive products should also be mentioned on the one hand, as well as the services and staff entitled to prescribe contraception, on the other hand. These regulations

[6] Due to the weight of international funding allocated to family planning programs, international expertise is still very much present, despite the fact that a growing number of national specialists have been trained.

[7] For example, this was the case in numerous African countries that had been colonized by the French as they had inherited the notorious law of 1920, intended to combat the fall in the birth rate in France which forbade any propaganda in favor of contraception.

TABLE 102–6 Distribution of Third World Countries According to the Legal Conditions for Induced Abortion in 1999

Legal conditions for induced abortion	Africa	Latin America	Asia	All N	All %
Illegal	0	2	0	2	2
To save the woman's life	25	14	10	49	46
To save the woman's life or protect her physical health	1	0	0	1	1
To save the woman's life or protect her physical and mental health	23	13	4	40	38
For the above reasons + for economic or social reasons	1	3	1	5	5
On request	3	2	4	9	8
Number of countries	53	34	19	106	100

Source: United Nations 1999; see also Appendix Table 102–2.

condition access to the various contraceptive methods available. In some countries, the distribution of the pill is very restrictive,[8] while elsewhere it is delivered widely, by ancillary medical personnel, through community-based distribution. The restrictions can also affect the type of contraceptive products receiving marketing authorization. The contragestive "morning after pill" (RU 486), considered as an abortive, is still rarely authorized. Sterilization, which is definitive, has been the subject of restrictions and regulation in many countries. In Brazil, for example, where the feminist movement vigorously denounced the improper recourse to caesarean sections and to the sterilizations that frequently followed, laws were adopted in 1998 to better control these practices (Berquó and Cavenaghi, 2002, 2003).

c. *Legislation on Induced Abortion*

In many countries, induced abortion remains an offense and is the object of repressive legislation or is restricted to certain situations, according to a grading of risks recognized as justifying induced abortion. A recent inventory[9] (United Nations, 1999; Appendix Table 102–2) lists the regulations in the developing countries of Africa, Asia, and Latin America. In each of the Third World continents, approximately half the countries apply very restrictive legislation to induced abortion, and the others apply relatively flexible criteria (induced abortion possible in order to protect the woman's physical or mental health).

Penal sanctions for induced abortion have been removed in a certain number of Third World countries. In Tunisia, a precursory country, induced abortion was authorized in 1965 for women with five children, which was a great innovation for a Muslim country. In China, within the context of the single child policy, induced abortion is not only permitted but is imposed in cases of unplanned pregnancies (Attané, 2000). The issue of legislation on induced abortion remains, during international conferences, a bone of contention that positions its supporters against conservative countries, among which is the United States, and representatives of religious authorities (see Chapter 35, Vol. I). The United States have notably withdrawn all support from the population programs of developing countries, which include induced abortion as a birth control method. Nineteen countries in Africa, 11 countries in Latin America, and 5 countries in Asia relaxed their legislation between 1994 and 1999 (see Appendix Table 102–2).

d. *Creation of National Population Commissions*

These rules and regulations are increasingly being adopted within the wider context of the definition of a population policy, highly encouraged by the United Nations authorities on population. In the 1990s the UNFPA supported the design and the adoption by governments of documents defining their policies with regard to population. Coordinating bodies, generally dependent on ministries in charge of planning, were established to supervise the implementation of fixed objectives. The centralized and hierarchical organization of the bodies responsible for applying the decisions regarding birth planning, at both national and regional levels, can also give rise to regulations and possibly to the adoption of laws. This is the case in China (Huang, 1992, quoted by Attané, 2000; see also Chap. 112), in India (see Chap. 113), and in Mexico (Cosio-Zavala, 1994; see also Chap. 114), but not in Brazil (Martine, 1996).

When the objective of fertility control is integrated into a population policy and legislative measures are

[8] The restrictions relate to those authorized to prescribe (doctors only?) and to the clients (married women but not young girls, for example).

[9] Communicated by Agnès Guillaume (Meeting on Population Policies, CEPED, October 17, 2003).

adopted, it remains for the governments to support (and sometimes impose) their application through information and persuasion programs, or incentive and restrictive measures.

3. Information and Communication Campaigns

The adoption of fertility control methods in societies where, more often than not, large families were considered as a resource had to be encouraged by campaigns of information and communication, and sometimes even pure propaganda. It was necessary to let the population know of the existence of contraceptive methods, give information on how to access them, and convince individuals (the campaigns were mainly aimed at women) of the justification of planning their births. All the tools of modern communication, posters, radio, and television were placed at the service of the campaigns in some countries. India, Indonesia, Bangladesh, the Philippines, China, and Mexico, to name only a few large countries, carried out large intensive poster, radio, and television campaigns (Faria and Potter, 1995). In Mexico, for example, the national family planning program was constantly supported by information campaigns regarding population that used various methods of communication: publicity posters, radio, television, meetings in work places, schools, and clinics. Awareness and practice of contraceptive methods became common, even in rural areas. In Brazil, the success of the *telenovelas* and the campaigns led by well-known singers lobbying in favor of birth spacing are well known. But in other countries, other media, appreciated locally (e.g., puppets, theater, and concerts) were also used. These campaigns are targeted according to sociological context. In Latin America the campaigns are often centered on the delay of entry into reproductive life and are aimed at young people. Elsewhere the theme of limiting fertility to two or three children was preferred. In China, however, coercion rather than communication was used.

In sub-Saharan Africa, birth spacing that makes a long and intensive period of breast-feeding possible is still the most frequently used method for controlling fertility. The practice of postpartum abstinence also exists in some societies. In Yoruba land, in Nigeria, for example, a female must abstain from any sexual relations with her husband while she is breast-feeding, most frequently for 18 months to two years. These associated practices explain why the interval between two births in Africa is usually between two and a half and three years. Faced with the little success of campaigns promoting the use of modern contraception, information campaigns have sometimes used these spacing customs, valued culturally, to develop their argument in favor of fertility control. These campaigns include lobbying for prolonged breast-feeding, which lessens the probability of a new pregnancy.

The cultural obstacles to contraception often are based on a population's religious beliefs. We are aware of the Catholic Church's traditional opposition to any form of birth control other than continence. In theory, Catholic couples can only resort to periodic abstinence, which has a high risk of failure. This is how some countries (such as Mauritius) developed information on this contraceptive method, which remains marginal in the progression of fertility control. In Muslim societies, where there is no such structured hierarchy as in the Catholic Church, the governments or those responsible for family planning programs have attempted to obtain declarations from the Imams, who have a well-established moral authority, in favor of birth control. This was achieved in Egypt, Iran, Algeria, and Tunisia, for example (Gueddana, 2001; Ladier-Fouladi, 2003; Ouadah-Bedidi, 2004).

The UNFPA has also encouraged governments to introduce modules titled "Education in Population and Family Life into school syllabuses", in which information on population issues and the control of population growth are provided. The information regarding fertility control is more or less explicit, depending on the culture.

These actions have indeed resulted in increased awareness of the existence of contraceptive methods within all the developing countries. They are essential to breaking down reluctance regarding a particular method and to ensuring adequate use by clients. But the gap between knowledge and use remains a characteristic of some programs (Curtis and Neitzel, 1996). It is particularly clear in sub-Saharan Africa, where contraceptive use is still very low despite widespread knowledge of the existence of these methods. The information is necessary but frequently insufficient to make families change their practices, especially when they are not accompanied by improvements in the standard of living, increased opportunities for children's education, and pensions for elderly people. In some programs, therefore, more direct incentive or restrictive measures have been used.

4. The Carrot and the Stick: Incentive Measures and Restrictive Measures

Incentive measures have mainly been aimed at those prescribing contraception (Cleland and Mauldin, 1991). Bonuses to family planning agents, depending on the results obtained, are given in a certain number of countries. In Mexico, for example, bonuses were distributed to doctors and family planning agents, according to their performance evaluated in terms of family planning clients recruited in health centers. The frequency of female sterilizations in this country, condemned many times by feminists, is undoubtedly not unconnected to the financial incentives given to the medical personnel (Potter, 1999). But incentives have also been aimed at people who have accepted to use contraception. In India, during the 1970s in the large-scale male sterilization campaigns, where clients were recruited for vasectomies around the large stations and then operated on in nearby buses, those who accepted the operation were given radio sets and a small payment (Mauldin and Sinding, 1993). More recently, advantages in terms of retirement pensions were provided to families that had limited their fertility (United Nations, 2002a).

Again in India, restrictive measures have been used in some States, even as far as adopting measures to terminate the employment of civil servants who had had a third child. This radical measure caused Indira Gandhi to lose her position as Prime Minister in 1977, and the measure was then revoked (see Chap. 113). In Indonesia, programs of reciprocal surveillance and denunciation of those who did not respect contraception objectives within local communities were put in place. There were also campaigns of distribution of contraceptives by military personnel, a barely disguised method of coercion.

But it is China that springs to mind when considering coercion carried out on the population to oblige fertility control. No other country had gone so far. The family planning policy, which in 1979 became the single child policy, was implemented by force. To achieve the fertility reduction decided in the 1970s, a complete battery of measures was adopted, at the level of married life as well as in the practice of contraception. In order to apply these measures, the Chinese government did not hesitate to resort to a drastic supervision of the population, to police control, and encouragement of denouncement. The responsibility of local administrative bodies in the achievement of the objectives of the population policy was clearly outlined. Increase in the legal minimum age at marriage and the frequent separation of spouses (e.g., due to individuals being sent to distant manufacturing units) reduced their cohabitation. Police surveillance of females of reproductive age was established in workplaces to check that they were not pregnant. In rural areas, political leaders were in charge of this surveillance and were sanctioned in cases of unplanned births in their administrative entities. The single child policy led to unprecedented coercive measures, to the extent that a couple must seek permission to try for a child, within the limits of the objectives assigned to each township or administrative entity. This control of reproduction upsets family life and contains the seeds of a demographic imbalance in terms of sex ratio and cohort size (see Chap. 112).

5. The Infrastructure Chosen for Family Planning Services

To intervene in fertility control, governments have defined extremely varied strategies. The administrative and political tradition of each country, the cultural sensitivity, the existing infrastructure, and the advice of international bodies of experts are all parameters to be taken into consideration in the implementation of fertility intervention programs. Two intervention strategies were jointly tested: entrusting the diffusion of fertility control methods to designated structures like the IPPF clinics (where it was their specific objective) and integrating family planning services into existing health services.

The first attempts to distribute family planning were frequently carried out by voluntary associations grouped into federations around the IPPF. This organization played an essential role in the initial experiments of establishing family planning services in developing countries. Based on voluntary participation and a network of volunteers associated with publicity campaigns, the national structures were created. They received constant financial support (from the IPPF, various other foundations, and the UNFPA) but also support in the form of training and supply of contraceptives. They were generally the fermenting agent that led to changes in the attitudes of frequently reticent political leaders that then resulted in the adoption of fertility control programs and policies. This has been the case particularly in Africa where, at the beginning of this establishment (in the 1970s), the health care system and a large section of medical personnel were reticent to accommodating family planning services.

The need for the population to clearly identify the objectives and be provided with adequate services by specifically trained personnel justified the establishment of programs via specific family planning channels. When family planning services are completely integrated into the public health infrastructure, there is a risk of conflict between the main objectives of public health and the provision of contraception. The risk is greater when there is a shortage of health resources. Moreover, international donors that play a major role in many of the developing countries' family planning programs fear that the funds provided for controlling fertility will end up in other sectors, which do not conform to their objectives. In each society, fertility control policies are based on decisions that sometimes support the preeminence of specific services with considerable autonomy, and sometimes the health ministries controlling the whole family planning services organization.

Mexico: An Example of an Ambitious Family Planning Program Provided by Effective Health Services

It was during the 1960s that Mexico became concerned with its high population growth. Indeed, population growth was slowing and the fertility level, over 6.8 children per woman in the 1950s, was resulting in a massive surge of young people into the employment market. In 1973, a general population law was adopted, attesting to the Mexican government's interest in population problems and specifically in fertility control. Official family planning programs were therefore established in order to provide families with information concerning contraceptive methods and to provide them with the means of using them (Cosio-Zavala, 1994, p. 213–214; see also Chap. 114).

In 30 years, the total fertility rate was reduced by half: from 6.8 children per woman in 1960–1965 to 3.1 in 1990–1995 (United Nations, 2003a). This rapid decrease was largely due to the National Family Planning Program, established in 1977. Several factors contributed to the success of this program. On the one hand, contraceptive methods were distributed by using a powerful network of health centers, clinics, and hospitals controlled by social security institutes and the Health Ministry. On the other hand, the population had confidence in its medical units, and the advice provided on family planning was quickly accepted (Gautier and Quesnel, 1993). In Mexico, where the medical system is mainly composed of large public establishments, the provision of modern contraceptive methods by health services was very effective. A special effort was made in the provision of family planning services by rural medical units and clinics based in small rural localities, particularly from 1980 onward (Cosio-Zavala, 1994).

New users were recruited while giving birth in the hospital, where medical and paramedical personnel offered them an IUD or sterilization, which are the most common methods in Mexico. Personnel received financial incentives according to the number of new family planning users that they recruited. Official programs played an important role in providing rural families and impoverished urban families with accurate information and contraceptive services. It must be stated that the frequent recourse to sterilization (31% of women of reproductive age in 1997), made possible by the considerable medicalization of family planning services, is certainly an effective means of controlling population growth but its nearly definitive effect on women's reproductive ability is severely criticized by those in favor of respecting reproductive rights.

6. Funding the Programs

The issue of funding Third World countries' fertility control efforts is a crucial one. Although the majority of these countries have given high priority to these policies, they are faced with a structural shortage of resources in every area. The implementation of legislative and statutory measures, the incentives and disincentives, the infrastructure (buildings, access to services, etc.), and local salaries are partly provided by the governments concerned. But the expansion of family planning services would not have been possible without the considerable mobilization of the international community. This consists, in more or less equal parts, of bilateral aid, multilateral aid (via UN organizations), and from international NGOs such as the IPPF, as well as some large foundations (e.g., Rockefeller, Ford, and Mellon). The training of staff, provision of contraception, monitoring and evaluation of programs, and research into contraceptive methods have been the subject of very large programs, in which international donors have contributed considerably.

Table 102–7 summarizes the contributions of the various donor countries to population programs in 1998. These programs include family planning, education regarding population, support for both population policies and demographic data collection, and programs for preventing the spread of sexually transmitted diseases and HIV/AIDS. Programs on fertility and contraceptive use are the main ones supported. The United Kingdom and the United States alone account for half of the aid provided. They are followed by the countries of Northern Europe who place a very high priority on population issues in their aid to developing countries. Finland, Sweden, and Norway devote more than 5% of their international aid to supporting population programs. France, Italy, and Japan, in contrast, place little priority on population issues in the support that they provide. Steven Sinding (2002), in a review of the effects of international funds on fertility decline, concludes that it is difficult to demonstrate a statistically significant link between the external funding of programs and fertility decline. Nevertheless, several examples confirm the effectiveness of international funding when it is given to governments that are greatly committed to program implementa-

TABLE 102–7 International Population Funding (in Millions of U.S. Dollars) and Proportion of Funding for Population Programs in International Aid Provided by Each Donor Country, 1998

Country	International population funding in 1998	% of population funding in international aid provided by each donor country
Australia	45	4.6
Belgium	10	1.1
Canada	39	2.3
Denmark	60	3.5
Finland	23	5.8
France	17	0.3
Germany	125	2.2
Italy	6	0.3
Japan	89	0.8
Netherlands	119	3.9
Norway	71	5.4
Sweden	78	5.0
Switzerland	18	2.0
United Kingdom	126	3.3
USA	620	7.1
Other countries	92	—
Total	**1,538**	**2.8**

Source: Seltzer, 2002, p. 38.

tion. This is the case of Bangladesh, Kenya (from the 1980s onward), and Indonesia (during the 1970s).

The external funding provided to programs can in no way substitute for internal efforts, but they can provide the supplement that makes it possible to reach the critical mass necessary to reinforce the effectiveness of fertility control policies. Without the incentive of international aid, some governments would have been less inclined to adopt fertility control policies.

IV. FORTY YEARS OF PROGRESS IN CONTRACEPTION

The majority of fertility control programs, most frequently called Family Planning Programs, were established from the 1960s onward. It was from this period that scientific progress in contraception resulted in the availability of contraceptive methods of unprecedented effectiveness like the pill and the IUD, which considerably facilitated the implementation of such programs. Initially supported by NGOs, notably the IPPF, these programs were progressively integrated as development planning tools by some young independent precursory states (India and Tunisia, e.g.). Stimulated by UNFPA, these programs grew in scale from 1970 and became the most important section of the population policies of most developing countries.

1. The Progressive Generalization of Contraception in Asia and Latin America, the Delay in Sub-Saharan Africa

Appendix Table 102–3 (Seltzer, 2002) makes it possible to measure the progress between the first observations of contraceptive practices (between 1970 and the end of the 1980s according to country), and the most recent ones (at the end of the 1990s, apart from some exceptions) for the majority of developing countries. According to this comparison, sub-Saharan Africa still stands out from the large-scale movement of the spread of modern contraceptive use, but in other regions of the world, contraception is high in the majority of countries. For example, let us take an arbitrary threshold of 25% of contraceptive prevalence among women aged 15 to 49 years. In Asia, only Cambodia, Laos, Pakistan, and Mongolia are below this threshold. In North Africa and the Middle East, only five countries are under the threshold (Sudan, Iraq, Yemen, United Arab Emirates, and Oman), and in Latin America only two (Haiti and Bolivia). On the contrary, in sub-Saharan Africa only a small minority exceeds this 25% threshold: South Africa, Botswana, Cape Verde, Kenya, Namibia, Mauritius, and Zimbabwe.

Figure 102–3 illustrates the progress in contraceptive prevalence among married women aged 15 to 49 in some countries of the three Third World continents, chosen either for their population size or because of the originality of their experience. The starting dates depend on available data.

a. Four Asian Countries: China, India, Bangladesh, Philippines

For Asia, four countries are represented (Figure 102–3a): China, India, Bangladesh, and the Philippines. The contrast between China and India, the two most populated countries of the world, is striking.

China

In China, following the adoption of the single child policy and the associated coercive measures, the overwhelming majority of women are obliged to use contraception. In 1983 contraceptive prevalence reached 70% and in 2000 it was over 80% (see Chap. 112).

India

On the contrary, although India's birth limitation program was adopted in 1952, and it had at certain

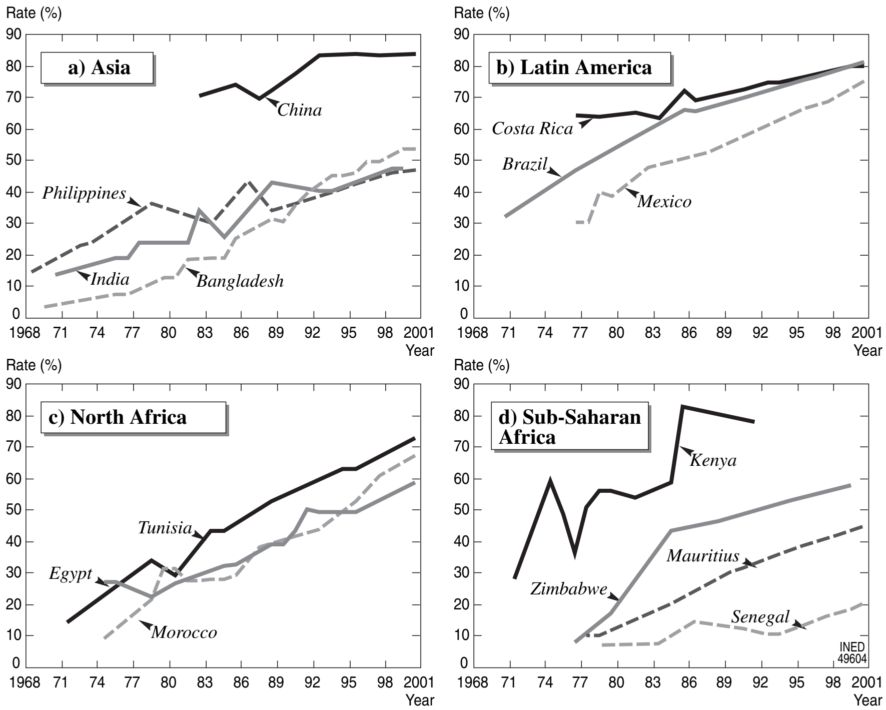


FIGURE 102–3 The trend in contraceptive prevalence among married women (aged 15–49), in some Third World countries; all methods combined.

periods experimented with coercive measures, contraceptive use has progressed slowly. The first data available show a contraceptive prevalence of 15% in 1970, and since 1989 when it reached 40%, its progress has slowed and is currently at 48% (women aged 15–44). Sterilization is the main method used (27% of married women in 1986, 36% in 1992). The programs' optimistic objectives, explicitly announced in successive plans, have never been reached at the forecast date (Bhat, 2002). India's birth limitation policy, although vigorous and sometimes coercive, has not had the anticipated effect (Srinivasan, 1998). Indeed, in this huge country, the differences are very marked between the advanced Southern states (notably Kerala and Tamil Nadu), and those of the North (Gujarat, Rajasthan, Utar Pradesh) (see Chap. 113).

Bangladesh

Bangladesh is one of the world's poorest countries, with a high infant mortality rate of 79 per 1,000 in 1995–2000, relatively low urbanization, and a population that is 90% Muslim and still values the tradition of the seclusion of women (*purdah*) inside the family home, and the prohibition of public contact with any unrelated males (Adnan, 2002). The median age at first marriage among women aged 25–49 is 15 years (Robinson, 2001). However, the rate of contraceptive prevalence has increased from 8% in 1975 to 54% in 2000 (Figure 102–3A), and fertility has been reduced by half between 1960 (7.1 children per woman) and 2000 (3.6 children per woman). Since the country's independence in 1971, the government has always placed a high priority on the objective of reducing fertility and on successive family planning programs. These have been significantly supported by international donors. It seems that the programs have not only made contraception available to couples requesting it, but they have also contributed to encouraging the population to adopt the limited family model (Cleland *et al.*, 1994).[10] The pilot studies and interventions in the Matlab zone demonstrate for the first time that providing quality family planning services in a rural area can result in a high increase in contraceptive prevalence and a significant decrease in fertility (Robinson, 2001). The main architects of this success were field personnel, area supervisors, family planning assis-

[10] Since 1977 Bangladesh has accommodated the famous Maternal and Child Health and Family Planning project, developed in Matlab. This project has been used as a laboratory for numerous health and family planning operations in developing countries. In particular, it has been demonstrated there that family planning services, even in an environment of great poverty, can be very effective provided that they are accompanied by close monitoring of the population, as is the case in this zone (Caldwell and Caldwell, 1992).

tants, and particularly, family assistants from the local area. Furthermore, the literacy of women has increased considerably, including in rural areas. The average age at first marriage has increased little, but the practice of induced abortion has played a certain role in fertility reduction (Islam *et al*., 1998).

However, the beginning of fertility decline preceded the setting up and the full development of experimental and governmental family planning programs. Factors external to the programs therefore encourage the population to reduce its fertility (Adnan, 2002). Massive changes in economic, social, and family structure appeared in rural Bangladesh in the second half of the 20th century. The level of women's education has improved, which tends to favor the decline in infant mortality. Economic difficulties have also contributed to changes in fertility preferences. The density of the population has increased considerably since the 1950s, increasing the number of farmers without land and favoring emigration to towns. Furthermore, poverty forced a large number of women to break away from traditional seclusion in order to look outside the home for a means of providing for their families. It is therefore a conjunction of constraining economic factors and an effective family planning program that explains the success. In spite of the maintenance of a very costly system for the wide distribution of contraceptives, the progress in family planning is now stationary and fertility was stagnating at a rate of 3.3 children per woman in 2000.

Philippines

In the Philippines, contraceptive use was, between 1970 and 1980, higher than in any of the other South Asian countries, such as India and Bangladesh, and this was partly due to the common practice of withdrawal and periodic abstinence, which are the only methods recommended by the Catholic Church. These two methods still represent, at the end of the 20th century, a third of contraceptive practice. Since the mid-1980s, the progress of contraception has been slower than in other Southeast Asian countries. Marilou Costello and John Casterline (2002) consider that there are not enough organizations that encourage contraceptive use and that a certain number of regulations continue to favor large families (maternal leave, lack of pensions, tax advantages only after the fourth child, etc.). It is also possible that the majority of the population, both Catholics and Muslims, are attached to a certain pronatalism. However, the experience of Latin America for Catholics, and of Iran and North Africa for Muslims, demonstrates that belonging to a pronatalist religion does not prevent the population from detaching itself from high fertility norms.

b. Three Latin American Countries: Mexico, Brazil, Costa Rica

For Latin America, we have selected two large countries, Brazil and Mexico, and Costa Rica, a smaller country that has had a very original population policy since the 1970s (Figure 102–3B). Latin America entered into the era of voluntary fertility control earlier than did the Asian countries presented above, but once again, we have three contrasting situations.

Mexico

In Mexico, the evolution of contraceptive prevalence was supported by clear political will declared in 1975 and implemented through a very dense network of public and private health services (Gautier and Quesnel, 1993). Based on the public health infrastructure which had the confidence of the public, very ambitious programs were adopted. They were particularly dependent on the cooperation of the medical staff, motivated by the attribution of advantages linked to the performance of the programs (Potter, 1999). Between 1977 and 2000, the prevalence rate was increasing regularly (from 30% to 77%), half of it by recourse to sterilization (see below) and minor but stable use of traditional methods. See Chapter 114.

Brazil

In Brazil, the transition occurred not through governmental programs or a clearly defined policy, but through a change in attitude regarding fertility, within the population as well as the medical profession, which was in favor of sterilization. Since around 1985, a rapid fertility decline has occurred, partly due to the production of contraceptives in the country itself and their distribution by the private sector and private family planning centers (Rutenberg *et al*., 1987), but particularly due to the practice of sterilization, which has become commonplace. Following caesarean sections, which are very common in Brazil, doctors tend to offer sterilization, and in Brazil this is the main method of limiting births. Women are prepared to adopt it as a convenient and inexpensive method of limiting fertility. In 1996, 77% of married women aged 15 to 49 used a form of contraception and among them, 52% had been sterilized, twice as many as were using the pill. Although induced abortion is illegal, it is still practiced in every sector of society. In 1990, there were 44 induced abortions for every 100 live births (Bozon and Enoch, 1999).

Costa Rica

As a consequence of firm political decisions and despite limited resources, Costa Rica has implemented very effective health policies that have resulted in a greater decrease in infant mortality than in other, better equipped countries (Caldwell, 1986). These decisions, the objectives of which were to reduce the inequalities between social classes, also indirectly had a beneficial effect on the adoption of contraceptive methods. With a contraceptive prevalence of 80% among married women, Costa Rica is the most advanced country in Latin America with regard to fertility control (Ghai, 1997).

c. Three North African Countries: Egypt, Tunisia, Morocco

In Africa, the differences between North Africa and sub-Saharan Africa are striking, both in the field of contraception as well as in their overall demographic situations (Figures 102–3C and 102–3D).

Egypt

Just like India, Egypt adopted a national family planning policy very early (1952), which the successive leaders of the country have continually supported. The activities to control population growth have associated programs for the direct distribution of contraception with activities to promote development (schooling, job creation, improvement of public health) (Kamran, 2002). Contraceptive prevalence has certainly increased to reach nearly 60% of married women in the year 2000, but at a slower rate than in other North African countries, even though they grasped the idea of controlling fertility through contraceptive practice later than Egypt.

Tunisia

Like Egypt, Tunisia adopted a policy for controlling population growth earlier than did its neighbors, based on a remarkable economic and social development effort (Vallin and Locoh, 2001a). Even though, for a long period, the increase in age at marriage affected fertility decline more than contraception did, contraceptive use increased regularly and had reached nearly 70% in 2000. Initially, the IUD was the main method used and has only recently begun to be replaced by the pill. Political will has constantly been reiterated in favor of the family planning program, and efforts to reach all sections of the population have been remarkable. One of the key reasons for this success is that 10 years before the adoption of a population policy, President Bourguiba had promoted the adoption of a modern personal status code, which was very favorable to the improvement of women's status. This essential document made it possible from 1965 for Tunisia to lift a number of obstacles to the adoption of a family planning program.

Morocco

Another experience is that of Morocco, which is in the process of achieving the same performance in terms of contraceptive prevalence after a delayed start. The pill is the main contraceptive used there (32% of women were using contraception in 1995). Indeed, up to 1975 fertility remained stable, with low contraceptive use (around 5%), despite a population policy that was favorable to a reduction in population growth. The rapid adherence of couples to contraception seems to be explained by economic factors, the start of a long-term crisis coupled with the spread of information regarding contraception by migrants returning from the northern countries, and the arrival of women in the labor market (Courbage, 2002). Algeria is another example of a delayed but rapid adhesion to the use of modern contraception (Ouadah-Bedidi, 2004).

d. Sub-Saharan Africa: Some Countries in Advance, the Majority Delayed

In sub-Saharan Africa, some countries resolutely committed themselves to active and effective policies. The most advanced is now South Africa, with a contraceptive prevalence of 61% in 1998 (Udjo, 2003) followed by Zimbabwe (nearly 60%) and Kenya (around 40%). They represent the countries, still quite few in sub-Saharan Africa, that are relatively advanced in their fertility transition (Sibanda, 1999). In Zimbabwe, the government has made special efforts to establish family planning centers in rural areas and to set up an effective contraceptive distribution network at the community level. The action of this infrastructure seems to explain part of the fertility decline.

With these exceptions, all the problems of underdevelopment combine to impede family planning programs in sub-Saharan Africa. The education of women is progressing slowly; children's health, after recording significant progress is now on the decline, in particular under the effect of AIDS; and the few resources of the states are insufficient to provide health services, and more so family planning services. It must be stated that, where fertility is declining, it is less as a result of the effective distribution of family planning than as a result of changes in the aspirations of families. The impoverishment of adults leads them to revise the traditional value placed on large families. And it is in the cities, more than elsewhere, that they have the desire and the means to reduce fertility. Where it is available, contraceptive practice is progressing rapidly. But there is also an increase in the number of induced abortions, in almost always disastrous conditions (Guillaume, 2000, 2003).

Kenya

Kenya was the first African country to adopt a national family planning program and was one of the countries with the highest fertility in the world until 1970–1975 (see also Chap. 115). It was only with the 1975–1979 five-year plan that a reduction in the rate of population growth was officially made an objective by the government. In accordance with a phenomenon frequently observed at the beginning of demographic transition, Kenya experienced a transitional fertility increase due to the improvement in health levels (decline in infertility) and the decline in traditional practices (particularly protracted breast-feeding). During the last quarter of the 20th century, this country then set out on the path of increased fertility control, with very rapid reductions in cities, but also significant decline in rural areas, verified by the results of three demographic and health surveys (1988, 1993, 1998). Contraceptive use in Kenya is one of the highest in sub-Saharan Africa. Nevertheless, family planning services are far from fulfilling the needs of spacing or limitation expressed by the women in the surveys.

A very considerable reduction in infant and child mortality and the rapid increase in education for young girls, associated with the establishment of good quality family planning services, have played a determining role in the acceleration of fertility decline. Furthermore, in 1976, a law set the minimum legal marrying age to 18 years for women and 20 for men (Sala Diakanda, 1991). In this country, the work of NGOs and Protestant organizations was a determining factor in the establishment of family planning programs, in parallel with the government's efforts (see Chap. 115).

Senegal

The evolution of contraceptive prevalence in Senegal (less than 15% in 2000) is, on the other hand, close to that which prevails in numerous sub-Saharan African countries: the late introduction of contraception, with varying degrees of reticence on the part of the authorities, the recent adoption of policies aiming to control fertility, weakness of programs, and the population's resistance to call into question the primacy of the value placed on procreation (Pison *et al*., 1995). Yet again, it is through spacing methods (long periods of breast-feeding, and in some societies, postpartum abstinence), far more than through contraception, that fertility control occurs in Africa, except in some countries, and in the cities (Locoh, 2002).

Mauritius

In Figure 102–3D, an atypical country has been included. The small island of Mauritius has for a long time been a pioneer in contraceptive practices, but with strong emphasis on so-called natural methods, already mentioned with regard to the Philippines. Here again, the influence of associations with the Catholic Church can be found at the beginning of the programs. The success of Mauritius in fertility control is also due to its more general development policy, which has resulted in a rapid increase in the population's standard of living, notably in their access to health services and education (Hillcoat-Nallétamby, 2002; Widmer, 2004).

2. Differentiated Contraceptive Choices, According to Continent

Fertility control in developed countries was implemented mainly through limited technical and often risky means (withdrawal, condoms and female condoms, and also the Ogino method); in many cases, these were unsatisfactory for sexual fulfillment (notably withdrawal). The frequent failures led to numerous induced, often illegal, abortions, with all the dangers that they involved for the physical and mental health of women, when induced abortion itself was not the main means for controlling natality as in the countries of Eastern Europe.

The scientific discovery of effective and reversible contraceptive methods radically changed the provision of family planning from the 1960s onward. Couples in developing countries benefited from these methods where they were available and encouraged. Certainly it was first of all necessary that the couples or the individuals wished to limit their fertility (with the exception of countries that coerce individuals, such as China), but once this bridge had been crossed effective contraceptive methods enabled them, in theory, to carry out their plans with fewer risks and with less constraining methods than were available to couples in the northern countries to reduce their fertility.

Today a broad range of contraceptive methods is offered by service providers to their clients, but in reality, political, economic, or cultural choices have led to each family planning program promoting some methods over others. Depending on each region in India, Africa, and Latin America, the provision of family planning is very different. The training received by family planning agents; the advice that they give, or which is provided by the media; the availability of methods or, inversely, shortages in stock; and the cost of procuring contraception are all factors that affect demand and individual preference. A summary of the methods used in the three large Third World regions, Asia, Africa, and Latin America, shows to what extent the choice of method differs from one to the other (Table 102–8 and Figure 102–4; see also Appendix Figure 102–1).

According to the latest available surveys, on average 6 out of 10 married women in developing countries use one form of birth control or the other, with more in Asia and Latin America (65 and 69% respectively), but many fewer in Africa (only 25%). The most popular method is sterilization (22% of married women). Nearly one in every three women in Latin America and nearly one in every four in Asia are sterilized, while in Africa this practice is completely marginal (2%). Then comes the coil (or the intrauterine device [IUD]), used by 16% of women and especially popular in Asia (20% of married women). In Latin America and in Africa, the pill is more popular than the IUD, but it is least likely to be used in Asia. Despite the intensive campaign in favor of condoms, carried out jointly by family planning programs and in HIV/AIDS prevention programs, only 3% of married women use them. This low prevalence rate can be compared to that of male sterilization (3%), which is only slightly higher (4.4%) in Asia, providing proof of the persistent reticence of men to take responsibility for fertility control. Among the other methods is the injection (Depo-provera), the prevalence of which could increase in Africa due to the prestige associated with the medical act of an injection and the possibility of using it without the knowledge of people around. Finally, the distribution of Norplant (subcutaneous placing of a contraceptive that lasts five years on average) is so far very low.

TABLE 102–8 Prevalence of contraceptive methods (%), among married women aged 15–49, by large regions (latest available surveys, from around 1997)

Method	Developing countries	Africa	Asia	Latin America
Pill	5.9	7.1	4.8	13.8
Injection	3.1	4.2	2.9	3.0
Intra-uterine device (IUD)	16.3	4.9	19.6	7.4
Female sterilization	22. 0	2.2	24.8	29.5
Male sterilization	3.6	0.1	4.4	1.6
Traditional methods	5.4	5.3	5.0	8.9
Vaginal barrier methods	0.2	0.1	0.2	0.3
Condom	3.1	1.1	3.4	4.2
Other modern methods	0.6	0.2	0.7	0.1
All methods	60.2	25.2	65.8	68.8
No method	*39.8*	*74.8*	*34.2*	*31.2*
	100	100	100	100

Source: United Nations, 2002a.

In the statistics, the so-called traditional methods include the so-called popular methods (based on knowledge of local pharmacopoeia and on customary practices) and the diverse variants of the periodic continence method (Ogino, temperature, etc.). They represent on average 5% of contraceptive practice in the Third World, with the exception of Latin America, where they make up 9%, a probable expression of the influence of the Catholic Church in certain circles of the continent. Figure 102–4 shows the distribution of modern contraceiption and traditional contraception (i.e., various pharmacopoeia, Ogino, etc.) for each country presented.

V. WHAT ARE THE RESULTS?

Some countries in Asia achieved their fertility reduction objectives to such an extent that their growth is not ensuring cohort replacement; other countries such as Tunisia are on the same path. Some have even adopted policies encouraging fertility and others could soon follow (see Chap. 103). Let us attempt rather to evaluate the part played by policies for controlling fertility reduction in the decline observed over the last 50 years.

Fertility has fallen in the majority of developing countries, mainly due to the generalization of contraceptive practices, and to a lesser extent due to the delay in women's first union formation. Fertility control policies were developed in the majority of countries during the last half-century, and, in parallel, through conferences, discussions, and program evaluations, the community of population specialists have so strongly insisted on the necessity of these policies that we are inclined to credit them with the successes obtained. Yet a study of the demographic and social history of some countries, a comparison of the paths that they have taken, and the arsenal of measures used could lead to questions regarding the specific role played by fertility control programs and policies. Were they the decisive factors in the reductions observed? Or were there other, less immediate but essential, economic, social, and cultural factors behind this behavioral revolution?

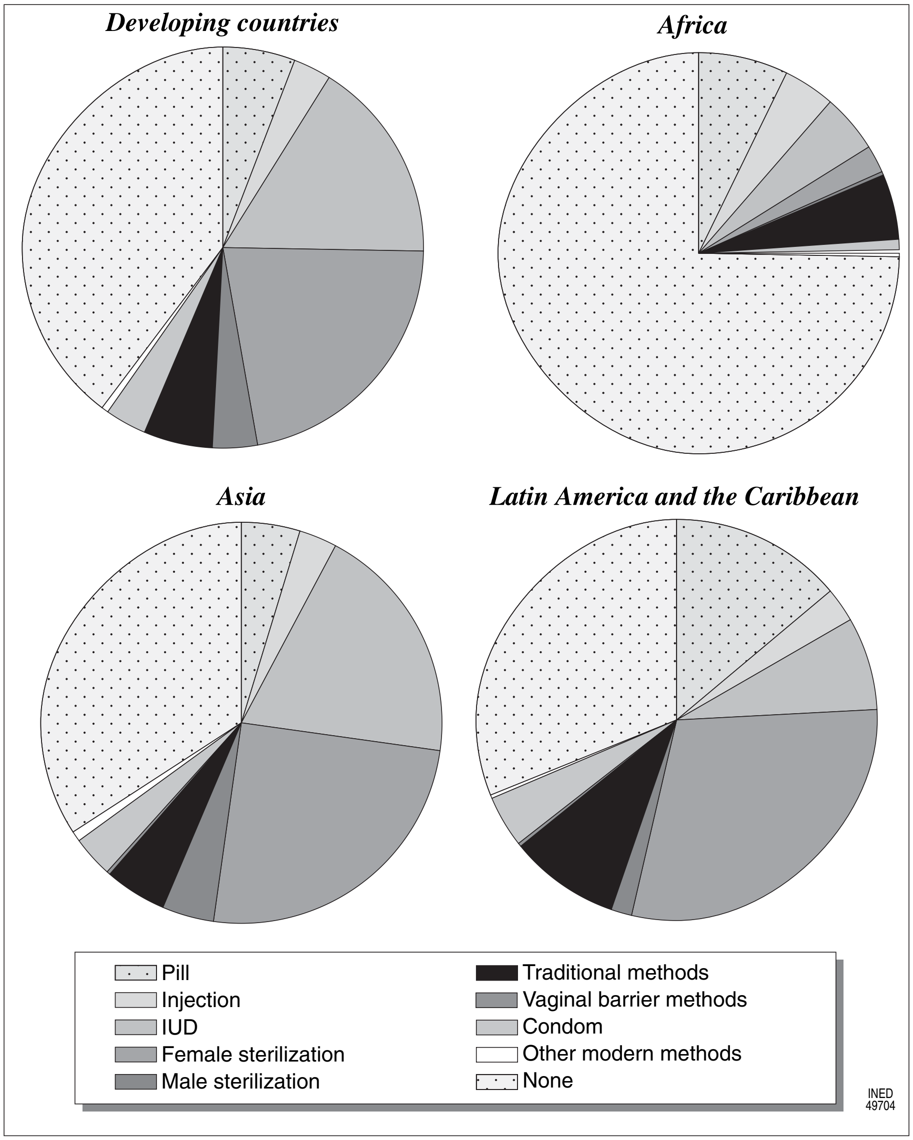


FIGURE 102–4 Prevalence of contraceptive methods, by large regions (latest available surveys, from around 1997). (Source: United Nations, 2002b.)

Historically, the efforts to encourage fertility control have sometimes considerably preceded the definition of governmental policies and the establishment of family planning programs. In Egypt, the first *fatwa* in favor of birth limitation was issued by the Egyptian religious authorities in 1937 (Fargues, 2002). In Bangladesh, the fertility decline preceded the establishment and development of the main governmental family planning programs (Adnan, 2002). In Tunisia, the personal status code promulgated by Bourguiba in 1956 had, before any concerns regarding controlling population growth, profoundly changed the status of women and thus prepared the ground for future population policies outlined only after 1965. The policies aiming to affect fertility generally proved themselves to be effective when they were on favorable ground, often prepared by economic progress, the education of women and their entry into the labor market, urbanization, and the development of health infrastructure which initiated the decline in infant mortality.

The meager results of population policies implemented in sub-Saharan Africa, despite the considerable resources, give the *a contrario* confirmation. In the majority of countries, progress at the economic and social levels is meager when it is not called into question. Infant mortality remains too high when it is not increasing due to the AIDS epidemic and the decrease in public health resources. Adult illiteracy and the lack of infrastructure for spreading contraceptive information and methods are still obstacles. While in the large cities fertility continues to fall rapidly, it is mainly because the economic crisis has convinced city dwellers that they can no longer ensure the well-being of large families and that family planning programs are available there. In rural areas, only some countries have been able to establish effective programs.

A comparison of national experiences will make it possible to demonstrate the difficulty in making a univocal diagnosis of the effect that fertility control policies have had.

Brazil and Mexico

All the countries that have succeeded in their fertility transition or that are in the process of doing so have not had fertility control policies on the same scale. Some of them, such as Brazil, have not had an explicit one. Comparing this large country with the other Latin American giant, Mexico, is instructive. Both currently have equivalent fertility levels (around 2.3 children per woman) and high levels of contraceptive use among married women (77% in Brazil, 70% in Mexico). Mexico applied governmental programs with significant incentives on the mobilization of medical personnel all over the country and on female sterilization as the main method of birth limitation. Brazil, without an explicit policy, and without a governmental program, achieved more or less the same result, and also made large-scale use of female sterilization (which is nevertheless forbidden). It did this through the spread of medical advice and promotion of sterilization after caesarean sections (see above). The entry of women into the labor market, and the already established tradition of education led women to reduce the size of their families. The motivation to have fewer children was present and was satisfied by private supply outside any specific policy.

Tunisia and Algeria

Other comparisons are essential, such as that of Tunisia and Algeria, the two neighboring North African countries. Tunisia had, from 1965, begun a fertility control program and is often quoted as an example of a successful fertility control policy (Vallin and Locoh, 2001a), but Algeria, which had, up until 1980, refused to commit to fertility reduction policies, and whose family planning programs have constantly suffered from the excuses of successive governments and insufficient infrastructure, currently has a fertility level that is very close to that of Tunisia (2 children per woman). The two countries have experienced a spectacular decrease in first union formation and, in Algeria, the mass education of women associated with the economic and political crisis has precipitated changes in behavior and the adoption of family ideals that are radically different from those of preceding generations.

Tunisia remains an example of good practices that have integrated population policies into development policies leading to the completion of the fertility transition. Economic and social progress has gone hand in hand with the development of programs for contraceptive distribution in the whole country. The first greatly contributed to the increase in age at first marriage, which has played a major role in fertility reduction; the second has given couples the means to control fertility within marriage. Although Algeria has taken a different path from that of Tunisia, it has arrived at the same point in terms of fertility, but not in terms of well-being. The existence of effective family planning programs in Tunisia has been a factor of well-being that Algerian women have not benefited from in the same way. It may also be true that without the example of Tunisia and its family planning programs, the changes in Algeria might have been slower.

Pakistan and Bangladesh

These two large countries, which were one until 1970, have followed very different paths. The poor performance of Pakistan in terms of fertility control (5 children per woman and 24% of married women practicing contraception) contrasts with the success of Bangladesh (3.6 children per woman and 54% of married women practicing contraception, see above), that nevertheless was handicapped by great poverty. Here again, the comparison pleads in favor of the effectiveness of population policies when social development policies are associated with them. In Bangladesh, these conditions were met and fertility control policies, carefully defined after a period of experimentation according to social context, coincided with an increase in female literacy and women's progressive entry into the labor market in employment outside of the family circle, providing them with relative autonomy. While the fertility control programs were impressive, they were not the only cause of the fertility decline. Social progress, demographic pressure on available land, and poverty among families strengthened the incentive measures of the programs.

During the same period, the situation was very different in Pakistan. Fertility decline was delayed. Zeba Sathar (2002) considers that the main reason for this was the "*ineffectiveness of the population policy and the State and the inability of the private sector to meet . . . the demand for good quality family planning services*" (p. 349). But she adds that the economic progress made has not set social changes in motion. Women's status is practically unchanged. For example, only one woman out of ten is literate. *"These factors block any rapid change in fertility desires and in women's motivation to control their fertility"* (Sathar, 2002, p. 349). While fertility control policies have been failing, development policies have also failed to support them.

And China . . .

We have not discussed the effectiveness of fertility policies in this final section. The case of China might be viewed as an example of a completely successful fertility control policy, but it was necessary to apply conditions that have not been used elsewhere and on such a scale and at a high cost to the population. The population's indoctrination, the permanent surveillance of their activities, and the application of sanctions to offenders were possible only through administrative and police control of the population, and a large-scale repressive arsenal. It must also be noted that substantial progress has been made in public health, and this has resulted in a rapid decline in infant mortality. Here again, programs other than fertility planning programs were necessary. In India, where fertility control policies also recommended very tough repressive measures, the population, through democracy, successfully resisted them, an act not possible in China.

All the experiences in fertility decline are unique. In examining the various national experiences of the past 50 years, it becomes obvious that policies aiming to reduce fertility were not the determining factor, but they did play a role in accelerating the process, in cases where the population was motivated to better control its fertility. By disseminating information, by making effective contraceptive methods available to family planning clients, and by giving legitimacy to a desire to control fertility, these policies became truly effective.

In some cases, fertility decline will probably be a consequence of a lack of economic and social policies, rather than the implementation of fertility control policies. This is observed in Africa, where fertility decline results more from the pressure of poverty than from the effects of family planning programs or other governmental fertility control measures. Here again, the stakeholders' motivation is essential. The economic crisis, which has become structural in the majority of countries, pushes men and women to change their fertility ideals.

Would the decline in fertility have occurred in the same manner if the international community had not seized the issue and had not designed, encouraged, or financed governmental policies pertaining to the subject? Undoubtedly. But it would have taken longer for the reduction to start, and the process would have been more difficult for those wishing to limit their family size. The most satisfactory results, at the macroeconomic level as well as that of the population's well-being were achieved when economic development and social policies were combined with and strengthened by fertility control policies.

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Adnan Shapan, 2002. Dilatation démographique, pression sur les terres et changement social. Une baisse de la fécondité en situation de pauvreté absolue: le Bangladesh, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde. Géants démographiques et défis internationaux*, p. 63–106. Paris, PUF, INED, 768 p. (Les cahiers de l'INED, no.149).

Arnold Fred, Kishor Sunita and Roy T. K., 2002. Sex-selective abortions in India, *Population and Development Review*, vol. 284, p. 759–785.

Attané Isabelle, 2000. *La planification familiale en Chine, pour ou contre la femme?* Paris, CEPED, 55 p. (Les dossiers du CEPED, no.60).

Attané Isabelle (dir.), 2002. *La Chine au seuil du 21e siècle. Questions de population, questions de société*. Paris, INED and PUF, 602 p. (Les cahiers de l'INED, no.148).

Banister Judith, 1987. *China's changing population.* Palo Alto (California), Stanford University Press, 488 p.

Ben Salem Lilia and Locoh Thérèse, 2001. Les transformations du mariage et de la famille, *in:* Jacques Vallin and Thérèse Locoh (eds.), *Population et développement en Tunisie. La métamorphose*, p. 143–169. Tunis, Cérès Éditions, 802 p.

Berquó Elza and Cavenaghi Suzanna, 2002. *The impacts of the voluntary sterilization law on reproductive rights of men and women.* Campinas, NEPO, 21 p. (Paper presented at the Population Association of America meeting, Atlanta, May 2002).

Berquó Elsa and Cavenaghi Suzanna, 2003. Direitos reprodutivos de mulheres e homens face á nova legislação brasileira sobre esterilização voluntária, *Cadernos de Saúde Pública*, vol. 19 (special issue), p. 441–453.

Bhat Mari P. N., 2002. India's changing dates with replacement fertility: a review of recent fertility trends and future prospects, *in:* Population Division (ed.), *Completing the fertility transition*, p. 376–391. New York, United Nations, Department of Economic and Social Affairs, 561 p. (ESA/P/WP.172/Rev.1).

Bongaarts John, 1978. A framework for analysing the proximate determinants of fertility, *Population and Development Review*, vol. 4(1), p. 105–132.

Bozon Michel and Enoch Etheline, 1999. Brésil: la transition démographique rapide d'un pays hétérogène, *Population et sociétés*, no. 345, p. 1–4.

Caldwell John C., 1986. Routes to low mortality in poor countries, *Population and Development Review*, vol. 12(2), p. 171–220.

Caldwell John C. and Caldwell Pat, 1992. What does the Matlab fertility experience really show? *Studies in Family Planning*, vol. 23(5), p. 292–310.

Campbell Cameron, 2002. Pretransitional fertility in China, *Population and Development Review*, vol. 28(4), p. 735–750.

CEA, 1984. *Program d'action du Kilimandjaro concernant la population africaine et le développement autonome.* Addis Abbaba, United

Nations, Economic Commission for Africa, 10 p. (Adopted during the Arusha [Tanzania] Conference, 9–13 January 1984).

Cleland John and Mauldin W. Parker, 1991. The promotion of family planning by financial payments: The case of Bangladesh, *Studies in Family Planning*, vol. 22(1), p. 1–18.

Cleland John, Phillips James F., Amin Sajeda and Kamal Ghulam M., 1994. *The determinants of reproductive change in Bangladesh: success in a challenging environment*. Washington DC, The World Bank, 187 p.

Coale Ansley J. and Hoover Edgar M., 1958. *Population growth and economic development in low income countries.* Princeton, Princeton University Press, 389 p.

Cosio-Zavala Maria-Eugenia, 1994. *Changements de la fécondité au Mexique et politiques de fécondité.* Paris, L'Harmattan and IHEAL, 256 p. (Recherche et documents Amérique Latine).

Costello Marilou P. and Casterline John B., 2002. Fertility decline in the Philippines: current status, future prospects, *in:* Population Division (ed.), *Completing the fertility transition*, p. 529–538. New York, United Nations, Department of Economic and Social Affairs, 561 p. (ESA/P/WP.172/Rev.1).

Courbage Youssef, 2002. Sur les pas de l'Europe du Sud: la fécondité au Maghreb, *in:* Population Division (ed.), *Completing the fertility transition*, p. 469–483. New York, United Nations, Department of Economic and Social Affairs, 561 p. (ESA/P/WP.172/Rev.1).

Curtis Sian L. and Neitzel Katherine, 1996. *Contraceptive knowledge, use and sources.* Calverton (Maryland), Macro International Inc., 92 p. (DHS Comparative Studies, no.19).

Davis Kingsley and Blake Judith, 1956. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol.4(4), p. 211–235.

Fargues Philippe, 2002. *Terminating Marriage.* Cairo, American University in Cairo Press, The New Arab Family, 30 p. (Cairo Papers in Social Sciences).

Faria Vilmar E. and Potter Joseph, 1995. *Television, telenovelas and fertility change in northeast Brazil.* Austin, The University of Texas at Austin, Texas Population Research Center, 34 p.

Fauveau Vincent, 1994. *Matlab: women, children and health.* Dhaka, ICDDR, 462 p.

Gautier Arlette and Quesnel André, 1993. *Politique de population, médiateurs institutionnels et régulation de la fécondité au Yucatan (Mexique).* Paris, ORSTOM, 114 p. (Études et thèses).

Ghai Dharam, 1997. *Social development and public policy: some lessons from successful experiences.* Geneva, United Nations Research Institute for Social Development, 48 p. (UNRISD Discussion Paper 1012–6511).

Gueddana Nabiha, 2001. L'expérience du program tunisien de planification familiale (1956–1996), *in:* Jacques Vallin and Thérèse Locoh (eds.), *Population et développement en Tunisie. La métamorphose*, p. 203–232. Tunis, Cérès Editions, 802 p.

Guillaume Agnès, 2000. L'avortement en Afrique: mode de contrôle des naissances et problèmes de santé publique, *La chronique du CEPED*, no. 37, p. 1–4.

Guillaume Agnès, 2003. *L'avortement en Afrique. Une revue de la littérature des années 1990 á nos jours*. Paris, CEPED, CD-Rom.

Guzman Jose Miguel, 1994. The onset of fertility transition in Latin America, *in:* Thérèse Locoh and Véronique Hertrich (eds.), *The onset of fertility transition in Sub-Saharan Africa*, p. 43–68. Liège, IUSSP, Derouaux Ordina Éditions, 308 p.

Hillcoat-Nallétamby Sarah, 2002. *La pratique contraceptive á l'Ile Maurice: politique nationale, pratiques individuelles.* Paris, L'Harmattan, 206 p. (Populations).

Huang Maochen, 1992. Zuhzhi guanli (Family planning organization), *in: Jihua shengyu guauli xue*, p. 143–175. Beijing, Zhongguo renkou chubanshe, 503 p. (China population publications).

IIPS and ORC Macro, 2000. *National Family Health Survey (NFHS-2), India 1998-99: Andhra Pradesh.* Mumbai (India) and Claverton (Maryland), International Institute for Population Science and ORC Macro, 310 p.

Islam Mazharul, Am Mamun Abdullah and Bairagi Radeshayam, 1998. Fertility and its proximate determinants in Bangladesh: evidence from the 1993/94 Demographic and Health Survey, *Asia-Pacific Population Journal*, vol. 13(3), p. 3–22.

Kamran Asdar Ali, 2002. *Planning the family in Egypt. New bodies, new selves.* Austin, Universtity of Texas Press, 233 p.

Ladier-Fouladi Marie, 2003. *Population et politique en Iran. De la monarchie á la république islamique*. Paris, INED and PUF, 356 p. (Les Cahiers de l'INED, no. 150).

Lebrun Marc, 1985. *From Margaret Sanger to Mercedes Concepcion, the IUSSP history/De Margaret Sanger á Mercedes Concepcion, l'histoire de l'UIESP*. Liège, IUSSP/UIESP, 45 p. (Bilingual opuscule).

Locoh Thérèse, 1997. Programmes de population et amélioration du statut des femmes, une convergence difficile, *in:* Jeanne Bisilliat (ed.), *Face aux changements, les femmes du sud*, p. 285–296. Paris, L'Harmattan, 366 p.

Locoh Thérèse, 2002. Fertility decline and family changes in Sub-Saharan Africa, *Journal of African Policy Studies*, vol.7(2–3), p. 17–48.

Locoh Thérèse and Hertrich Véronique (eds.), 1994. *The onset of fertility transition in sub-Saharan Africa.* Liège, IUSSP, Derouaux Ordina Éditions, 308 p.

Macura Milos, 1986. The significance of the United Nations international population conferences, *Population Bulletin*, (19–20), p. 14–25.

Martine George, 1996. Brazil's fertility decline, 1965–95: a fresh look at key factors, *Population and Development Review*, vol. 22(1), p. 47–76.

Mauldin W. Parker and Segal Sheldon J., 1988. Prevalence of contraceptive use: trends and issues, *Studies in Family Planning*, vol. 19(6), p. 335–353.

Mauldin W. Parker and Sinding Steven W., 1993. *Review of existing family planning policies and programs: lessons learned*. New York, The Population Council, 50 p. (Policy Research Division. Working Papers, no. 50).

Nations unies/United Nations, 1994. Synthèse des conférences régionales sur la population convoquées dans le cadre des préparatifs de fond de la conférence internationale sur la population et le développement, *Bulletin démographique des Nations unies*, no. 37/38, p. 3–22.

Ouadah-Bedidi Zahia, 2004. *Baisse de la fécondité en Algérie: transition de développement ou transition de crise?* Paris, Institut d'études politiques, 710 p. (Doctoral thesis).

Ouadah-Bedidi Zahia and Vallin Jacques, 2000. Maghreb: la chute irrésistible de la fécondité, *Population et Sociétés*, no. 359(July–August), p. 1–4.

Page Hilary J. and Lesthaeghe Ron (eds.), 1981. *Child-spacing in tropical Africa: traditions and change*. London and New York, Academic Press, 332 p.

Pison Gilles, Hill Kenneth H., Cohen Barney and Foote Karen A. (eds.), 1995. *Population dynamics of Senegal*. Washington, D.C., National Academy Press, 254 p. (Committee on Population, Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy of Sciences).

Potter Joseph, 1999. The persistence of outmoded contraceptive regimes, *Population and Development Review*, vol. 25(4), p. 703–739.

Robinson Warren, 2001. Common beginnings but different outcome: the family planning programs of Pakistan and Bangladesh, *in:* Zeba Ayesa Sathar and James F. Phillips (eds.), *Fertility transition in south Asia*, p. 347–385. Oxford, Oxford University Press, 434 p.

Rutenberg Naomi, Ochoa Luis Hernando, and Arruda José Maria, 1987. The proximate determinants of low fertility in Brazil, *International Family Planning Perspectives*, vol. 13(3), p. 75–80.

Sala Diakanda Daniel Mpembelé, 1991. De l'émergence de politiques de population en Afrique, *Politique africaine*, no. 44, "Politique de population," p. 37–50.

Sathar Zeba, 2002. Une forte croissance démographique au Pakistan, mais avec les prémisses d'une baisse de fécondité, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde. Géants démographiques et enjeux internationaux*, p. 341–354. Paris, INED and PUF, 768 p. (Les Cahiers de l'INED, no. 149).

Seltzer Judith R., 2002. *The origins and evolution of family planning programs in developing countries.* Santa Monica, Arlington and Pittsburgh, Rand, 184 p. (Population Matters).

Sibanda Amson, 1999. Reproductive change in Zimbabwe and Kenya: the role of the proximate determinants in recent fertility decline, *Social biology*, vol. 46(1–2), p. 82–99.

Sinding Steven, 2002. The role of international funding in future fertility declines among intermediate fertility countries, *in:* Population Division (ed.), *Completing the fertility transition*, p. 86–90. New York, United Nations, Department of Economic and Social Affairs, 561 p. (ESA/P/WP.172/Rev.1).

Srinivasan Krishnamurthy, 1998. Population policies and programs since independence (a saga of great expectations and poor performance), *Demography India*, vol.1(January–June), p. 1–22.

Udjo Eric O., 2003. A re-examination of levels and differential in fertility in South Africa from recent evidence, *Journal of Biosocial Science*, vol. 35(3), p. 413–431.

United Nations, 1995. *Population and development. Vol.1: Program of action adopted at the International conference on population and development, Cairo, 5–13 September 1994*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 100 p.

United Nations, 1999. *World abortion policies.* New York, United Nations, Department of Economic and Social Affairs, Population Division. (ST/ESA/E+SER.A/178).

United Nations, 2000. Levels and trends of contraceptive use as assessed in 1998. New York, Department of Economic and Social Affairs, Population Division, 196 p.

United Nations, 2002a. Views and policies concerning population growth and fertility among governments in intermediate fertility countries, *in:* Population Division (ed.), *Completing the fertility transition*, p. 222–262. New York, United Nations, Department of Economic and Social Affairs, Population Division, 561 p. (ESA/P/WP.172/Rev.1).

United Nations, 2002b. *World contraceptive use 2001.* New York, Department of Economic and Social Affairs, Population Division. (Chart, E.02.XIII.7).

United Nations, 2002c. *World population monitoring 2002. Reproductive rights and reproductive health: selected aspects.* New York, Commission on Population and Development, 191 p. (Draft, thirty-fifth session, 1–5 April 2002).

United Nations, 2003a. *Fertility, contraception and population policies.* New York, Population Division, Department of Economic and Social Affairs, United Nations Secretariat, 37 p.

United Nations, 2003b. *World population prospect. The 2002 revision. Highlights.* New York, United Nations, Department for Economic and Social Affairs, Population Division, 21 p. (http://esa.un.org/unpp).

United Nations, 2003c. *Population, education and development: the concise report*. New York, Department of Economic and Social Affairs, Population Division, 56 p.

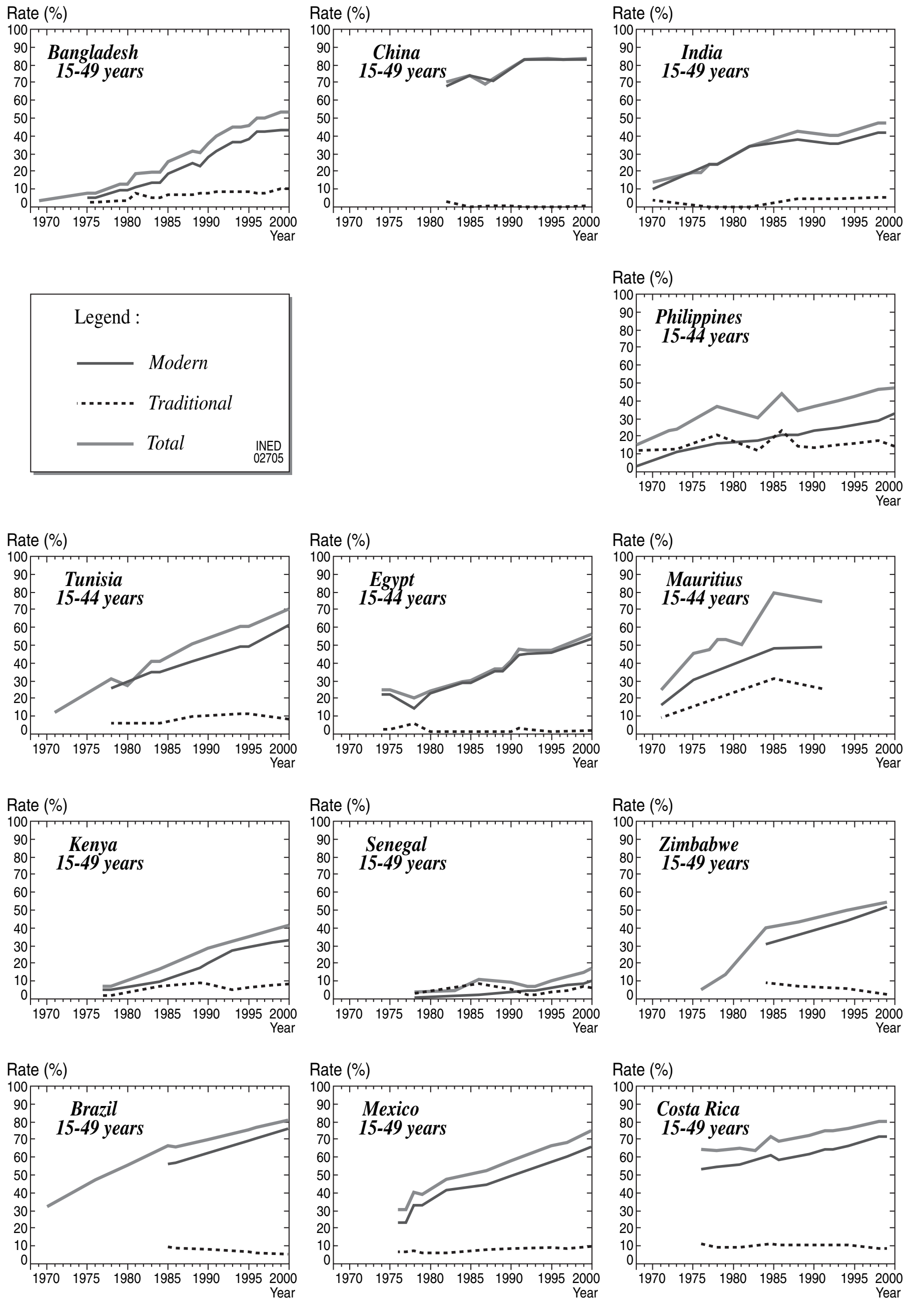
Vallin Jacques and Locoh Thérèse (dir.), 2001a. *Population et développement en Tunisie. La métamorphose.* Tunis, Cérès Éditions, 802 p.

Vallin Jacques and Locoh Thérèse (dir.), 2001b. Les leçons de l'expérience tunisienne, *in:* Jacques Vallin and Thérèse Locoh (dir.), *Population et développement en Tunisie. La métamorphose*, p. 569–580. Tunis, Cérès Éditions, 802 p.

Westoff Charles F., 2003. *Trends in marriage and early childbearing in developing countries.* Calverton (Maryland), ORC Macro, 53 p. (DHS Comparative Reports, no. 5).

Widmer Isabelle, 2004. *La Réunion et Maurice. Parcours de deux îles australes des origines au 20e siècle*. Paris, INED and PUF, 292 p. (Les Cahiers de l'INED no.155).

Zhao Zhongwei, 2002. Fertility control in China's past, *Population and Development Review*, vol. 28(4), p. 751–758.



APPENDIX FIGURE 102–1 Trend in contraceptive practices, various developing countries, 1968–2000.

APPENDIX TABLE 102–1 Trend in Total Fertility Rate between 1970 and 2000, by Large World Regions

Regions	1970–1975	1980–1985	1985–1990	1990–1995	1995–2000
Africa	**6.7**	**6.3**	**6**	**5.6**	**5.3**
East Africa	7	6.9	6.7	6.3	6.1
Central Africa	6.3	6.5	6.6	6.5	6.4
North Africa	6.3	5.6	4.8	4.1	3.6
Southern Africa	5.5	4.9	4.1	3.5	3.3
West Africa	7	6.7	6.7	6.4	5.9
Sub-Saharan Africa	6.8		6.4	6.1	5.8
Asia	**5.1**	**3.7**	**3.4**	**2.9**	**2.7**
East Asia	4.5	2.4	2.4	1.9	1.8
South Central Asia	5.6	5	4.4	4	3.6
South East Asia	5.5	4.2	3.7	3.2	2.8
Western Asia	5.6	5	4.7	4.2	3.9
Europe	**2.2**	**1.9**	**1.8**	**1.6**	**1.4**
Latin America and the Caribbean	**5**	**3.8**	**3.4**	**3**	**2.7**
Caribbean	4.4	3.1	3.1	2.7	2.5
Central America	6.4	4.6	3.9	3.4	3
South America	4.7	3.7	3.2	2.8	2.6
North America	**2**		**1.9**	**2**	**2**
Oceania	**3.2**	**2.6**	**2.5**	**2.5**	**2.4**
Australia/New Zealand	2.6		1.9	1.9	1.8
Melanesia	5.8		4.9	4.8	4.4
Micronesia	4.8		3.8	4.1	4.3
Polynesia	5.5		4.1	3.7	3.2
In the world	**4.5**	**3.6**	**3.4**	**3**	**2.8**

Source: United Nations, 2003a.

APPENDIX TABLE 102–2 Legal Conditions for Recourse to Induced Abortion in Africa in 1994 and 1999

Legal conditions for induced abortion	1994	1999
Illegal	Central African Republic, Djibouti, Egypt, Mauritius, São Tomé and Príncipe	
To save the woman's life	Angola, Benin, Chad, Côte d'Ivoire, Democratic Republic of the Congo (formerly Zaire), Gabon, Guinea- Bissau, Kenya, Lesotho, Libya, Madagascar, Mali, Mauritania, Mozambique, Niger, Nigeria, Senegal, Somalia, Tanzania, Togo Sudan[a]	Angola, Benin, **Central African Republic**, Chad, *Democratic Republic of the Congo*, Côte d'Ivoire, **Djibouti**, **Egypt**, Gabon, Guinea-Bissau, Lesotho, Libya, Madagascar, *Malawi*, Mali, Mauritania, **Mauritius**, Niger, São Tomé and Príncipe, Senegal, Somalia, *Swaziland*, Togo Sudan[a]
To save the woman's life or to protect her physical health	Burkina Faso, Burundi, Comoros, Democratic Republic of the Congo, Eritrea, Equatorial Guinea, Ethiopia, Malawi, Morocco, Rwanda, Uganda Cameroon[a] Zimbabwe[a,b]	Zimbabwe[a,b]
To save the woman's life or to protect her physical and mental health	Gambia, Guinea, Sierra Leone, Swaziland, Uganda Algeria[a] Botswana[a], Ghana[a,b], South Africa[a,b], Liberia[a], Namibia[a,b], Seychelles[a,b]	*Algeria*, **Burundi**, **Comoros**, **Equatorial Guinea**, **Eritrea**, **Ethiopia**, Gambia, Guinea, **Kenya**, **Morocco**, **Mozambique**, **Nigeria**, **Rwanda**, Sierra Leone, **Tanzania**, Uganda **Cameroon**[a] Botswana[a,b] , **Burkina Faso**[a,b], Ghana[a,b], Liberia[a,b], Namibia[a,b], Seychelles[a,b]
For the above reasons and for economic or social reasons	Zambia[b] Cape Verde[a,b]	Zambia[b]
On request	Tunisia	**Cape Verde**, **South Africa**, Tunisia

Note: The countries where legislation has been relaxed since 1994 are marked in **bold**, and the countries where legislation has become more restrictive since 1994 are marked in *italic*.

[a] In cases of rape or incest.

[b] In cases of fetal malformation.

Sources: Guillaume, 2003; United Nations, 1999.

APPENDIX TABLE 102–2 (*continued*)

Legal conditions for induced abortion	**1994**	**1999**
Illegal	Chile, Colombia, Dominican Republic, Honduras Bhutan, Nepal, Philippines	Chile, *El Salvador*
To save the woman's life	Antigua, Barbados, Dominica, El Salvador, Guatemala, Haiti, Nicaragua, Paraguay, Suriname, Venezuela, Brazil[a], Mexico[a] Afghanistan, Bhutan, Cambodia, Indonesia, Laos, Myanmar, Sri Lanka, Yemen	Antigua, Barbados, **Colombia**, Dominica, **Dominican Republic**, Guatemala, Haiti, **Honduras**, Nicaragua, Paraguay, Suriname, Venezuela Brazil[a], Mexico[a] Afghanistan, Bangladesh, **Bhutan**, Indonesia, Laos, Myanmar, **Nepal**, **Philippines**, Sri Lanka, Yemen
To save the woman's life or to protect her physical health	Argentina, Costa Rica, Peru Bolivia[a], Ecuador[a] Thailand (a)	
To save the woman's life or to protect her physical and mental health	Bahamas, Grenada, Guyana, Jamaica, Saint Kitts and Nevis, Saint Lucia, Trinidad and Tobago, Uruguay Panama[a] Saint Vincent and Grenadines[a] Pakistan Malaysia[b]	Bahamas, **Costa Rica**, **Ecuador**[a], Grenada, Jamaica, **Peru**, Saint Kitts and Nevis, Saint Lucia, Trinidad and Tobago Uruguay[a], **Argentina**[a], **Bolivia**[a], Panama[a] *Malaysia*, Pakistan **Thailand**[a], *South Korea*[a]
For economic or social reasons	Barbados[a,b], Belize[a,b] India[a,b], South Korea[a,b]	*Belize*[b], **St Vincent and Grenadines**[a,b], Barbados[a,b] India[a,b]
On request	Cuba North Korea, Vietnam, China	Cuba, **Guyana** North Korea, **Cambodia**, China, Vietnam

APPENDIX TABLE 102–3 Rate of Contraceptive Prevalence, Oldest and Most Recent Year of Observation, for Married Women, 15–49 Years

Country	Year of observation		Rate of contraceptive prevalence		
	Oldest	Most recent	All methods observed:		Modern methods in the most recent year of observation
			Oldest	Most recent	
Asia					
Bangladesh	1976	2000	8	54	43
Cambodia		1995	—	13	7
China	1982	1997	71	84	83.5
Hong Kong	1972	1992	50	86	80
India	1970	1999	14	48	43
Indonesia	1976	1997	18	57	55
Laos	—	1993	—	19	15
Malaysia	1974	1988	33	48	31
Mongolia	—	1994	—	61	25
Myanmar	1992	1997	17	33	28
Nepal	1976	1996	3	29	26
Pakistan	1975	1994	5	18	13
Philippines	1968	1998	15	47	28
South Korea	1974	1994	35	77	67
Singapore	1973	1982	60	74	73
Sri Lanka	1975	1993	43	66	44
Taiwan	1980	1992	69	82	74
Thailand	1975	1996	33	72	69
Vietnam	1988	1997	53	75	56
Latin America					
Bolivia	1983	1988	26	48	25
Brazil	1970	1996	32	77	70
Colombia	1976	2000	43	77	64
Costa Rica	1976	1999	64	80	71
Cuba	—	1987	—	70	67
Dominican Republic	1975	1996	32	64	59
Ecuador	1979	1999	34	66	52
El Salvador	1978	1998	34	60	54
Guatemala	1978	1999	18	38	31
Haiti	1983	2000	7	28	22
Honduras	1981	1996	27	50	41
Jamaica	1976	1997	38	64	61
Mexico	1976	1997	30	68.5	60
Nicaragua	1981	1997	27	60	57
Panama	1976	1984	54	58	54
Paraguay	1979	1998	36	57	48
Peru	1978	2000	31	69	50
Trinidad and Tobago	1977	1987	52	53	44

Source: United Nations, 2000.

(continues)

APPENDIX TABLE 3 (*continued*)

Middle East/North Africa					
Algeria	1987	1995	36	52	49
Egypt	1974–1975	2000	25	56	54
Iran	1977	1994	36	70	51
Iraq	1974	1989	15	14	10
Jordan	1976	1997	25	53	38
Kuwait	—	1987	—	35	32
Libya	—	1995	—	40	26
Morocco	1974	1997	7	58	49
Oman	1988	1995	9	22	19
Sudan	1989	1993	9	10	7
Syria	1978	1993	20	40	28
Tunisia	1971	1994	12	61	49
Turkey	1978	1998	38	64	38
United Arab Emirates	—	1995	—	27	24
Yemen	1979	1997	1	21	10
Sub-Saharan Africa					
South Africa	1976	1998	37	62	61
Benin	1981	1996	9	16	3
Botswana	1984	1988	28	33	32
Burkina Faso	1993	1999	8	12	5
Burundi	—	1987	—	9	1
Cameroon	1978	1998	2	19	7
Cape Verde	—	1998	—	53	46
Central African Republic	—	1994	—	15	3
Chad	—	1997	—	4	1
Congo	—	1991	—	8	2
Côte d'Ivoire	1980	1998	3	15	7
Eritrea	—	1995	—	20	7
Ethiopia	—	1990	—	4	3
Ghana	1980	1998	10	22	13
Guinea	1993	1999	2	6	4
Kenya	1978	1998	7	39	31
Lesotho	1977	1992	5	23	19
Liberia	—	1986	—	6	6
Madagascar	1992	1997	17	19	10
Malawi	1984	1996	7	22	14
Mali	1987	1996	5	7	5
Mauritania	1981	1990	1	3	1
Mauritius	1971 (15–44)	1991	25	75	49
Mozambique	—	1997	—	6	5
Namibia	—	1992	—	29	26
Niger	1992	1998	4	8	5
Nigeria	1982	1990	5	6	4
Rwanda	1983	1992	10	21	13
Senegal	1978	1997	4	13	8
Sudan	1979	1993	5	8	7
Swaziland	—	1988	—	20	17
Tanzania	1992	1999	10	25	17
Togo	1988	1998	12	24	7
Uganda	1989	1995	5	15	8
Zambia	1992	1996	15	26	14
Zimbabwe	1976	1999	5	55	52

Sources: Mauldin and Segal, 1988; Seltzer, 2002.

C H A P T E R

103

Toward a Policy Turnaround in Low-Fertility Developing Countries?

MAGALI BARBIERI

Institut national d'études démographiques (INED), Paris, France

Apart from the West and Japan—a special case in Asia because of its early industrialization—the countries or territories whose fertility reached replacement level (2.07 children per woman) by the 1980s were in East and Southeast Asia (with the exception of Cuba and a few other sparsely populated Caribbean islands). To be more precise, these countries are Singapore, South Korea, Hong Kong, Macao, and Taiwan. Whereas the governments of the last three countries appear only mildly concerned by the socioeconomic consequences of their low fertility levels, Singapore and South Korea are sufficiently worried to have discontinued the strict population control policy instituted 30 years earlier. Despite their low demographic weight (4 and 47 million inhabitants, respectively, in the year 2000), the way in which the policies adopted by these two countries have evolved deserves special attention. Culturally acceptable for Asian populations, the route they take paves the way for many other countries in the region that have already caught up with them (China, Thailand, and Indonesia) or that will soon catch up (North Korea and Sri Lanka). Policies chosen by these pioneer countries may also present a more attractive alternative for other regions of the South, likely to reach low fertility levels at some point, than those developed in the West.

Faced with low fertility, Singapore and South Korea share similar concerns. The solutions chosen by each to rectify the situation are polar opposites, however: whereas Singapore has taken a resolutely pronatalist line, albeit with characteristics that clearly distinguish it from family policies implemented in the West, South Korea, instead of attempting to influence reproductive behavior, has developed policies essentially designed to lessen the socioeconomic consequences of fertility decline. These two options will be examined in turn, after a brief reminder of each country's previous policy features.

I. SINGAPORE

From the time of independence (1965), the government of Singapore viewed demographic growth control as a crucial element of its development policy. Population growth had reached an all-time high between the two preceding censuses, with an annual average of 4.3% between 1947 and 1957 (Figure 103–1). A stringent policy was adopted in 1966. Aiming to reduce the number of large families, it was linked to the slogan: "Stop at two!"; abortion was legalized, and became widely available at very low cost; female sterilization was encouraged after the second birth, and women who went ahead with it were offered a broad range of fiscal and social advantages (priority access to housing, and health services and education for themselves and their children), as opposed to couples with more than two children who, on the contrary, were penalized in these same areas.

This policy was set up within a highly favorable context. Economic growth was strong and sustained. It reached an average of 7% annually during the 1960s and 1970s, and generated a gradual improvement in living standards, such that in 1987, the average income per person was similar to that of Japan, and was superior to those of Spain and Italy. In addition, health

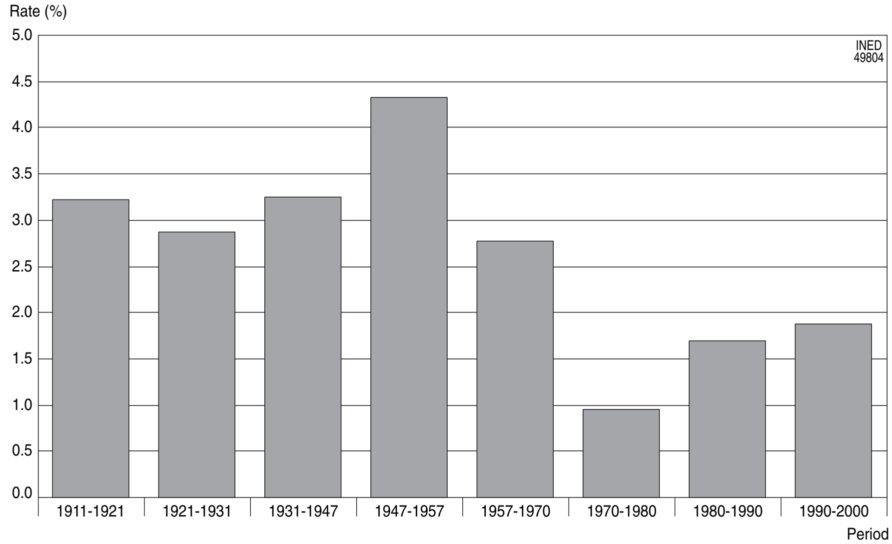


FIGURE 103–1 Intercensal average annual growth rate for the Singapore population, 1911–1921 to 1990–2000. (Source: SDS, 2001.)

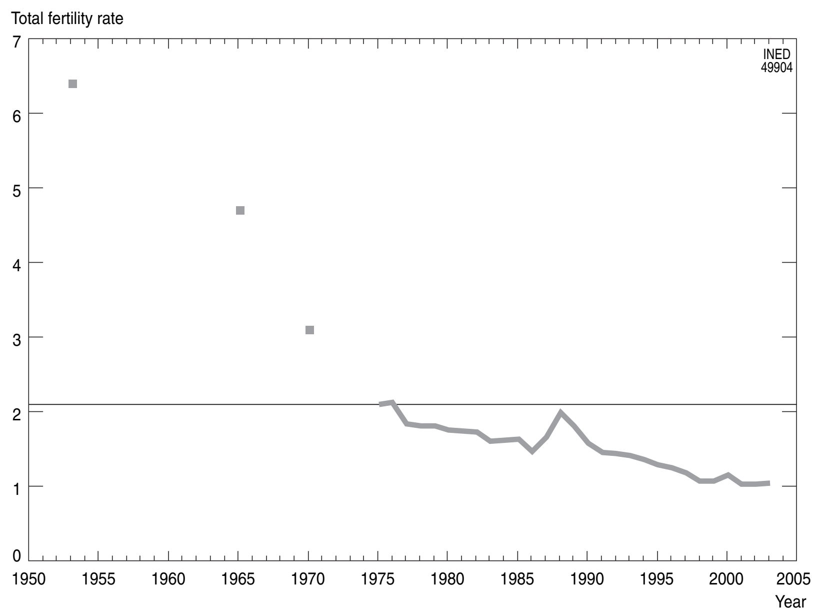


FIGURE 103–2 Fertility evolution in Singapore since 1950. (Source: International Data Base, U.S. Census Bureau. Retrieved June 20, 2004, from http://www.census.gov/cgi-bin/ipc/idbsprd)

improvements were considerable: life expectancy at birth reached 73 years in 1987, and infant mortality was lower than in the United States, the United Kingdom, and West Germany. This general situation encouraged very rapid fertility decline, and the total fertility rate (TFR) plummeted from 6.5 children per woman in the mid-1950s to 1.4 in 1986, after crossing the 2.1 threshold for the first time in 1975 (Figure 103–2).

From the mid-1970s onward, the government has been concerned by the fertility decline's effect on population age structure, especially on the ratio between the working population and the elderly. For the first time, in 1978, the demand for labor exceeded the offer. This prompted deliberations within the government that culminated in setting aside birth limitation policies and adopting a series of strictly pronatalist measures from 1984 onward.

This new policy's goal was to stabilize fertility above replacement level through an increase in three-child families. Some measures are similar to those introduced in the West: tax incentives, housing subsidies, ease of access to the best state schools, and so on, but they were implemented alongside other original policies. Thus, for example, as fertility decline was partly attributed to an increase in age at marriage and the growth of female celibacy, the government developed programs designed to help young people meet each other and form unions through activities worthy of the best marriage bureaus.

Its discriminatory nature, however, was the most striking aspect of this new policy. In 1984, the Singapore prime minister's address inaugurating the new population policy emphasized the importance of encouraging the most educated women to increase their fertility (Lee *et al.*, 1991; Teo, 1995). At the start, therefore, university-educated women were the only ones to benefit from the advantages associated with the arrival of a third child. Conversely, the least educated women (who had not finished secondary education or whose spouse had not) continued to be offered economic benefits if they agreed to be sterilized before having a third child. It soon became clear, in political discourse, that these measures were designed to encourage population growth within the dominant ethnic community, that of the Chinese, at the expense of Indians and Malaysians. These were not the groups who expressed their opposition to the government's family policy, however; it was rather women with some education (completed secondary school) who demanded the same advantages as the most educated women.

Their demands were satisfied by the adoption of a new program that substituted income level for education level as the condition of admissibility, and dropped the deterrents placed on the least privileged couples. "Have three or more if you can afford it!" became the slogan for this program. The economic and social incentives were generalized, and a special effort was made to encourage early births (additional fiscal advantages for women having a second child before the age of 28 years). Finally, all women with fewer than three children who applied for an abortion or sterilization were compelled to attend a counseling session especially aimed at dissuading them.

Up to the present day, this pronatalist policy does not appear to have born fruit: as the 21st century dawns, the TFR is fluctuating around 1.4 to 1.5 children per woman, the average age at reproduction is steadily increasing (from 27 years in 1980 to almost 31 years in 2000) (Singapore Department of Statistics [SDS], 2002a), and the proportion of third or higher births is still falling (from 55.5% of births in 1980 to 16% in 2000) (SDS, 2002a). Bolstered by its experience during the 1960s and 1970s, however, the Singapore government remains confident in its capacity to modify reproductive behavior in the end, and is continuing this course of action, even introducing a new program in 2000 to reinforce the measures developed since the end of the 1980s. This policy approach is made all the more interesting by the fact that it is being implemented in a context of very strong population pressure, with densities today reaching more than 5,000 persons per square kilometer (Figure 103–3) and an annual population growth rate close to 2%, two-thirds of it from sustained immigration (SDS, 2001).

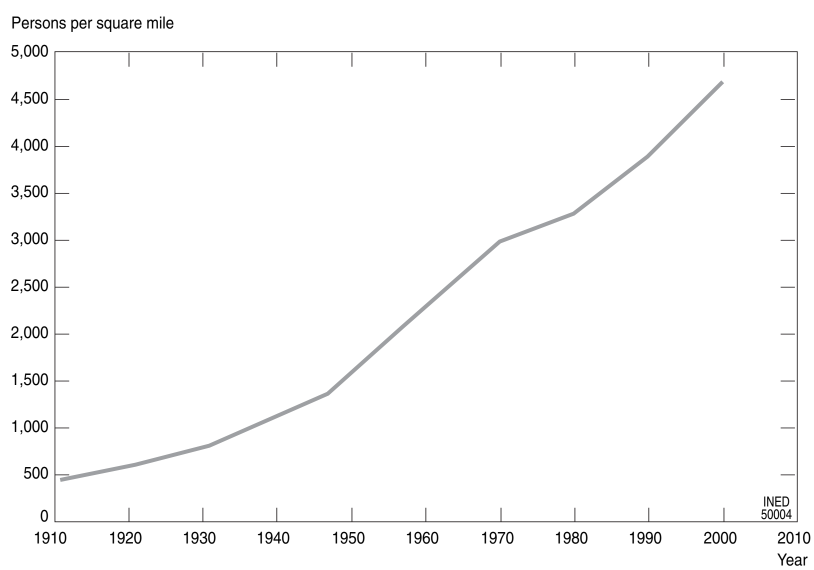


FIGURE 103–3 Evolution of population density in Singapore since 1911. (Source: SDS, 2001.)

II. SOUTH KOREA

Like Singapore, South Korea's current population policy marks a complete break with that of the 1960s, but the approach taken is an entirely different one. In 1961, South Korea was the third country, after Pakistan in 1953 and India in 1958, to develop a birth control policy to facilitate the country's economic development. Initially, the program focused on family planning education and the promotion of modern contraceptive methods, using local health professionals specifically trained in these areas. In 1981, family planning program developments were accompanied by new measures, articulated around a set of fiscal and social incentives designed to encourage birth limitation, similar to those introduced in Singapore. The government committed substantial funding to support this program, multiplying by 10 the budget allocated to it (Kwon, 2001).

From the 1960s onward, sustained economic growth reached an average annual rate of 8% and remained at this level for 30 years. At the same time, mortality was falling steadily: between 1960 and 1995, life expectancy at birth rose from 55 to 73 years and the infant mortality rate dropped from 56 to 13 per 1,000. In 1988, population policy had achieved its original goal: the TFR, as high as 6 children per woman in 1960, had fallen below replacement level (Figure 103–4), and the rate of contraceptive use was up to 80%, compared with 25% in 1971 (Kwon, 2001). Within the government, therefore, voices were heard lamenting the negative consequences of the population policy implemented during the three previous decades.

Various types of issues were raised, including some related to changing population age structure (Figure 103–5), associated with a rapid increase in the dependency ratio and of the proportions of elderly (the proportion of children under 15 years halved between 1960 and 1995, whereas that of the population aged 65 years and over more than doubled and is set to double again between 1995 and 2010), and others concerning reproductive health, with the disturbing prevalence of selective abortion by sex (Cho, 1996). The sex ratio at birth, which had remained at about 105 boys for 100 girls until 1982, began to rise, reaching 116 in 1993 (Figure 103–6). More troubling still, this increase was particularly strong for lower birth orders, increasing from 108 to 117 for second births, and from 129 to 191 for third births in the space of only five years (between 1985 and 1990), despite criminal lawsuits against doctors practicing selective abortion (three years imprisonment and fines of up to 12,500 American dollars).

In 1996, a new policy strategy was adopted, breaking with that implemented 35 years earlier. In contrast to the approach developed in Singapore, however, the Korean government decided to invest in counteracting the negative economic and social consequences of low fertility rather than attempting to modify reproductive

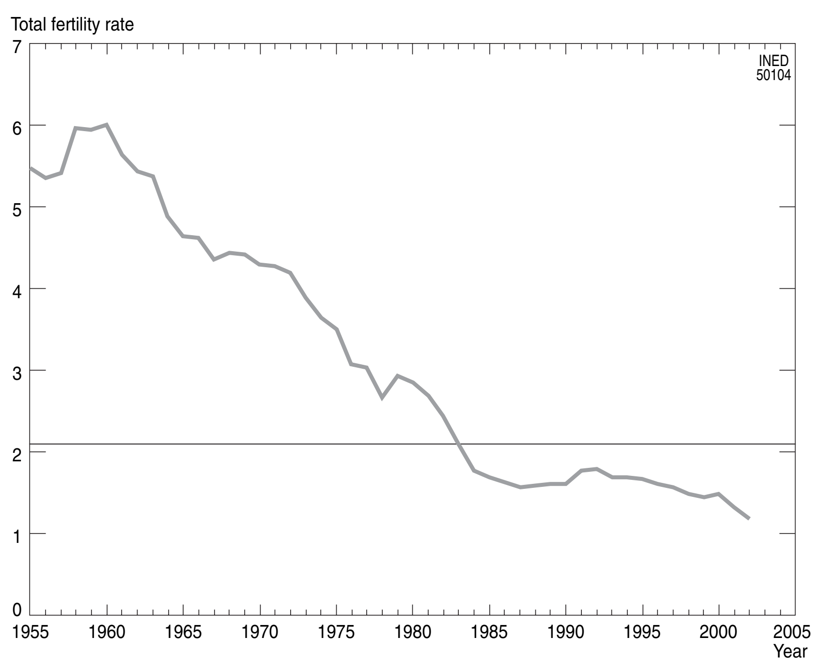


FIGURE 103–4 Fertility evolution in South Korea since 1955. (Source: Before 1975: Coale *et al.*, 1980; after 1975: Korea National Statistical Office. Retrieved June 22, 2004, from http://www.nso.go.kr/eng.)

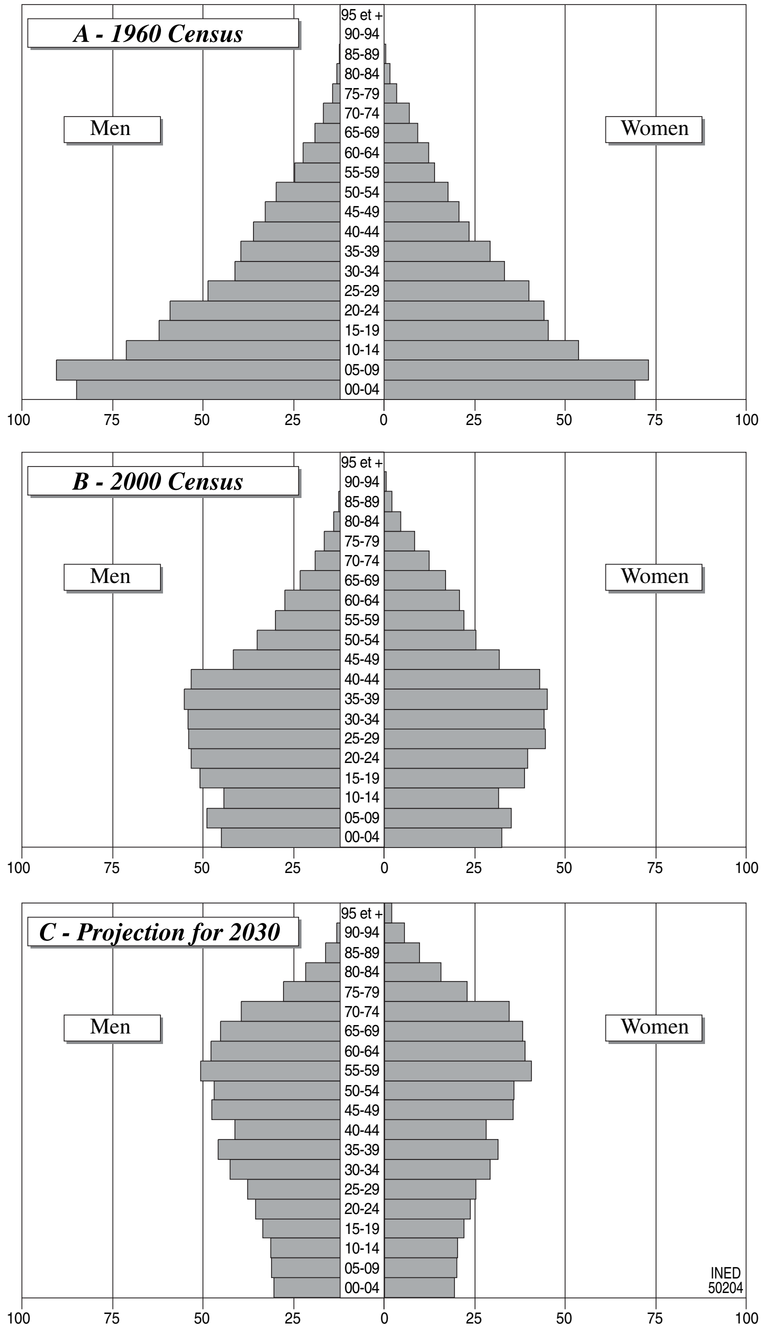


FIGURE 103–5 Population pyramids for South Korea in 1960, 2000, and projected for the year 2030. (Source: National Statistical Office: census results and projections. Retrieved June 22, 2004, from http://www.nso.go.kr/eng).

behavior. Supporters of this new perspective believe the efforts made to raise fertility levels by Western and other Asian countries, such as Singapore, Taiwan, and Malaysia, to be in vain. The strictly demographic aspect of Korean policy is thus limited to restructuring the family planning program and adopting legislation designed to fight discrimination against girls, whereas the new policy essentially operates within the area of employment and in the creation of social support policies for the elderly.

The family planning program was reformed with the aim of improving the quality of care and diversifying the contraceptive methods available; until then, the main investment was in sterilization—the only

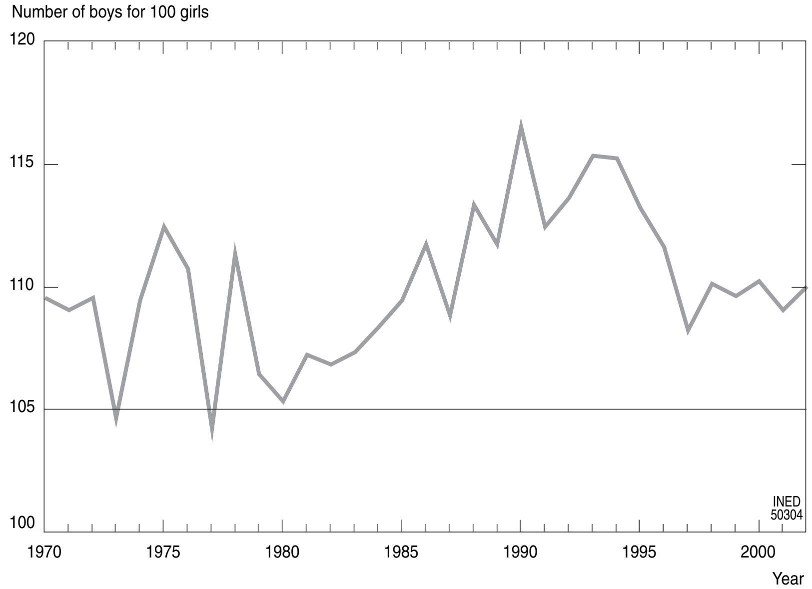


FIGURE 103–6 Evolution of the sex ratio at birth since 1970, South Korea. Source: *Ibid.*, Figure 103–4.

practical low-cost method. The goal was no longer to increase the contraceptive prevalence rate but to reduce the number of abortions, whose proportion had risen from around 25% of births in 1970–1974 to almost 50% in 1985–1991, and to improve coverage for the benefit of economically disadvantaged women and adolescents by extending family planning services to all aspects of reproductive health, especially those concerning sexually transmitted diseases. The program shared one point in common with Singapore's discriminatory policies, however: the poorest women were encouraged to be sterilized after their second child and became almost the only ones to benefit from fertility control measures. For the rest, family planning services were left increasingly to the private sector, although contraceptives continue to be reimbursed by Korean social services.

In years to come, South Korean population policy may turn to more resolutely pronatalist measures. For the moment, pressure on the labor force is certainly weaker in South Korea than in Singapore; the active population is still growing, although more slowly; and its proportion of the total population is now stable, after a steady rise between 1960 and 1995 (from 55% to 71%). According to the estimates, this proportion should nonetheless drop by 5 to 6 percentage points in the course of the next 20 years, at the same time as the demand for labor increases, producing a deficit of over a million jobs in 2020. Insofar as the Korean government does not wish to encourage immigration for social and political reasons, altering reproductive behavior may become a government priority once again.

III. OTHER LOW-FERTILITY COUNTRIES

Although other low-fertility Asian countries are showing fewer signs of discarding earlier birth control policies than Singapore or South Korea, there is nonetheless evidence of a change in direction. Certainly, China and Indonesia, the largest of them, are not yet ready to reverse their policies in favor of higher fertility, as demographic inertia leaves them concerned about the socioeconomic consequences of population growth in the medium term. Although characterized by the lowest fertility rates in the world (0.9 children per woman, according to the most recent estimates) but politically part of China, the territories of Macao and Hong Kong also show no interest in supporting any upward reproductive trend. Elsewhere, however, the fact that governments are discarding the strictest aspects of birth control programs, and are cutting spending on national family planning programs more generally, can be interpreted as the first signs of a turnaround in demographic policy (Mason, 2001). These types of changes are visible in Malaysia, Thailand, the Philippines, and especially in Taiwan where they are the most striking.

From the 1960s onward, Taiwan introduced many noncoercive measures to reduce fertility. The TFR fell below 2.1 as early as 1984 (Figure 103–7), and new programs were created in 1989, and reinforced in 1992, which moved away from the earlier policy approach. The two-child family continued to be encouraged, but the government developed a set of measures promoting

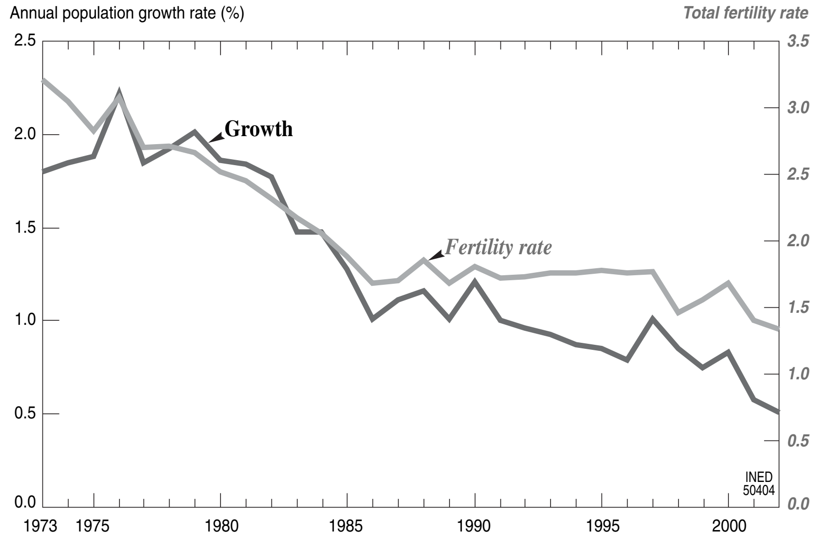


FIGURE 103–7 Population growth and fertility in Taiwan, 1973–2002. (Source: Republic of China, Statistical Yearbook of the Republic of China, 2003, Table 8, p. 11 and Table 15, p. 28.)

family values and programs supporting marriage among young people, rather like Singapore. With a population growth rate cut by four in 20 years (Figure 103–7) and considered satisfactory (Zhenghua *et al*., 1994), the goal was not so much rapid population increase as the gradual stabilization of fertility around replacement level. As the fertility curve illustrates, these measures could not reverse the TFR's steady decline. The most recent official statements, however, which simply reiterate their support for marriage, show that the government is still cautious about adopting pronatalist measures (Retrieved May 25, 2004, from http://www.moi.gov.tw/english/PopulationAffairs.asp).

In Taiwan as elsewhere, this type of policy is accompanied by the transfer of family planning services to the private sector or by their integration into the health services. More and more frequently, rather than reducing population growth, policy stresses the need to diversify and improve the quality of family planning services. Like Taiwan, one after the other—Malaysia, the Philippines, and Thailand—have abandoned quantitative goals with respect to fertility or demographic growth, and new concerns are emerging in spheres that, earlier, were at best considered peripheral but were most often completely absent from population policy.

The United Nation's International Conference on Population and Development held in Cairo in 1994 signaled a break in this respect, promoting concepts such as women's status and improving gender relations and reproductive health, which Asian countries increasingly take into account in their population policies. From that point on, these programs, previously directed exclusively toward married women, attempted to include adolescents and single women of all ages. Indonesia is a perfect example of this. During the first thirty years of its existence, its national family planning program ignored adolescents and unmarried women; still today, however, whereas the fertility level has just fallen below replacement, a profound change is emerging in government priorities. Aiming not only to make contraception available to all women, irrespective of marital status, it also intends to involve men, ambitiously aspiring to reach a level of male contraceptive use of 10% by the year 2010. Moreover, the fact that Asian governments have taken up the reproductive health concept has led to a growing integration of family planning programs into health services; it has also made it possible to direct resources previously devoted uniquely to reducing fertility levels to other areas, the first of which is the battle against sexually transmitted diseases in general, and the HIV/AIDS epidemic in particular.

CONCLUSION

The Singapore government's dynamic pronatalist policy is an exception among low-fertility Asian countries as a whole. Most are still cautious, not only about engaging on this extreme course, but even about relaxing their antinatalist policies established during the period of strong demographic growth. The historical and political context accounts more for this situation than do either the current demographic characteristics or the predicted course of population and components. Singapore does not differ significantly from South Korea, Hong Kong, or Taiwan in terms of its fertility levels, the present or future proportions of its elderly population, the population pyramid as a whole, or the estimated date at which total numbers will start to fall. It is understandable, on the other hand, that governments are loathe to envisage a radical policy reversal in countries where demographic inertia is still strong—where, in other words, the population's past evolution means that it will continue to increase strongly in future decades despite low fertility levels, as in Thailand (25% increase projected between 2000 and 2030)—or simply where the numbers involved, even if only a small proportion of total population, represent a huge demographic mass, as in China (more than 180 million persons between 2000 and 2030) or Indonesia (more than 70 million).

Outside Asia, countries whose fertility is likely to reach replacement level in the short term are still scarce. With the exception of many Caribbean islands and others in the Pacific, where geographic isolation lends a positive perception of the negative population growth predicted for 2030, the few countries where TFR will drop below 2.1 children per woman by 2010–2015 are distributed between Latin America (Costa Rica and Brazil), the Middle East (Lebanon and Iran), and North Africa (Tunisia). These are the candidates for an about-turn in population policy. Elsewhere, changes in reproductive behavior have been slow or late enough that the proportion of persons aged 65 years and over and the projected dependency ratios for the next half-century remain favorable for economic growth.

Acknowledgment

This chapter was translated from the French version by Heather Juby.

References

Coale Ansley, Cho Lee-Jay and Goldman Noreen, 1980. *Estimation of recent trends in fertility and mortality in the Republic of Korea*. Washington, National Academy of Sciences, Committee on Population and Demography, 77 p.

Cho Namhoon, 1996. *Achievements and challenges of population policy development in Korea*. Seoul, Korea Institute for Health and Social Affairs, 100 p.

DGBAS, 2003. *Statistical yearbook of the Republic of China 2003*. Taipei, Directorate General of Budget, Accounting and Statistics, 274 p.

Kwon Tai-Hwan, 2001. The national family planning program and fertility transition in South Korea, *in*: Andrew Mason (ed.), *Population policies and programs in East Asia*, p. 39–64. Honolulu, East-West Center, 157 p. (East-West Center Occasional Papers no. 123).

Lee Sharon M., Alvarez Gabriel and Palen J. John, 1991. Fertility decline and pronatalist policy in Singapore, *International Family Planning Perspectives*, vol. 17(2), p. 65–69, 73.

Mason Andrew (ed.), 2001. *Population policies and programs in East Asia*. Honolulu, East-West Center, 157 p. (East-West Center Occasional Papers no. 123).

SDS, 2001. *Singapore population*. Singapore, Department of Statistics, 21 p.

SDS, 2002a. *Twenty-five years of below replacement fertility: implications for Singapore*. Singapore, Department of Statistics, 12 p.

SDS, 2002b. *Singapore's current population trends*. Singapore, Department of Statistics, September, 6 p. (Statistics Singapore Newsletter).

Teo Peggy, 1995. Population planning and change in Singapore, *Population and Environment: A Journal of Interdisciplinary Studies*, vol. 16(3), p. 237–252.

Jiang Zhenghua, Zhang Lingguang, Cheng Chang Ming, Hong Moon Sik and Hayashi Kenji, 1994. Fertility policies in China, Taiwan, Republic of Korea and Japan, *in*: Korean Institute for Health and Social Affairs (ed.), *Low fertility in East and Southeast Asia: issues and policies*, p. 267–288. Seoul, Korean Institute for Health and Social Affairs, 380 p.

CHAPTER

104

Pronatalist Policy in Industrialized Nations

GUSTAVO DE SANTIS

Facoltà di Scienze Politiche, Università di Messina, Messina, Italy

INTRODUCTION: WHAT IS PRONATALIST POLICY?

Pronatalist policy is a complex and coherent system of measures designed to promote fertility. Essentially two avenues can be taken to portray pronatalist policy: start with government stated objectives or measure the empirical effects of selected policies. Both approaches, however, have significant drawbacks.

In Gérard Calot's opinion (Chap. 105), France is probably the only developed democratic country to have adopted explicitly pronatalist policies with some continuity through time, before significantly moderating its position on the issue in recent years. In 1995, however, Miroslav Macura, Mitchell Eggers, and Thomas Freijka (1995) created a more extensive list of pronatalist countries for the period 1976–1991; along with France, at least part of the time, were Bulgaria, the former Federal Republic of Germany, Greece, and Hungary.[1] Jonathan Bradshaw, John Ditch, Helen Holmes, and Peter Whiteford (1993) included Luxemburg, Norway, France, and Belgium in the group of countries showing the greatest generosity toward large families.[2] Shortly afterward, however, the same (or almost the same) group of authors extended the list to include Germany (Ditch *et al.*, 1998b). Anne Gauthier and Jan Hatzius (1997), in contrast, only mention France and Belgium as explicitly pronatalist countries, whereas Didier Blanchet and Olivia Ékert-Jaffé (1994) listed Austria, Hungary, Czechoslovakia, Bulgaria, and France.[3] Since then, Anne Gauthier (1999, 2002) has proposed yet another classification, and such examples could be further multiplied. In a word, it is not easy to classify countries in terms of their degree of support for fertility. For instance, governments respond to questionnaires on the topic sent regularly to them by the United Nations (United Nations, 2002) in ways that do not always correspond to what is actually done, and the traditional typology of industrialized countries (see, e.g., Esping-Andersen, 1990; Reher, 1998) does not always make it possible to group together countries with similar fertility policies.

Besides, an examination of the panoply of measures actually implemented in different countries, and a comparison of them with the declared intentions,

[1] It is true that this list includes countries which, at least during the period concerned, were not "developed and democratic," as the condition used by Gérard Calot in Chapter 105 would imply; a good part of them are nonetheless developed countries.

[2] According to Anne Gauthier (1996b), this classification is based on observed results, not on stated objectives.

[3] The criterion used here to identify pronatalist countries is the concentration of measures on families with a second or third child. The same authors, however, suggest an additional criterion—the economic relevance of this assistance, which, at that time in the countries cited reached between 10% and 19% of the average wage. In a second group of countries (Germany, the Netherlands, Norway, Poland, Austria, Japan, USSR), financial assistance to families also increases with the number of children, but to a lesser extent (from 1% to 11% of the average wage), thereby indicating, according to these authors, objectives that fall between pronatalism and redistribution (see below).

shows that pronatalism is in any case scarcely mentioned and that it is perhaps not even the main objective and is just one among many other goals, such as women's health protection, promotion of mother's labor force involvement, support for living standards in relatively large families, and so forth.[4] In order to classify these projects within a pronatalist framework, the relationship of cause and effect must be identified and measured. Already an extremely difficult goal to achieve within the social sciences in general, however, it seems virtually unattainable in this particular case, given the problem of defining the independent variable (population policy) and of establishing the criteria of analysis of dependent variables.

According to Jean-Claude Chenais (1985), four indicators could be used to achieve this:

- Cohort total fertility rates
- The period total fertility rate (TFR), also recommended by Gérard Calot (2001)
- The rate of natural increase
- Population age structure

From a strictly demographic viewpoint, cohort female reproduction (as defined in Chapter 19) is probably the ideal dependent variable. Despite this, studies rarely take it into account for a number of reasons, one of which being purely practical: as making policy decisions and evaluating their consequences are generally short-term undertakings, the long period of time needed to measure cohort fertility is hardly compatible with the brief duration in which policies are generally of relevance (including the population policies of interest to us here). The possible alternatives fulfill different criteria. The last two, moreover, take account of the need to provide results that are easily understood by the general public which, in democratic systems, elects the government and guides their policies, including those in the population sphere. One consequence of the problem of choosing between the possible indicators, at least while the four of them are not evolving in the same direction, is to provide arguments for government leaders who prefer to postpone their actions.

Writing about the various approaches, Anne Gauthier and Jan Hatzius (1997) distinguish four methods of investigation, according to whether they are based on

[4] For measures currently implemented in Europe, see *The Clearinghouse on International Developments in Child, Youth, and Family Policies* (http://www.childpolicyintl.org/). Sleebos (2003) provides a synthesis of these measures and of their possible effects on fertility (see Palomba and Menniti, 1994, for a similar effort, ten years before). It should not be forgotten, however, that many family policies evolve rapidly and not necessarily in the same direction, either in time, or from one country to another (Gauthier, 1999, 2002).

1. Opinion polls ("In a given context, what type of population policy is acceptable, or desired, or effective?" "Would you change your demographic behavior if the government took this particular measure?") (Moors and Palomba, 1995);
2. Quantitative intercountry comparisons that have not been formalized within a statistical model that tests hypotheses, checks correlations, or the like (Chap. 105; Chenais, 1985, 1996);
3. Formal quantitative comparisons[5] (Ékert, 1986; Blanchet and Ékert-Jaffé, 1994; Gauthier et Hatzius, 1997);
4. Individual data (Blanchet, 1987).[6]

The first method often focuses on a particular policy measure, while the three others have generally been used up to now only to evaluate the global results of the whole system operating in one country or another. Moreover, as it collects no data on real variations in demographic behavior following the adoption of new laws, the first approach can go no further than informing us of respondents' opinions and intentions in the face of new events that are purely hypothetical ("Would you have, or would you have had, a [another] child if . . .?") (De Santis and Breschi, 2003).

In fact, to identify specifically (potentially) pronatalist measures, or the most important ones at least, the different authors frequently choose what is probably the only feasible approach in the current state of knowledge: to restrict themselves to those that appear logically most closely linked to reproductive behavior. It is important to remember, however, that this choice carries a strong risk of losing sight of potentially important elements. It could be argued, for example, that all governments always implement demographic policies to some extent, though often implicitly rather than explicitly, in the way the fiscal system treats families (according to the number, the work situation, and kinship relationships of their different components, for instance), in the provision of free or subsidized services (child care, education, medical care, etc.), and the like (Chap. 105). Besides, attempts have often been made to study and quantify the financial impact of current legislation on families by composition and

[5] Supposedly a far superior method to the preceding one. As already mentioned, however, it should not be forgotten that the dependent variable is not self-evident and that population policy is composed of a wide variety of courses of action, often difficult to measure and intrinsically impossible to compare (economic support, career support, leave, provision of various services with different territorial or social coverage, etc.): translating all these measures into quantitative terms can sometimes give the researcher a false image of objectivity.

[6] Which nonetheless often employ simulation methods. Other studies based on microdata have been evaluated in Gauthier, 1996a.

income (Franco and Sartor, 1990) and, from the slightly different angle of intergenerational accountancy, the age profile of economic exchanges between individuals and the State, also an indirect indicator of the protection offered to minors and their families (Jägers and Raffelhüschen, 1999; Sartor, 2004a).

At times, macroscopic effects have been underscored[7]: in Chapter 105, for instance, Gérard Calot cites two classic examples of strong variations in marriage associated with the current economic climate: one observed in Austria in 1987 (linked to a modification of family taxation) and the other in Sweden in 1989 (linked to new retirement pension reversion laws). Jean-Claude Chenais (1996, 1998) also holds the policy context responsible for fertility differences in Sarre between the years spent under French administration (1945–1977) and the period that followed, after uniting with Germany; or those between East German and West German fertility up to reunification in 1990, and so on.

The case of Sweden's recent fertility revival (followed by its decline) also seems to belong to the realm of unexpected demographic consequences of political economy policies. In Sweden, working women receive a maternity allowance equivalent to a certain percentage of their most recent salary. Therefore, if a woman reduces her workload (e.g., by going part-time) when her first child is born and becomes pregnant a second time, her new allowance will start from a lower salary level. To correct this shortcoming, from 1984 onward women with successive births have been guaranteed an allowance at least equal to the preceding one, as long as the subsequent birth occurs within 30 months of the one before (Näsman, 1991; Rønsen, 2004). The resulting effect on fertility in the early 1990s, however, when the TFR rose above replacement level of 2.1 children per woman, appears to have almost disappeared today; since 1996, Swedish fertility has hovered between 1.6 and 1.5 children per woman.

Measures adopted to protect certain social categories, like single mothers in the ex-GDR in 1976 (Vukovich, 1993), or in the United States today (Gauthier, 1996a), provide interesting cases of unexpected and perverse consequences. In this type of situation, a fear, partly reinforced by the facts, is that concentrating assistance on specific groups will encourage individuals to place themselves voluntarily in protected categories in order to take advantage of these legally sanctioned benefits. Thus, in the concrete case mentioned here, the worry is that such measures will encourage women to have children outside legal marriage (and, especially in the American case, without looking for work).

Thus, pronatalist policies are extremely diverse in the approach they take and in their content. Without wishing to deal with all these aspects, I will attempt in the following text first to clarify the relationship between pronatalist and social policies, and, in a second section, to classify pronatalist policies essentially in terms of their content while also taking into account several other complementary criteria.

I. PRONATALIST POLICY, SOCIAL POLICY

Ignorance of the real cause of fertility decline discourages the adoption of pronatalist measures that, at least in terms of long-term fertility levels, are costly and only moderately effective (see, e.g., Olivia Ékert, 1986; Didier Blanchet and Olivia Ékert-Jaffé, 1994; Anne Gauthier and Jan Hatzius, 1997; Rønsen, 2004) or virtually ineffective (Demeny, 1987, 2003; Ditch *et al*., 1998a). Results are slightly more visible, on the other hand, over a shorter duration (see, e.g., Näsman, 1991) or when it involves reducing disparities between different social classes (Ékert-Jaffé-Jaffé *et al*., 2002).

In addition to doubts about their effectiveness is the public's poor appreciation both of what is really happening (despite insufficient fertility levels, the population in industrialized countries is generally still growing) and of the possible consequences which, not perceived as completely negative (as a decrease in population numbers can reduce pressure on resources and the environment), are therefore seen as insignificant, remote, and not requiring urgent action (Frinking, 1985; Vukovich, 1993). There is a potential lack of motivation, therefore, at the base of the rare explicitly pronatalist policies implemented.

Arising more generally out of the arguments usually used to justify public action, there is also some ambiguity about how to confront the problem (Sartor, 2004b).[8]

[7] There have probably also been many other cases where the effects, less evident or more diffuse in time, have gone almost unnoticed or, at least, for which it is more difficult to find clearly identifiable causes.

[8] I will not discuss here the question of knowing the extent to which public actions are decided through the power exerted over collective preferences by a minority. This interpretation is basically the one suggested by Samuel Preston (1984), who provides a plausible explanation for the inadequate support given to fertility by modern States: that the elderly, now the dominant group, use up resources that they do not wish to see redistributed to younger generations and their families (for a review, see Gauthier, 1999, 2002). Similarly, Peter McDonald (2002) expresses his astonishment that the young do not create pressure groups to obtain more state resources for family formation, and Olivia Ékert-Jaffé *et al.* (2002) consider that it is possible to assess the effects of social policy on the fertility not only of the whole population, but also for population subgroups affected differently by the various measures.

On the one hand, there are "market deficiencies": most individuals wish to behave in one way (in this case, to have an adequate number of children), but individual behavior does not conform to this ideal for various reasons. For instance, although surveys indicate that approximately two children is the ideal in almost all developed countries (although this has perhaps started to fall more recently: see Goldstein *et al*., 2003), the actual TFR is most often significantly below this level. This gap between ideals and actual behavior could be attributed to the operation of the market economy which, by encouraging short-term interests, inadequately compensates those who, by having children, are on the contrary investing in the long term (McDonald, 2002).

At an entirely different level, the state may instead be perceived as a protector, aware of, and concerned by, the long-term consequences of individual behaviors whose repercussions are not necessarily clear to the individuals themselves. This is the justification for compulsory pension plans, for example: left to themselves, individuals would not accumulate sufficient savings during their working life and would find themselves unexpectedly impoverished in their old age. The same could also apply to fertility: individuals prefer to have few children (or none at all) because they do not clearly appreciate the effect that low fertility will have on their quality of life during old age (Rendall and Bahchieva, 1998). This is also made all the more true by a predicted population composition imbalance: superficially, having children no longer seems necessary to guarantee comfortable retirement years, as alternative social protection systems now exist. These can only function, however, with a relatively regular succession of cohorts and therefore if the collapse in fertility is not generalized.

This basic ambiguity about the state's role in fertility, in turn influenced by the bad memories left by the pronatalist policies of many totalitarian governments, is also responsible for the lack of clarity concerning legitimate ways of curbing fertility decline. What seems sure is that such measures can never be presented in the form of punishment for those who fail to do certain things (e.g., the tax on celibacy introduced by the fascist regime) (Ipsen, 1997), but only as advantages for those who do them (e.g., cheap loans for young couples or facilitating access to housing); these advantages must be carefully justified, especially by invoking the need to cover part of the cost of more expensive lifestyles, thereby truly giving individuals the freedom to make choices for which they would otherwise be penalized.

Intervention through prohibitions is also becoming increasingly rare and politically delicate (prohibiting contraception, abortion, divorce, etc.). It is therefore preferable, and even essential, to give priority to incentives, encouragement, and guidance, while completely respecting individual liberty. Concretely, this means making certain individual choices more attractive than others, whether by using financial incentives or other mechanisms.

Beyond the simple question of numbers, family and society are also interested in ensuring that future generations do not fall below a certain threshold, particularly with respect to the quality of health care and education. This often translates into the provision of certain free or subsidized services, such as medical visits, schooling, and the like. Legislation aiming to reach this qualitative goal, however, also demands a heavy parental investment (through extending the duration of schooling, e.g.) and may lead to the interiorization of these values by families themselves (Moors and Palomba, 1995). In some cases, therefore, this legislation may end up reducing fertility excessively, because, when limited resources are available, investing more in each child means cutting down on the number of children.

Added to this potential conflict between quality and quantity, highlighted at the family level by Gary Becker (1960, 1981), is the controversial question of the family's role and women's position in society: do children thrive more if they are cared for, at least during early childhood, by the family (especially the mother) or in an appropriate institutional setting[9]? In the first option, mothers are encouraged to reduce their outside commitments, at least during the first years after the birth, although this creates a potential conflict with their professional aspirations (Chenais, 1998; McDonald, 2002); in the second, substantial investments are made in nurseries, child care centers, and other early childhood facilities. More often however, countries oscillate between the two options and move somewhat in both directions.

The population question is subtly linked to that of social policy, which has reached sizeable proportions in all developed countries (absorbing between 18% and 33% of gross domestic product [GDP]); this often generates some confusion between the two spheres, pronatalist and social, in terms of means and goals. Thus, families with children are often granted economic and administrative benefits, at least as long as they do not exceed a certain income threshold, beyond which these advantages diminish or disappear completely. However, as mentioned, social policies are

[9] Linked to this is the following controversial issue: are the best institutions private (as in Anglo-Saxon countries) or public (as is generally the case in continental Europe)?

often the camouflage used by governments to have fertility support measures accepted by a sceptical public opinion: having children is expensive and leads to a reduction in living standards, at times below the poverty line and, officially, the state intervenes only to prevent this happening.

Finally, it is just as important to consider the individuals toward whom population policy may be directed. The traditional family of the past was ordinarily represented by a family head (male, economically active) who benefited from any services offered by the State or his employer (such as family allowances, medical services for himself and his family, etc.). With the disintegration of this family form and the multiplication of alternative lifestyles that do not exclude having children, defining the beneficiaries, and even the typology of interventions, has become difficult since, as already mentioned, they often implicitly conceal a bias in favor of certain family types. Paying a salary to housewives, or at least to mothers who stay at home, for instance,[10] tends to exclude working mothers; if it is a fixed amount, a monthly allowance allocated at the birth of a child tends to penalize working women (who have to give up their salary, at least for a time); if it varies according to the woman's salary, on the other hand, it tends to benefit the wealthiest social classes. This explains at least part of the extreme care taken when adopting different measures, their variability through time and space, and perhaps also the caution over introducing index-linked measures. This is a sort of recognition of the need for perpetual revision of the bases and criteria of public intervention in fertility.

For European Union countries, a final dimension is worth bearing in mind. The European Union has no direct experience with family policy issues; nor does it have a shared definition of the family. But it is competent in many areas closely linked with the family and fertility, like that of labor force activity (which, e.g., interacts with maternity in terms of maternity leave) and of sexual equality, which motherhood should not threaten. The relationship between national governments and membership of the European Union is certainly a complex one, and if national policies in this domain converge much less than might have been expected, or do not converge at all, this is not by chance (Gauthier, 1999, 2002). In the future, however, it is possible that the problems shared by all European Union countries—the inadequacy of intergenerational exchange first among them—will end up homogenizing their policies to some extent.

[10] Which can occur, even implicitly, in terms of fiscal policy: for example, by granting a tax reduction or allowance for dependent wives.

II. A POSSIBLE CLASSIFICATION OF PRONATALIST POLICIES

A detailed analysis of fertility support measures adopted in different countries is impossible in the space of a few pages because of the frequent changes of direction, the complexity of different legislations within which specific actions simultaneously pursue a variety of goals (demographic, social, spatial redistribution of resources, etc.), and the lack of comparability between contexts (diversity of per capita income, variations in women's full- or part-time labor force involvement, etc.). (For a more complete account, see the sources cited in note 4 and the references at the end of the chapter.)

Table 104–1 brings together the elements of a framework and calendar for pronatalist practices, in the broader sense of support or protection for motherhood. As can be seen, the earliest interventions go back to the end of the 19th century. Up to the 1960s, measures gradually became more generalized (e.g., from a few groups of women to all women) and were reinforced, extending the period of compulsory or optional abstention from work and increasing economic support. Since then, trends have been less well defined, particularly for financial constraints, imposed by the already precarious state of many public budgets. In general, however, parental leave (often unremunerated) has been extended, allowing one

TABLE 104–1 Main Historical Stages of Family Policy Measures in Industrialized Countries

(1) 1870–1929	First paid maternity leave Prophylactic health measures for mothers and children Financial support for mothers in need
(2) 1930–1944	Allowances for workers with minor children Pronatalist policies in several countries
(3) 1945–1959	Universal family allowance Extension of maternity leave Other measures (preventative, health, and housing)
(4) 1960–1974	Higher family allowance for the poorest, means tested
(5) 1975–1995	Further extension of maternity leave and other benefits for working parents Articulated and integrated family policies (in several countries) Development of preschool child care services Public financial assistance for the poorest families

Source: Gauthier, 1996b, p. 193.

TABLE 104–2 Child Benefits, of Various Types[a], in Several European Countries (ca. 2000)

Country	First law	Conditional on means	Maximum age of child[b]	Amount (Euros per month, per child)[c]	Birth allowance
Austria	1948	no	19	105/145	no
Belgium	1930	yes	18 (25)	68/185	yes
Denmark	1952	no	18	91/126	yes
Finland	1948	no	17	90/120	yes
France	1932	yes[d]	20	104/133	yes
Germany	1954	no	18 (27)	128/179	no
Greece	1958	no	18 (22)	6/12	no
Ireland	1944	no	16 (19)	40/53	(yes)
Italy	1937	yes	18	10/137	no
Luxemburg	1947	no	18 (27)	139/233	yes
Netherlands	1939	no	18	49/97	no
Norway	1946	no	16	113/183	no
Portugal	1942	yes	16 (25)	n.a.	no
Spain	1938	yes	18	18	no
Sweden	1947	no	16 (20)	105	no
United-Kingdom	1945	yes	16 (19)	64/100	yes

[a] Periodic allowances, tax deductions, etc.
[b] The maximum age in particular cases (e.g., school, unemployment) are given in brackets. In cases of disability, age restrictions do not usually apply.
[c] Minimum/maximum (variations may depend on age, birth rank, family income, etc.).
[d] Benefits start at the second child.
Sources: The Clearinghouse on International Developments in Child, Youth and Family Policies (http://www.childpolicyintl.org); MISSOC (http://europa.eu.int/comm/employment_social/missoc/missoc_info_en.htm).

parent to spend time with the children as long as they have not reached a specified age, often 8 years old. Measures taken in this sphere are somewhat less linked to the economic and political climate; although they increased more slowly than other types of social spending during the 1960s and 1970s, they also proved more resistant to spending cuts during the 1990s, especially in the countries subject to the budgets restrictions imposed by the Maastricht Treaty.

All industrialized countries have strictly economic policies supporting families with children, although these diverge widely in their intensity and characteristics; some of these are presented in Table 104–2. Under the classic hypotheses of perfect competition (whose main assumptions include the rationality of parents/consumers, perfect prediction of the future, and the existence of a perfect captial market with no liquidity constraint), most of these actions should give perfectly equivalent results. In particular, allowances at birth or when the child reaches a given age,[11] monthly benefit payments, tax exemption, and the creation of a retirement credit for mothers can all be translated into a given sum of money expressed in current monetary value, taking account of the frequency of events and their distribution through time. However, the classic theses of marginal economics do not adequately represent the reality of industrialized countries today, for the manner in which the allowances are paid is of consequence for both the lawmakers and their beneficiaries (De Santis and Breschi, 2003).

In addition, many possible secondary regulations complicate any evaluation of potential benefits; among these, and not mutually exclusive, may be found the family's socioeconomic circumstances,[12] the wife's occupation (or sometimes the husband's, if the wife does not work), the children's birth order, the presence of two parents, the use of certain public services,[13] the existence of maximum and minimum transfer limits, and so on.

The presence of children (or of family responsibilities) is often a reason for tax exemptions. In the past, these exemptions often took the form of reduction in

[11] Among the European Union countries, this type of payment is made in Belgium, Denmark, Finland, France, Ireland, and the United Kingdom.

[12] The exclusion or reduction of benefits for the better-off is now an almost universal criterion, especially because of growing restrictions on budget balancing.

[13] Sometimes offered in place of cash transfers. For instance, mothers in Germany receive a child care allowance during the first year if they do not send their child to a state nursery.

taxable income. Because of progressive tax rates, however, this solution was more advantageous to the wealthiest families. Although justifiable from one point of view,[14] it was difficult to defend politically, especially during a period of reduction in public spending. Nowadays, transfers are most often made as fixed amounts and, usually, via tax credits only available to taxpayers.

The question of maternity leave and payments is particularly delicate both for compulsory and optional periods (Table 104–3). Most frequently, maternity benefit rates are linked to salary levels before the birth, although sometimes there are preset inferior and superior limits. Introduced in order not to discourage fertility among working women, this measure has various potential drawbacks. In particular, working women draw the greatest benefits from these measures, though some measures may discriminate among them at times by offering greater advantages to employees, notably those in the public sector.

The overall result, therefore, is to render the assessment of transfers rather obscure, linked probably to the absence, even in France, of a complete picture of the system. It is almost impossible for a family to calculate ex-ante (and even often ex-post) the total value of transfers received. To make meaningful international comparisons, therefore, estimates for specific family types are often employed; the problems remain, however, especially because countries' ranking changes, sometimes radically, with the passage from one family type to another, or with taking into consideration a more limited or broader range of services. As Table 104–3 suggests, for example, countries vary enormously in terms of the provision of child care services; for the youngest age group (0–2 years) these services cover from 2% to 60% of the population concerned, and in the following group (3–5 years), between 28% and almost 100%. The price itself is also very variable: free of charge in France and Belgium, services have to be paid for in the other countries. As costs generally change with family income, calculating the average cost is not easy, although estimates suggest that it may reach 36% of the average salary in Greece (and exceed 20% in Ireland and the United Kingdom).

These figures do not reveal all the differences, however—such as in spatial coverage and timing (with or without flexibility), in the quantity and quality of services provided by the institutions (such as meals), in preferential conditions of admissibility, and such like. All these variables are hard to quantify, but extremely relevant in terms of real support for families.

An additional reason for the poor comparability of family policies (among which we should look for potentially pronatalist measures) between industrialized countries (or even only between the European Union countries described here) is associated with housing policy. Such policies are usually aimed toward the most underprivileged population groups but, as family size enters into the assessment of needs, the largest families are often recognized as the most impoverished and therefore become the main recipients of such policies. These measures, however, are in turn relatively complex, and are often linked, in addition to the family's economic status assessment, to such factors as housing type (e.g., only social housing) or geographic situation (e.g., only zones with substantial pressure on housing).

Taking all these types of action into account as far as possible, Table 104–4 gives the list of European countries, ordered according to their level of family support (or more precisely, to 20 two-parent families with one to three children, not necessarily representative of families of this type). Although subject to the arbitrary nature of many choices, the gap between countries appears enormous in some cases: Luxemburg has a clear lead, followed by France and Belgium, who are practically equal. At the opposite extreme are Greece, Italy, Spain, and the Netherlands (with a net deduction on families instead of a transfer in their favor).

It should not be forgotten, however, that the variables researchers tend to examine do not include hidden transfers from the state to families through the provision of two free (or subsidized) services, namely health and education. Integrating these two elements into the calculation leads to the conclusion that even the least pronatalist countries (such as Italy) actually make positive rather than negative net transfers to families, as shown in Table 104–4 (Sartor, 2004a). This does not mean, however, that the table order would be upended if these advantages were also taken into account, as European countries are relatively homogeneous when it comes to taking responsibility for health and education expenses. The considerations about costs related to dependent children for which families actually remain responsible (for both direct expenses and opportunity costs) remain valid, but these costs would be far higher if families also had to pay for the medical expenses and schooling provided for their children.

[14] Despite their shortcomings, studies on the cost of children seem to suggest an increase in cost with the increase in family living standards (Ékert-Jaffé, 1994) and that of mother's income (opportunity costs; Gauthier, 2002).

TABLE 104–3 Maternity and Paternity Leave, and the Proportion (%) of Children in Nurseries or Preschool in Europe (ca. 2000)

Countries	Leave			Percentage (%) of children in nursery or preschool	
	Beneficiary[a]	Duration	Amount as a % of salary	0–2 years	3–5 years
Austria	M	16 weeks	100%	3	80
	P	30–36 months	Flat rate		
Belgium	M	15 weeks	75%–80%	30	97
	P	3 months per parent	Flat rate (low)		
	F	3 days	—		
Denmark	M	19 weeks	90%	58	83
	P	10 weeks + 52 weeks (×2) sick leave	60%		
	F	2 weeks	100%		
Finland	M	18 weeks	65%	48	73
	P	26 weeks + leave of absence and right to part-time employment	Flat rate		
	F	18 days			
France	M	16 weeks	100%	29	99
	P	3 years	Flat rate (means-tested)		
	F	11 days	100%		
Germany	M	14 weeks	100%	5	85
	P	3 years	Flat rate (means-tested)		
Greece	M	17 weeks	50%	(n.d.)	(n.d.)
	P	3.5 months per parent	Unpaid		
Iceland	M/P/F	3 months (×2) + 3 months (×1)	80%	(n.d.)	(n.d.)
	M	18 weeks	70%	2	55
Ireland	P	14 weeks + 3 days/year family leave	Unpaid		
Italy	M	5 months	80%	6	95
	P	10 months (+5 days/year family leave)	30%		
Luxemburg	M	18 weeks	100%	(n.d.)	(n.d.)
	P	6 months + 2 days/year family leave	Flat rate		
Netherlands	M	16 weeks	100%	8	71
	P	6 months (each parent) + 12 days/year family leave	Unpaid		
	F	2 days	100%		
Norway	M/P	52 weeks	80%	(n.d.)	(n.d.)
	P	52 weeks	Flat rate		
	F	4 weeks	Flat rate		
Portugal	M	6 weeks	100%	12	28
	P	24 months + 30 days/year family leave	Unpaid		
Spain	M	16 weeks	100%	5	84
	P	Until child's third birthday	Unpaid		
	F	2 days	100%		
Sweden	M	14 weeks	80%		
	P	18 months + 6 months (unpaid)	80%	48	79
United-Kingdom	M	18 weeks	6 weeks 90%, then flat rate	2	60[b]
	P	13 weeks	Flat rate		

n.d. = no data.
[a] M = maternity, P = parental; F = father only.
[b] 3–4 years (compulsory schooling starts at 5 years).
Sources: The Clearinghouse on International Developments in Child, Youth and Family Policies (http://www.childpolicyintl.org); Kamerman, 2000.

TABLE 104–4 Value of Global Subsidies for Children Relative to the Average for All Countries (Includes Child Care Expenditure and Housing Subsidies)

Luxemburg	841
France	254
Belgium	238
Germany	115
Austria	42
United Kingdom	27
Denmark	15
Sweden	–13
Finland	–57
Ireland	–75
Portugal	–81
Netherlands	–135
Spain	–245
Italy	–327
Greece	–600
Average	0

Source: Ditch *et al*., 1998a, p. 65.

CONCLUSION

Despite countless analyses and multiple attempts at explanation, the real causes of fertility decline still remain obscure. On the basis of these facts, some believe that any attempt to intervene to change these effects may be useless, and possibly counter-productive (Demeny, 1987).

European countries seem to have chosen to act in all reasonable directions: economic intervention, creation of child care services, career support for mothers, and so forth. But partly because of limited means, partly because of confusion surrounding goals (much more often social or redistributive that pronatalist), and partly through erratic application through time, attempts made up until now have not met with great success apart from a few fertility rebounds linked to specific situations. TFRs react only mildly, if at all, to the efforts governments make to influence them. Part of the explanation may be that the support received by families, even in the most favorable systems, covers only a small fraction of the expenses they face.

To successfully invert the downward trend of fertility over the last several centuries, if indeed that could be achieved, cultural, organizational, and financial inputs must reach far in excess of those that Western nations have agreed to up until now.

Acknowledgments

This chapter was translated by Heather Juby from Jacques Vallin's French translation of the original Italian text.

The author would like to thank Nicole Malpas for her collaboration in the bibliographic search.

References

Becker Gary S., 1960. An economic analysis of fertility, *in:* Gary S. Becker (ed.), *Demographic and economic change in developed countries*, p. 209–240. Princeton (New Jersey), Princeton University Press, National Bureau of Economic Research, 356 p.

Becker Gary S., 1981. *A treatise on the family.* Cambridge (Mass.), Harvard University Press, 288 p. (1991 revised and supplemented edition, 424 p.).

Blanchet Didier, 1987. Les effets démographiques de différentes mesures de politique familiale: un essai d'évaluation, *Population*, vol. 42(1), p. 99–128.

Blanchet Didier and Ékert-Jaffé Olivia, 1994. The demographic impact of family benefits: evidence from a micro-model and from macro-data, *in:* John Ermisch and Naohiro Ogawa (eds.), *The family, the market and the state in ageing societies*, p. 79–103. Oxford, Oxford University Press, 239 p.

Bradshaw Jonathan, Ditch John, Holmes Helen and Whiteford Peter, 1993. A comparative study of child support in fifteeen countries, *Journal of European Social Policy*, vol. 3(4), p. 266–271.

Caldwell John C., 1982. *The theory of fertility decline.* London, Academic Press, 386 p.

Calot Gérard, 2001. Mais qu'est-ce donc qu'un indicateur conjoncturel de fécondité? *Population*, vol. 56(3), p. 325–327.

Chesnais Jean-Claude, 1985. Les conditions d'efficacité d'une politique nataliste: examen théorique et exemples historiques, *in:* IUSSP (ed.), *Proceedings of the IUSSP international population conference, Florence, 1985*, vol. 3, p. 413–425. Liège, IUSSP, 525 p.

Chesnais Jean-Claude, 1996. Fertility, family and social policy in contemporary Western Europe, *Population and Development Review*, vol. 22(4), p. 729–740.

Chesnais Jean-Claude, 1998. Below-replacement fertility in the European Union (EU-15): facts and policies, 1960–1997, *Review of Population and Social Policy*, no. 7, p. 83–101.

Cigno Alessandro, 1991. *The economics of the family.* Oxford, Clarendon Press, 212 p.

Demeny Paul, 1987. Pronatalist policies in low-fertility countries: patterns, performance and prospects, *in:* Kingsley Davis, Mikhail S. Bernstam, and Rita Ricardo-Campbell (eds.) *Below replacement fertility in industrial societies: causes, consequences, policies*, p. 335–358. New York, Population Council, 360 p. (supplement to *Population and Development Review*, vol. 12).

Demeny Paul, 2003. Population policy dilemmas in Europe at the dawn of the twenty-first century, *Population and Development Review*, vol. 29(1), p. 1–28.

De Santis Gustavo and Breschi Marco, 2003. Fecondità, costrizioni economiche e interventi politici, *in:* Marco Breschi and Massimo Livi Bacci (eds.), *La bassa feconditàin Italia tra costrizioni economiche e cambio di valori*, p. 189–210. Udine, Forum, 216 p.

Ditch John, Barnes Helen and Bradshaw Jonathan (eds.), 1998a. *Developments in national family policies in 1996.* York, University of York, 326 p. (European Observatory on National Family Policies, European Commission).

Ditch John, Barnes Helen and Bradshaw Jonathan and Kilkey Majella, 1998b. *A synthesis of national family policies 1996.* York, University of York, 78 p. (European Observatory on National Family Policies, European Commission).

Ékert Olivia, 1986. Effets et limites des aides financières aux familles: une expérience et un modèle, *Population*, vol. 41(2), p. 327–348.

Ékert-Jaffé Olivia, 1994. Chiffrer une évolution du coût de l'enfant? *Population*, vol. 49(6), p. 1389–1418.

Ékert-Jaffé Olivia, Joshi Heather, Lynch Kevin, Mougin Rémi, and Rendall Michael, 2002. Fécondité, calendrier des naissances et milieu social en France et en Grande-Bretagne: politiques sociales et polarisation socio-professionnelle, *Population-F*, vol. 57(3), p. 485–518.

Esping-Andersen Gøsta, 1990. *The three worlds of welfare capitalism.* Cambridge, Polity Press, 246 p.

Franco Daniele and Sartor Nicola, 1990. *Stato e famiglia.* Milan, Franco Angeli, 223 p.

Frinking Gérard, 1985. L'identification des politiques de soutien de la fécondité et méthodes d'évaluation de leurs effets, *in:* IUSSP (ed.), *International population conference*, *Florence*, *1985*, vol. 3, p. 379–389. Liège, IUSSP, 525 p.

Gauthier Anne-Hélène, 1996a. The measured and unmeasured effects of welfare benefits on families: implications for Europe's demographic trends, *in:* David Coleman (ed.), *Europe's population in the 1990s*, p. 297–331. Oxford, Oxford University Press, 346 p.

Gauthier Anne-Hélène, 1996b. *The state and the family: a comparative analysis of family policies in industrialized countries*. Oxford, Clarendon Press, 231 p.

Gauthier Anne-Hélène, 1999. Historical trends in state support for families in Europe (post-1945), *Children and Youth Services Review*, vol. 21(11–12), p. 937–965.

Gauthier Anne-Hélène, 2002. Family policies in industrialized countries: is there convergence? *Population*, vol. 57(3), p. 457–484.

Gauthier Anne-Hélène and Hatzius, 1997. Family benefits and fertility: an econometric analysis, *Population Studies*, vol. 51(3), p. 295–306.

Goldstein Joshua, Lutz Wolfgang, and Testa Maria-Rita, 2003. *The emergence of sub-replacement family size ideals in Europe*. Vienna, Vienna Institute of Demography, 27 p. (European Demographic Research Papers, no. 2).

Ipsen Carl, 1996. *Dictating demography: the problem of population in fascist Italy.* Cambridge, Cambridge University Press, 281 p.

Ipsen Carl, 1997. *Demografia totalitaria: il problema della popolazione nell'Italia fascista.* Bologne, Il Mulino, 393 p. (Translated from English 1996).

Jägers Thomas and Raffelhüschen Bernd, 1999. Generational accounting in Europe: an overview, *European Economy. Reports and Studies*, no. 6, p. 1–16.

Kamerman Sheila B., 2000. Early childhood education and care: an overview of developments in the OECD countries, *International Journal of Educational Research*, vol. 33(1), p. 7–29.

Livi Bacci Massimo, 1991. Politica, politiche e demografia, *in:* Antonio Golini, Alain Monnier, Olivia Ékert-Jaffé, Rudolf Andorka, Jacqueline Hecht, Elisabet Näsman, Raimondo Cagiano de Azevedo, Gian-Carlo Blangiardo, Ron Lesthaeghe, Guy Moors, Heins Moors, Santo Koesoebjono, Rosella Palomba, Gérard Calot and Massimo Livi Bacci, *Famiglia figli e società in Europa. Crisi della natalità e politiche per la popolazione*, p. 343–359. Turin, Fondazione G. Agnelli, 362 p.

Macura Miroslav, Eggers Mitchell, and Frejka Thomas, 1995. Demographic change and public policy in Europe, *in:* Hein Moors and Rossella Palomba (eds.), *Population, family and welfare.* Oxford, Clarendon Press, 297 p.

McDonald Peter, 2002. Sustaining fertility through public policy: The range of options, *Population*, vol. 57(3), p. 423–456.

MISSOC, 2002. Family benefits and family policies in Europe, *Missoc-Info*, no. 1, p. 88. (http://europa.eu.int/comm/employment_social/missoc/missoc_info_en.htm).

Moors Heins G, and Palomba Rosella (eds.), 1995. *Population, family and welfare. A comparative survey of European attitudes.* Oxford, Clarendon Press, 297 p.

Näsman Elisabet, 1991. Il caso svedese, *in:* Antonio Golini, Alain Monnier, Olivia Ékert-Jaffé, Rudolf Andorka, Jacqueline Hecht, Elisabet Näsman, Raimondo Cagiano de Azevedo, Gian-Carlo Blangiardo, Ron Lesthaeghe, Guy Moors, Heins Moors, Santo Koesoebjono, Rosella Palomba, Gérard Calot and Massimo Livi Bacci, *Famiglia figli e società in Europa. Crisi della natalità e politiche per la popolazione*, p. 169–208. Turin, Fondazione G. Agnelli, 362 p.

Palomba Rosella and Menniti Adele, 1994. Genitori e figli nelle politiche familiari, *in:* Antonio Golini (ed.), *Tendenze demografiche e politiche per la popolazione. Terzo Rapporto IRP*, p. 79–100. Bologne, Il Mulino, 376 p.

Preston Samuel H., 1984. Children and the elderly: divergent paths for America's dependents, *Demography*, vol. 21(4), p. 435–457.

Reher David Sven, 1998. Family ties in Western Europe: persistent contrasts, *Population and Development Review*, vol. 24(2), p. 203–234.

Rendall Michael S. and Bahchieva Raisa A., 1998. An old age security motive for fertility in the United States? *Population and Development Review*, vol. 24(2), p. 293–307.

Rønsen Marit, 2004. Fertility and family policy in Norway. A reflection on trends and possible connections, *Demographic Research*, vol. 10, p. 265–286.

Sartor Nicola, 2004a. I rapporti finanziari tra Stato e famiglia, *in:* Massimo Livi Bacci (ed.), *Atti del convegno La bassa fecondità tra costrizioni economiche e cambio di valori*, p. 167–195. Rome, Accademia dei Lincei, 568 p.

Sartor Nicola, 2004b. Le ragioni del sostegno pubblico dei figli: un riesame critico, *in:* Massimo Livi Bacci (ed.), *Atti del convegno La bassa fecondità tra costrizioni economiche e cambio di valori*, p. 547–566. Rome, Accademia dei Lincei, 568 p.

Sleebos Joëlle E., 2003. *Low fertility rates in OECD countries: facts and policy responses.* Paris, OECD, 62 p. (Social, Employment and Migration Working Papers, no. 15).

United Nations, 2002. *National Population Policies 2001.* New York, United Nations/Nations unies, 425 p.

Vukovich Gabriella, 1993. Review of population policies: developed countries, *in:* United Nations (ed.), *Population policies and programmes*, p. 85–91. New York, United Nations, 267 p.

CHAPTER

105

The Effect of Pronatalist Policies In Industrialized Countries

GÉRARD CALOT

Institut national d'études démographiques (INED), Paris, France

During the 20th century, few industrialized countries implemented policies that had the explicit objective (or the explicitly declared objective) of influencing fertility. If dictatorships and other totalitarian regimes[1] that resorted to coercive methods of varying severity are excluded, the list is especially short. By confining ourselves to democratic countries which, over a long period, created legislation whose explanatory memorandum constantly and expressly mentioned, among other concerns, worries regarding a sufficient level of fertility, France is about the only example (see Chapter 119). Even then, the strength of pronatalist motivation, which was very marked at the end of World War II, has become very blurred in this country during the last two or three decades.

It is necessary, however, to remember that some countries—through what is called family policies, presented as targeting exclusively social objectives—adopted legislation which, in many cases, is not fundamentally different, in their content or their methods, from those presented elsewhere as targeting both social and demographic objectives. Indeed, even though some deny it, all the countries implicitly apply family policy measures, even if it is only in the way that the number of children is taken into account with regard to obligatory deductions (income tax, in particular), and social benefits, or in regard to education, care for little children (crèches, nurseries), or the budget on child health.

[1] Such as that of Nazi Germany, for example, which in the 1930s developed a pronatalist policy and succeeded in achieving some increase in fertility (see Chap. 100).

The general issue discussed here is the effect of policies on fertility: whether they are policies which at least partly aim for such an effect or policies with purely social objectives which can indirectly affect fertility. Beforehand, I will make some remarks on the possibility of assessing demographic effects. I will then deal with the issue raised by using the example of the increase in French fertility after World War II (Sect. 2), the case of Sarre (Sect. 3), a comparison of West and East Germany (Sect. 4), and a more general comparison of the trends observed in Western Europe (Sect. 5).

I. THE DIFFICULTY OF LINKING CAUSES AND EFFECTS IN SOCIAL SCIENCE

In demography, as in the other fields of social science, the impossibility of organizing provoked experiments makes it difficult to irrefutably connect causes and consequences. While analysis allows, exceptionally, for presumptions to be established, it is clear that causal certainty is, and will always remain, beyond reach.

Generally, it is on the basis of a temporal analysis that attempts to estimate effects are made: such a cause appears at a given date or is likely to exert its effects on a given scale from such a date; the study of the variation of this scale before and after the pivotal date can shed light on any possible effects. The diagnosis established is based, implicitly or explicitly, on the gap between the trend actually observed of the indicator considered and that which would have been recorded, but which is by definition not observable, in the absence of the cause studied.

The trend of the considered indicator that would have been recorded in the absence of the cause studied can be estimated under three specific circumstances.

1. Before the pivotal date, the indicator varies little, but over quite a short time interval, after the pivotal date, the indicator begins to increase or decrease significantly, and is then maintained without any significant variations near its new level. The shorter the interval is with the highest variations and the closer it is to the pivotal date, the more authority we have to attribute the recorded change in level to the considered course. In other words, this favorable circumstance is that for which the effect is immediate and massive on a scale that varies relatively little before and after sudden and significant variations in levels.
2. One case that is similar to the previous one is that of regular variation before the pivotal date in the indicator (e.g., in an appreciably linear manner); then over quite a short time interval, after the pivotal date, it increases or decreases significantly and then follows a regular trend comparable to that which prevailed before. The effect is therefore estimated by the gap between the extrapolation of the previous trend and the new trend.

   These two situations correspond to the case where the indicator considered has a considerable inertia before and after the appearance of the cause studied, and where the effects produced are limited to a change in level without a change in the rate of the chronological trend.
3. Another case, different from the preceding ones, is that in which a second demographic entity exists which has not been submitted to the cause studied and which, apart from that, is homogenous to the first entity. After the appearance of the cause in the first entity, a study of the trend in the considered indicator in the second entity provides an estimate of the trend that would have occurred in the first entity in the absence of this cause. This favorable circumstance occurs when a control group serves as a reference in the absence of a cause.

It is rare to be in one or the other of these situations. The exceptional historical and geographical specificities must be taken into account in order to estimate the effects. Note that with regard to the medium- or long-term effects on fertility, only the third situation is likely to provide some indication.

As an illustration, let us consider the following two examples:

1. The effects of legislative changes in Austria (changes to the law regarding income tax) and in Sweden (changes to the law regarding reversion pensions), which resulted in sudden variations in the number of marriages (Figure 105–1);
2. The effect of World War II on fertility in France and Switzerland on the one hand (Figure 105–2), and Sweden on the other hand (Figure 105–3).

Considering the scale of the variation recorded over a period of only a few months, the curve of the monthly number of marriages in Austria and Sweden, after correcting for seasonal changes, is significantly invariant (Figure 105–1). Moreover, at both ends of this period, the levels are identical, which indicates the very specific nature of the cause (in both cases, the advantage of being married before December 31). We are therefore in the first situation with a particularly transitory cause.

If the war period of 1940–1945 is excluded, the comparison of trends in the total fertility rate (TFR) in France and Switzerland between 1930 and 1955 shows a relatively significant parallelism, with the French indicator exceeding the Swiss indicator by 0.4 children per woman (Figure 105–2).

In both countries, fertility was decreasing slowly between 1930 and 1935. The slight increase observed in France from 1936 and which was accentuated in 1939 was abruptly interrupted by World War II and the invasion of the country. Although a certain increase was observed from 1942, the spectacular jump that marked the beginning of the phenomenon that became known as the baby boom did not occur until 1945. Swiss fertility stopped declining in 1937 and increased slowly up to the middle of 1940, followed by a rapid and massive increase until the end of 1945. Then, in parallel to the trend in French fertility, a reflux occurred, with the Swiss rate returning to between 2.62 and 2.30 children per women between 1946 and 1955, while the French rate was between 2.98 and 2.67.

This is how during the war, Switzerland, spared the hostilities, experienced its baby boom, which occurred on exactly the same scale as that of France: the difference between the TFR of 1946 and 1939 was 0.82 children per woman in Switzerland and 0.83 in France. But the French baby boom, prevented by the war, only occurred after the liberation. The great difference between these two countries is that the rate observed in 1946 was achieved by an exceptional increase in the French rate in the space of one year (the TFR increased from 2.15 in May 1945 to 3.14 in April 1946), while in

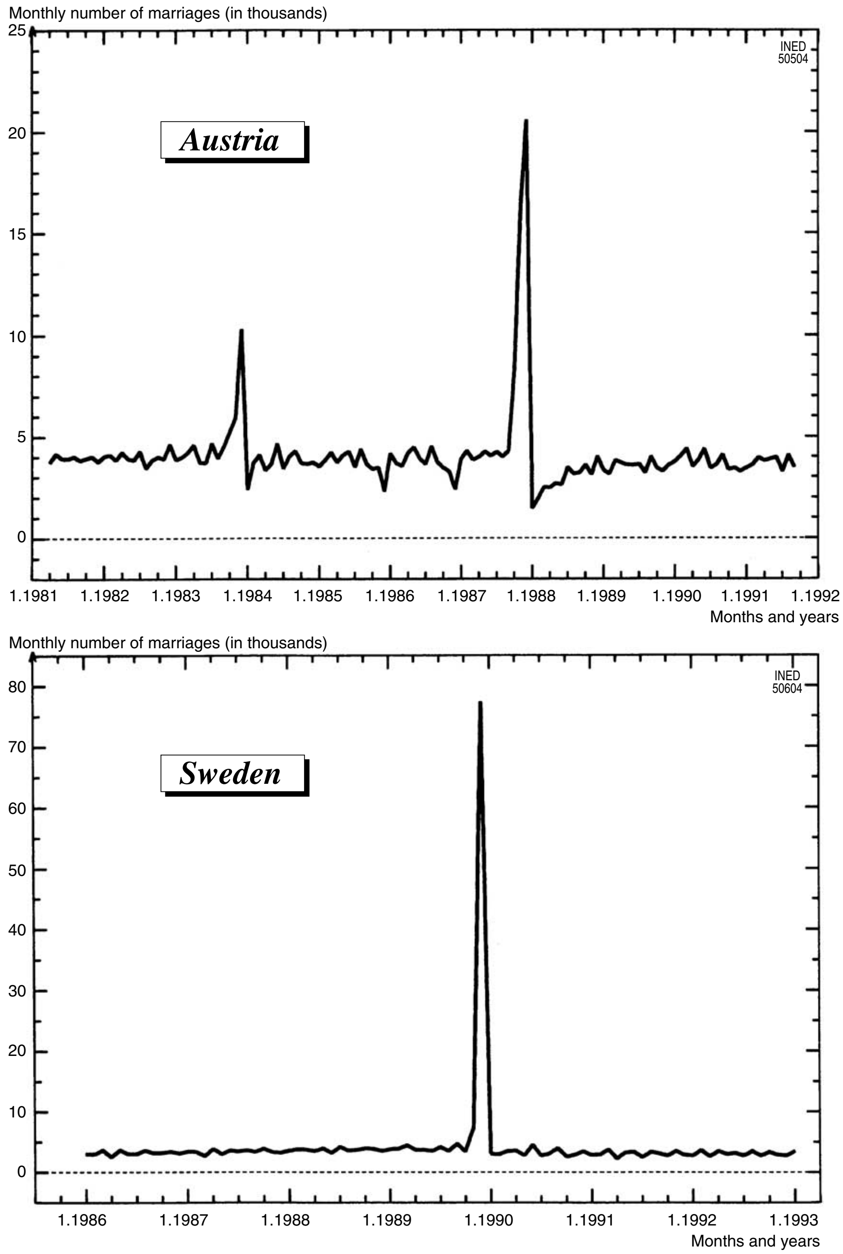


FIGURE 105–1 Monthly number of marriages corrected for seasonal variation: Austria 1981–1991 and Sweden 1986–1992. (Source: Population database of the Observatoire démographique européen [ODE], sardon@ined.fr)

Switzerland it was the result of a regular progression over the course of 6 years.

We are therefore led to believe that the Swiss fertility trend provided a pertinent reference to what would have been the progress of French fertility if the war had not taken place. In these conditions, the vigorous increase in fertility that occurred at the end of the war, in France as well as in the majority of countries that took part in the conflict, could well have been influenced by the war only in its timing but not in its nature, and not even in its scale: it corresponds to a trend, which certainly no contemporary had foreseen, but which could probably also have occurred earlier and spread over a longer period if peace had been maintained.

This presumption is corroborated by examining the fertility trends of Sweden, the other European country that was spared World War II (Figure 105–3).

We observe that the parallelism between the French curve and the Swiss and Swedish curves after 1945 correspond to a greater difference, around 0.15 children per woman, than during the 1930s (the gap

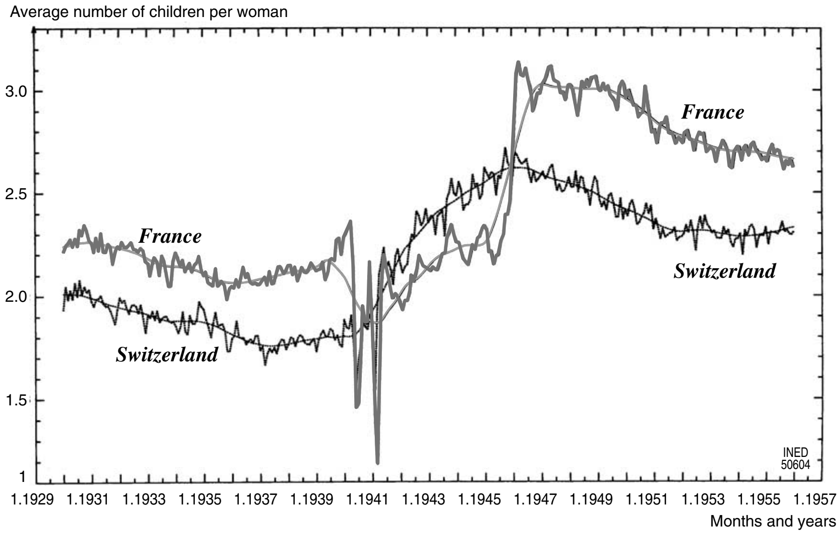


FIGURE 105–2 Trend in monthly TFR in France and Switzerland from 1930 to 1955. Data corrected for seasonal variation and data smoothed by moving average graduation (Hoem and Linnemann, 1988). (Source: Population database of the ODE)

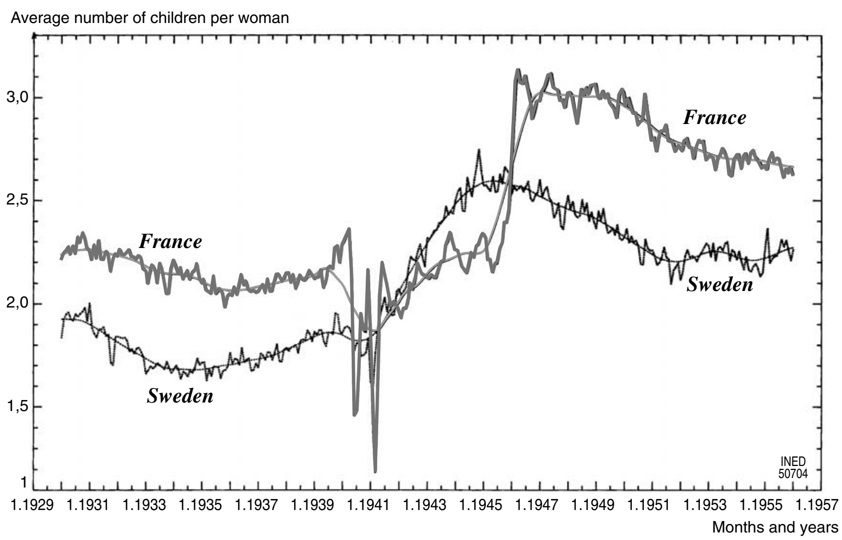


FIGURE 105–3 Trend in monthly TFR in France and Sweden from 1930 to 1955. Data corrected for seasonal variation and data smoothed by moving average graduation (Hoem and Linnemann, 1988). (Source: Population database of the ODE)

between the French rate and the Swiss one was on average 0.29 children per woman during the 1930–1939 period, and 0.44 during the 1946–1955 period; compared to the Swedish rate, the gap is 0.38 and 0.52, respectively).

II. THE RISE IN FRENCH FERTILITY AFTER WORLD WAR II

The principles of a particularly active policy in favor of families was defined in a decree established in 1939, known as the *Code de la Famille* (The Family Code) (Dalloz, 1995), with the clearly defined objective of maintaining fertility. But it was not possible to implement the policy until the liberation, in the context of a social security system made up of three branches: illness, old age, and the family. It consisted of the provision of considerable benefits to families. The funding of the family branch relied on employer contributions and was based on a 9% rate of deduction from salaries.

From 1945, all salaried employees benefited from these services. But it was not until the end of the 1950s

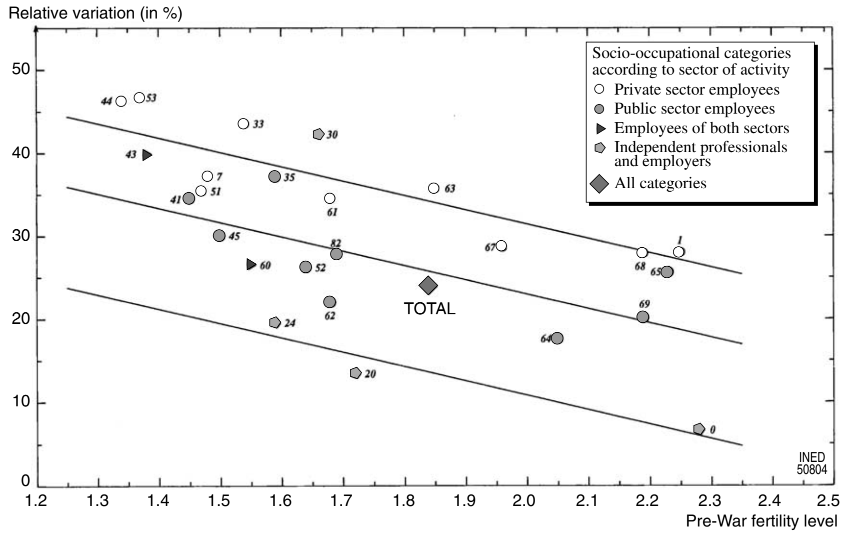


FIGURE 105–4 Variations in fertility between socio-occupational categories in France according to the 1962 family survey. Correlation between the relative variation of fertility between the pre-war period and the post-war period and the pre-war level. Fertility measured by the average number of children after 10 years of marriage. Fertility level corresponding to the marriage cohorts of 1920–1934, and 1940–1951 respectively. The code numbers in the figure refer to INSEE's classification of occupational categories. (Sources: Febvay, 1959; Deville, 1972; Lery, 1972)

that the system of family benefits was extended to nonsalaried workers: farmers, tradesmen and craftsmen, employers, and independent professionals. This very specific historical occurrence makes a differential analysis of fertility according to socio-occupational categories possible (data obtained from family surveys carried out during the censuses of 1954 and 1962). Furthermore, the establishment of a family benefits system represented a greater improvement in income for private sector employees (industry and trade), who only received modest benefits before the war, than for state and public sector employees, who already received family benefits before the war, but at a lower level than before 1945.

Undoubtedly, all the socio-occupational categories recorded an increase in fertility between the pre-war period and the post-war period. But at equal pre-war fertility levels, to what extent was this increase influenced by the implementation of this new family policy? More precisely, at equal pre-war fertility levels, was the increase virtually identical within each of these three groups (private sector, public sector, and nonsalaried employees) and significantly different from one group to another?

In Figure 105–4, a point for each socio-occupational category represents the correlation between (**A**) the relative variation of fertility between the pre-war and post-war period, and (**B**) the pre-war fertility level. The fertility levels, before and after the war, of each of the socio-occupational categories was measured by the average number of children after 10 years of marriage in the marriage cohorts of 1920–1934 and 1940–1951, respectively (Febvay, 1959; Deville, 1972; Lery, 1972).

Within each of the three groups, the points of the graph are ordered[2] according to a linear trend, with the three trend lines being virtually parallel. As expected, at a given pre-war fertility level the private sector employees recorded the highest fertility increase. They are followed by the public sector employees, then the nonsalaried workers.

The gap between the adjustment line corresponding to the private sector and that corresponding to the public sector is around 9 points, the gap between the latter and that corresponding to nonsalaried workers is around 12 points. If we could attribute the gap between the two extreme groups to the difference between the legislation applied (this naturally constitutes a hypothesis), we come to the conclusion that the effect of implementing the 1945 family policy was to increase fertility by around 20% from what it would have been without the family policy, that is, by 0.4 children per woman for a fertility level that would otherwise have been 2.0 children per woman. This

[2] Liberal professionals, a particular socio-occupational category made up of nonsalaried professionals, demonstrated a fertility increase between the pre-war and the post-war periods that was exceptionally high, considering their pre-war level. This anomaly could be due to the high standard of living of this socio-occupational category, which would make them less sensitive to the effects of the family policy.

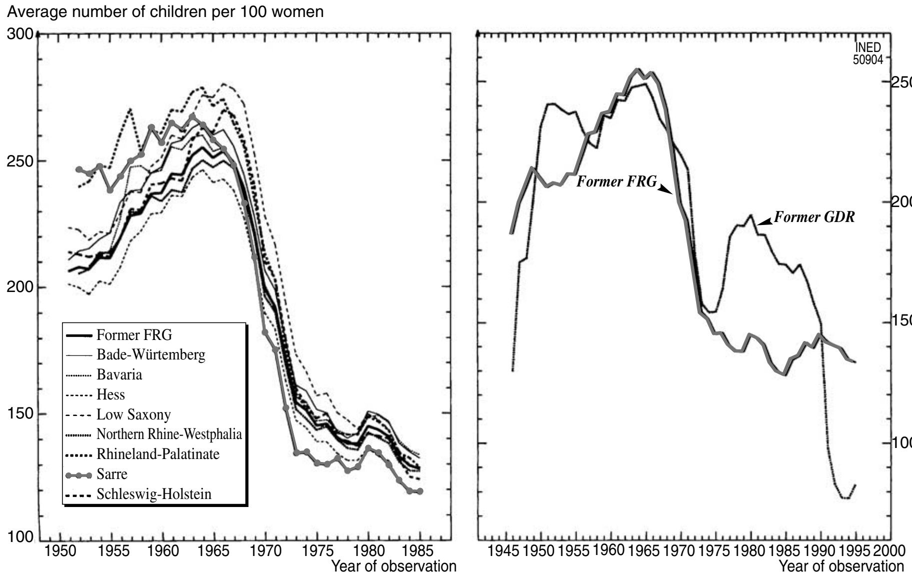


FIGURE 105–5 Trend in TFR in Germany, in the *Länder* of the West from 1951 to 1985 (graph on the left) and in both East and West Germany from 1946 to 1995 (graph on the right). (Sources: for Sarre, Schwartz, 1978; for both East and West Germany, Monnier, 1989; for updating these data and for the other *Länder*, Population database of the ODE)

conclusion is obviously weak: for example, one could object that the fertility changes that occurred within a given socio-occupational group could also be due to changes in the social recruitment of this category.

III. THE FERTILITY TREND IN SARRE

After World War II, Sarre remained under French control until a referendum was held in 1955 which led to its entry into the Federal Republic of Germany as a Landon January 1st 1956. For ten years, the French family policy was applied to the population of Sarre which, following a transition period, became subject to the legislation of Federal Germany which was less favorable to families. It is also interesting to compare the fertility trend in Sarre with that of the other Länderof West Germany (left part of Figure 105–5). If the French family policy implemented in 1945 had an effect on fertility, it should be possible to observe this effect on the demography of Sarre.

Figure 105–5 shows the fact that at the beginning of the 1950s, Sarre had one of the highest fertility levels among the eight *Länder* considered[3] and that from the 1970s it had one of the lowest levels, and even the lowest in 1985.

[3] The three entirely urban *Länder* of the former Federal Republic of Germany (Berlin, Bremen, and Hamburg) have been excluded from the analysis.

This result does not establish the certainty of the effectiveness of the French policy, but it supports the presumption of its effectiveness.

IV. THE FERTILITY TREND IN THE FORMER GDR AND THE FORMER FRG

The graph on the right of Figure 105–5 shows the trend of the TFR in the former Federal Republic of Germany (FRG), and in the former German Democratic Republic (GDR). Undoubtedly, in many aspects, the two halves of Germany were separate up until 1989, in their political and socioeconomic system above all, but also in the population's standard of living, the rate of female employment, the care of young children, the situation with regard to housing, and so forth. On many levels, the fertility of the two countries was very different, particularly with regard to age at motherhood and the proportion of births outside marriage. However, we cannot ignore the striking similarity in the fertility levels of both countries between 1957 and 1973. Furthermore, around 1970, these two countries were similar in the fact that their TFRs were the lowest in the world.

In 1975, the government of the GDR created a group of social measures explicitly aimed at increasing fertility; thereafter, the two rates separated quickly, that of the East surpassing that of the West by 0.4 to 0.5 children per woman from 1977 to 1985. Then, the fall of

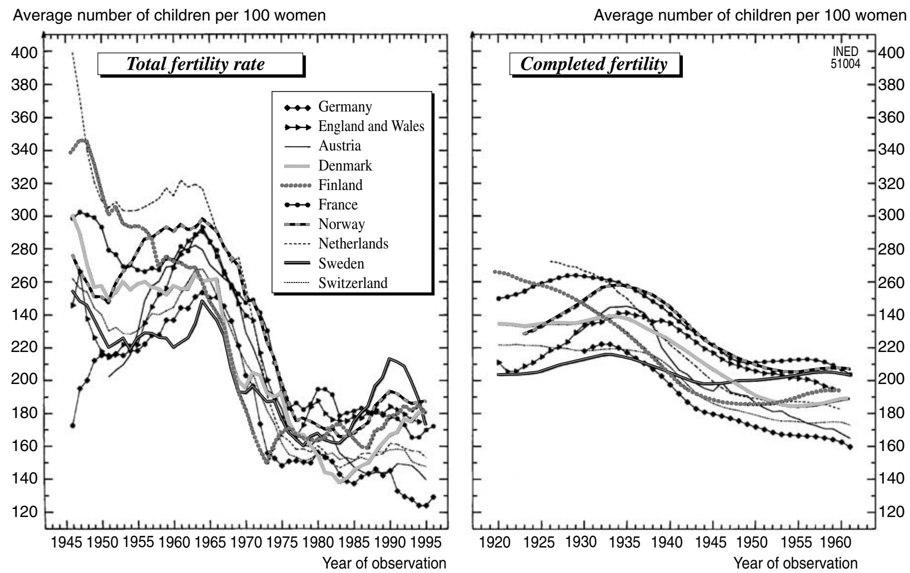


FIGURE 105–6 Trend in TFR (left graph) and the completed fertility of cohorts (right graph) in Northwestern Europe from 1946 to 1995. (Source: Population database of the ODE)

the Berlin Wall, the collapse of the Communist system, and the unification of the country completely overturned the situation.

While it is clear that analogical reasoning certainly does not make it possible to assert that the FRG would have recorded a trend similar to that of the GDR if it had adopted the same measures in the same period, it seems less dubious to suggest that the GDR would probably have experienced a similar trend to that of the FRG in the absence of the measures taken in 1975.

V. THE FERTILITY TREND IN NORTHWESTERN EUROPE

Since World War II, the countries of Northern and Western Europe have reached an increasing degree of similarity in population trends: the most striking trend inflections appear simultaneously and according to increasingly comparable terms. In particular, the massive decline in the TFR, which occurred between the mid-1960s and the mid-1970s, is very similar from one country to another (left side of Figure 105–6). Consequently the peculiarities in the trends, notably the exceptions to the parallelism, at least when the exceptions are quite significant, reveal national specificities that we need to identify and then attempt to explain.

The long-term demographic effectiveness of the French family policy appears, vaguely but probably, in the fact that in the TFR and, more especially, in the completed fertility from the cohort born in 1920 to that born in 1960, France had one of the highest levels in Europe (Figure 105–6). It is naturally impossible to quantify this effect on the basis of Figure 105–6 alone.

Close examination of Figure 105–6 shows, on the other hand, that the four Scandinavian countries (Denmark, Finland, Norway, Sweden) are distinguished by a clear increase in their TFR since the beginning of the 1980s. The most striking case is that of Sweden where the period of increase (1983–1990), which raised the rate above replacement level, was replaced by a period of decline of a virtually identical scale, from 1991. In terms of completed fertility, the Scandinavian peculiarity results in a lesser decrease of the curves, and even a near-invariance for Sweden: the Swedish curve crosses the curves of nearly all the other countries, which shows a decrease of at least 0.5 children per woman between the female cohorts born in 1930 and those born in 1960. While Swedish women born around 1930 were the least fertile females on the continent, their daughters, with a fertility level similar to that of their mothers, have one of the highest fertility levels in Europe. This is probably due to the indirect pronatalist effect of social policies, particularly those aiming to facilitate the compromise between professional life and family life.

CONCLUSION

At the end of this brief discussion of some remarkable changes in fertility in Europe, let us add some indications of general interest on the effects of policies that support fertility. As was mentioned at the beginning of this chapter, coercive measures, particularly those that consist of reducing fertility control by limiting access to contraception or induced abortion have been excluded from this discussion.

First of all, each time that the opportunity arises (exceptional, it is true) of evaluating the effect on fertility of policies aiming to improve the living conditions of families (regardless of the reasons, partly demographic or exclusively social), the resulting conclusion invariably includes a positive effect on fertility levels.

Secondly, this effect is generally modest. While the TFR could react with some intensity to such and such a change in legislation, cohort completed fertility would not have changed considerably. The maximum effect that can be expected from an active family policy over the long term is probably of a few tenths of a child per woman. Earlier we provided an idea of the scale of the effect of the French family policy implemented after the liberation. Econometric work carried out in France by Olivia Ékert (1986), came to the conclusion that the total compensation of the cost of a child could increase completed fertility by around 0.5 children per woman.

According to our previous considerations, it appears that no country could have avoided the two considerable variations in fertility observed during the 20th century (the baby boom of the 1940s and the baby bust of the 1965–1975 period) through appropriate policies. It is only on the fringes that changes could have been made.

However, for the female cohorts born around 1960, the difference that separates the completed fertility of the cohorts from replacement level, which can be taken as a reasonable demographic objective (maintaining the adult core of the population at its current number, which implies a slight population increase and a certain aging, due to the increase in average life span), generally remains in Europe within the scope of the possible effects of family policies. It is therefore possible to obtain simple replacement fertility in Western Europe.

It is true that many of our contemporaries are repelled by a policy for stimulating fertility, either because they consider that governments should not intervene in this most personal of issues, that of family choice, or because they condemn any objectives that are similar to those of previous totalitarian regimes. It is possible, however, that the changes that affect the balance between ages, particularly with regard to social security and particularly to pensions, will result in the disappearance of some of these objections.

Insofar as an active family policy is inevitably expensive for the public purse, what needs to be determined is whether the benefits that we might obtain from higher fertility are worth the cost. But that, in truth, is another discussion.

Acknowledgment

This chapter was translated from the French by Accenta Ltd.

References

Dalloz (ed.), 1995. *Code de la santé publique, code de la famille et de l'aide sociale*, 11th ed. Paris, Librairie Dalloz, 58 p. Ékert Olivia, 1986. Effets et limites des aides financières aux familles: une expérience et un modèle, *Population*, vol. 41(2), p. 327–348.

Deville Jean-Claude, 1972. *Structure des familles. Enquête de 1962*. Paris, INSEE, 344 p. (Les Collections de l'INSEE, Série Démographie et emploi, no. 13–14).

Febvay Maurice, 1959. Niveau et évolution de la fécondité par catégorie socio-professionnelle en France, *Population*, vol. 14(4), p. 729–739.

Hoem Jan M. and Linnemann Per, 1988. The tails in moving average graduation, *Scandinavian Actuarial Journal*, vol.4, p. 193–229.

Lery Alain, 1972. L'évolution de la fécondité avant et après la dernière guerre, *Économie et statistique*, no. 37, p. 21–33.

Monnier Alain, 1989. La conjoncture démographique: l'Europe et les pays développés d'outre mer, *Population*, vol.44(4–5), p. 901–923.

Schwartz Karl, 1978. La baisse de la natalité en Allemagne fédérale, *Population*, vol. 33(4–5), p. 999–1017.

PART

III

HEALTH POLICIES AND THEIR DEMOGRAPHIC IMPACT

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

From man's appearance on the Earth up to the 18th century AD—that is, for tens of thousands of years—humankind never escaped the clutches of a stark mortality pattern: over a quarter of newborns did not survive infancy, and the mean length of life barely exceeded 25 years.

Less than three centuries have brought immense progress and unprecedented social change. The new mortality regime gives current populations a life expectancy two to three times longer than their predecessors, and this is probably a continuing trend. Some countries are plainly well in the lead, while others are still lagging, but all have moved forward and the general trend is extraordinary. And the last half-century has been the most decisive period of recent centuries. The great leap forward was a post-war achievement, brought about by the spread of antibiotics, the general introduction of social security systems in the developed countries, and the launching of large-scale health improvement programs in Third World countries.

Have health policies delivered? They aim to, so should they be credited with the outcomes achieved? Plainly, they are far from the only contributory factor to the recent mortality decline, and it would be foolhardy to claim them as directly and uniquely responsible. First, without the startling medical progress of the last 150 years, the situation today would still be very far short of where it actually is. Admittedly, this progress itself can be credited to health policies in the broad sense through investment in research, and the gains of that progress have been so great only because of health policies that spread the benefits to the general population. But it would be rash to conclude from this that merely taking medical research and current health policies forward will necessarily bring all populations of the world rapidly up to the level of the current trailblazers.

But, what do we mean by a "health policy"? This is not the place in which to rehearse the definitions of health,[1] nor dwell at length on the simple fact that a health policy is the deployment of society's resources to improve public health. But this simple approach can go down many different roads (Pradeau, 2004). More especially, it can focus on very different degrees of involvement by the State (or other public institutions).

A strict economic liberalism approach might consist in a policy based solely on the idea of a market for health populated by players who were production (pharmaceutical firms, etc.), service (hospitals, private clinics, fee-for-service health professions), or retail (pharmacies, etc.) undertakings, and consumers (people with health problems, or healthy individuals concerned with prevention). The purpose of health policy would then be, as in other public policy spheres, to regulate that market in order to maximize system efficiency by optimum management of resources that can never be equal to the vast needs for improvement of public health. The limits of such an approach need little explanation.

Still in terms of the basic functions of the State (Montesquieu, [1758]; Constant, 1815) that are compatible with free enterprise, health policies may include the efforts of government bodies to regulate certain occupations that have a responsibility to society, which can include the health professions. Government will then decide what skills are required to carry on these occupations and the legal framework in which they can be exercised. It may also order the different levels of occupation hierarchically, or prescribe specializations and, finally organize how they should be coordinated or work together.

The basic functions of the State may also require some aspects of the prevention or treatment of some diseases to be seen as a public policy imperative identical or similar to national defense, for which public health policies must be framed. It was one of the earliest approaches of this kind that helped to contain most effects of the major epidemics through the enactment of sometimes quite harsh public injunctions, like isolation or the establishment of quarantine lines to tackle the plague or cholera. The concept of public health is also what lay behind the late 19th- and early 20th-century political concerns with hygiene, followed by the introduction of compulsory vaccination schemes or sectoral policies to tackle specific diseases, like tuberculosis or syphilis, considered as social evils which the State or, more generally, the authorities, had a duty to tackle.

Taking public involvement one step further, health policy can be spelled out in terms of concrete objectives and deliverables. With regard to infectious diseases, for example, the idea rapidly gained ground that a policy objective could be not just to tackle, contain, and limit the consequences of a disease, but quite simply, to eradicate it entirely. So, the World Health Organization (WHO) set, and eventually achieved, the objective of eradicating smallpox.

Health measures may also aim to provide more specific protection to groups at particular risk. Special maternal and child welfare programs, and statutory protection for workers against work-related accidents or occupational diseases, are cases in point. The same concept was even extended to maximize the benefits when, for example, informed

[1] Which can be found in various chapters of the Treatise, in particular Chapters 40 ("Health, Illness and Death," by Catherine Gourbin and Guillaume Wunsch), 41 ("Measuring the State of Health," by Catherine Sermet and Emmanuelle Cambois), and 80 ("Lengthening of Life and the Population Health Status," by Jean-Marie Robine and Carol Jagger).

by the Chinese barefoot doctors scheme, the WHO's Alma Ata conference adopted the primary health care strategy designed to redefine health policy priorities for the benefit of the poverty-stricken populations of the Third World.

Public involvement goes yet a step further when the objective becomes more all-embracing, as in the health section of some countries' population policy statements: setting target dates and percentages for cutting the infant mortality rate, or even hard targets for life expectancy gains. This kind of objective has even been included in some WHO programs under the "health for all 2000" initiative.

But the surprisingly unrealistic title of this initiative itself was merely a partial reflection on a worldwide scale of the major development made in the concept of health policy by social democracy and more so by communist regimes. It is a concept that requires the State—either directly, as in the United Kingdom and communist countries, or indirectly, through special purpose welfare agencies, as in France and the Scandinavian countries—to pay the full cost (or at least a very big share) of health care for the entire population in order to minimize social inequalities in illness and death while maximizing the general health and life expectancy of the population.

Finally, at the maximum degree of public involvement, a health policy may be based on the legal recognition of a real right to health for all, although this involves little practical commitment given the many factors in individual health that are largely outside political control and policy action. A constitutionally entrenched right to health will remain mere lip service unless legislation or regulation lays down the dividing line by time and circumstances of what is and what is not possible, in order to define where the duty of the state and authorities really lies, which is little different from setting more or less achievable targets.

These various concepts that can underpin health policies have only gradually emerged during a history of health that often receives only a passing mention in attempts to trace the history of population policies (including in the opening chapters of this volume). In Chapter 106 of this third part, therefore, Jacques Vallin and France Meslé briefly review the history of health policies to inform an understanding of current issues. In Chapter 107, the same authors then consider the strategies deployed in recent decades, and in Chapter 108, they review the measurement of real health policy outcomes. This assessment can obviously not be wholly satisfactory, as measurement in this field suffers from a lack of data, aggravated by the growing complexity of the issue in our modern societies. In Chapter 109, therefore, Viviana Egidi and Vittoria Buratta discuss the monitoring systems that need to be put in place before health policy outcomes can be properly assessed.

References

Constant de Rebecque Benjamin, 1815. *Principes de politique, applicables a tous les gouvernements représentatifs et particulièrement a la constitution actuelle de la France*. Paris, A. Eymery, 321 p.

Montesquieu de Secondat Charles, [1758] 1995. *De l'esprit des lois*, 2 vol. Paris, Éditions Gallimard, 1628 p. (Folio Essais. 1758 version, final one revised by Montesquieu, spelling modernized, published by Laurent Versini).

Pradeau Francis, 2004. *Les politiques de santé. Politiques de santé publique et politique conventionnelle*. Poitiers, CHU de Poitiers, 13 p. (Notes for a course on health policies, retrieved July 5, 2004, from http://membres.lycos.fr/pradeau/cours/pol_sante).

C H A P T E R

106

Origin of Health Policies

JACQUES VALLIN AND FRANCE MESLÉ

Institut national d'études démographiques (INED), Paris, France

In its broadest sense, health policy may be defined as a collective action to improve a population's health status, irrespective of the type of collective entity responsible for the action, and without awaiting direct central-government involvement. Even if we accept this definition, there are two obvious prerequisites to such action: (1) the population (or at least some individuals) must be, or believe themselves to be, capable of combating disease and death; (2) society must be sufficiently organized to produce institutions capable of and willing to make a collective contribution to bettering the population's health status. Both conditions seem to have existed at the very birth of the great civilizations whose writings have come down to us. Indeed, the first prerequisite was undoubtedly met at a rather early stage of prehistory, since an examination of skeletons has shown that, by the Neolithic, humans knew how to reduce fractures by immobilizing broken bones and preserving their alignment (Sournia, 1997).[1] As to whether the second precondition was met, we must let our imagination run. However, as far back as the first known writings, medicine already had an effective presence; it was sufficiently structured for us to conclude that, from earliest antiquity, most societies had begun to sketch out a health policy, if only through the organization of the medical profession. True, this embryonic arrangement probably never led to the implementation of a health policy (in today's full sense of the term) before the 19th century. No doubt it was necessary to wait until health issues attained their full magnitude, and access to medical care became sufficiently effective to overcome major natural constraints. Nevertheless, current health policies are the culmination of a long maturing process that moved in step with the construction of human societies.

[1] "But as some overlaps between the two fragments remain, we can conclude that [man] did not know how to exert a traction at both ends that would have restored a perfect alignment. We are even more hesitant about assessing cranial trepanations, which were partly healed in the individual's lifetime. [. . .] Was their significance religious or rather therapeutic? [. . .] The question will probably never be answered"(Sournia, 1997, p. 10).

I. ANTIQUITY, MEDICINE, AND RELIGION: FROM IMHOTEP TO GALEN

1. Egypt

The oldest medical text in the world is undoubtedly the Edwin Smith papyrus, written at the start of the Old Egyptian Empire, that is, in about 3000 BC. It is a true classification of external pathologies methodically arranged from head to foot. Each chapter described a clinical case with exploratory methods, statement of diagnosis, and therapeutic recommendations. This form of medicine (in which gynecology[2]

[2] "Egyptian doctors could identify uterine prolapse (for which pessaries were applied), metritis, vulvitis, and uterine cancer, fought by means of local injections, purges, and aromatic fumigations whose composition is unfortunately not known." They could also establish "an early diagnosis of pregnancy by observing the comparative growth of two plants, one irrigated with water, the other with the urine of the woman assumed to be pregnant" (Sournia, 1997, p. 26). This role of gynecology in the medical concerns of the time was certainly linked to the relative sexual freedom in ancient Egypt, where women enjoyed considerable autonomy and played a major part in social, political, and religious life.

and surgery featured prominently) was, to some degree, effective, although it did not hesitate to rely heavily on the divining art. We have relatively full information on the subject in other papyri, most of which are copies of older ones. The corpus indicates that medical science was already very elaborately organized. Its best-known figure is unquestionably Imhotep, scribe, sage, poet, astrologer, architect (designer of the stepped pyramid of Saqqarah), and priest and vizir of Zoser (2630–2611 BC), as well as of three other pharaohs, but, most importantly, a physician whose thought long influenced the entire ancient world.

The position occupied by this practitioner at the highest level of the State is not entirely accidental. At the time, and throughout Egyptian antiquity, doctors belonged to the civil-service elite. They were members of a highly organized, hierarchic corporation comprising physicians-in-chief, chiefs of physicians, inspector-physicians, physicians-in-chief of the North and South, court physicians, court inspector-physicians, and, at the very top, physician to the Pharaoh. They were paid a fixed salary, practiced free of charge, and were assisted by a series of auxiliaries of different rank (nurses, nursing aides, etc.). This organized profession, embedded in the core of government, presumably focused its energies not on the populace but on a set of individuals whose status was reflected by the rank of the caregiver. Nevertheless, it had a very visible presence among miners and workers at the giant construction sites set up to build temples and pyramids, who risked many accidental traumas.

In sum, Egypt had (1) a rather extensive, well-codified medical science handed down from generation to generation, and (2) an organized medical profession incorporated into the machinery of State and used (at least in part) to dispense care to workers in major construction projects. These features constitute, in essence, an embryonic health policy. As in many other areas of social life, Egyptian society tightly controlled health care resources and codified their use. So much so that, recounting these facts, Jean-Charles Sournia (1997) asks: "Why is it that minds capable of writing, in 3000 BC, a sensible and useful treatise on small surgical operations were succeeded by lackluster heirs, capable only of recopying manuscripts handed down by their fathers?" (p. 31). Was the immobility of minds not imposed by the monopoly exercised in this sector, as in so many others, by an omnipotent State, nearly frozen inside structures resting on an equally static society? The author carefully avoids settling the issue, arguing that the reverse scenario is just as possible. In our view, human mental processes are such as to make the first possibility more likely.

2. Mesopotamia

Mesopotamia, too, saw the very early advent of a rather sophisticated and organized form of medicine, but one less directly subordinated to political authority. By contrast, illness was chiefly viewed as punishment for a transgression, and the main purpose of a medical examination was to discover the sin and identify the spirit responsible for the ailment. Yet, as Jean-Charles Sournia notes (1997), "moral transgressions were compounded by causes of physical impurity: putting one's feet in dirty water, [. . .] touching an unclean body, etc.," (p. 19), which could sometimes produce more realistic diagnoses. Therapies based on such foundations were naturally of uneven effectiveness, but, as in Egypt, surgery was the prime option (reducing fractures, dressing wounds, cataract treatments, urethral probe, etc.); there was also medication with plants and other substances, mineral or organic. Physicians were not functionaries, but the State paid their fees, as indicated on the famous Hammurabi Stele, which sets rates for different procedures. Many tablets also attest to the abundance of periodically updated lists and classifications. At the very least, therefore, we can conclude that there were official efforts to regulate health care activities.

3. Ancient China

"His main concern was that the populace should have enough to eat and that the rites of mourning and sacrifices should be properly observed. By the greatness of his soul, he won the support of the people; by his honesty, their trust; by his diligence, their successes; by his justice, their happiness." Thus did Confucius (551–479 BC) (1981, quoted by Hoizey and Hoizey, 1988, p. 33) pay tribute to King Wu, founder of the Zhou dynasty (12th century–221 BC). Confucius does not explicitly list health among the King's concerns in regard to his people, but he probably could have, for as in the Pharaohs' Egypt and Hammurabi's Mesopotamia, medicine was already an organized profession in Zhou China. There were *jiyi* (physicians for ailments) and *yangyi* (physicians for wounds) while the *shiyi* (food doctors) kept watch on nutrition. The *Yi King*, written in the fourth century BC, was an already very detailed medical treatise gathering knowledge partly recorded in older texts. Most important, principles of hygiene and prevention were being formulated as early as this period. It is hard to determine the precise extent to which the authorities were involved in organizing medicine and applying its principles to the well-being of the populace. But we do know, for example, that during the era of the Warring Kingdoms (fourth and third centuries BC) the growth of towns led to the construction of sewers (Hoizey and Hoizey, 1988, p. 44). Sanitation was thus already an issue in public health policy.

4. Greece

Everyone has heard of the Hippocratic Oath, still taken today by medical graduates when they begin practicing. Whether Hippocrates (460–370 BC) actually wrote the text himself is of little importance.[3] Unless it had antecedents unknown to us, it marked a new step in the implementation of a health policy, insofar as it may be regarded as the first outline of a code of medical ethics. While produced by the profession itself, the code clearly sought to control the actions of the profession in the collective interest, even if compliance relied entirely on an appeal to practitioners' moral sense.

Hippocrates was so successful that he influenced medical education in Europe until the 19th century and was almost universally recognized as the father of medicine. Yet his reputation undoubtedly rests less on his own discoveries and the 60-odd texts attributed to him than on his genius as a teacher and the popularity of his famous aphorisms, which summarize in a short, proverb-like sentence each principle suited to different pathological conditions. The aphorisms had to be learned and, most important, recited by heart, particularly in the schools of Cos and, later, Thessaly, founded by Hippocrates at a time when classroom materials consisted merely of parchment for teachers and wax for students. The method was hugely successful:

> Not only were Hippocrates's aphorisms studied until the 18th century, but many teachers wrote a host of treatises more or less closely adapted from those of Hippocrates. Until the 19th century, teachers prepared phrases that were easy to remember, like those of the old master of Cos. (Sournia, 1997, p. 39)

The founding of medical schools, distant ancestors of our modern universities, was also a vital asset for introducing health policy.

5. Rome

The Romans did not have a high opinion of the medical profession, whose practice they largely entrusted to slaves (often Greek, as the Romans held Greek medicine in high esteem). Nevertheless, the Romans left traces in the long march toward the introduction of health policies. The most important, in all probability, is not the work[4] of their best-known physician, Galen.[5] This is despite the fact that, in the second century, he dominated ancient medicine, showed great skepticism toward many treatments (by advocating the notion that nature was the supreme healer), and developed a remarkable synthesis of philosophical doctrines concerning medicine. Setting aside the Chinese example mentioned earlier, the Romans' chief contribution was to found a sector of critical importance in any public health policy: sanitation. The planning rules for all cities in the Empire included highly efficient sewage systems, public latrines, and fountains for distributing drinking water. Some cities appointed and paid doctors (known by the Greek name of *archiatri*) to care for the poor and assist them during epidemics; campaigning armies enlisted the services of physicians and surgeons. Health care establishments called *valetudinaria* (the first known hospitals) were set up for veterans and the infirm.[6] The existence and proper functioning of such buildings, structures, and services required a powerful and well-managed administration. Their disappearance in the ruins of the Roman Empire is yet another indication of their role as health-policy components.

Lastly, even if its economic concerns (annuities) were rather far removed from health care issues, we must give a mention here to Ulpian's table.[7] It was an ancestor (admittedly a very imperfect one) of today's life tables,[8] a key instrument for tracking changes in mortality and hence the effects of health policies.

Despite the Byzantines' efforts to preserve the heritage of the Ancients, in our area of interest as in many others, the West lost track of it almost entirely. The West did not rediscover its roots until the Arab conquest and the shock of the Crusades.

[3] Jean-Charles Sournia (1997, p. 40) regards this as highly unlikely.

[4] *De anatomicis administrationibus, De usu partium corporis humani, De sanitate tuenda, De temperamentis et facultatibus simplicium medicamentorum, De locis affectis.*

[5] As it happens, Claudius Galen was Greek. He was born in Pergamon in AD 131 and died in 201.

[6] However, Jean-Noël Biraben (1990) argues that the first true health care institution for the ill (*nosocomium*) in the West was founded in Rome in 380.

[7] Ulpian (170–228), jurisconsult, pretorian prefect under Alexander Severus.

[8] *"It is uncertain whether the figures in Ulpian's table are intended to represent the expectation of life or the present value of an annuity of 1. However, it is worth noting that the figure at birth, namely 30, is about the level previously cited for the duration of life in ancient Greece and Rome"* (Dublin *et al.*, 1949).

II. MIDDLE AGES AND RENAISSANCE: FROM AVICENNA TO AMBROISE PARÉ

1. The Arab-Muslim Contribution

In the seventh and eighth centuries, the Arabs conquered all the southern regions of the ex-Roman Empire, and Asian lands far beyond. It is the scholars of the Muslim world, at its apogee after the conquest, who were largely responsible for recovering the ancient medical heritage and transmitting it (with enhancements) to the Medieval and early Renaissance West. The main figure in this transmission was unquestionably Avicenna.[9] His Canon (*Quanum fit'tibb'*) is a systematic review of all the diseases known at the time. Admittedly, Jean-Charles Sournia sees it as an "obscure hodge-podge from which we can deduce nothing of use to patients" and, while expressing his esteem for Avicenna, he saves his admiration for Avicenna's far less famous predecessor Rhazes (850–923). Yet Avicenna remains the emblematic and still very widely celebrated embodiment of the role played by Arab-Muslim civilization in the development of medicine and public health. His contribution was twofold. First, it served as the scientific link between the ancient world and the medieval gestation period that heralded the Western Renaissance. Second, it enriched the concept of hospital inherited from the Roman *valetudinaria* by linking it closely to the teaching of medicine.

By the eighth century, the caliphs, followed by emirs and sultans, were opening hospitals in their cities. The Caliphate of Cordoba alone seems to have had 40 hospitals. These often-spacious establishments were divided into wards, including one for the insane, and were equipped with pharmacies and libraries, in itself, a considerable improvement on ancient models. Most important, in 932, Caliph Al-Muqradir instituted a compulsory examination as a prerequisite for practicing medicine. Students could prepare for the exam by completing an internship with a teacher to whom they paid tuition, or by attending a teaching hospital. The second option soon became very popular. Students were required to examine patients and then hand them over to more qualified assistants; the teacher would confirm the diagnosis and prescribe the treatment. Thus, for four centuries, "the Muslim world had teaching hospitals, which the Christian West did not establish until the eighteenth century" (Sournia, 1997, p. 80). However, this crucial component of all health policies did not involve the authorities except at the regulatory level. Most of the resources came from public charity. Strongly encouraged by the Koran, which prescribed almsgiving as a duty, such donations provided hospitals with generous endowments.

2. In the West: from the Founding of Universities to the Discovery of the Human Body

Founded on the ruins of the Roman Empire, the Christian West was fragmented, politically and administratively destructured, and ruralized. It would no doubt have lost all memory of ancient learning if the monasteries had not carefully preserved some traces of it in their libraries. But the elements of medical science salvaged by the monks were confiscated by the Church. Despite the brilliant and innovative experience of the Salerno School,[10] the Church clung jealously to its monopoly of knowledge, virtually refusing to make it available for treating patients and, even less so, for promoting public health.

Neither the founding of the first universities[11] nor their vigorous flowering in the 13th century nor the growing number of hospitals changed this situation to any significant extent. The universities took too long to distinguish between faith and reason, and to condescend to comparing theory and reality. As for the growing number of hospitals, often run by the Church, they never set any missions for themselves other than religious: these included charitable functions (taking in the poor) and (even more often, and for a long time) the organization of pilgrimages.[12] The hospitals provided no care and no instruction. In case of need, the only solution was to summon a doctor from outside. Paradoxically, the only specialty to record some progress in this period was surgery. From the treatment of wounds to the extraction of stones and even

[9] Abu Ali ibn Abdellah ibn Sina, known as Avicenna, was born in 980 in Bukhara (in present-day Uzbekistan) and died in 1037 in Ispahan (Iran), "exhausted by his work and the joys of the flesh" (Sournia, 1997, p. 77). In almost 200 works, he recorded his extensive knowledge of subjects as varied as astronomy, mechanics, acoustics, music, optics, philosophy, and medicine.

[10] A unique school in Christian Europe, Salerno accepted students of all faiths; its faculty consisted of lay practitioners who taught nothing but medicine, separately from philosophy. For two centuries (11th and 12th), Salerno was the only glimmer in the long deep night that shrouded scientific medical analysis at the time.

[11] Bologna and Montpellier contend for seniority, as both were likely founded in the 12th century. Valencia, Oxford, Paris, Naples, Padua, Cambridge, Toulouse, and Salamanca followed soon after, in the 13th century.

[12] The words *hospital* and *hotel* derive from their function of welcoming "hosts," whether pilgrims or not.

trepanation, the methods, instruments, and indications became more sophisticated, while the first steps were taken toward the use of anesthetic gases. Held in contempt by the medical profession and entrusted to barbers (of whom the "short-robed surgeon-barbers" and "long-robed surgeon-barbers" formed the intermediate cadre and the elite, according to their degree of training), surgery advanced thanks to the relative intellectual freedom of its practitioners and their proximity to the realities of the human body.

However, the Late Middle Ages did witness the enactment of some public measures to cope with the plague in Italy. These anticipated the more significant gains achieved in the field during the Renaissance. In 1348, the Doge of Venice appointed three civil servants *pro conservatione sanitatis* to combat the spread of the plague (Bourdelais, 2001b, p. 7). "In 1374, Genoa and Venice refused entry in their ports to ships from infected localities and, in 1377, Ragusa actually decreed one month's isolation, soon extended to 40 days. Quarantine [. . .] was born" (Bourdelais, 2001b, p. 8). A half-century later, in 1423, the quarantine system was strengthened in Venice by the establishment of the Lazaretto: one of the islands in the lagoon (the island of Santa Maria di Nazareth, whose name was distorted into *lazaretto*) was set aside to accommodate persons at risk of contamination (Biraben, 1976).

Starting in the 15th century in Italy and the 16th century throughout the rest of Western Europe, the Renaissance finally awakened the scientific spirit of a world in which it had long slumbered, stifled by dogma, tradition, and prohibitions. In medicine as in other fields, religious dogma was questioned; while respect for the Ancients persisted, their ideas were assessed, criticized, and judged. One of the most important transgressions for medical progress concerned the human body. Artists paved the way when Michelangelo dared to depict a fully naked Christ, but doctors went even further by finally daring to dissect corpses (a procedure hitherto regarded as sacrilegious) and obtaining permission to perform dissections in public for teaching purposes. Anatomical knowledge took a spectacular leap forward, its dissemination aided by the talent of the greatest masters (Michelangelo, Dürer, and Leonardo da Vinci) and the brand new invention of printing.

Paradoxically, surgeons were initially excluded from this change, as they were forbidden to dissect. Their art nevertheless continued to advance, most notably by developing techniques to address an entirely new challenge: firearm wounds. The most famous surgeon in France was Ambroise Paré (1509–1590), thanks to his abundant writings: beyond his own contributions, they encompass just about all of the surgical knowledge of his time.

However, neither physicians nor surgeons could fully exploit the fresh knowledge of human anatomy, not only because they knew virtually nothing about the functioning of the organs they were discovering, but also because their operations were hampered by restrictions due to the risk of infection.

III. THE 17TH AND 18TH CENTURIES: CONTAINMENT OF MASS EPIDEMICS AND DECLINE OF FAMINE

In contrast with this scientific development (which was as spectacular as its practical effects were limited) public-health measures were being introduced by the Late Middle Ages, as described above. They became increasingly specific during the Renaissance and, most of all, the 17th and 18th centuries. Despite their lack of clearly stated theoretical foundations, these measures proved ever more effective in stifling mass epidemics, while government action in non-health areas largely contributed to the decline of famine.

1. From the War on Plague to the Organization of Public Health

The 16th century saw a generalization of the measures introduced by some Italian towns to combat the plague in the previous century. The succession of plagues in Europe during this period led governments to enact health regulations and establish the means to apply them, namely, the creation of a health police, the establishment of special health institutions, and the development of health statistics.

With the accumulation of experience gained from specific measures implemented in various localities, full-scale plague regulations of general import were devised in Western Europe in the early 16th century; this new strategy spread to Northern Europe in the second half of the century. At the same time, the war on plague shifted from the local level to the regional or national level. In France, for example, the provincial *parlements* (judicial assemblies) began by codifying the regulations of the main towns before extending them to the entire province (Biraben, 1976). Provisions typically included isolating persons who were contaminated or who risked being contaminated (this went as far as building specialized hospitals to confine plague victims without, however, providing them with medical care); the radical cleansing of infected places (incineration of furniture and clothing, etc., and even demolishing roofs in order to "aerate" houses); more general measures for urban sanitation, such as street-sweeping; the establishment of professions specializing in handling patients and corpses, including the notorious *corbeaux* ("crows," i. e., undertakers); and strict rules for separate burial. These plague regulations are the ancestors of all the health regulations later promulgated to combat epidemics, food risks, and, most recently, pollution hazards.

Health offices with nearly dictatorial police powers were set up to enforce these regulations. By 1504, in Milan, a Special Health Council had life-and-death powers. France, in different periods, introduced health captains and health provosts with archers under their command. "The health offices, vested with dictatorial powers, did not play as exclusive a role as Jules Guiart believed (1933), but they certainly made an important contribution to eradicating the plague" (Biraben, 1976, p. 143).

The plague also contributed indirectly to the birth of health statistics. Sometimes, it prompted the authorities to conduct censuses ahead of (or in the wake of) an epidemic.[13] More importantly, it provided the opportunity to underscore the value of a continuous, systematic registration of deaths by administrative offices. At first, the registration was designed to count plague victims; it was thus confined to the latter and took place sporadically. But by the 16th century, in London, it became a systematic, recurrent enterprise that produced the "bills of mortality" from which John Graunt (1662) compiled his celebrated *Observations*, the documents that marked the birth of demography and epidemiology.

The history of the last outbreak of the plague in Western Europe, halted in Marseille in 1722, is emblematic of the success of these public health policies. Owing to a lax enforcement of the plague regulation, the port was contaminated by the arrival of a ship whose name has remained famous: the *Saint-Antoine*. But the epidemic was quelled thanks to the enactment of draconian measures, culminating in the total blockade of more than a hundred towns and villages of Provence and Languedoc, including all of Marseille. At the time, the key decisions were taken at the national level. On September 14, 1720, the Regent himself banned all travel into and out of Marseille, and even ordered the disinfection of mail.

[13] Two examples: "In late 1587, under the threat of plague, the town of Apt ordered a census of mouths to feed"; "On a larger scale, in 1720, as soon as the town was threatened by the Marseille plague, Chancellor Ranchin ordered a census before the outbreak of the disease and another in 1722, after the epidemic, for assessing the extent of its toll" (Biraben, 1976, p. 108 and p. 109).

2. Toward a More Extensive State Control of the Health of the Population

In the 18th century, the authorities became increasingly concerned about the health of their populations, albeit in rather different ways from one country to another. In the Germanic countries, the State acted directly by setting up national institutions in charge of enforcing regulations on the medical professions, handling epidemics, administering hygiene measures, and disseminating new ideas about health. This arrangement was called "medical police" by Johann Peter Frank (1745–1821) (1779), whose views about the need to apply coercive regulations to achieve a healthy population were revived by the Third Reich (Hick, 2001). However, its role was not confined to policing. A special cadre of medical functionaries was in charge of the following:

> supervising medical workers in the administrative territory under their control, examining the health of the inhabitants, [. . .] training midwives, providing free care for the poor, dispensing emergency care and reporting outbreaks of epidemics, conducting surveys and recommending hygiene remedies, [. . .] and performing topographic, climatological, and nosological analysis of the locality. (Hudemann-Simon, 2000, p. 153)

In England, this move toward health policy took a very different form, which relied on the already old foundation of the 1601 Poor Law. First, it was the towns that were vested by the royal authorities with responsibility in the area; second, rather than setting up a health policy, the main goal was to offer the poor a basic access to health care in tandem with purely economic assistance. (In particular, their employment in public works was paid for in kind or in cash.) The influence of Malthusian ideas led to a radical revision of this strategy in the 19th century (Robinson, 2002).

3. The Role of the Authorities in the Decline of Famine and Massacres

Even if this diverts us momentarily from the concept of health policy per se, we cannot discuss 18th-century Europe without at least briefly mentioning the role of organized States in the decline of famines and massacres.

At the margin of health policies themselves, governments largely contributed to the decline of famine, without necessarily working toward this goal in explicit fashion. Local, regional, and national authorities sought to contain the plague, but they also sought to reduce the uncertainties of food supply for the populace. To prevent food shortages, local and provincial authorities, followed by central governments themselves, fought the speculation that often aggravated the early stages of famine by requisitioning grain stocks and controlling their distribution in the markets (Cabourdin *et al.*, 1988).

Particularly in this period, officials contributed significantly to "abating the fear of wheat crises" (Chaline, 1998, p. 25) by enacting policies to stimulate agricultural production and food imports. Meanwhile, the improvement of roads and the gradual abolition of internal customs barriers played an important role in reducing swings in local and regional food supply.

The paradox of the age was that, even as warfare became increasingly deadly and involved ever larger armies, the formation of central States considerably reduced the number of local disputes, and especially that of their direct and indirect civilian victims. Without minimizing the scale of the horrors that marked the 19th and 20th centuries, the consequences on mortality rates remained "far milder than the utter demographic disaster caused, in their time, by the Hundred Years War for France or the Thirty Years War for the [Holy Roman] Empire, with their combination of massacres, epidemics, and famines" (Chaline, 1998, p. 28).

This digression has taken us a fairly long way from health policies, but the direct and indirect role of governments in the period was sufficiently significant for its impact on mortality and health status to be recalled here.

IV. THE 19TH CENTURY: HOSPITALS FOR THE SICK AND HYGIENE FOR ALL

The two major public-health innovations in 19th-century Europe consisted in (1) finally steering hospitals toward caring for the sick and toward the teaching of medicine, and (2) emphasizing the virtues of hygiene, both public and private.

1. From the Paupers' Hospital to the Medical Hospital

For centuries, hospitals, which were essentially ecclesiastical institutions, only provided accommodation, most notably for travelers (pilgrims), the poor, the infirm, and the sick (in particular, the insane), without ever dispensing medical care. Their charitable functions for the poor expanded over time, but so did their social-protection functions. When Louis XIV established the *Hôpital Général* (consolidating *La Salpêtrière*, the *Grand Hôpital de Bicêtre*, the *Hôpital de la Pitié*, and the *Hôpital Scipion* in Paris), he was pursuing a threefold objective:

> health-related (preventing the sick poor from spreading their diseases by begging), civic (putting back to work unemployed people who, in their idle state, had taken to begging), and religious (teaching them the principles of religion they did not know, and thus enabling them to win salvation). (Biraben, 1990, p. 26)

Until the late 18th century, the hospital's overriding function was merely to shelter the indigent and the mentally ill. Some prominent exceptions did begin to emerge, such as the French Hôtels-Dieu and, most significantly, England's voluntary hospitals (Hudemann-Simon, 2000).

It was practically not before the late 18th century that the Western hospital finally became a health care facility and that medical treatment gradually became its prime mission, with the teaching of medicine as an important related activity. The first hospital in the modern sense of the term, the *Allgemeine Krankenhaus*, was founded in Vienna by Emperor Joseph II in 1784, drawing on the English and French experience. The institution was chiefly a care facility, built and administered in conformity with medical standards: its architecture included ventilation and natural lighting; patients were grouped by diagnosis; separate rooms were provided for contagious patients, convalescents, and healthy individuals; and there was a single patient per bed. The intake was no longer restricted to paupers, and a category of fee-paying patients was instituted. The hospital also became the university's teaching center for medical training (Hudemann-Simon, 2000, p. 86–87). This model was soon copied all over Europe.

In France, the Revolution abruptly destroyed the old system just as it was starting to reform. The Convention's decree of 22 Messidor Year II (July 11, 1794) nationalized hospitals, deprived them of income from farms and real estate, and put them up for sale "to terminate the charity hospital for good" (Imbault-Huart, 1996, p. 56). In so doing, however, it halted the first attempts to introduce medicine in the hospital. At the same time, all academies and schools of medicine, surgery, and so forth were abolished. However, by 14 Frimaire Year III (December 2, 1794), a decree by the Thermidorian Convention, inspired by Antoine Fourcroy, reorganized medical education by associating it with the hospital. According to Jean-Charles Sournia (1989), this decision was the true starting point of the process that transformed "an institution dedicated to misfortune into a scientific instrument." This is somewhat of an exaggeration. In reality, the shift occurred very slowly. Not until the end of the 19th century did the hospital as health care facility prevail over the hospital as poorhouse in France, which lagged far behind England. There, the transition was an uninterrupted and far swifter process that began with the voluntary hospitals.

While paupers were obviously not excluded from hospitals, they were increasingly admitted for the sole purpose of receiving medical care. On the other hand, in the same period but in other European countries as well, under the influence of economic liberalism and Malthusian ideas, the poor lost the benefit of financial relief and even medical assistance at home that had been gradually instated in earlier centuries. In England, for example, the Poor Law Amendment Act of 1834 reduced relief and tightened the eligibility criteria, which led to a sizable reduction in the number of "physicians for the poor" (Hudemann-Simon, 2000).

2. Jenner and the First Vaccinations

The 18th century had imported into Europe the Chinese technique of "variolization," which consisted in protecting humans by inoculating them with smallpox pus. Wang Tang, prime minister of the Song emperors in circa 1000, cited the use of variolization for the dual purpose of reducing mortality and preserving women's beauty (Hannoun, 1999, p. 11). The practice, which had spread along the Silk Road to Constantinople, was brought back to England in 1715 by Lady Mary Montague, wife of the English Ambassador to Turkey. However, it met with scant success outside the United Kingdom, where William Black had even developed a plan for general inoculation. In France, Charles-Daniel de La Condamine (1754) wrote "smallpox decimated us, inoculation brands us with a vintage."

The variolization method encountered some success in Parisian circles (Hannoun, 1999), but its rationale was hotly contested on account of the risks incurred, and its use remained rather limited. Inoculation was not officially accepted until 1763, and the first inoculation clinic was not established until 1799, just when Jenner's vaccine was arriving from England. In 1796, Edward Jenner (1749–1823) discovered vaccination, which provided the same result with less risk through inoculation of vaccine (cowpox) (Jenner, 1798). The 19th century thus had at its disposal a less aggressive, more effective method that may be regarded as the true ancestor of all modern vaccines. It met with swift success in all countries. In England, 6,000 people had already been vaccinated by 1800; the following year, the number reached 100,000 (Jenner, 1801). The first vaccinations on the Continent took place in Geneva and Vienna in April 1799 (Faure, 1994). Talleyrand, the French Minister of Foreign Affairs, allowed an English doctor to introduce the stocks in France, and the first vaccination was administered in 1800. Ministers and leading politicians joined the propaganda committee founded by the Duc de la Rochefoucault-Liancourt. "Pinel,[14] Guillotin,[15] the new École de santé in Paris, and General Bonaparte were its main advocates. Entire schools and regiments volunteered to be vaccinated" (Sournia, 1997, p. 214). According to the Interior Ministry, the number of vaccinations rose from 150,000 in 1806 to over 750,000 in 1812 (Biraben, 1979). After a relative decline, the figure stabilized at 500,000 to 600,000 a year (Faure, 1994). The phenomenon was a milestone in the history of health policies, as it marked the first preventive-medicine campaign. However, because of significant accidents and the persistence of smallpox, it also sparked the public's first rejections of prevention. "This first successful preventive-medicine operation created a lasting climate of incomprehension and reciprocal distrust between doctors and the populace" (Faure, 1994, p. 134). We shall return to the actual effectiveness of this first experiment in vaccination, which became the focus of the scientific debate on the role of medical progress as the driver of the decline in mortality (see Chap. 108).

In any event, vaccination became compulsory in most European countries, starting with Greece in 1825, joined by Germany in 1874, England in 1883, Italy in 1888, and Austria-Hungary in 1891. Only France awaited the 20th century, passing legislation in 1902 to introduce mandatory vaccination (Rollet-Échalier, 1990).

3. The Age of Hygienists

In the late 17th and early 18th centuries, well-known doctors from several European countries had recommended specific actions informed by the Hippocratic tradition, which held environmental conditions responsible for endemic and epidemic diseases. The measures included draining marshes and peat bogs, ensuring the circulation of water, cleaning houses, and changing the air in dwellings and places of assembly (Bourdelais, 2001). In the 19th century, with the rapid industrialization and urbanization of European countries, but also with the outbreak of cholera in 1832, these ideas were vigorously revived and began to influence government policy. As contagion mechanisms had not yet been identified, hygienists blamed epidemics on "poisons" in the

[14] Philippe Pinel (1745–1826) was appointed chief physician at Bicêtre Hospital by the Convention in 1793.

[15] Joseph-Ignace Guillotin (1738–1814), professor of anatomy and philanthropy, was inspired to invent the guillotine for humane motives.

environment, and recommended the reduction of "miasmas" (contaminated atmospheres) in water, air, dwellings, and elsewhere. By acting on these living conditions, governments could therefore take steps to improve the survival and health of the populace, especially the poor. These goals motivated men such as the French physician Louis-René Villermé (1782–1863) (1840), cofounder of the *Annales d'hygiène publique et de médecine légale*, and the British lawyer Edwin Chadwick (1800–1890) (1842), a very influential adviser to the British government.

These ideas led to the widespread construction of sewage and potable-water networks. In Paris, the first covered sewers date from the 14th century, but the construction of a full-scale, citywide underground system began in 1850. In London, cesspits gave way to continuous wastewater drainage thanks to the construction of the first sewers in 1844.

In England, Edwin Chadwick secured passage of the Public Health Act in Parliament in 1848. The legislation established the Central Board for Health; it required boroughs (and not just the main cities) to provide drainage, potable-water supply, waste collection, street paving, and other services, and it required the opening of health boards in all localities with a mortality rate exceeding 23 per 1,000.

In every country concerned, however, this policy met with stiff resistance, particularly with regard to raising the necessary funds, and when it ran counter to the interests of businesses and wealthy private individuals (Szreter, 1993). But the movement triggered by hygienists without rock-solid scientific arguments found its fulfillment in the Pasteurian revolution.

V. 1880–1945: FROM PASTEUR TO THE WELFARE STATE

Apart from the initial attempt at vaccination, health policies until the end of the 19th century could rely on little more than relatively questionable scientific foundations and medical techniques. The cholera outbreak even rekindled fears of the great epidemics of the past, and underscored the lack of knowledge of the causes of diseases. It was only in the 1880s, with the discoveries of Louis Pasteur, that the resources of modern medicine were truly established, paving the way for increasingly effective policymaking. These new weapons for combating disease and death also fortified the notion that the authorities should bear some responsibility for equalizing the odds of survival by making the benefits of the new technologies available to the largest number of people.

1. Louis Pasteur: Victory over Infection and the Advent of Prevention

The first studies that triggered this revolution were unquestionably those of the physicist Louis Pasteur (1822–1895). He began by examining beer yeast in the 1860s; then he investigated a silkworm disease whose cause he identified as a microscopic organism. In 1878, the year after his discovery, Pasteur published his *Théorie des germes* ([Theory of Germs] Pasteur, 1878); that same year, Charles-Emmanuel Sédillot coined the term *microbes* to denote them (1804–1883). In quick succession, Pasteur developed anthrax vaccine (Pasteur *et al.*, 1881) and rabies vaccine (Pasteur, 1885). Contemporaneously, Robert Koch (1843–1910) discovered the tuberculosis bacillus in 1882 (Koch, 1882) and the cholera vibrio in 1884 (Koch, 1886). And this was only the start of a long series of discoveries that, in the 19th and 20th centuries, led to the production of vaccines against many infectious diseases (Table 106–1).

These discoveries immediately won the support of hygienists, who largely helped to popularize both the new underlying ideas and their value for improving public health. The hygienists argued their case before

TABLE 106–1 Chronology of Diseases for Which a Vaccine Was Developed

Eighteenth century		**Twentieth century**	
Smallpox inoculation[a]	1715	Pneumococcus	1911
Vaccine	1796	BCG (tuberculosis)	1921
		Diphtheria	1923
		Whooping cough	1926
		Tetanus	1927
Nineteenth century		Yellow fever	1932
		Typhus	1941
Rabies	1885	Influenza	1944
Cholera	1892	Japanese encephalitis	1954
Typhoid	1896	Poliomyelitis	1955
Plague	1897	Poliomyelitis (attenuated)	1962
		Measles	1963
		Mumps	1967
		German measles	1969
		Meningitis	1973
		Chickenpox	1974
		Brucellosis	1979
		Hepatitis B	1981
		Tick encephalitis	1991
		Hepatitis A	1992

[a] Introduced in Europe.
Source: Adapted from Hannoun, 1999, p. 29.

the medical and political communities. Health policies could now resolutely engage in two programs previously foreshadowed by hygienists and henceforth scientifically recognized: prevention through vaccination and asepsis procedures. The first considerably reduced the incidence of infectious diseases; the second enabled surgery to reach full efficiency at last. Germ theory eventually led to the discovery, in the mid-20th century, of sulfonamides followed by antibiotics, further strengthening the arsenal of medicine and health policies against infectious diseases.

Vaccination soon became a major issue in public health policies, which typically sought to combine information or incentive campaigns with mandatory measures. The concept of compulsory vaccination had already been invented, as noted earlier, in connection with Jenner's vaccine. France was one of the very last European countries to adopt the method, by passing the 1902 Act. The concept was revived to protect the population from a number of infectious diseases as the vaccines against them were successively developed. However, in a country like France, public reluctance was fairly strong. In France, for example, the distribution of vaccines was encouraged, but the only new vaccine made compulsory before World War II (against diphtheria) did not become so until 1938 (Rollet-Échalier, 1990, p. 206); in Britain, the only official action to disseminate it was an incentive campaign (Parish, 1965).

Even before the advent of antibiotics, Pasteur's discoveries also confirmed the presuppositions of the Hungarian obstetrician Ignaz Semmelweis (1818–1865) (1847) and the English surgeon Joseph Lister (1827–1912) (1867) on the means of protecting their procedures from the danger of infection by applying a maximum of antiseptic measures. Pasteurian sterilization enabled asepsis to replace antisepsis. The ensuing renewal of surgery was all the more consequential because of the discovery, in the mid-19th century, of safer, more effective methods of anesthesia. Freed from the risks of infectious complication and the impossibility of sparing the patient from excessive pain, surgery took an unprecedented leap forward between the wars.

The absolute weapon against infectious diseases seemed to have been discovered at last with the introduction of sulfonamides in the 1930s and, even more decisively, antibiotics during World War II. In 1932, Gerhard Domagk found a substance capable of selectively killing pathogen bacteria; this led to the market release of the first sulfonamides in 1935. But, as early as 1928, Alexander Fleming (1881–1955) (1929) had discovered penicillin, whose production did not actually begin until 1942.

2. From the First Provident Funds to Social Security

In traditional societies, some institutions (most notably the Church in the West, more seldom the civil authorities) consistently devoted special attention to the poor. But their health was never the main issue. It scarcely elicited the interest of governments except insofar as they sought to exercise a health-police function aimed more at protecting society than at treating the poor. In the late 19th century, the conjunction of rising socialist ideas and the new medical weapons born of Pasteur's discoveries fostered the emergence of a new health strategy. Its ultimate stage was the coverage of health care costs by the community. Previously, apart from some aspects involving public health, individuals were alone responsible for their health care. They had to bear the cost themselves, and were almost entirely dependent on family solidarity in case of illness or accident.

With the expansion of the salariat and the rise of a collective consciousness of the working masses fueled by the dissemination of Marxist ideas, new forms of solidarity developed. Individually, manual and clerical workers lacked the income, and, even more, the capital, to cope with health contingencies or the usually disastrous consequences of a disabling accident that deprived them of work. But the risks could be substantially cushioned if their burden was managed collectively. These circumstances led to the early forms of social insurance on the margins of (or incorporated into) the nascent labor organizations in the 19th century. By the mid-19th century, many mutual assistance societies were already active. France, for example, had nearly 4,000 in 1860, covering more than 500,000 members (Comiti, 1997). Meanwhile, the action of labor unions had led employers, spontaneously or under State coercion (including court rulings and new laws), to recognize their liability for occupational accidents and diseases. To protect themselves against the risks arising from this liability, employers, in turn, set up special insurance funds.

The State intervened fairly quickly to regulate, sometimes to slow down, and more often to hasten the march toward the collective coverage of illness and accident risks. But the Welfare State was not attained until the second half of the 20th century. To measure the road traveled, we can, for example, read the works of Louis-Adolphe Thiers (1850), speaking on behalf of the *Commission de l'assistance et de la prévoyance publique* (Commission on Assistance and Public Provident Insurance):

> A healthy, hard-working man, endowed with ordinary intelligence, who does not indulge in every vice, can, through his

> labor, in our civilized societies, meet his various obligations, barring fortunately accidental circumstances such as illness, an unemployment spell, a hailstorm, or a flood. Society's effort to assist the healthy man in these accidental occurrences is a noble concern that it should pursue in its providential function, and that its virtue demands; the degree of success will be equal to the esteem that [society] deserves. But, apart from these exceptional occurrences, the society that would seek, to whatever extent, to take charge of the fate of a group of its members would turn them into idlers, rabble-rousers, and trouble-makers, at the expense of all hard-working and peaceful citizens. (Thiers, 1850, p. 31–32.)

Thus the important point, for Louis-Adolphe Thiers, was that the State should control the mutual assistance societies to which it provided support, so as to avoid such slippages.

The true foundations for modern social-protection systems were laid in Germany in the 1880s, under the influence of Chancellor Otto von Bismarck. An initial law introduced the first compulsory sickness-insurance system in 1883. This was followed in 1884 by an occupational-accident protection system, and, in 1889, by a law on old-age and disability insurance. However, unemployment insurance was not established until 1927 under the Weimar Republic, just two years before the Great Depression.

France took much longer to move from mutual assistance societies to a more general system at the national level. Admittedly, ever since Colbert, who had established a sailors' insurance fund under Louis XIV, the State had intervened directly to set up specific funds for civil servants, military personnel, miners in 1894; rail workers in 1909–1910; and female factory workers and female farm workers in 1910. But *assurances sociales* (social insurance) was not introduced until 1928–1930, followed by legislation on family benefits in 1932. Even so, it should be stressed that, in 1936, the medical insurance established a few years earlier covered just over one-third of the population, and the reimbursement rates were very low. The factory workers and farm workers collecting pensions under the new system represented a mere 20% of the population aged over 60 (Comiti, 1997). It was only in 1945 that the government began to extend the system with a view to introducing a universal right to health care.

Sweden was far ahead on this issue, since, as early as the 1910s, it laid the foundations of a social-protection system for the entire population (Olsson Hort, 2001). By contrast, during the same period, the United States continued to apply a very free-market approach to health policy.

In many sectors, World War II led European societies toward reform and even reconstruction. One such sector was health. The 19th century had separated health policy from social policy, transformed the hospital for paupers into a hospital for care. The second half of the 20th century blended these two concerns into a single quest for the reduction of social inequalities and, in particular, the unfairest of all: inequality in the face of illness and death. Europe adopted the broader framework of the Welfare State, responsible for providing all citizens with access to decent living conditions, from cradle to grave, in all areas: education, housing, work, leisure, retirement, and so on. Health care accordingly became an entitlement, and the State now had to ensure its availability to all.

In the health sector, two new systems were born from the ruins of war. The first was not a genuine innovation in conceptual terms, as it was directly inspired by the Bismarckian model. In fact, it was a generalized version of the latter. It rested on the establishment by the State of compulsory sickness-insurance funds ("regimes"), managed by the "social partners" (employers' organizations and labor unions) and financed by social contributions, which are assessed mainly on earned income. The regimes sign agreements with health care provider partners in the public and private sectors, such as hospitals, doctors in private practice, and pharmaceutical firms. In consultation with the State, they define compensation rules and the medical fees charged to patients, as well as procedures for their total or partial reimbursement. The State sets the rules and ensures the system's overall balance, but does not intervene in its management unless exceptional circumstances require it. Citizens are granted a right to health in the form of a health care entitlement, but patients are free to choose their providers. This is the system adopted in Austria, Belgium, France, Germany, and the Netherlands, among others (Letourmy, 2000).

The other system is modeled on the Beveridge Plan,[16] implemented in the United Kingdom in 1946–1948 (Harris, 2001). It rests on the primacy of a public health care sector, largely funded through taxes (the private sector is not, however, eliminated: it remains open to patients who can afford it). Citizens have access to it free (or nearly free) of charge, but the choice of entry point is restricted. The government directly controls its financing, supplied in the form of fiscal allocations to institutions and fees to physicians, which are proportional to the number of persons enrolled on their lists. The authorities also control medical care, provided in public-sector facilities. This British model has been reproduced in Ireland, the Nordic countries, Italy, Spain, Portugal, Greece, and

[16] Named after Sir William Beveridge (1879–1963), author of the eponymous report filed in 1942 and published in 1943.

elsewhere. In fact, it is not too different from the totally State-controlled system set up in the Soviet Union before the war and by all the Communist regimes of Central and Eastern Europe from 1945 onward.

By contrast, the United States still does not have a universal protection system, although efforts to help some population categories have been made in recent years (most notably, the introduction of Medicare[17]). Private insurance or, at best, insurance agreements between employers and employees are the only way to pool health care risks in the United States.

In contrast, the notion of a universal right to health (or at least of the need to reduce the huge disparities at the global level) largely informed the creation of the World Health Organization (WHO) and shaped its program goals. This process occurred even as the colonial powers themselves were beginning to show concern about the health of indigenous populations (the plundering of resources was partly giving way to the development of local production, which required a healthy workforce). The spectacular medical progress since the late 19th century had raised the hope that industrial countries would be able to make their citizens equal in the face of death. Likewise, it seemed realistic to espouse the goal of offering the disinherited inhabitants of the Third World the benefits of the cheapest new technologies. This would enable them, if not to attain the highest life expectancies, at least to substantially narrow the gap with the richest countries, without waiting for the Third World to reach the same level of economic development.

For 30 years—the period from 1945 to the first oil shocks of the early 1970s, known in France as "Les Trente Glorieuses" (Fourastié, 1979)—the ideals of social security and third-worldism have forged ahead, largely inspiring national and international health policies, as long as the extended economic prosperity after World War II lasted. With the global economic crisis of the 1970s, the comeback of economic liberalism, the fall of the Soviet Empire, and the imbalance of age distributions generated by the decline in European fertility well below the replacement rate, those major advances have reached their limit. The Welfare State is pilloried in the northern countries, while drastic structural adjustments are forced on the southern countries. In the northern countries, health care systems are, if not openly questioned, at least subject to ever tighter budgetary constraints. In France, for example, a supply-control policy (involving the rationalization of budgetary choices, health care planning, and the definition of a "health care map") has given way to an even harsher policy of spending control (cutbacks in benefits) to avoid an excessive rise in contributions, which are increasingly contested. In some southern countries, particularly in sub-Saharan Africa, tottering health services have been dealt a fatal blow with the eruption of HIV/AIDS.

These developments make an overall reappraisal of health policies more necessary than ever.

[17] Medicare was established by an act on July 30, 1965, under Lyndon Johnson's administration (Corning, 1969). The system offers free hospital care to retired workers over 65 who have worked for more than 40 quarters, and to their spouses (Card*et al.,* 2004).

Acknowledgment

This chapter was translated from the French version by Jonathan Mandelbaum.

References

Beveridge William H., 1943. *Social insurance and allied services.* London, Allen and Unwin, 300 p.

Biraben Jean-Noël, 1975. *Les hommes et la peste en France et dans les pays européens et méditerranéens. Vol. I: La peste dans l'histoire.* Paris, Mouton, 459 p.

Biraben Jean-Noël, 1976. *Les hommes et la peste en France et dans les pays européens et méditerranéens. Vol. II: Les hommes face à la peste.* Paris, Mouton, 416 p.

Biraben Jean-Noël, 1979. La diffusion de la vaccination en France au XIX[e] siècle, *Annales de Bretagne et des pays de l'Ouest*, vol. 86(2), p. 265–276.

Biraben Jean-Noël, 1990. Histoire des hôpitaux de Paris (VII[e]–XVIII[e] siècles), *Revue de la Bibliothèque nationale*, no. 36, "L'hôpital et la ville," p. 20–28.

Bourdelais Patrice (ed.), 2001a. *Les hygiénistes. Enjeux, modèles et pratiques. XVIII[e]–XX[e] siècles*. Paris, Belin, 544 p.

Bourdelais Patrice, 2001b. Les logiques du développement de l'hygiène public, *in*: Patrice Bourdelais (ed.), *Les hygiénistes. Enjeux, modèles et pratiques. XVIII[e]–XX[e] siècles*, p. 5–26. Paris, Belin, 544 p.

Cabourdin Guy, Biraben Jean-Noël, and Blum Alain, 1988. Les crises démographiques, *in*: Jacques Dupâquier (ed.), *Histoire de la population française. Vol. II: De la Renaissance à 1789*, p. 175–220. Paris, Presses universitaires de France, 602 p.

Card David, Dobkin Carlos, and Maestas Nicole, 2004. *The impact of nearly universal insurance coverage on health care utilization and health: evidence from Medicare.* Cambridge (Massachusetts), National Bureau of Economic Research, 73 p. (Working Paper 10365; http://www.nber.org/papers/w10365).

Chadwick Edwin, [1842] 1965. *Report on the sanitary condition of the Labouring population of Great Britain.* Reprint. Introduction by M. W. Flinn. Edinburgh, Edinburgh University Press, 457 p.

Chaline Jean-Pierre, 1998. Une Europe transformée, *in*: Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. Vol. II: La Révolution démographique 1750–1914*, p. 19–36. Paris, Librairie Arthème Fayard, 648 p.

Comiti Vincent-Pierre, 1997. *Histoire sanitaire et sociale.* Paris, Presse universitaires de France, 128 p. ("Que sais-je?" series).

Confucius, 1981. *Entretiens de Confucius.* Translation, introduction, notes, maps, and chronology by Anne Cheng. Paris, Éditions du Seuil, 180 p.

Corning Peter A., 1969. *The evolution of Medicare . . . from idea to law.* Washington, US Department of HEW, 151 p. (http://www.ssa.gov/history/corning.html).

Dublin Louis I., Lotka Alfred J., and Spiegelman Mortimer, 1949. *Length of life: a study of the life table*, rev. ed. New York, The Ronald Press Company, 321 p.

Faure Olivier, 1994. *Histoire sociale de la médecine (XVIII^e–XX^e siècle)*. Paris, Anthropos, 272 p.

Fleming Alexander, 1929. On the antibacterial action of cultures of penicilium, with special reference to their use in the isolation of *B. influenzae*, *British Journal of Experimental Pathology*, vol. 10(3), p. 226–236.

Fourastié Jean, 1979. *Les Trente Glorieuses ou la révolution invisible de 1946 à 1975*. Paris, Fayard, 288 p.

Frank Johann Peter, 1779. *System einer vollständigen medicinischen Polizey*, 6 vol. + suppl. Vienna, Schaumburg, 7,000 p.

Graunt John, [1662]. *Natural and political observations made upon the bills of mortality.* London, printed by T. Roycroft for J. Martin, J. Allestry, and T. Dicas, 150 p. (+ U. S. ed. by W. F. Wilcox, Baltimore, Johns Hopkins Press, 1939 + French ed. by Éric Vilquin, Paris, INED, 1977).

Grmek Mirko D. (ed.), 1995. *Histoire de la pensée médicale en Occident. 1. Antiquité et Moyen-Âge.* Paris, Seuil, 380 p.

Grmek Mirko D. (ed.), 1997. *Histoire de la pensée médicale en Occident. 2. De la Renaissance aux Lumières.* Paris, Seuil, 386 p.

Grmek Mirko D. (ed.), 1999. *Histoire de la pensée médicale en Occident. 3. Du romantisme à la science moderne.* Paris, Seuil, 428 p.

Guiart Jules, 1933. Pourquoi la peste nous a quittés. Histoire de la peste en France au XVI^e siècle et au XVII^e siècle, *Revue de l'Université de Lyon*, no. 1, p. 53–91.

Hannoun Claude, 1999. *La vaccination.* Paris, Presses universitaires de France, 128 p. ("Que sais-je?" series).

Harris Jose, 2001. Le cas de la Grande Bretagne, *in*: Comité d'histoire de la Sécurité sociale (ed.), *Un siècle de protection sociale en Europe*, p. 19–26. Paris, La documentation française, 286 p.

Hick Christian, 2001. "Arracher les armes des mains des enfants": La doctrine de la police médicale chez Johann Peter Frank et sa fortune littéraire en France, *in*: Patrice Bourdelais (ed.), *Les hygiénistes. Enjeux, modèles et pratiques. XVIII^e–XX^e siècles*, p. 41–59. Paris, Belin, 544 p.

Hoizey Dominique and Hoizey Marie-Joseph, 1988. *Histoire de la médecine chinoise.* Paris, Payot, 294 p. ("Médecine et sociétés" series).

Hudemann-Simon Calixte, 2000. *La conquête de la santé en Europe. 1750–1900.* Paris, Belin-De Boeck, 176 p.

Imbault-Huart Marie-José, 1996. Médicalisation et fonction sociale de l'hôpital: convergences et contradictions, *in*: Musée de l'Assistance publique-Hôpitaux de Paris (ed.), *Depuis 100 ans, la société, l'hôpital et les pauvres*, p. 53–86. Paris, Douin, 236 p. (Catalogue of exhibition at Musée de l'AP-HP, October 24, 1996–March 2, 1997).

Imbert Jean, 1977. *Guide du chercheur en histoire de la protection sociale. Vol. II (1789–1914).* Paris, Association pour l'étude de l'histoire de la sécurité sociale, 254 p.

Jenner Edward, 1798. *An inquiry into the causes and effects of the* variolae vaccinae, *a disease discovered in some of the western counties of England, particularly Gloucestershire, known by the name of the cow pox*. London, Sampson Low, 75 p.

Jenner Edward, 1801. *The origin of the vaccine inoculation.* London, D. N. Shury, 12 p.

Koch Robert, 1882. Die Aetiologie der Tuberkulose, *Berliner klinische Wochenschrift*, vol. 19(15), p. 221–230.

Koch Robert, 1886. *Die Cholera auf ihren neuesten Standpunkte.* Berlin, Verlag von Marlin Hampel.

La Condamine (de) Charles-Daniel, 1754. *Mémoire sur l'inoculation de la petite vérole lu à l'Académie royale des sciences.* Paris, Durand, 95 p.

Letourmy Alain, 2000. Les politiques de santé en Europe, vue d'ensemble, *Sociologie du travail*, vol. 42(1), p. 13–30.

Lister Joseph, 1867. On the antiseptic principle in the practice of surgery, *The Lancet*, vol. 90(2), p. 353–356 and 668–669.

Musée de *l'Assistance Publique-Hôpitaux de Paris* (ed.), 1996. *Depuis 100 ans, la société, l'hôpital et les pauvres.* Paris, Douin, 236 p. (Catalogue of exhibition at Musée de l'AP-HP, October 24, 1996–March 2, 1997).

Olsson Hort Sven E., 2001. La société civile, l'État et la sécurité sociale en Suède: centralisation et décentralisation dans le modèle scandinave, *in*: Comité d'histoire de la Sécurité sociale (ed.), *Un siècle de protection sociale en Europe*, p. 49–58. Paris, La documentation française, 286 p.

Parish Henry J., 1965. *A history of immunization.* Edinburgh and London, E. & S. Livingstone, 356 p.

Pasteur Louis, 1878. *La théorie des germes et ses applications à la médecine et à la chirurgie. Lecture faite à l'Académie de médecine par M. Pasteur en son nom et au nom de MM. Joubert et Chamberland, le 30 avril 1878*. Paris, Masson, 23 p.

Pasteur Louis, 1885. Méthode pour prévenir la rage après morsure, *Comptes rendus de l'Académie des sciences*, vol. 101(17), p. 765–774.

Pasteur Louis, Chamberland Charles-Édouard, and Roux Émile, 1881. De la possibilitéde rendre les moutons réfractaires au charbon par la méthode des inoculations préventives, *Comptes rendus de l'Académie des sciences*, vol. 92, March 21, p. 662–665.

Robinson Warren C., 2002. Population policy in early Victorian England, *European Journal of Population*, vol. 18(2), p. 153–173.

Rollet-Echalier Catherine, 1990. *La politique à l'égard de la petite enfance sous la III^e République*. Paris, INED, PUF, 594 p. ("Travaux et documents" series, Cahier no. 127).

Semmelweis Ignaz F., 1847. Höchst wichtige Erfahrungen über die Aetiologie der in Gebäranstaltle epidemischen Puerperalfiebern, *Zeitschrift der Gesellschaft der Aerzte in Wien*, vol. 4, p. 242–244.

Sournia Jean-Charles, 1989. *La médecine révolutionnaire.* Paris, Payot, 306 p.

Sournia Jean-Charles, 1997. *Histoire de la médecine.* Paris, La découverte, 358 p. ("Sciences humaines et sociales" series, no. 41).

Szreter Simon, 1993. The idea of demographic transition and the study of fertility change: a critical intellectual history, *Population and Development Review*, vol. 19(4), p. 659–701.

Thiers Louis-Adolphe, 1850. *Rapport général présenté par M. Thiers au nom de la Commission de l'assistance et de la prévoyance publique dans la séance du 26 janvier 1850*. Paris, Paulin, Lheureux et Cie, 156 p.

Villermé Louis René, 1840. *Tableau de l'état physique et moral des ouvriers employés dans les manufactures de coton, de laine et de soie.* 2 vol. Preface by Jean-Pierre Chaline and Francis Démier. Paris, Jules Renouvard et Cie, 448 p., 452 p. ("Études et documentation internationales" [EDI] series).

CHAPTER

107

Health Policies:

What Strategies? For Whose Benefit?

JACQUES VALLIN AND FRANCE MESLÉ

Institut national d'études démographiques (INED), Paris, France

Infirm of constitution, prey to constant pain, he shunned external help from medicine to follow instead the regime he had thought out for himself; he liked to say that his health, his longevity, was his work.

Victor Delbos (1971) on Immanuel Kant (1724–1804)

Health gains throughout history have been dictated by advances in two conjoined directions: the development of increasingly efficient new technologies, and their spreading use. The former requires investment that is scientific (in medical, pharmaceutical, and technological research) and industrial (to produce medical equipment of all kinds). The latter is more exacting still, requiring not just investment in training (of health professionals and the public) and infrastructure (care and preventive health facilities, health administration), but ensuring that all citizens have access to health provision and benefit from appropriate preventive measures. History has also shown that market forces clearly cannot deliver both with optimum outcomes. There are at least minimum aspects of health policy that are basic functions of the State—public health protection/enforcement, compliance with codes of practice, organization of prevention, and so on—and this minimum level of collectivist care is almost certainly nowhere near enough. Much could be said on the issue of whether the State is best placed to deliver an objective like the "health for all" objective called for by the World Health Organization (WHO), assuming that objective is not wholly unrealistic per se. In any event, the State is not the only relevant actor. Arguably, it may be more helpful at this point simply to review the different objectives and strategies of current health policies, and the actors and policy instruments involved.

I. OBJECTIVES AND STRATEGIES

A recurrent theme of the preceding chapter's historical review is that public health policy objectives may be very wide-ranging. A treatise on demography could start by distinguishing objectives with demographic impacts (mainly via mortality, and incidentally reproductive health as the determinant of fecundity and fertility) from those that are only remotely (e.g., for diseases or disabilities with no survival effect) or wholly (e.g., purely economic cost control or containment) unconnected with population dynamics, unless they negatively impact survival gains by restricting health resources. On that basis, this chapter would address only the former, treating the others as off-topic. Arguably, the dividing lines between these categories are too vague to be strictly demarcated, although the main focus will obviously be on the former.

The objectives of a health policy (or program, or measure) could also feasibly be distinguished by sphere of action (public health protection/enforcement, medical prevention, care provision, etc.) to consider the scale and type of aims (protecting the general population against certain risks, closing glaring health gaps, or meeting all individual health needs collectively by across-the-board health resources funding). Not only are these two criteria too inextricably linked to completely unravel the individual strands, whether starting from sphere of action or degree of involvement, but also any attempt to give even a halfway logical presentation of them means moving seamlessly between the two. At the risk of setting what may at times be somewhat arbitrary or vague limits, therefore, this chapter will endeavor to move linearly from issues around the general protection of a population from collective external risks to those around the curing of acute diseases or transient injuries, then improvements in the quality of years gained, to conclude with the reduction of certain glaring inequalities, and finally the collective payment of the total cost of health for all.

1. Public Health Protection/Enforcement

Granted that 20th-century society no longer needs protecting from many of the health risks present in earlier centuries, there remains an array of specific and general measures to protect the population against various health risks, some of which place mandatory requirements on individuals.

a. Protection from Criminal Assaults

One aim of enforcement action has a connection with the topic of this chapter, in that one of its functions is to protect citizens from criminal assault, and the demographic issues of such protection are less marginal than might be thought. Not only is the significant decline in exposure to this type of risk since the establishment of organized States a matter of record (Chesnais, 1976, 1981, 1985), but also the violence to which some populations are still exposed (e.g., those populations in countries where the State is either weak or disregards human rights) argues strongly for the importance of this objective. For example, Colombia's nonaccidental violent death rate in 1988 was 16 times that of Costa Rica (Ruiz and Rincón, 1996), while in 2000, its homicide mortality was still 10 times higher than that of the United States (Chesnais, 2003). The most exposed countries are also often those where reliable statistics are most lacking. Policies may differ widely even between long established democracies. The rules on carrying weapons are a case in point, and in this respect, human life is surely far better protected in most western European countries than in the United States.[1]

b. Supervision of Prostitution

Vice squads also partially pursue health objectives when, for example, conducting checks on the health of sex workers to protect that of the clients. This was a major concern in Europe before treatment was available for syphilis and other venereal diseases. In particular, this concern was behind the early debates on brothels and the various ways they were regulated in different countries at different times (Nor, 2001). It is very much on the agenda today in countries affected by HIV/AIDS. In Thailand, for example, the government has made condom use compulsory in brothels under threat of closure (P. Caldwell, 1995). In sub-Saharan Africa, wherever HIV/AIDS has wrought most havoc (which is also where the dividing lines of what constitutes prostitution are sometimes very unclear and varied in practice), such enforcement activities are far more difficult to carry out.

c. Epidemiological Surveillance at the Borders . . .

One key branch of modern public health protection/enforcement is also the product of tackling major epidemics down the ages. Health surveillance at national borders, which is its direct descendant, although now increasingly low-key, is still very real. Travelers must prove that they have been vaccinated against diseases that vary with their country of origin. The WHO keeps an updated database on countries of the world by reference to the presence and scale of contagious endemic diseases.[2] This database effectively operates as a sort of international public health protection/enforcement measure, since for three diseases at least (cholera, plague, and yellow fever), each country must report every case diagnosed to the

[1] Homicide mortality is much higher in the United States than in western European countries (Fingerhut and Kleinman, 1990; Chesnais, 2003), and most homicides in the United States are gun-crime killings; there is a strong probability that this high firearms-related component is directly related to their very widespread prevalence in the population (Shahpar and Li, 1999).

[2] This systematic international tracking system, carried out by the Communicable Disease Supervision and Response Unit, currently covers 15 diseases (anthrax; avian influenza; Crimean Congo, Dengue, and Ebola hemorrhagic fevers; hepatitis, influenza; Lassa fever; meningococcal disease; plague; Rift Valley fever; SARS [severe acute respiratory syndrome]; smallpox; tularemia; and yellow fever). The WHO maintains a website with continually updated information (http://www.who.int/csr/disease/en), as well as world maps showing the current status of each disease (http://www.who.int/globalatlas).

WHO.[3] It is this information, possibly supplemented from other sources, that individual countries use to decide for which diseases a certificate of vaccination is required to enter their territory. An unvaccinated traveler may be forced to submit to immediate vaccination or other appropriate preventive measures (e.g., taking medicine). But these routine measures may be tightened up at any time to address a particular threat, as was recently seen with the emergence of SARS (WHO, 2003).

d. . . . and Inland

But epidemiological surveillance does not stop at the border; it also accounts for much of the administrative activity of public health agencies. In France, for example, while it may have been largely supplanted by other tasks, the main function devolved on county health and social action authorities (DDASS) and the special staff of medical health officers that run them was to monitor for emerging epidemic risks and take all necessary steps to stop them from spreading. The general use of vaccination and effective antibiotics have considerably reduced the priority given to this function, but the lineup of legal and administrative measures remains in place, ready to be reactivated if need be. Here too, public enforcement activity relied heavily and still relies on the information collected through compulsory notification systems: not just the obvious cause-of-death notification (in France, cause-of-death certificates are sent first to the DDASS medical officers), but also diagnosed cases of certain infectious diseases. In France, for example, 22 infectious diseases (plus, recently, childhood lead poisoning) are still subject to compulsory notification (Table 107–1) (Antoniotti *et al.*, 2002). In 1998, the *Institut de veille sanitaire* (approximately equal to the United Kingdom's Health Protection Agency) was set up to alert the authorities to emerging threats to public health (Bonnici, 2003).

TABLE 107–1 List of Compulsorily Notifiable Diseases in France

Infectious diseases	
Botulism	Meningococcal cerebrospinal meningitis and meningococcaemia
Brucellosis	Locally contracted malaria
Cholera	Imported malaria in the overseas departments
Diphtheria	Plague
African hemorrhagic fever	Poliomyelitis
Typhoid and paratyphoid fever	Rabies
Yellow fever	Suspected Creutzfeldt-Jakob disease and other human transmissible subacute spongiform encephalopathies
Acute hepatitis B infection	Tetanus
HIV infection at any stage	Mass food poisoning
Legionella infections	Tuberculosis
Listeria infections	Exanthematic typhus
Other diseases	
Childhood lead poisoning	

Source: Decree No. 99–363 of 6 May 1999 prescribing the list of diseases for which individual data must be notified to the health authority and amending the Public Health Code (Official Gazette of 13 May 1999) and Decree No. 2001-910 of 5 October 2001, amending the list (Official Gazette of 6 October 2001).

e. Tackling Legal and Illegal Drug Abuse

Drugs are another area of public health protection/enforcement. Narcotics use was seen as a public health issue from quite an early date, and attempts were made to prevent people accessing them by making the manufacture, sale, and even consumption of them illegal. Dealing in narcotics was made illegal in 1916 in both Britain (Defence of the Realm Act, Regulation 40b) and France (Act of 12 July 1916 "on the import, sale, possession and use of noxious substances, in particular opium, morphine and cocaine"), and very strict controls were put in place on their issue on medical grounds (Bachmann and Coppel, 1989). Penalizing users was clearly the softest option, as they are the weakest link in the chain and hence were most accessible to enforcement action. But it also very quickly proved to be the least effective starting point, and trying to stamp out drugs through a crackdown on users alone was an endless task. Most antidrug policies therefore now focus on producers and traffickers: going back up the chain, punishing the middlemen and top men if possible, and destroying production sites. The problem soon transcends the national borders of the country concerned. Some countries, like the United States, feel sufficiently powerful to try and mount enforcement and production site destruction operations alone (or to force the States concerned to do so), while others can only rely on international police cooperation or enabling action by international organizations like the WHO.

One major strategic issue here is whether all substances currently classed as drugs should be banned,

[3] New International Health Regulations (IHR) adopted by the World Health Assembly in 1969, replacing the 1951 regulations, were revised in 1973 and 1981 (WHO, 1983). They are now being reworked to take account of the new international trade situation (OMS, 1999a).

or only the most dangerous ones. There is a growing call for liberalization of so-called soft drugs like cannabis or marijuana and their derivatives like hashish, khat, and so forth (Nadelmann, 1994) on the grounds that they are not lethal, not ruinous to health, and only very rarely lead to physical dependence (Roques, 1999). The argument that they are still dangerous in that they gradually lead users to hard drugs does not seem to be borne out in practice (Jauffret-Roustide and Simmat-Durand, 2004). Some countries have in fact at least partly relaxed their prohibitions (Désesquelles, 1997).[4] By contrast, they remain strictly prohibited wherever the idea prevails that they are harmful, with some authors hypothesizing a "hard use of soft drugs and soft use of hard drugs" (Jauffret-Roustide and Simmat-Durand, 2004; Roques, 1999). Some, although admittedly fewer, voices advocate legalization as the best way to kill off drug trafficking by ruining producers and dealers through the price collapse that would ensue (Nadelmann, 1994). No country to date, however, has ventured to test this theory out.

Contrast this with the case of tobacco. There is a credible argument that, but for its early introduction in Europe, and especially its mass distribution to soldiers in the trenches during the Great War (Ravenholt, 1990),[5] and not least in many cases its state-controlled production and marketing (as in France, with the creation of SEITA),[6] tobacco would have inevitably been banned in the same way as any other drug, and as likely classed as a hard rather than a soft drug (Meslé, 1990). Neither total freedom of movement of tobacco, and probably still less state control of its production and distribution, has stopped smoking becoming a leading cause of death in many countries through the phenomenal increase in cancer of the respiratory system throughout the 20th century, let alone the rise in ischemic heart diseases. Health policies, however, have had to come at it from different angles. It is nevertheless possible to bring within the purview of public health protection/enforcement, such steps as the WHO's political condemnation of the practices of major producers who "use their sophisticated and usually secret research to ensure that just the right dose of nicotine gets delivered to smokers for addiction to occur and be maintained [. . . and who] targeted children whom they need to fill consumer charts as smokers die" (Yach, 2000, p. 3). At the other end of the chain, passing recognition may also be given to those lawyers who claim (and sometimes win) large sums in financial compensation for their sick clients or their families.

Alcohol is a different issue from tobacco, in that the health damage it causes appears only after a certain level of consumption, while below that level, it may even have certain beneficial effects (INSERM, 2001, p. 21). But alcohol has been the target of public health protection/enforcement actions that are often as spectacular as they are ineffective. The Prohibition Era in the United States, the reknowed inspiration for so many police films, springs to mind.[7] More recently, Gorbachev's celebrated anti-alcohol abuse campaign, which ran from 1985 to 1988 and stimulated the then slowly declining Soviet life expectancy, was short-lived, soon giving way to a mortality surge (Meslé *et al.*, 1994; Shkolnikov and Nemtsov, 1997). In France, likewise (Berger *et al.*, 2000), public health protection/enforcement measures were taken against alcohol production, in particular phasing out grower-distiller rights (the French government abolished inheritance of grower-distillers' right on August 30, 1960), and against the consumption of farm-distilled alcohol by such measures as licensing of drinking establishments (Sournia, 1986; Fillaut, 1999). More of the decline in alcohol abuse has been due to economic policies (fixed selling prices and levies) and health education policies rather than to public health protection/enforcement proper, however.

f. Food Safety

After drugs, tobacco, and alcohol naturally comes food. Over time, the food sector has been brought under increasingly detailed, far-reaching and strict regulation, at least in the developed countries, where consumption of homegrown food has become marginal, and even meal preparation is increasingly done outside the home. That it is by no means a side issue is easily exemplified by recalling the acute infant mortality crisis that occurred in France in 1945 as a result of the disruption in the milk distribution chain caused

[4] Even the most liberal countries maintain some restrictions (see Désesquelles, 1997, for a comparative table of legislation in 11 European countries; see also Frydman and Martineau, 1998, for their table comparing legislation in 15 European countries).

[5] Generally speaking, wars have been key factors in tobacco adoption: the American Civil War, Wellington's campaign against Napoleon in Spain, and the Crimean War for the British, but, for Western countries as a whole, the Great War was the most decisive step. Annual average per capita cigarette consumption doubled during World War I, for example (Ravenholt, 1990).

[6] SEITA was actually a cigarette and match production department of central government before becoming a national corporation in 1984, and then merging with the Spanish *Tabacalera* to constitute the European limited company Altadis in 1999.

[7] Prohibition was introduced in 1919 by the 18th Amendment to the Constitution and required a new amendment, the 21st, to bring it to an end in 1933.

by the Liberation (Henry, 1948; Norvez, 1990). Not only are farming and imports closely supervised in developed countries today, but basic foodstuffs are increasingly industrially processed through an entire agri-foodstuffs production process, the health aspects of which are increasingly microregulated, supervised, and controlled, to prevent the risks of biological contamination or to ensure that the increasingly varied range of inputs used in the finished product are safe to eat. Likewise, strict checks are operated on the marketing network (storage, transport, covered and open markets, supermarkets and small shops) and catering (from urban restaurants to school and company canteens). Food safety concerns now approach paranoia levels. France's listeriosis scare in the early 2000s is just one example of the media hype over a handful of deaths (fewer than in previous years) into a scandalous death toll.[8] The mad cow crisis was worse, but the scale of the measures taken on the basis of often misinterpreted if not questionable scientific evidence (Vallin and Meslé, 2001) and ultimately on the basis of the now-sacrosanct precautionary principle, speaks volumes for the demands on safety (and public health protection/enforcement) which our societies now make.

Public health protection/enforcement is also the starting point for another aspect of food safety: dietetics. Policies in this area naturally focus on health information and education campaigns (see below), but notwithstanding the steadfast opposition of powerful lobbies, some countries are also starting to take regulatory control measures. In the United States, for example, the federal Food Drug and Administration made labeling of processed fats in food ingredients compulsory in 2004, while in 2003 alone, 179 state government bills or resolutions were passed to cut consumption of fatty foods (Cérisola and Mistral, 2004). The French government's public health bill tabled in 2004 included proposals to prohibit placing soft drinks dispensers in schools, but the soft drinks industry lobby is still trying to prevent the bill's implementation.[9]

g. Road and Other Transport Safety

As for serious offenses and murder, protecting the public against road and other transport risks is the job of law enforcement in the ordinary sense. The growth in road traffic, especially in the aftermath of World War II, turned road safety into a major public health issue, firstly in the developed countries (Vallin and Chesnais, 1975; Chesnais and Vallin, 1977), but also nowadays in many developing countries (OMS and Banque mondiale, 2004). Granted, regulations and the police are not the only means of tackling this scourge—improved roads, tracks and vehicles, information and education campaigns, organized emergency services for traffic accidents also play their role—but law enforcement action in the ordinary and broader sense, including public health protection/enforcement, are in the first line. The minimum measure is to draw up a highway code, mark out public highways with signs to give it organized, practical application, and to have a compulsory driving license scheme whereby drivers can be required to display an ability to control their vehicle and a minimum knowledge of the rules that they must obey. The basic requirements have slowly been put in place in the different countries (the driving license was established in France in 1893, for example, at least for the Paris region,[10] but became compulsory in Belgium only in 1965). But above all, these provisions have advanced considerably with progress not only in road and vehicle technology, but also that of the means of control (police mobility, radar speed traps, etc.), and a heightened awareness of the danger. The British Road Traffic Act 1930 is probably the first detailed general regulation put in place here (Chesnais, 1985). Road safety measures made a big advance in many countries in the 1970s, partly as a result of the 1973 oil crisis:[11] lowering of speed limits, wearing of seat belts, increased visibility of vehicles (daytime use of headlights for motorcyclists), blood alcohol testing, etc. It also saw a rise in motrobike accidents, against which Japan, for instance, swiftly introduced draconian but effective measures.[12] Claude Levy (1980) reports that Japan banned the use motorcycles with engines over 700 cc (but not their construction for export), and introduced a new driving test for motorbikes from 400 to 700 cc engines that was so difficult that in the first year (1976) that only 4% of those who took it passed.

[8] The so-called soaring toll of 2000 resulted in 9 deaths and one abortion, compared to the average 20 to 40 deaths a year in the 1980s and 1990s (http://www-causfra.ined.fr)

[9] *Le Monde*, 16 February 2005.

[10] The ordinance of August 14, 1893, regulating the movement of "mechanized motor vehicles other than those used to operate licensed railways" stipulated that "driving shall not be permitted [. . .] except by those holding a certificate of aptitude issued by the chief police officer."

[11] Sharp oil price rises in many countries made drivers more aware of the economic as well as the health downsides of speeding (Vallin and Meslé, 2001).

[12] Claude Levy (1980) reports that Japan banned the use motorcycles with engines over 700 cc (but not their construction for export), and introduced a new driving test for motorbikes from 400 to 700 cc engines which was so difficult that in the first year (1976) only 4% of those who took it passed.

It is not only to users that safety measures apply. Vehicle manufacturers, too, are subject to a growing number of increasingly stringent standards, be it on driver vision and vehicle signaling, in-car protection of passengers (seat belts, anchor points for child seats, etc.), shock absorption, or speed (cruise control), not to mention antipollution standards whose objectives clearly go beyond road safety alone.

On a more general note, transport policy clearly has an impact on road mortality. Putting road before railway development (or vice versa), or even (which more or less amounts to the same thing) public before private transport, increases or curbs road traffic and, ceteris paribus, the accident rate. But regulation of the type of vehicles or goods carried offers a range of intermediate measures (no through road for flammable materials, piggyback rail-road transport compulsory under certain tunnels, etc.).

The Scandinavian countries and the United Kingdom took the earliest steps to restrict and check this modern scourge (Chesnais, 1981, 1985). France makes periodic attempts to make up for its late start. Recent new efforts have been made to step up road safety. Paradoxically, it is in the world's least developed countries, where there are fewest vehicles on the roads, that the road death tolls are highest for want of sufficient resources to put in place high quality road safety (WHO and World Bank, 2004).

The first attempts to harmonize traffic safety rules at the international level date back many years. The United Nations and WHO were behind the adoption of two international conventions in 1949 (on road traffic, and on signs, signals, and markings). The Vienna Conventions currently in force date from 1968 (Nations unies/United Nations, 1968a, 1968b). A World Road Safety Day was held in 2004 (WHO and World Bank, 2004).

Other means of transport are also restrained within the confines of regulatory safety measures: the railways, aviation, and waterborne transport (marine or waterway, commercial shipping, or pleasure boating) are very tightly regulated. Enforcement of these regulations is greatly facilitated by the fact that most relate to public transport. For one thing, there is much more public clamor for high safety and surveillance standards (as far as zero-risk level), while for another, there are many fewer transport units to be controlled, and responsibility for their use is much more centralized.

h. Civil Emergency Preparedness: Fire, Floods, Natural Disasters, Terrorism

Another basic function of the modern State is to protect the populace against fire, floods, natural disasters, and the like. This is obviously an age-old concern, at least at the local government level, but the public has been increasingly demanding in this area. So-called civil defense services against fire are long established; for example, in France, the organized fire service was first established during the Revolution (Dalmaz, 1998). In most cases fire-fighting services are part of the armed forces. They are today highly organized and well equipped in the developed countries, where regulations to minimize risks (from the prevention of forest fires to electrical and gas safety standards, etc.) are also well developed. Protection against natural disasters is also largely a matter of regulatory control measures, especially in building standards (earthquake precautions, prohibition on building on floodplains, etc.), let alone the lessons recently learned in France from the consequences of the summer 2003 heat wave, which killed about 15,000 people, provoking a large political debate on the health services' capacity to face this type of event.

i. Tackling Pollution

Regulation and enforcement are also essentially the means by which authorities seek to protect populations against the health-damaging effects of air, water, soil, and other kinds of pollution. The rules are obviously directed first and foremost at the businesses responsible for these problems (from 19th-century factory smoke to modern industrial emissions, pollution by agricultural inputs, and the problems of radioactive waste), but also at the public (domestic sewage disposal, selective sorting of domestic refuse, partial or total driving bans in towns on specific high air pollution days, tackling river and beach pollution, oil spills, etc.). These modern antipollution policies clearly follow the efforts made throughout history (mentioned earlier in Sect. 1 of this chapter), to protect populations from the risks of infection from wastewater or drinking water (sanitation, drinking water supply). While these efforts have now largely paid off in most developed countries, much still remains to be done in many developing countries. It also bears pointing out that the protection of populations from pollution far transcends the sphere of public health protection/enforcement alone and is a major economic issue.

j. Specific Protection of Particular Population Groups

Public health protection/enforcement also includes specific protective measures for particular population groups. A whole area of this policy has gradually formed in Europe in reaction to the damaging effects of the industrial revolution on the health and survival of workers. Under pressure from trade unions and humanitarian organizations, the State has laid down an increasingly complete corpus of protective rules in this area on night work, working hours, health and safety at work, heavy work, and so forth, often making specific provision for children and women, in particular pregnant women. At the same time, special corps of labor inspectors have been established. Such rules often still fall well below these protective standards in developing countries, where child labor is still common practice, to the detriment of both their health and education. In India, for example, the issue is a major talking point, with child laborers themselves increasingly making a vocal stand, and receiving international media coverage in recent years.

On a more general note, specific public health protection/enforcement measures have been taken for mothers and children in most countries, in particular pregnant women, newborns and infants.

k. Pharmaceutical and Medical Safety

Public health protection/enforcement also aims to protect the populace against the various possible ways in which medical activity itself may go awry. Not only does it set out in general terms to contain the dangerous activities of charlatans of all kinds[13] by regulating access to the different professions (doctors, pharmacists, paramedical occupations, midwives, nurses, and health care assistants) and imposing on them, either directly or through likewise regulated professional organizations, appropriate codes of professional conduct, it also sets out to control the validity of procedures and clamp down on fraud. The pharmacovigilance system is particularly complex but clinical procedures are themselves increasingly subject to at least ex post control through the justice system to which discontented patients are increasingly resorting. In the United States in the 1970s, compensation amounts awarded by the courts to the victims of medical negligence (malpractice as much as maltreatment) were so high as to necessitate a substantial increase in insurance premiums (Robin and Nativi, 1987). Disproportionate as court cases may at times be, especially where politics are involved, as happened in the infected blood supplies affair in France in the 1990s (La documentation française, 1993), the issue is a live one. This case clearly underscores the key role that the authorities can and must play in ensuring that medical professionals exercise their medical skills properly. Conversely, by deliberately disregarding the known consequences of traditional commercial blood collection for spreading HIV/AIDS, China's political authorities gave this killer disease an opening that should never have been allowed to happen (Beach, 2001; Shan *et al.*, 2002).[14]

2. Medical Prevention

At the intersection of public health protection/enforcement and medical activity proper lies an entire sector of public health action—that of medical prevention. This can be distinguished between primary prevention, directed toward preventing the occurrence of diseases (vaccination, combating disease-carrying vectors, reducing risk factors) and secondary prevention, directed toward early detection and elimination of diseases before they become full-blown (screening). Both primary and secondary prevention can equally well address infectious and chronic diseases: for example, vaccination against measles, eradication of malaria-carrying mosquitoes, and tackling the risk factors for heart attacks fall within the sphere of primary prevention, while screening for tuberculosis or cancer are secondary prevention concerns.

a. Vaccinations

Vaccination is and always has been a major instrument of health policy ever since the work of Jenner and the discoveries of Pasteur. Most of the diseases that can be prevented by vaccination are under control in the developed countries, but vaccination policies remain a huge issue for the developing countries, where infectious and parasitic diseases remain major causes of death (whether they are vaccinable diseases like measles, or others, like malaria and HIV/AIDS, which cannot yet be immunized against), as well as in developed countries where a fall-off in vaccination uptake could have far-reaching consequences.

Compulsory Vaccinations and Vaccination Campaigns

This is an area where public health protection enforcement strategies (e.g., in countries where vaccination is made compulsory for all or for specific population groups) and enabling policies for medical prevention are inseparable.

Vaccinations against some diseases are compulsory in most countries. In France, for example, this is currently the case for diphtheria, tuberculosis (BCG), tetanus, and poliomyelitis (compulsory vaccination against smallpox was dropped in 1982 after the WHO officially declared the disease to have been eradicated). The list of compulsory vaccinations varies

[13] Or at least those which at a point in time are considered as such, for the lessons of history are that in certain circumstances, the quacks of one period may be recognized as another period's saviors.

[14] See, e.g., an article published in *Le Monde* of 2 December 2003.

from country to country, as do the means of enforcement. In France, failure to vaccinate is a criminal offense, but rarely do the authorities go so far as to prosecute. Compliance is ensured mainly by the fact that vaccination is done in early childhood and is required for enrollment in school. The natural counterpart is that it is offered free of charge in various public and private offices.

But as well as compulsory immunization, vaccination information and promotion campaigns play a very great role everywhere. In the influenza period, for example, vaccination is offered free of charge for the individuals over the age of 65 in many developed countries (especially Britain, Australia, New Zealand, and Canada) (Hannoun, 1999; see also the websites of the different health ministries).

The Eradication of Smallpox

The history of the eradication of smallpox puts the strategic issues of a vaccination policy clearly in context. The pre-Jenner treatment of variolization carried a high risk, but the devastating effects of the disease were such that the individual or collective risk was worth taking. Jenner's vaccine inoculation and its subsequent improvements reduced the risk to almost nothing, gaining full acceptance for the compulsory vaccination strategy. The situation was very different when the WHO decided to launch its eradication campaign in 1966. Except in those very few regions where the disease remained endemic, most countries no longer experienced more than the odd case or small sporadic imported outbreaks, such that in Europe and the United States, for example, even the very low vaccination risk had become less tolerable (i.e., the risk of post-vaccinial encephalitis could be higher than that of contracting smallpox), and public opinion found increasingly less justification for compulsory vaccination whose only remaining purpose was to protect travelers and prevent the recurrence of epidemics (Hannoun, 1999). The WHO therefore sought a new justification for its mass eradication campaign: that of attempting to stamp out all risk of infection for all humankind forever, thereby rendering vaccination superfluous. Whereas there had been an estimated two million-odd cases of smallpox in 1967, highly specific identification of disease areas and the systematic vaccination of all the populations concerned, especially in India and Africa, brought the annual number of cases rapidly down until the very last reported case in 1976. To satisfy itself of the total disappearance of the disease, the WHO offered a $1,000 bounty to anyone reporting a new case. No real case was ever reported, and the WHO declared smallpox officially eradicated in 1980.

The Expanded Program on Immunization

Building on this first great success, the WHO set up the Expanded Program on Immunization (EPI) in 1980 with the aim of immunizing every child in the world against six diseases responsible for most developing country mortality: polio, measles, diphtheria, whooping cough, tetanus, and tuberculosis (Jones *et al.*, 1985). Vaccination of pregnant women against tetanus, to protect their unborn children from neonatal tetanus, was soon added to the list. The aim was to achieve 100% vaccination coverage worldwide by 1990. Unlike smallpox, however, these diseases remain endemic or fiercely epidemic (measles) in large areas of the world, and eradication any time soon remains fairly unlikely, apart from polio. Eradicating this disease, where major gains have already been made, is one of the official goals to be achieved by 2005 (see, in particular Table 107–3). Major resources have been put into this, not just by the WHO itself, but also by UNICEF,[15] the UNDP[16] and the World Bank, as well as through many bilateral cooperation programs and the work of countless non-governmental organizations (NGOs). The outcomes are far from negligible (Desgrées du Loû, 1996; Moulin, 1996), and vaccination policy remains of great strategic importance, not just for the diseases that can be vaccinated against,[17] but also and perhaps especially, for those that cannot.

b. Tackling Vectors

Some nonvaccinable diseases are transmitted to humans by vectors, and action can be taken here to avoid infection. This is particularly so with many tropical infectious or parasitic diseases, foremost among which is malaria.

Malaria—Seemingly Ineradicable

Here again, an attempt can be made to reduce the incidence of, if not eradicate, the disease by drastically reducing the vector population (or, to the same effect, reducing that population's contact with humans). If the parasite cycle can only complete itself by being harbored in the human organism, it will be enough to stop it remaining in situ for a sufficiently long period of time, without having to completely eradicate the vector population. That is more or less what is being done with malaria, even though it seems that one of

[15] United Nations International Children's Emergency Fund, established in 1946 and renamed United Nations Children's Fund in 1953, retaining the same acronym. It is headquartered in Geneva.

[16] The United Nations for Development Fund, set up in 1965 and based in New York.

[17] There is, in particular, renewed interest in BCG to tackle opportunistic diseases in AIDS.

[18] The major marshland reclamation schemes carried out prior to

the main vector species in Africa, mosquitoes of the *Anopheles funestus* genus, attack various animal species as well as humans (IRD, 2000). With no hope of a vaccine, the argument for a WHO-led world eradication strategy was given added weight by the success achieved in the 1940s by European countries where malaria was endemic. These countries moved from basic control of infected areas by marshland reclamation to eliminating the disease through the use of DDT.[18] In fact, the success of the WHO's first attempt in this direction produced hopeful results: in Sri Lanka, after a mass campaign to eradicate mosquitoes by systematic DDT spraying, life expectancy made a spectacular leap from 46 years in 1945–1947 to 60 years in 1954. Notwithstanding that we now know that the decline in malaria was not the only factor in this spectacularly sharp mortality decline[19] the disease had then almost vanished from the island (Molineaux, 1985). But, to an even greater extent than for vaccination, attempts to repeat the operation on anything like a large scale met with much less success, especially in Africa. Not only do the different species of *Anophele* (the vector) or *Plasmodium* (the parasite) not have identical patterns of behavior, but some mosquitoes are becoming DDT-resistant, and some mutant parasites have become insensitive to the preventative action of quinine. Despite the great hopes of the 1980s, there is still no vaccine in sight (Molineaux, 1985).

Other Hopes and Attempts

After malaria, schistosomiasis is surely the second most rife serious parasitic disease in the world. Here, too, WHO-backed systematic campaigns were mounted against the disease vector (a small aquatic snail of the planorb genus). These campaigns obviously went hand in hand with environmental clean-up and treatment for those infected, but also included molluscicide application in the water. The success of such campaigns can have a big impact on the health and mortality of developing world populations, especially in Africa. As for malaria, however, the aim of eradication is still far from being achieved.

Many more examples can be given of attempts to eradicate various tropical diseases in the same basic way as the attack on malaria in Sri Lanka. Few have had any notable success. Aside from the same technological problems as for malaria, one main factor in all the failures is the perennial problem of mounting large-scale operations of this kind in what is not necessarily a conducive economic, social, and cultural context. This strategy of large-scale vertical sectoral programs, indeed, was largely discredited at the WHO Alma Ata conference in the late 1970s.

c. Tackling Risk Factors

There is a fourth important aspect to medical prevention: action on risk factors. The idea is to prevent disease by attacking the factors that contribute to its development. This category obviously includes all road accident prevention measures beyond the public health protection/enforcement mentioned earlier (road quality, vehicle roadworthiness, transport policy, etc.). Here, primary prevention of cancer and cardiovascular diseases is considered.

Cancer

Where cancer is concerned, a distinction must be made between risk factors that derive from the organization of the economy and society and act on groups, and individual lifestyle-related ones (Flamant, 1976: Hatton *et al.*, 1985), even though the two often interact. Occupational risks are an area where policy action is paramount. They are also quite extensively covered by a combination of some of the public health protection/enforcement measures referred to earlier in this chapter and employer liability. But more generally, prevention here is also the focus of ongoing negotiations between employers, the state, social welfare agencies, and trade unions in light of the evolving frontiers of knowledge. Asbestos is a recent example, which also goes beyond the strictly work-related setting, since this versatile substance has found many household (not least in buildings) as well as workplace uses.

The individual factors are directly related to individual lifestyle, especially consumption patterns (of food, alcohol, tobacco, medicines, etc.). Aside from the food safety measures and the regulations on smoking and drinking mentioned earlier, the health strategy here is essentially about educating the public.

Cardiovascular Diseases

Secondary prevention of cardiovascular diseases, sometimes directed toward the same risk factors as cancer (smoking and drinking, e.g.) is traditionally a major plank of developed countries' health policies. Lester Breslow (1985) distinguishes programs directed at isolated factors from multifactorial programs: tack-

1940, such as the Pontine Marshes just outside Rome, had helped reduce the incidence of malaria. It was not really possible to eradicate it until DDT came on the market in 1943. The very last reported case in Europe, however, was in Macedonia in 1976 (retrieved July 20, 2004, from http://www.chez.com/demosolidarites/adiconfindex.html).

[19] In 1960, Harald Frederiksen went so far as to claim that the life expectancy gain achieved in Sri Lanka was due only to rising living standards, an argument easily refuted by Peter Newman (1965).

[20] René Laënnec (1781–1826), French physician, inventor of the

ling smoking, controlling cholesterol levels, and monitoring and controlling high blood pressure. Multifactorial programs seek to reduce all the main risk factors at once (the three just cited as well as others, like obesity). In either case, there are two possible strategies: either using various parameters to identify the most at-risk share of population toward which to gear the entire program (medical approach), or gearing the program toward the population as a whole, wagering on the knock-on effect that universality can produce (community approach), as with the North Karelia program, which is considered more at length below. Another case in point is the program to tackle cerebrovascular diseases launched in Japan in 1969 (Goldman and Takahashi, 1996), focused in particular on a reduction in salt intake.

d. Medical Screening

Alongside vaccines or other means (like tackling vectors) of preventing disease transmission, screening for medical conditions can help prevent a fatal outcome or relatively chronically incapacitating complications. This is often referred to as "secondary prevention." It is another major approach of preventive medicine, and many community health measures are based on this strategy.

Secondary prevention is a strategy that was deployed to particular effect in Europe's recent past in tackling tuberculosis. From Laënnec's stethoscope[20] to the controlled use of x-rays[21] and, finally biological tests (skin test), increasingly early and accurate diagnoses became possible, enabling early treatment of patients and a greater likelihood of recovery with fewer serious after-effects (especially since the discovery of antibiotics). Systematic screening campaigns were organized (in particular through occupational, school, and armed forces health services). It was this strategy much more than the BCG vaccine, which reduced the incidence of tuberculosis to the marginal position it now occupies in developed countries (Vallin and Meslé, 1999).

Screening is now a major weapon in the preventive medicine armory against many medical conditions. Three particularly important areas today are cancer, cardiovascular diseases, and congenital disorders.

Cancer

Anticancer policy is the most important example because, let alone the difficulty of acting on risk factors, there is almost no way of preventing the disease (in the strictest sense). What makes screening particularly beneficial is the close correlation between the success of treatments and how early the diagnosis is made. Also, the importance assumed by cancer in developed country mortality, especially since the huge decline in cardiovascular diseases, has made it a major issue of public health policies in developed countries. All have adopted more or less comprehensive cancer screening programs: screening for cervical cancer through cervical smears, breast cancer through mammography, and prostate cancer through rectal examination and blood PSA (prostate-specific antigen) levels is most commonly found in age- and sex-specific targeted programs. Systematic screening for stomach and colon cancer is still an unresolved issue because of the cost and the more invasive nature of the procedures. Screening for such diseases tends to be restricted to at-risk groups. In Japan, for example, systematic screening campaigns for stomach cancer have been running since the 1960s in certain high-incidence areas (Kitagawa and Hishami, 1970; Oshima *et al*., 1979, cited in Hatton *et al*., 1985).

Cardiovascular Diseases

There is also a varied range of screening programs, some more or less systematic, others targeted to at-risk categories, for cardiovascular diseases, especially those with the highest mortality or invalidity rates. Coronary angiography means that atherosclerosis can be screened for and treated, or a shunt fitted. Still in the field of coronary diseases, ultrasound scans can predict aortic aneurysm, while for cerebrovascular diseases, Doppler ultrasound can detect carotid artery damage.

Perinatal Diseases

The perinatal period also offers wide scope for screening for many disorders, and in all developed countries perinatal screening has been a major contributor to the sharp decline in infant mortality and a reduction in congenital disorders. This is much of what the compulsory antenatal and postnatal visits in many countries are designed to do. In extreme cases, antenatal diagnosis enables therapeutic abortion to be carried out, but also, in the same way as neonatal screening, it allows timely treatment for disorders that otherwise would be fatal or severely incapacitating. In France, for example, screening at birth for phenylketonuria (Guthrie test), introduced in the 1960s, has helped prevent a large number of disabilities simply by the use of a controlled diet. The lineup of means also includes a focus on Rhesus factor incompatibilities (by means of a parental blood test), the risks of

stethoscope and indirect auscultation.

21 Radioscopy and radiography came into large-scale medical use from the first World War (Bordry *et al*., 1992).

22 A reform undertaken as part of the care in the community

German measles, toxoplasmosis or HIV/AIDS in pregnant women, and social security reimbursement of amniocentesis for pregnant women over the age of 38 to detect Down's syndrome.

Medical Check-ups

More generally, age-group-specific check-ups for a comparatively wide range of possible diseases are organized in some countries.

Unfortunately, such a screening policy is much costlier and harder to organize than vaccination and tackling certain infectious disease vectors. In particular, it requires medical and laboratory facilities that most developing countries cannot provide for the entire population, and can only be implemented to the extent permitted by the general economic and social development of each country.

3. From Prevention of Disease to Treatment of Patients

Prevention is better than cure, it is said. And so it is, but when the disease is present, an attempt must be made to find the means to cure it. Should health policies that are closely interlinked—not just in establishing and enforcing health rules, but also in the implementation of medical prevention programs—also be involved in the treatment of diseases (other than by laying down rules of professional conduct)? It is a supremely philosophical and ideological question. Any modern state, however liberal its foundations, can easily accept full responsibility for public health protection/enforcement, and can recognize without too much difficulty that it must be responsible for at least a part of medical prevention, for its foremost duty is to the community, which does not necessarily always coincide with the individual interests of the persons involved. The public authorities have the right to decide on measures of public interest, and they must also bear the cost. Arguably, treatment of disease may be a different matter altogether, being expressly in the interest of the patient (or their family), and for which only the person concerned is competent and responsible: he is free to choose whether or not to seek treatment, and whether or not to resort to medical services, but he must also pay the cost, either directly or indirectly by taking out an insurance policy. But nowhere in the world is there not at least some collective provision to foot part of the bill for the treatment of diseases.

In theory, moreover, that can be done imperceptibly, at no added expense where, for example, the State intervenes only by creating a compulsory insurance system. There is then no need to levy a charge on the public budget. But, where a compulsory insurance system is in place, the premiums paid by the insured persons may not be entirely of the same nature as the taxes levied by the tax authorities, but can be equated to them, and increase the sum total of what are increasingly called compulsory levies. The value and method of funding of the cost of treatment can therefore be said to be a public policy decision. A timely definition is also needed of the risks to be covered, the premiums that can be reasonably charged, and the rates that service providers can charge. The increasing tendency is for the authorities to regulate all these areas in detail, even if in consultation with the social partners.

Beyond protection against collective risks (which is the justification for public health protection/enforcement and medical prevention), authorities have also been drawn into making provision for the treatment of patients to address at least, three different types of objectives: (a) to contain, reduce, or eradicate what are considered priority diseases; (b) to ensure the best possible health for a specific population group; and (c) to reduce inequalities. Multidisciplinary programs exist to address these objectives, covering both patient treatment and disease prevention. A few examples follow.

a. Tackling Priority Diseases

Tackling specific diseases may be seen as a priority for a variety of political, social, ethical, and other reasons. Anything that may reduce the scope of disease, including treatment provision, then becomes a public health issue. Such was the case in the past with infectious diseases like tuberculosis, and AIDS is a more recent, broadly comparable issue. Among degenerative and chronic diseases, some aspects of mental illness and cancer also may have relevance here.

Tuberculosis Then and Now

Tuberculosis emerged as a societal issue fairly early on, in the second half of the 19th century. A social disease, not merely as being chiefly one of poor families living in unsanitary housing and/or enduring poor physical working conditions, but also because these different harmful conditions were closely tied to the social changes wrought by the industrial revolution, urbanization, and free-market capitalism. Also, as tuberculosis was a contagious, stigmatizing disease, all possible means had to be used to minimize its incidence. This is why, in addition to the preventative steps mentioned earlier, national and local authorities in most European countries intervened quite early on in care provision, such as by building sanatoriums that offered free (or almost) care and treatment to sufferers (Faure, 1994, p. 217).

AIDS

More recently, AIDS came similarly onto the official agenda soon after appearing in the 1980s. Some of the reasons were different, not least the shock of forecasts of an unprecedented explosion and doubts about how quickly an effective vaccine could be developed. But others were of the same kind, like the opprobrium attached to a sexually related disease which itself produced a backlash political reaction in favor of people living with HIV/AIDS coupled with the seeming imperative of preventing infection by sufferers belonging mainly to fairly marginal social groups (homosexuals, drug abusers, blood transfusion recipients) spreading to the population at large. There again, the main focus of effort had to be not just prevention but also (if possible) curing sufferers. Huge funding was allocated to research, and once the (very costly) first drugs became available, they were issued in various ways either free of charge or subsidized to sufferers in many developed countries. One tragedy for the developing countries, and especially sub-Saharan Africa, is that these medicines are so costly and the sufferers so numerous that this strategy is not only beyond their reach, but also beyond the reach of many international programs. The other misfortune is that the huge research effort put forth by the rich countries has focused more on medicines than a vaccine, which would have been far more suited to the problem faced by widely infected countries. A glimmer of hope has emerged from the recent legal battle that has enabled some countries like Brazil, India, and Thailand (ONUSIDA, 2004) to start producing antiretroviral drugs without having to pay royalties to the pharmaceutical companies that developed them or others, like South Africa, to import them cheaply. Nevertheless, continuous lifetime treatment of all people with HIV/AIDS remains beyond the financial reach of most of the worst affected countries.

Severe Mental Disturbance

Very early on, the treatment of mental illness was seen largely as a public responsibility—less at the start in the interest of the sufferer (for whom treatment mostly comprised incarceration), than in that of society, which had to be kept safe from the dangerously ill (or simply unwanted). But just as imprisonment gave way to psychiatric hospitals, these gradually evolved into care provision, and this new function eventually came to substantially outweigh the old. In many countries, because of historical and more modern reasons (inability of the psychiatrically disturbed to cope, the unfair burden that their illness places on their families, etc.), psychiatric care has become a more or less full public responsibility. That principle remained intact when in the 1970s and 1980s many countries followed the Italian lead and completely rethought the ways people with psychiatric disorders were treated by switching the focus away from psychiatric hospitals onto treatment in the community (Meslé and Vallin, 1981), or doing away with them entirely, as in Italy, to retain only slimmed-down specialized provision, or ordinary psychiatric wards within general hospitals.[22]

In France, whether treated in the community or in residential care, severe mental disturbances (psychoses, severe personality disorders) form part of the list of chronic and costly illnesses whose treatment is always fully reimbursed by social security, and where prolonged sick leave arrangements are made.[23]

Cancer

Cancer also occupies a special place in modern societies. Most often the cause of prolonged and chronically debilitating diseases, it was long regarded as incurable, or to be treated by painful and costly procedures of uncertain outcome. But the rapid decline in infectious diseases left it as one of the two main causes of death in all high life expectancy countries, second only to cardiovascular diseases. The sharp drop in the latter from the 1970s made it more imperative still to tackle a disease that strikes a chill in contemporary society. Treatment in some countries became a public responsibility to varying degrees. In France, for example, not only do all cancer sufferers qualify for the chronic illness scheme, but new, specialized cancer

movement inspired by Franco Basaglia, culminating in Act 180 of 1978.

[23] More simply called the "chronic illness" scheme, this in fact goes far beyond psychiatric illnesses, in providing free hospitalization and medical care (no self-pay component) for officially prescribed illnesses (currently 30), or "any disorder requiring prolonged and particularly costly treatment," while daily benefits for work incapacity lasting more than six months are maintained (Vallin, 1979b). The 30 prescribed diseases are incapacitating cerebral vascular accident, medullary insufficiency and other cytopenias, chronic arteriopathies, complicated schistosomiasis, severe cardiac insufficiency, active chronic diseases and cirrhosis of the liver, severe primary immunodeficiency and HIV infection, types 1 and 2 diabetes, neurological and muscular disorder and epilepsy, hemoglobinopathies and hemolyses, hemophilia and severe blood-clotting disorders, acute high blood pressure, coronary diseases, acute chronic respiratory disorder, leprosy, Parkinson's disease, hereditary metabolic diseases, cystic fibrosis, chronic renal disease and nephrotic syndrome, paraplegia, polyarteritis nodosa, lupus and scleroderma, severe rheumatoid arthritis, psychosis or acute personality disorder, hemorrhagic rectal colitis, multiple sclerosis, severe ankylosing spondylitis, complications of organ transplant, progressive tuberculosis, malignant tumor or hematopoietic disorder.

[24] "Cancer is a real national tragedy that demands a big effort in

hospitals were built more or less on the tuberculosis sanatorium model, the most renowned of which is surely the Institut Gustave Roussy at Villejuif in the Paris region. Very recently again, the priority focus put on this disease was reaffirmed at the highest level of government in an official statement by the President of the Republic.[24]

b. Guaranteeing the Best Possible Health to Specific Population Groups

For a variety of reasons, it also became seen as a government responsibility to deliver the best possible access to health, including curative treatment provision, to specific population groups. Typical of these groups are workers, pregnant women, children, and the elderly.

Workers

Since the 19th century, the health of workers (essentially manual workers at the time; the preferred present-day term would be *employees*) has been a priority on government agendas, either from concern to preserve the labor force which underpinned the industrial revolution, or from ultimately yielding to workers' demands, especially in regard to the risks run simply by working. In France, for example, the introduction of a right to compensation for a work-related accident or occupational disease brought into being the first instance of comprehensive coverage of certain health care provision. Although it had existed in theory since the Revolution, based on the very general Civil Code provision on personal liability,[25] this right of workers against their employer had to be clarified and strictly regulated by an entire corpus of specific legislation to be really exercisable, and became universal only with the development of compulsory insurance systems. At the end of the day, therefore, it is because of a strong and continual public policy commitment that employees in many countries have full coverage of all work accident and occupational disease treatment costs, plus specific benefits for disruption of employment, and temporary or permanent work incapacity. Alongside the development of this special legislation, the State, local authorities, and mutual insurance organizations also established different occupational health services that play an important preventive role, but that may also be involved in actual care provision.

Pregnant Women

Pregnant women are another population group that benefits in many countries from specific legislation creating an entitlement not just to preventive measures for them and their unborn children (the key preventive role played not just by maternity leave but also special labor law rules merits a passing mention), but also to certain care and treatment provision. Even before pregnancy occurs, medical care (e.g., infertility treatment, in particular medically assisted reproduction) is provided free by the health system in many countries; likewise treatment to bring a high-risk pregnancy to term, or the treatment of certain women's diseases that may affect the child. Finally, specific provision has been made for childbirth and the postpartum period in many countries. This special policy on pregnancy bears a clear relation to policies to support fertility in countries where it is felt to be too low, France being a textbook example. But, in developing countries especially, it is also closely connected with the recent rise in reproductive health programs under the impetus given to this sector by the UNFPA[26] since the Cairo conference. The Cairo conference referred to here is the International Conference on Population and Development (ICPD).[27]

Children

Special provision also exists for children in many developed and developing countries, going well beyond prevention alone. In developed countries, a range of specific policies have been and still are being pursued, including the provision of free medical care or dietary supplements to combat malnutrition (especially through free or subsidized school meals). Most typical of the recent period, however, are the perinatal programs, to which most of the measures referred to in the previous paragraph could have been tied in. These programs were largely instrumental in bringing infant mortality down to levels that were unthinkably low, even a few short decades ago, through an unprecedented reduction in neonatal mortality. They rely heavily on the preventive measures mentioned earlier, but also most often include free treatment of the diseases, disorders, and abnormalities detected by

terms of research, prevention, screening, and treatment, including psychological counselling" (Jacques Chirac, statement of 14 July 2002).

[25] "Any act whatever of man, which causes damage to another, obliges the one by whose fault it occurred, to compensate it" (*Civil Code*, article 1382).

[26] The United Nations Fund for Population Activities, UNFPA, headquartered in New York, was set up in 1969. It is now called the United Nations Population Fund.

[27] International Conference on Population and Development (ICPD) Cairo, 1994 (see, in particular, United Nations, 1998/United Nations, 1995).

[28] Finding that the Ministry for Public Health was working only

screening. This is more easily achieved since this type of treatment is often required as a matter of course in the immediate postnatal period, when the newborn is still in a hospital environment (maternity clinic/ward) and in many developed countries, the period of hospitalization for childbirth is fully paid for by the health service, albeit increasingly foreshortened.

In the least-developed countries, childbirth is seldom so medicalized, and the most urgent infant mortality problem remains post-neonatal mortality, where infection and malnutrition play a bigger role. It is on these aspects of child health that governments, as well as international bilateral and multilateral cooperation have focused. The major WHO and UNICEF campaigns on prevention of early childhood infectious diseases have already been mentioned. The early childhood programs developed by these agencies, along with others like the UNDP (United Nations for Development Fund), the FAO (Food and Agriculture Organization, a United Nations agency, set up in 1945 and headquartered in Rome), and the WFP (World Food Program, set up in 1963), also include many curative aspects, like treating diarrhea by administration of oral rehydration salts (Victora *et al*., 2000), and combating malnutrition (Taylor *et al*., 1968; Kielmann and Taylor, 1969; Beghin and Vanderveken, 1985).

The Elderly

The health of the elderly is another area where policy action in some countries has gone well beyond prevention. In developing countries, where this age bracket still makes up only a very small share of the population, and where traditional intergenerational ties remain very strong, governments tend not to step in, leaving most elder care provision to families. In developed countries with older populations, by contrast, the matter has become a far-reaching societal issue. One point of interest in this connection is the exception made for the elderly in the United States, in an otherwise advanced general free-market approach to payment for health care. As a general rule, this is a cost to individuals and families, who are clearly free to insure against it or negotiate insurance with their employers. But, in an exception to this sacrosanct principle, Medicare was introduced in 1965 for individuals over the age of 65 and, in some cases, people under 65 with disabilities, and those on kidney dialysis (Card *et al*., 2004). Care provision under Medicare is neither complete (only hospital expenses are automatically covered, and Medicare offers only the possibility to insure the costs of medical care on a voluntary basis), nor reciprocity-free (it is available only to individuals with at least 40 quarters' worth of paid-up contributions, and their family members), but it is the only large-scale case in which the United States has created a compulsory insurance system and set the rules and rates. It illustrates how high the health of its older population is on the country's agenda.

Assistance to Victims

Finally, there is the issue of assistance to victims of accidents or attacks. First aid is generally provided by increasingly well-equipped emergency preparedness services. The consequences of road traffic accidents are generally provided for by insurance, broadly within the framework of compulsory insurance. Public help for the victims of natural disasters and attacks has recently developed considerably, even extending with growing frequency to the provision of psychological counseling.

c. Reducing Inequalities

The third big reason for increased policy involvement beyond prevention and into the sphere of actual treatment is the desire to reduce inequalities, be it international inequalities between populations in different countries of the world, or internal disparities within each country. Domestic efforts have mainly been made in two directions: geographical disparities and social disparities.

Reducing Internal Disparities

Geographical disparities in mortality have narrowed significantly in developed countries (see Chapter 55), basically as the result of policy action across all spheres with no overt health objective in many cases. This is because the State, whether centralized or not, generally seeks to extend its economic, social, and cultural development action in as harmonious a way as possible territorywide, thereby laying the foundations for a reduction of all kinds of geographical inequalities that contribute in some way to inequalities in health. Countrywide networks of medical provision more explicitly directed toward, and coupled with, public health protection/enforcement and preventive policy action that are likewise deemed to benefit the population as a whole, clearly contribute to deliver that aim. But some countries have gone further. Finding that some regions were still suffering from specific problems resulting in what was judged to be abnormally high excess mortality, some governments pursued a specific policy to bring the levels down. So, in 1956, the French Ministry for Public Health and Population commissioned INED to survey the causes of the abnormally high infant mortality rate recorded in the Nord and Pas-de-Calais departments (Girard *et al*., 1960), and then mounted a campaign toward local communities, business, and local authorities to try and stem the rise. A more defining case still is that of North Karelia in Finland (Puska *et al*., 1981, cited by Breslow, 1985). Toward the end of the 1960s, this 180,000-strong community had what may have been the highest coronary heart disease mortality in the world (Breslow, 1985). In the early 1970s, a wide-ranging program was launched to reduce blood pressure, cholesterol, and smoking, and to encourage early treatment of people with cardiovascular diseases, involving the entire organized voluntary social service movement and the network of health agencies. Five years later, in 1977, coronary heart disease mortality in the 30–64 age bracket had dropped by 13% among men and 31% among women (Breslow, 1985).

Social inequalities in mortality are also a focus of concern leading to specific public policy action on health. Having moved on from the time when the poor, like the sick, were perceived and treated more as a threat to society than a group deserving care and attention, they were offered first assistance, then specific health programs. France, for example, brought in statutory free medical assistance for the poor in 1893. The establishment of social security systems after World War II did not render these provisions obsolete, because many of the poor fell outside the scheme. They were simply reworked in 1953 into free medical help (AMG) providing free care to anyone living in France who could not afford to pay for it. But even when the universal health coverage (CMU) scheme was created in 1999, benefiting some 150,000 to 200,000 people living in France without social security entitlements, plus the 550,000 previously voluntarily insured, a significant number still fell through the net, as the entitlements were only for persons lawfully resident in France for more than 3 months. A state medical assistance scheme (AME) was therefore kept, and is the only means by which undocumented migrants and their families can access the health care they need. In 2003, 180,000 people were covered by AME.

Obviously, extreme poverty is not the only social inequality in health and mortality, and in the latter half of the 20th century, it was the rest of these inequalities that policy sought to address in many developed countries, by setting up near-universal social security systems. These will be considered further below.

Closing the Gap between Rich and Poor Countries

Policies to reduce international inequalities are of more recent date. They were mostly brought in after World War II, with the creation of the WHO, the rise of the Third World movements, and the establishment of international programs to reduce excess mortality in developing countries. Notwithstanding undeniable spot successes, however, the first tentative efforts based either on the roll-out of large-scale "vertical" programs to tackle specific high-incidence diseases (one of the WHO's main self-appointed missions), or the construction of major referral hospitals intended to energize the country's entire health delivery system (a key, but not always disinterested, strategy in bilateral cooperation measures), soon exhausted the possibilities. Generally, the efforts benefited only a small, privileged, mainly urban share of the population, leaving the bulk of the rural masses by the wayside. At the 1978 Alma Ata conference, drawing on the impressive Chinese experiment with "barefoot doctors," introduced during the Cultural Revolution,[28] the WHO took a radically different tack toward the primary health care strategy (WHO, 1978).

This new strategy was less about trying to eradicate a specific disease more or less completely than attaining decisive all-around health improvements for populations that had the least access to them. The aim was to "address the main health problems in the community, providing promotive, preventive, curative and rehabilitative services accordingly," while the Declaration of Alma Ata specifies that this includes, at the very least, health education, food supply and proper nutrition, supply of safe water and basic sanitation, and maternal and child health care that included family planning, immunization against major diseases, control of locally endemic diseases, treatment of common diseases and injuries, and provision of essential drugs.

Governments are responsible for refocusing their health policies on this new strategy, while international organizations and others in the cooperation community must provide technical and financial support in redeploying the assistance they provide to the health policies of poor countries toward these new objectives.

As obvious and essential as it may seem, this new strategy has encountered a series of obstacles on the ground. However "primary" they may be, these forms of health care are many and various, and costly when

for 15% of the country's population, Mao Tse Tung called in 1965 for a focus on medical work in rural society, as a result of which 150,000 medical workers were given basic training, and left the towns to exercise their skills in rural communities. These "barefoot doctors" were healthcare personnel who had received a few months of basic medical training, followed by advanced training courses. They were to provide stop-gap services to the maximum number of people, including prevention, basic treatment, screening, and referral (of difficult cases) to doctors or hospitals. They were either assigned to villages or formed part of mobile outreach units covering more sparsely populated areas (Selden, 1972).

[29] Except in three countries (Germany, the Netherlands and Ireland) where those on the highest wages were kept outside the

the target group concerned is in fact the bulk of the population. But also, both the main beneficiaries and the actors in the previous system were fiercely—if only passively—opposed to any major redeployment of existing resources. Finally, the rich countries gave no substantial new financial pledges at Alma Ata. Most governments naturally brought in (often through legislation) health policies that endorsed the primary health care strategy, but any attempts to really implement them were few and far between, and almost never produced a real shift in priorities, which might have produced the looked-for results. And yet it was this strategy (whether framed before or after Alma Ata) that has enabled countries such as Costa Rica (Rosero-Bixby, 1985), Kenya (Mosley, 1985), and the State of Kerala in India (Nayar, 1985), not to mention China, referred to earlier, to increase their life expectancies well beyond what their economic development might suggest.

A slew of initiatives have been taken in a bid to give fresh impetus to this strategy, one of the main ones being the Bamako initiative (Brunet-Jailly, 1997; Soucat *et al*., 1997).

4. All for All

In fact, the WHO's adoption of the primary health care strategy can be seen as the international reflection of a growing idea at the time that public policy at all levels should be directed toward delivering access to health to all. The priority focus on primary health care was essential only because the rural masses who made up most of the Third World populations had no access to standard health care provision. The basic aim was to offer a minimum access to health to everyone, without waiting on economic development to maximize access to traditional services. It was a first step toward "health for all by the year 2000," soon to be adopted as a WHO slogan.

From an entirely different angle, this same concern to see public policy guarantee health for all by all possible means was what prompted many industrial countries (especially the European ones) to shift away from leveling the playing field by helping the poorest social groups to access health care toward a collective assumption of responsibility for the health of all within a welfare state, where public health is only one of many elements.

a. Developed Countries: the Welfare State

The welfare state is a concept that extends well beyond health matters to the collectivist assumption of all major social risks (unemployment, cost of family obligations, old age) and needs (education, leave, etc.). We are concerned here, of course, only with health, which is nevertheless exemplary. Disregarding the Soviet system, which no longer prevails in any developed country, health expenditure is by and large paid for in three different ways, ranging from the "social contractual" protection (Dumont, 1998) applied in the United States, via the French-style social security system, to the State-funded system established in Britain after the adoption of the Beveridge Plan (1943).

The "Social Contractual" System

No further consideration needs be given to the American social contractual system, of which Medicare was cited above as the only exception to its free-market principles, other than to say that it is based on the freedom of individuals to insure themselves with private insurers against the risks of disease and accident, either through company-established or mutual programs operated by employers or mutual insurance organizations. It is a model very far-removed from that of the welfare state.

Social Security

The French-style social security system evolved as an extension of the scope of the social insurance plan devised in Bismarck's Germany in the 1880s, rapidly taken up in Austria (1882), Denmark (1891–1892), Belgium (1894), Luxembourg (1901), and finally France (1930). The big difference from the American free-market system is that social insurance is compulsory, although it was initially limited to employees on low incomes (below a statutory maximum). The big change made in the aftermath of World War II in most European countries that had the system was to extend its benefits to all employed workers.[29] In almost all countries, too, it was then progressively extended to farmers, shopkeepers, craft workers, and others. The system is thus based on across-the-board coverage by state-ordained compulsory social insurance, funded out of earnings-related contributions (the cost burden being shared by employer and employee), and jointly run by the social partners (employers' organizations and trade unions) under State superintendence.

Different specific social insurance systems may often be layered onto a so-called general scheme. In France, for example, the general scheme coexists with

system. In Germany, however, since 1970, high wage-earners can opt-into the system, which now therefore covers 92% of employees compared to 70% in 1969, while Ireland has recently moved onto a UK-style universal protection system (Dumont, 1998).

[30] WHO document WHA30.43, available at the item "Resolutions and Decisions" of http://www.who.int/governance/en.

[31] On which, for example, see "Health for all in the 21st century"

a civil service scheme and other category-specific schemes (agricultural, in particular). The health insurance funds contract with public or private care delivery partners (hospitals, fee-for-service practitioners, pharmaceutical companies, etc.), and, in consultation with the State, set the methods of remuneration and cost of medical treatment rates that patients must pay, as well as full or partial cost reimbursement procedures. The vast majority of citizens thus have a recognized right to health in the form of a right of access to medical care, leaving patients free to use the health care providers of their choice (Letourmy, 2000). "Vast majority" does not mean all, however. A share of the population still remains excluded. In France, for example, with one of the highest coverage rates, nearly 1% of the population remains outside the system (Dumont, 1998, p. 2). However, as mentioned earlier, safety nets marginally attached to the social security scheme exist for excluded groups (AME, CMU).

Universal Protection

The most radical change occurred in the United Kingdom in the aftermath of World War II, with the introduction of a universal protection system covering the entire population from the very outset, following the adoption of the Beveridge Plan and the setting-up of the National Health System (NHS) in 1946–1948. All citizens were automatically covered, without having to prove employment or a contribution record. Another feature of this system, which also covers the other fundamental social risks or needs, is that pension, unemployment, and other benefit amounts are in no way earnings related. The British National Health Service provided a template for the systems subsequently adopted in Denmark (1960), Sweden (1962) and, more recently, many others, including Ireland, Spain, Italy, and Greece. It is based on an essentially hypothecated tax-funded, predominantly public, care sector (but with a private sector component for those who can pay). It is free (or almost) to all citizens, but there is restricted choice of points of entry. In the United Kingdom, for example, patients must register on the list of a referring doctor, and can only change doctors in certain conditions. The referring doctor decides on the need for referral to a specialist or hospital treatment. The public authorities have direct control over funding, through budgets allocated to hospitals and fixed fees paid to doctors according to the number of patients on their lists. They also control the medical services provided in public hospitals.

Scheme Coverage May Be Deceptive

A compact publication by Ariel Béresniak and Gérard Duru (1997) gives detailed descriptions of the schemes in force in each of the 15 European Union countries, with a flow chart for each to clarify their maze of complexities. Figure 107–1 shows the comparative systems for the United Kingdom and France.

Notwithstanding the wide differences in the ways they work, these two major systems now provide near-universal coverage of the population for the main health care requirements (especially in-patient care), as Table 107–2 shows. This is the result of major progress made by countries that initially established only limited social insurance schemes toward either extending coverage to almost all the population, or becoming a universal state system. Only the United States stands out for the wrong reasons among OECD countries.

Universal (or near-universal) coverage of the population is not, however, necessarily the same as universal coverage of needs. Not all expenditures are fully noncontributory. Hospital expenses are best catered for, with more than 90% of the costs being paid for in European countries in 1998, barring the odd exception; out-patient care is less so, with rates often between 80% and 90%, and the cost of medicines still less, ranging in many countries from 50% to 75% (Dumont, 1998). Nor do recent trends show any improvement, as problems of funding the social protection system as a whole, and especially disease protection, have recently begun to throw the canons of the welfare state itself into question. The theoretically total coverage offered by the NHS in the United Kingdom should not be taken at face value. Government spending on the NHS is obviously not limitless, and the system's drawbacks—not least long waiting lists for treatment—are enough to drive many toward a very costly, but flourishing private sector.

b. Developing Countries: "Health for All"?

To take WHO founders at their word, it could be reasonably argued that the WHO was set up in 1946 with the task of delivering health to all, since the preamble to its Constitution states, "The enjoyment of the highest attainable standard of health is one of the fundamental human rights of every human being without distinction for race, religion, political belief, economic or social condition" and "The health of all peoples is fundamental in the attainment of peace and security and is dependent upon the fullest cooperation of individuals and states" (WHO, [1946] 1989). There could be no clearer affirmation that access to health for all was a matter of urgency and that this engaged the responsibility of States, and hence international organizations.

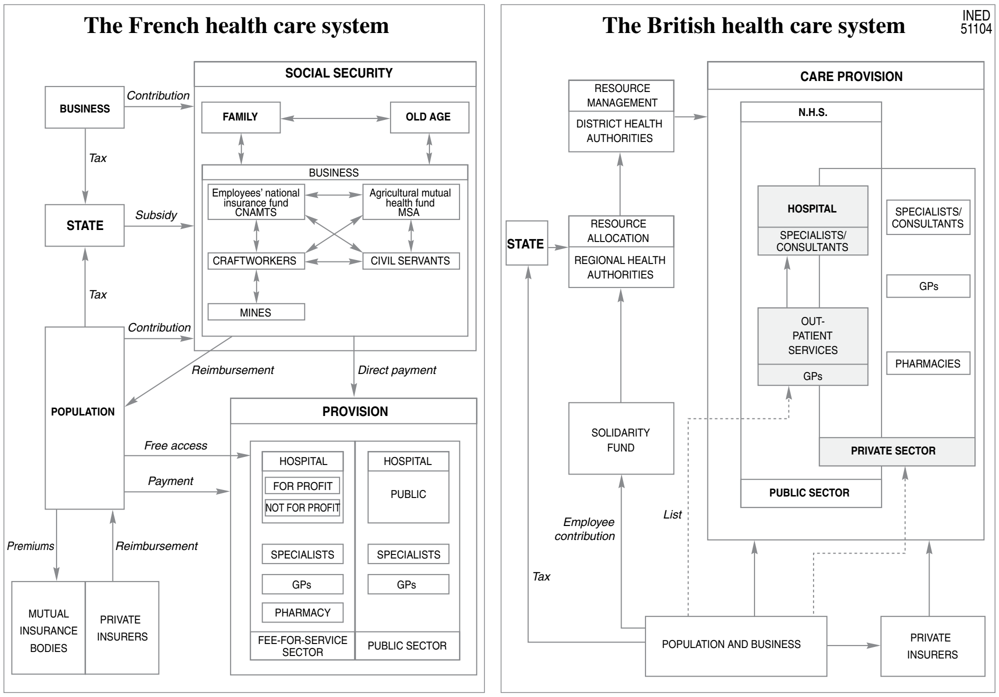


FIGURE 107–1 Flow charts of the British and French health care systems. (Source: Béresniak and Duru, 1997.)

TABLE 107–2 Share (%) of Population Covered for Hospitalization and Health Goods in OECD Countries, 1960–1998

Country	Hospitalization		Medical products	
	1960	1998	1960	1998
Europe				
United Kingdom	100	100	100	100
Sweden	100	100	100	100
Luxembourg	100	100	100	100
Finland	100	100	0	100
Denmark	95	100	95	100
Italy	87	100	87	100
Portugal	18	100	18	100
Greece	30	100	25	100
Switzerland	72	100	72	100
Ireland	85	100	30	40
Spain	50	99	50	99
Austria	78	99	78	99
France	80	99	76	98
Belgium	58	98	58	93
Germany (ex FRG)	84	92	84	92
Netherlands	71	73	71	62
Non-Europe				
Japan	88	100	88	100
Canada	68	100	5	34
United States	20	44	3	12

Source: OECD cited by Dumont, 1998.

Health for All by the Year 2000?

But the noblest intentions of this kind officially proclaimed at the highest level of the world's governing bodies are mere words until the practical programs needed have been decided and the means to carry them out allocated. In its earliest days, the WHO thought to take a "fire-fighting" approach through programs to stamp out major eradicable diseases with basic means, far from being able to deliver "the highest attainable standard of health" for all! For that reason, 30 years later, at the Alma Ata conference, the WHO sought to move closer to its aim with a more direct assault on the problem of poor rural communities denied access to existing health care provision, but also realistically by repositioning health in the more general context of the basic needs of life, and the more modest one of basic health needs. The decision of the 30th World Health Assembly, which was behind the Conference, set the aim of "the attainment by all the citizens of the world by the year 2000 of a level of health that would permit them to lead a socially and economically productive life."[30] This aim was a much less lofty ambition than the highest attainable state of health and indeed failed to deliver all the hopes it had

by the WHO Director General in the "History" section of the "Health for all" page on the WHO website (consulted on 23 July 2004).

raised. Even so, a very ambitious slogan gained popular currency from the 1980s in an attempt to move matters forward—"health for all by the year 2000" (OMS, 1981)—from which time all the WHO's activities would be carried on under that banner.

Or Health for All in the 21st Century?

By the late 1990s, however, the inescapable fact was that health for all would be nowhere near achieved by the end of the millennium. The idea was taken up again in more realistic terms through wide-ranging consensus between the WHO, the United Nations, the IMF (International Monetary Fund, established in 1945), the World Bank, the WTO (World Trade Organization, created in 1995 and based in Geneva), and others. Health for all was still a central concern, but the call was now "health for all in the 21st century," which left plenty of time. It is now a long-term strategic process intended to result in incremental improvements in the health of populations, without being an endpoint. Health for all by the year 2000 was only the start of that process.[31]

The gap between the goal initially set and the resources allocated was too wide. If even in rich countries, whose most comprehensive systems of health care free at the point of delivery had still failed to even up survival probabilities, let alone the span of healthy life, it is hardly surprising that the WHO should never have really been able to implement a strategy that lived up to its noble ambition of equalizing opportunities on a world scale. Too many opposing interests—political, obviously (national self-interest, ideological liberalism), but also economic, technical, and human—were in play. Not to mention that, fundamentally, the WHO lacked the financial resources for such an undertaking, and could only appeal to public or private good will.

As part of its strategy on health for all in the 21st century, the WHO Ninth General Programme of Work, covering the period 1996–2001, lays down 11 public health goals for countries, setting one or more specific targets for each goal to be achieved before the end of the period, mostly in the form of quantified indicators (Table 107–3). For example, the goal of "increasing the span of healthy life" is to be delivered by meeting two targets: life expectancy at birth must not be less than 60 years in any country in 2001, and life expectancy at birth differentials within all countries must be reduced by 50% compared to 1990 (OMS, 1994).

[32] The World Bank is a group of financial institutions in the United Nations system, most centrally the International Bank for Reconstruction and Development (IBRD), set up in 1945. Others in

Naturally, while their representatives at the World Health Assembly may have adopted the program in principle, countries still have the final say on whether or not to implement it and are also responsible for resourcing it on top of international assistance that always falls short of the needs.

More recently, however, all 191 United Nations member countries reaffirmed their commitment to hard economic and social development targets, several of them expressly health related, in the MDGs (Millennium Development Goals) (Annan, 2000; United Nations, 2000). These included reducing the under age 5 mortality rate by two-thirds and the maternal mortality ratio by three-quarters, and halting the spread and beginning to reverse the incidence of HIV/AIDS by 2015.

II. ACTORS AND RESOURCES

Health policy in the broad sense addressed here clearly involves a wide range of actors. It also involves deploying resources, the main problem being that these are usually woefully lacking. A tentative short outline of actors and resources follows.

1. A Wide Range of Actors

Policymaking is clearly the purview of those in power, but few policies can be successfully conducted without the active support of both those responsible for executing them and those on the receiving end. That is obviously true in democratic countries, but equally so in less-democratic ones. A distinction must be drawn between public and private actors, therefore.

a. Public Actors

The State

The State is the fundamental entity of the modern organization of human societies. It is still today and will doubtless remain for some considerable time the main center of decision-making and policy action, notwithstanding the present trend toward the development of supranational entities and decentralization to the regions.

Governments everywhere have a ministry/central department for health, although in many cases its jurisdiction will extend well beyond health strictly speaking (e.g., a social affairs ministry). Clearly, the definition and implementation of health policies will be primarily that ministry's responsibility, but most of the decisions required go beyond its strict remit. As formulated, for one thing, they may impinge on other

TABLE 107–3 Goals and Targets of the WHO Ninth General Program of Work for the Period 1996–2001

Goal	Targets
1. To increase the span of healthy life	Life expectancy at birth not less than 60 years Reduce disparities by 50%
2. To ensure universal access to essential health care	85% of the population must have access within an hour's walk to treatment of common diseases, essential drugs and vaccines, basic biological products and blood products 70% of people with STDs to be examined and treated All countries to launch initiatives for the elderly
3. To ensure survival and healthy development of children	Infant mortality rate not to exceed 50 per 1000 Under-five mortality rate not to exceed 70 per 1000 90% of children immunized against 6 diseases Proportion of low birth weight reduced to less than 10%
4. To improve the health and well-being of women	Maternal mortality rate reduced to 50 per 1000 All pregnant women to have access to prenatal care and trained attendants during childbirth Disability-free life expectancy for women to increase by 15%
5. To ensure healthy population development	Everyone to have access to family planning services
6. To eradicate, eliminate or control major diseases	Eradicate poliomyelitis and dracunculiasis Eliminate leprosy (reduce prevalence rate to less than 1 per 10 000), neonatal tetanus (less than 1 per 1000), and measles (reduce mortality rate by 95% and the incidence rate by 90%) Control malaria, TB and hepatitis B Reduce coronary heart disease mortality among the under-65s by 15% Reduce cancer mortality among the under-65s by 15%
7. Reduce the risk of HIV transmission	Achieve a rate of 50% of people having casual sexual contact using a condom
8. To reduce avoidable disabilities	Significantly reduce disabilities from sensory, locomotor and mental impairment among underserved population groups
9. Improve nutritional status	Reduce the prevalence rate for severe malnutrition by 50% Eliminate vitamin A and iodine deficiencies Reduce the prevalence of iron deficiency anemia in women by 33%
10. To enable universal access to safe and healthy environments and living conditions	Ensure access to drinking water for at least 85% of the population Provide at least 70% of the population with access to sanitary excreta disposal facilities or services Substantially reduce the risks of violence
11. To give all people access to healthy lifestyles and healthy behavior	All people to have access to information

Source: OMS, 1994.

ministries' areas of responsibility. For example, no public health protection/enforcement action can be taken without involving the central authorities responsible for public order (police, army). Health ministries may not even be involved in enacting some of these provisions (road traffic regulations, tackling crime, etc.), while other ministries may also take protective measures for specific population groups under their responsibility (concerning education, housing, work, the environment, etc.) which have far-reaching direct or indirect consequences for health. More often still, the resources to be deployed go beyond what the health ministry alone can allocate, be it financial resources (direct or indirect) or legal measures. In short, most health policy decisions are multidisciplinary in nature and require consultation across departments, not to mention, in many cases, the enacting of new legislation requiring the intervention of parliament and the deployment of democratic debate.

Implementing decisions is a matter for the different administrative authorities concerned. Their remit, and especially that of the ministry responsible for health, is very much a function of the collective health care funding system in each country. It is clearly much greater in the United Kingdom than in the United States. Between these two extremes, in countries with systems based on compulsory social insurance, State policy measures are put into action by semi-public organizations jointly managed by the social partners. In France, for example, the general social security scheme is run by the national health insurance fund (CNAM), a tripartite body of the State, employers' organizations, and trade unions. Finally, to ensure nationwide coverage, these government authorities or organizations have local offshoots. France's health ministry has a district health and welfare authority (DDASS) in each department, while the CNAM acts through a dense network of local social security funds.

Specific governmental authorities are often set up to handle specific areas of health policy implementation. In the United States, for example, the all-powerful Food and Drug Administration (especially its Center for Drug Evaluation and Research) not only enforces compliance with food health regulations, but also licenses the marketing of medication. This agency's influence stretches well beyond its national borders, being a quite proper compulsory vetting point for any foreign-manufactured foodstuffs or medications imported into U.S. territory. In France, the *Agence française de sécurité sanitaire des produits de santé* has jurisdiction only over medication, while foodstuffs are under the control of the *Agence française de sécurité sanitaire des aliments* (Bonnici, 2003). Similar kinds of official agency exist in other areas, like the French *Agence française du sang*, set up in 1992 after the contaminated blood supplies scandal, or the *Service central de protection contre les rayonnements ionisants* (SCPRI) for ionizing radiation (Ceccaldi, 1997).

Local Government

Existing supra- and subnational political structures are often involved in the health sphere. Care must be taken not to confuse the local outposts of central government with decentralized local authorities. In federal states like the United States of America, Canada, India, Brazil, Germany, and others, it is clear that the individual states act independently of the federal State within the limits of their constitutional responsibilities, which tend to be quite wide-ranging in health matters (especially care provision). But even such a highly centralized country as France has several tiers of decentralized local authorities (regions, departments, municipalities) run by elected councils with decision-making powers independent of the State in a wide range of spheres. All have health-related activities.

Not all measures pursued by States and local authorities are purely domestic. They also play a significant role in assisting health improvement in poor countries. This was and largely remains a mainly central government concern, but is increasingly diversifying in many directions. At the macro level, bilateral cooperation is increasingly carried out through multipartnership projects involving either a number of national organizations or a combination of national and international organizations (so-called multi- or bilateral cooperation). At the micro level, transnational cooperation between local authorities has also developed in recent decades. So, French regions and departments have recently struck up agreements with, especially African, partners for health development projects. Individual municipalities are even increasingly involved in this kind of activity, both through traditional twinning and in other ways. Finally, the last 10 years have seen the emergence of what is known as South-South cooperation, where a developing country that has acquired a level of expertise sets up cooperation arrangements with other less-advanced developing country partners. Tunisia is a case in point for reproductive health. Countries such as Brazil, India, and Thailand are others; having developed the means of producing antiretroviral drugs, they entered agreements with other AIDS-hit countries like Mozambique to help them do likewise (ONUSIDA, 2004).

International Organizations

Finally, above nation-state level stand the international organizations. At the world level, within the United Nations system, the main role obviously lies with the WHO. That was the reason for setting it up in 1948; it succeeded the Health Organization set up in 1920 within the League of Nations, which, after World War I, had itself replaced the Office International d'Hygiène Publique (OIHP), created in Paris in 1907.

WHO is tasked with implementing the policies laid down by the World Health Assembly, which is to the WHO what the United Nations General Assembly is to the United Nations. As reaffirmed in its General Programme of Work 2002–2005 (WHO, 2001), its four strategic goals are to (1) reduce excess mortality, morbidity, and disability, especially in poor and marginalized populations; (2) promote healthy lifestyles and reduce risk factors to human health that arise from environmental, economic, social, and behavioral causes; (3) develop health systems that equitably improve health outcomes; and (4) enable the creation of an institutional environment for health and promote an effective health dimension to social, economic, environmental, and development policies. It plans to deliver these goals by focusing on (a) increasing awareness among the different actors in health; (b) managing information, assessing trends, comparing performance, and stimulating research and development; (c) giving technical and policy support to national action on health and international cooperation; (d) negotiating national and global partnerships; (e) setting, validating, and monitoring the implementation of standards; and (f) stimulating the testing of new technologies. Its action is carried on in various major world regions either by the organization's own WHO regional centers (Africa, Europe), or in cooperation with independent regional institutions (e.g., Pan-American Health Organization [PAHO] in America). The WHO also has local representative offices in most countries.

In addition to the WHO, other big international institutions, with which the WHO also has strategic agreements, also intervene in the health policy sphere: FAO, ILO (International Labor Organization, set up in 1919 as part of the League of Nations, which it outlived to become a United Nations agency in 1946; its Permanent Secretariat is the Geneva-based ILO), UNESCO (United Nations Educational Scientific and Cultural Organization, set up in 1945 in Paris), IAEA (International Atomic Energy Agency, set up in 1957 in Vienna), UNIDO (United Nations Industrial Development Organization, set up in 1966, with headquarters in New York and Geneva), and the World Bank.[32] Other institutions with which the WHO has agreements are UNFPA and UNICEF (referred to earlier for their key roles in reproductive health, and maternal and child health) and specialized agencies such as UNAIDS (Joint United Nations Program on HIV/AIDS, Geneva), HCR (United Nations High Commissioner for Refugees, Geneva), and the WFP. Other regional organizations also play a key role in framing concerted programs on health: the OECD (Organization for Economic Cooperation and Development)[33] for Western countries, and the OAU (Organization of African Unity, set up in 1963 in Addis Ababa), which became the African Union in 2002. It is difficult to say exactly where the European Union fits in this lineup, for while it is not yet entirely a State, it is already much more than a simple international organization, exerting a growing influence on the health policies of its 6, then 9, 12, 15, and now 25 member countries.

the group are the International Development Association (IDA), founded in 1960; the International Finance Corporation (IFC), set up in 1956; the Multilateral Investment Guarantee Agency (MIGA), established in 1988; and the International Centre for Settlement of Investment Disputes (ICSID), created in 1966.

[33] Originally set up in 1948 as the Organization for European Economic Cooperation (OEEC) under the Marshall Plan, it reinvented itself as the OECD in 1961; it is based in Paris.

[34] Set up in New York in 1987, Act Up rapidly spread elsewhere, particularly to major cities such as San Francisco, Philadelphia, and Paris.

b. Private Actors

Many private actors are involved in the framing and implementation of health policies. At the base of the pyramid, obviously, are health professionals and patients. But a brief word must also be said about the other societal shapers most involved in health issues—business, the trade unions, and mutual insurance organizations. Add to these the general public itself, and the various NGOs working in the health sphere.

Doctors and Patients

Micro health decisions are taken in the private consultation between doctors and their patients. This is a necessary stage in any modern health policy, at least as far as curative medicine goes. But it is not the only one. From it arise fundamental problems concerning issues of responsibility and freedom of choice which may be addressed in different ways from country to country. Doctors have a responsibility to treat their patients to the best of their abilities, but do they have total freedom of choice in the treatment? Should it be circumscribed by public interest constraints, like the balance between national spending on health and other matters? Likewise, the patient's main concern is to secure the best treatment from his doctor, but is his freedom of choice (of doctor, how the proposed treatment is carried out, etc.) itself circumscribed by limits imposed by the public interest? These questions are high on the agenda today, with the problems of funding total health spending. But they also include philosophical, ethical, and ideological aspects that sporadically come to the foreground of public debate. Proof of that, if it were needed, is given by the fierce controversies around reproductive health issues that flared up in the 1990s around the Cairo conference, setting religious and sectarian groups at odds with the proponents of individual liberties. The attitudes and behaviors of doctors and patients themselves obviously very largely inform policymaking. Indeed, the word *doctor* is too restrictive here. More broadly, it is all health professionals that are involved: specialists in the different branches of medical and pharmaceutical practice, and the paramedical professions (physiotherapists, nurses, midwives, etc.) alike. One of the most pressing issues in the doctor-patient dialogue is that of patients accessing information on their illness and its treatment, and having a say in the decisions. One of the practical repercussions of the developing trend in this area is, for example, the laying-down of legal conditions on which individuals can access their medical records.

But health professionals and patients do not step into this debate only through their individual actions, but also through collective voluntary or even semi-public institutions. Two types of professional organizations are found: (a) purely private trade unions and associations and (b) professional regulatory bodies. The latter have certain official powers to regulate the activities of the profession that they represent. Recent decades have also seen the development of what can be a highly active patients' self-advocacy movement. The highest-profile example must surely be the AIDS support organizations that have been set up since the disease burst onto the scene. This new development reflects a quite radical shift in the patient-doctor relationship. Individually and collectively, patients want more of a say in how their illness is managed by the health system. The activity of an association like Act Up is wholly indicative of this new mindset.[34] Other patient and family support groups for other diseases and disabilities (e.g., paralysis, epilepsy, cancer) are also highly vocal self-advocates for their rights and a say in the decisions that affect them.

Companies

The companies most active in the health field are obviously those engaged in the production of medical resources: pharmaceutical firms, medical equipment manufacturers, hospital builders and developers, and the like. Little would be gained from a lengthy description of their often-considerable influence over health policymaking. Their function may be fundamental to the improvement of health, but their strictly economic interests may also on occasion distort the policy approach. They are not, for example, completely free of implication in the relative failure of the primary health care strategy, or in the eternally protracted delay in developing a vaccine against AIDS or malaria, as neither primary health care nor AIDS in Africa are particular money-making propositions at this time. Other kinds of company, moreover, may play an outright negative role when they form lobbies to thwart the introduction of policies to tackle alcohol abuse, smoking, road transport hazards, for example. Conversely, firms of all kinds who see the health of their employees as a valuable asset to be preserved have and still do positively inform the approach and implementation of health policies. The likely success or failure of any occupational health policy will obviously be shaped by their attitude. Finally, business, or more precisely, employers' organizations, directly inform the framing and implementation of health policies in countries with a compulsory insurance-based social security system.

There is also a very singular type of undertaking, set up to provide development assistance generally, which often includes health development assistance. These are the foundations, usually Anglo-American in origin. Whether big, long-established foundations like the Rockefeller Foundation (founded in the United States in 1913), the Ford Foundation (founded in the United States by Henry Ford in 1936), the Wellcome Trust (founded in Britain in 1936), or more recent ones like the Bill & Melinda Gates Foundation (set up in January 2000), they have all at some point funded large-scale health projects, especially in the developing countries.

Trade Unions and Mutual Insurance Organizations

The same can be said of trade unions, but they may also play a stand-alone health role. In fact, this function has been inseparably linked to the creation of the mutual health insurance organizations, which now mostly operate independently of the unions. In countries where most basic health costs are free at the point of delivery, mutual insurance organizations remain of varying importance in funding health costs that are not free (e.g., the co-payment charge, or patient's fixed contribution per day of hospitalization,) or non-reimbursable treatments. Their role is clearly more important still in countries dominated by the free-market philosophy, in competition, of course, with private insurance companies.

The Public

The public itself is a major actor in health policies, be it simply through the sum total of individual behaviors, or via organized voluntary movements. We are concerned here less with the behavior of patients, already referred to, than of healthy individuals, which will clearly inform prevention (lifestyle, hygiene, risk behavior, etc.), but also screening (attention to symptoms and use of health services). The importance of these factors to health policies has been highlighted in particular by the analysis of certain mortality differentials, like the sex-specific differential (see Chapter 54), social inequalities in mortality (Vallin, 1979a), and the health crisis in the communist regimes (Meslé *et al*., 1998; Meslé and Vallin, 2003). In each of these three examples, one factor of differentiation is the different capability to change individual attitudes, thereby optimizing new ways forward to improvements in health. This is particularly important in bringing about changes to existing behaviors in order to tackle cardiovascular diseases (Meslé and Vallin, 1993). Be it in changing individual behaviors or public pressure brought to bear on governments or local authorities to achieve a specific public health measure, health poli-

[35] *Médecins sans Frontières*, or Doctors without Borders, was set up in 1971 during the Biafra war, and has since been a seminal influence in France and many other countries.

cies are also shaped by voluntary organizations working proactively in this area. It is an area of rapid growth in the number of local voluntary organizations with a wide range of aims, especially in developed countries where public demands are increasingly exacting in this particular sector, as in consumer matters generally.

NGOs

Charities have a long and venerable history. Historically in Europe, they played a key role in health, especially through church activity. Today, what we now call NGOs play a major role in many developing countries, and some have a national, international, or even global reach. The main and also the oldest must surely be the International Committee of the Red Cross with its extensive network of national Red Cross organizations, and its counterpart in the Muslim countries, the Red Crescent. It was established in 1863, initially to see to the war wounded, the conditions of treatment of prisoners of war, and prevention of the risks incurred by civilian populations as a result of wars. One of its first achievements was to secure the signature of the International Convention of 1864 in Geneva after the Crimea War for "the amelioration of the condition of the wounded and sick in armed forces in the field" (see the Red Cross website).

Revised and expanded many times since, this is the root document of current international humanitarian law. It rapidly expanded its peacetime functions and its interventions. It is particularly proactive in prevention and relief in violent situations of all kinds from ordinary road traffic accidents to emergency situations and major disaster prevention and management. More recently, a number of charities specializing in health development assistance to the developing world have been set up and expanded. Prominent in the specific field of health are, for example, *Médecins sans Frontières*,[35] *Médecins du Monde*, and *Pharmaciens sans Frontières*.

But health is not about medicine alone, and the contributions of activist groups working against hunger and for development (OXFAM,[36] CARE,[37] ATD Quart Monde[38]), for children (Save the Children[39]), or for human rights observance (Amnesty International, established in England in 1961, Human Rights Watch,[40] etc.) are no less important. These large or small voluntary organizations—which allow themselves less and less to be co-opted without insisting on certain ethical criteria—have awoken the national and international public authorities to an awareness of the utility of a new concept, humanitarian intervention, which allows them easier access to populations at risk of famine or violence and the resulting health consequences, notwithstanding the principles of national sovereignty. A right to take action, albeit still very restricted, is starting to develop in this sense, especially since the Yugoslav and Rwandan conflicts.

2. Resources

From people to budgets, via provider organizations, instruments, and other specific goods, the resources needed to implement a health policy are many and various and clearly cannot be exhaustively treated in a chapter such as this. The human resources have already been touched on in the section on actors, moreover. Suffice it here, then, to expand on that aspect with a review of some data on the diverse situations of those human resources. Other than that, the focus will be on four other aspects, which arguably cannot go ignored: health care facilities, scientific research and training for health professionals, public information and health education, and health budgets.

a. Practitioners

The number of doctors per capita is a traditional indicator of a population's health care coverage. International comparisons must be approached with caution, as the definition of "doctor" varies between countries, in particular with respect to the level of qualification. China's barefoot doctors, in particular, do not meet current Western criteria, and a legacy of lower qualifications remains in the medical profession in ex-Soviet European countries. Even so, the statistics

[36] The Oxford Committee for Famine Relief was set up in 1942 to deal with the consequences of World War II, especially famine. It now pursues a wide range of development-related activities, including in the health sphere.

[37] Founded in the United States in 1945 to provide relief to populations who suffered in World War II, CARE has developed into an international confederation of voluntary organizations focused on tackling hunger.

[38] Set up in France in 1964 from a concern about entrenched extreme poverty in the industrial countries, *Aide à toutes détresses-Quart-Monde* has since expanded its work to include the developing world.

[39] Save the Children (UK) was established in England in 1919 to provide aid to young survivors in post-WWI war-ravaged Vienna. Save the Children (US) was founded in New York in 1932 to help Appalachian children reduced to poverty by the recession. Save the Children now focuses mainly on Third World countries.

[40] Founded in the United States in 1978, initially with the name Helsinki Watch, to monitor the application of the Helsinki Agreement. HRW now deals with human rights violations generally.

[41] Obstetrics was one of the final areas of traditional medical activity to be nearly entirely hospitalized. In France, for example, it is only since the late 1950s that most births have been hospital births (in maternity clinics).

compiled by the WHO are revealing as to the differences currently found in terms of human resources allocated to health. The number of doctors per 100,000 of the population varies from less than 4 in some sub-Saharan African countries (Burkina Faso, Eritrea, Gambia, Liberia, Niger, Somalia, Chad) and in Nepal, to more than 400 in Spain and Norway, as well as Belarus, Georgia and Russia, and even more than 500 in Italy and Cuba. It is 300 in France, 350 in Germany, but less than 170 in the United Kingdom and 190 in Japan, as shown in Table 107–4.

It might be assumed that the developing country deficit might be at least partly offset by a greater number of nurses. Not so. While the number of nurses per 100,000 of population is above 1,000 in Belgium, and even 2,000 in Finland, it is below 100 in more than 60 developing countries (including China), and even less than 10 in such countries as Nepal, the Central African Republic, Liberia, and Haiti. Figure 107–2 clearly shows, for countries with populations over 15 million, that higher medical practitioner density in the industrial countries is associated with an equally higher density of the nursing profession. This is clearly one explanatory factor of the relative failure of the primary health care strategy.

If the nurse-to-doctor ratio can be seen as a crucial issue in primary health care strategies in developing countries, the general practitioner-to-specialist ratio within the medical profession itself is equally so for any health policy. The development of specialties is

TABLE 107–4 Doctors and nurses per capita in countries with populations over 15 million for which information is available, around 1997

Country	Numbers per 100,000 of population of:		Population (million)	Country	Numbers per 100,000 of population of:		Population (million)
	Doctors	Nurses			Doctors	Nurses	
Western developed countries				**Latin America**			
United States	279.0	972.0	285	Brazil	127.2	41.3	172
Japan	193.2	744.9	127	Mexico	186.4	86.5	99
Germany	350.0	957.0	82	Colombia	116.0	48.3	42
France	303.0	497.0	59	Argentina	268.4	76.8	37
United Kingdom	164.0	497.0	59	Peru	93.2	115.2	26
Italy	554.0	296.0	58	Venezuela	236.3	64.4	24
Spain	424.0	458.0	41	Chile	110.3	47.2	15
Canada	229.1	897.1	31	**Asia**			
Australia	240.0	830.0	19	China	161.7	98.6	1275
Netherlands	251.0	902.0	16	India	48.0	45.0	1017
Former Soviet countries				Indonesia	16.0	50.0	212
Russia	421.0	821.0	1456	Pakistan	57.0	34.0	143
Ukraine	299.0	736.0	50	Bangladesh	20.0	11.0	138
Poland	236.0	527.0	39	Vietnam	48.0	56.0	78
Uzbekistan	309.0	1011.0	25	Philippines	123.0	418.0	756
Romania	184.0	409.0	22	Turkey	121.0	109.0	68
Kazakhstan	353.0	649.0	16	Iran	85.0	259.0	66
Africa				Thailand	24.0	87.0	61
Nigeria	18.5	66.1	114	Myanmar	29.7	26.1	48
Egypt	202.0	233.0	68	South Korea	136.1	291.2	47
Congo (DR)	6.9	44.2	489	Nepal	4.0	5.0	24
South Africa	56.3	471.8	44	Iraq	55.0	236.0	23
Tanzania	4.1	85.2	35	Malaysia	65.8	113.3	23
Sudan	9.0	58.0	31	North Korea	297.0	180.0	22
Kenya	13.2	90.1	31	Saudi Arabia	166.0	330.0	22
Algeria	84.6	297.8	30	Afghanistan	11.0	18.0	21
Morocco	46.0	105.0	29	Sri Lanka	36.5	102.7	19
Uganda		18.7	23	Yemen	23.0	51.0	18
Ghana	6.2	72.0	20	Syria	144.0	189.0	17
Madagascar	10.7	21.6	16				
Côte d'Ivoire	9.0	31.2	16				
Cameroon	7.4	36.7	15				

Data are for different years according to country. Most date from 1996, 1997, or 1998.
Source: WHOSIS database (http://www.who.int/whosis), retrieved July 26, 2004.

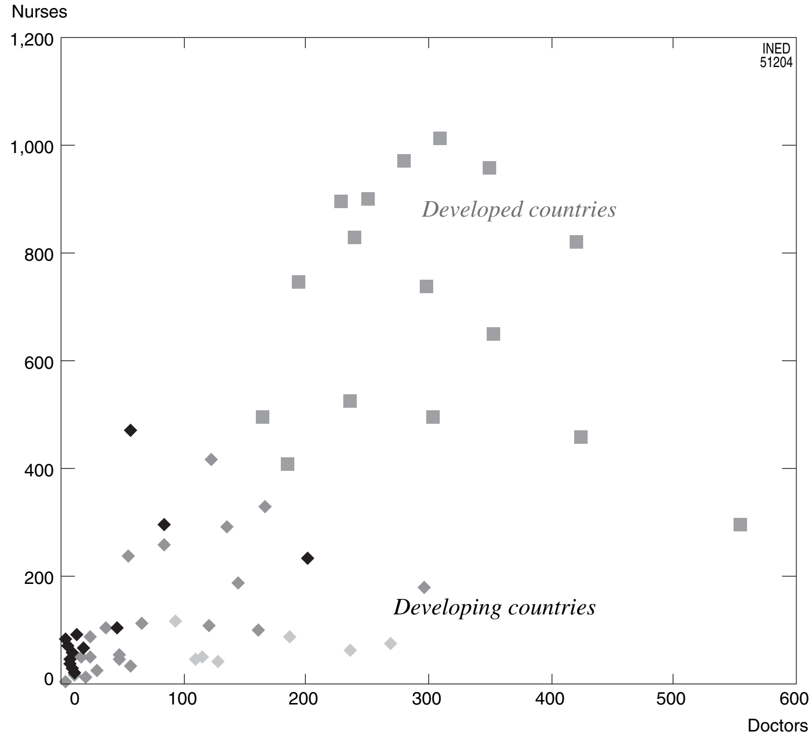


FIGURE 107–2 Relation between number of nurses and doctors per 100,000 of population in countries with populations over 15 million for which information is available, around 1997. (Source: Table 107–4.)

undoubtedly a big contributor to health improvement, but in managing existing resources, health policies cannot disregard the way in which patients access specialists. Should they have unrestricted direct access, or should there be compulsory referral by the general practitioner? The answer to this question has clear social, medical, and budgetary implications.

b. Health Care Facilities

Chapter 106 referred to the importance of the role played by hospitals in the history of medicine and health policies, in particular the turning point in Europe when the paupers' hospital became a care and treatment hospital. From that point, medicine in all industrial countries refocused itself entirely on the hospital, turning it into a linchpin of health policies. Not only were all major care and treatment activities gradually integrated into hospitals,[41] up to death itself (with the recent development of free palliative care provision), but the hospital has also become the referral point for all practitioners, both non-hospital and hospital doctors, as well as for medical research and training, as will be seen in the following point, and also actively informs prevention policies.

But "hospital" is a general concept. In concrete terms, organized hospital provision is a complex network ranging from big hospitals (general or specialized) capable of treating a variety of conditions and having the most advanced facilities available, to the local health clinic carrying out basic procedures. This provision is also underpinned to varying degrees according to country and health policy by a linkage between a public sector and a private sector, the dividing line between which may not be clear, especially in countries where compulsory health insurance predominates.

In France, for example, public hospitals have certain public service obligations (be it in teaching or care) that set them apart from private hospitals, normally called clinics, and for which they receive specific public funding. But the State can also place such obligations on private hospitals, which then fulfill a public hospital role. In any event, both are almost wholly funded from semi-public funds through the health insurance funds that reimburse patients for the cost of treatment or pay hospitals directly under agreements made with them. For patients, finally, the main difference between public hospitals and private hospitals derives from the regionalization to which the

[42] The first telethon, organized in New York by Jerry Lewis for the Muscular Dystrophy Association in 1966, was the first time that a million dollars was received in pledges during a television broadcast. The idea has since been taken up in many countries.

former but not the latter are constrained. This means that the public hospital must treat any presenting patient in its area, whereas the private hospital may but need not do so.

In developed countries, this hospital network structures all medical activity and works relatively well because of the scale of resources allocated. Large referral hospitals are established for all large geographical areas as the hub for many second-rank satellite hospitals or clinics of varying sizes and functions, as well as extensive non-hospital medical provision supported by a network of pharmacies, analysis and medical imaging laboratories, and other local diagnostic support provision, let alone nurses, physiotherapists, and other non-hospital paramedical professions. In other words, patients wherever they live are never far from appropriate medical provision and can always be referred to more specialized providers if need be. Nor must we overlook the extensive, highly specialized adjunct services to this basic general provision, like occupational health services, school health services, maternal and child welfare centers, mobile accident units, or home medical care services, to cite only a few key examples.

Imported on a large scale into developing countries, although less during the colonial period than post-independence, this model could not deliver the same instant benefits. How many ultra-modern hospitals have been built, whether more as a result of prestige- than health-for-all-conscious national policies, or rich countries or private foundations focused mainly on giving a high profile to their humanitarian investments, with no consideration either of the real use of the services they would provide or the inordinate operating resources they would demand? Most of these financial black holes, indeed, have fallen rapidly into decay for lack of upkeep before even being shunned by patients when they have become fee-for-service providers under structural adjustment programs. This grossly exaggerated picture may not be true in every case, but the misgivings are there, and they are not new: it was largely in reaction to this confusion that the WHO launched its primary health care strategy in the late 1970s. This is not to deny the value of hospitals nor the ultimate need for developing countries to develop similar provision to that which prevails in developed countries. Even the Chinese barefoot doctors were in touch with hospitals (either for continuing vocational training or for referring cases beyond their skills). But, in a context of finite resources, the construction of large modern hospitals was arguably not the top priority, since the vast majority of the population would never access them, and overwhelmingly suffered from problems that could have been addressed simply and cheaply in the community. This clash between European-style hospital systems and the WHO's primary health care strategy has surely been among the big health policy issues of recent decades in terms of resource allocation trade-offs, and the debate is far from over.

Capital funding trade-offs for medical facilities are crucial in developing countries, but are also an issue in developed countries. Whenever a major technological advance is based on the use of sophisticated and costly equipment, decisions are required on priorities as to where, when, and for whose benefit it should be installed. Even in countries where the decision is theoretically left to private initiative, the government may have to bring a series of assessments into play (external trade balance, special authorizations for medical facilities, etc.), but in those countries where social security systems and state-run health care predominate, the authorities also have to make substantive trade-offs. The issues around the first scanners in the 1980s, then that of MRI (magnetic resonance imaging), are cases in point.

c. Research and Vocational Training

In health no less than other sectors, two kinds of investment are particularly decisive for continued progress: scientific research and training for professionals.

Research

Who pays for research? Where is it done? The answers vary widely according to the country. Most is done in developed countries, which means that in many cases its broad thrust only very imperfectly addresses the needs of developing countries. The delay in developing an AIDS vaccine because of the priority focus on treatment is only one of many examples. Increasingly, however, developing countries are themselves investing significantly in research, and to that extent, improvements are beginning to filter through.

Also, research within each country may be done in universities or specialized research bodies like INSERM (*Institut national de la santé et de la recherche médicale* [National Institute of Health and Medical Research]) in France, or commercially, especially in pharmaceutical laboratories. The former are mainly engaged in basic research, the latter in applied research, although there is no hard-and-fast rule about it, because some aspects of basic research offer more practical spin-off opportunities than others, and much also depends on who first patents the results, which is the main income-generating outcome of the research effort. That in turn is closely tied to the answer to the question: where does the money come from?

In fact, research draws on three main sources of funding: the public coffers, private finance, and public donations. Governments and local authorities in all countries contribute to the research effort, probably but not always somewhat more in countries less prey to free-market ideology. A large amount of public funding is allocated to medical research in the United States, for example, especially from that powerful funding provider the National Institute of Health (NIH) within the U.S. Department of Health and Human Services. The main difference between France and the United States from this angle is due less to the importance of public financing in France than that the NIH allocates its budget on invitations to tender to a wide range of small and medium-sized research centers, especially universities, whereas INSERM itself conducts a key share of medical research (often, admittedly, in cooperation with universities and hospitals) out of its own budget. In all countries, most drug research, and more generally research that leads fairly directly to the design of saleable medical products, is private sector funded. Finally, charities attract levels of donations for research that are increasingly significant for the sectors that benefit from them. In developed countries especially, there is increasing public interest in research advances. A growing number of single-issue voluntary organizations focused on specific diseases manage to raise substantial research funding. Cancer organizations in many developed countries, and more recently, AIDS groups, are cases in point. The latter case is accompanied by a new trend, referred to earlier, where patients and their families are a driving force. Telethons,[42] for example, are most often staged in France by muscular dystrophy support groups like the *Association française contre les myopathies*.

As well as public, business, and voluntary funding, it must also be borne in mind that foundations can also be major sources of research funding. This includes both general-purpose foundations, as well as those with a more specific health research focus, like cancer research.

Training

Since the rediscovery of the tradition of antiquity and the Arab-Muslim contribution to European medicine in the 19th century, medical training in the industrial countries, not to say the world over, has followed approximately the same model, combining university-based theoretical training with hospital-based experiential learning. Almost all faculties of medicine throughout the world have relatively close links to a so-called teaching hospital, which is itself in most cases involved in research activities (Puzin, 1996). Not only do students learn their future profession in the hospital, but their teachers are also attending physicians or researchers in that hospital. Here again, however, it is not just doctors as such that are concerned by this dual theory/practice training, but all medical professions, regardless of level or specialism. This symbiosis between training and clinical practice has been a decisive factor in progress. But it still has its downsides. The functions and conditions of the exercise of hospital medicine are clearly not all-points identical to those of non-hospital medicine, such that medical training in its current form may not necessarily be adequate preparation for the latter's particular requirements.

In this area, however, funding policies are as chalk and cheese. In some countries, like France, where education, including at the university level, is generally paid for by the State or local authorities, medical training is free (or nearly so). In many others, the reverse is true—it is paid for by students themselves or their families, although the cost burden may be partly alleviated by study grants for needy students offered by specialized organizations (including public bodies) or universities themselves.

In free education systems, the mass democratization of access to university education generally, and students' particular attraction to medical career opportunities, has sometimes prompted the public authorities to question the value to the community of training a surplus of doctors. Medical professional bodies themselves have seen in this a danger of impoverishing their profession by fuelling fiercer competition on the market, which they see as incompatible with the dispassionate and stable practice of medicine. In France, for example, faculties of medicine were the first to be authorized to apply selective entry to their students (whereas previously, any secondary school completer could theoretically enroll in the university course of their choice). This measure itself proved insufficient, and an admission quota was introduced to limit the number of students allowed to enter medical training each year, with specific quotas for each discipline. So strictly and indiscriminately was this policy applied that there are now shortages in certain fields, a fact recently aggravated by the growing fear among potential candidates of finding themselves sued for medical malpractice.

[43] French readers may recall Pierre Mendès France's celebrated and widely reported gesture of drinking a glass of milk in public in 1954 to symbolize the importance of changing behavior and finding alternatives to drinking wine and spirits (Collovald, 2000). It was the

Another important aspect of medical training is keeping theoretical knowledge updated during working life. While there is no lack of continuing medical training resources in teaching hospitals, the same is not true of non-hospital medicine. There are two significant resources here: for theoretical knowledge, professional regulatory bodies stage regular symposiums or specialized professional training courses; while on the material and financial front, pharmaceutical laboratories and medical equipment and instrument manufacturers often support this kind of activity. The leverage this gives them over medical practice and health policies should not be underestimated.

d. Health Education

Without delving into the health grounds sometimes adduced to explain various dietary prohibitions or taboos and the ritual ablutions found in nearly all religions, it must be conceded that all cultures contain various health-related dictates. Most of these tend to relate to hygiene and food. In traditional societies, these norms are learned primarily within the family. With the spread of public education, they are increasingly transmitted in the classroom. And schooling has played and still plays a major role in the improvement and maintaining of health—not only through the specific health education that may form part of the curriculum, but also quite simply by raising general educational levels. In France, for example, the debate as to whether Pasteur, with his discovery of the role of bacteria, or Jules Ferry, with compulsory free education, did most to improve the health of the population is meat for endless discussion; the one certainty is that both were major contributors. Likewise, in developing countries, the role of rising female educational levels in infant mortality decline has been abundantly emphasized (J. Caldwell, 1979, 1986). Not only does education make mothers more apt to seek information on optimum dietary requirements, preventive health, and treatment for their children, but it also gives them access to more independence and decision-making power within the family, assets that they can leverage to obtain the means to do the best by their children.

But the development of modern medicine and the accelerating pace of advances in knowledge and techniques mean that early learning is no longer sufficient to permanently imprint the healthiest behaviors and attitudes. Health education campaigns directed at both adults and children have become a major instrument of health policies (J. Caldwell, 1999). They tend to be targeted directly on health-related behavior or to popularize new health instruments whose collective effectiveness depends on their mass uptake.

The idea of health education campaigns may be either to discourage health-damaging behavior or to encourage healthy behavior. The "discouragement" side can be illustrated by one of the biggest mass education campaigns in developing countries with very high diarrhea-induced child mortality rates, where mothers were taught that children were at risk from dehydration and that the widespread practice of withholding fluids from the child was a recipe for disaster. Coupled with campaigns to promote oral rehydration salts, education was a big contributory factor in bringing down child mortality levels. Similar examples abound in the developed countries as well, such as campaigns on the ingrained belief in the mythical benefits of alcohol among the French,[43] Russians, and Poles, or against the health traps of drug abuse or the hyped-up "macho man" culture of speeding and other risk-taking behavior. On the health-promoting side, countless campaigns on hygiene—the point of which was already grasped by 19th century-hygienists (see Chap. 106)—have often drawn heavily on the discovery of the germ theory of disease (the basis of which is now clearly understood) to strengthen or refocus age-old practices.

In all these areas, health education campaigns have been a potent means of action for health policies. In one renowned example, it may even be said that knowledge was advanced by the campaign. This was the celebrated survey of British doctors on smoking and the incidence of respiratory tract cancer symptoms among their patients, the first result of which was to raise awareness and reduce smoking among doctors themselves even before the survey findings could be used to persuade the authorities of the value of an anti-smoking campaign (Royal College, 1962). This was the springboard for the preventive health drive that brought down tobacco consumption in Britain and led to a decisive decline in respiratory system cancer-related mortality in the United Kingdom (Vallin and Meslé, 2001). Health education campaigns are also behind the attempts to reduce or control cholesterol

starting point of a long campaign that began to deliver results in a falling alcohol-related death rate from the mid-1960s. Conversely, while Mikhail Gorbachev's draconian ban on alcohol sales to the public gave an immediate surge in Russian life expectancy in 1985, a pattern of decreasing life expectancy soon reasserted itself when the measures were relaxed (Meslé *et al.*, 1994), largely due to a failure in having prepared and supported the measures by a mass information and education campaign.

levels following the discovery of its connection with ischemic heart diseases.

Health education campaigns are also a key way to promote the use of new health instruments developed for prevention or treatment. Vaccination is surely the prime example, where robust mass information and education campaigns were behind the large-scale dissemination of vaccines and gaining public acceptance for making some compulsory. The WHO Expanded Program on Immunization (EPI) in particular was, and still largely is, based on this means of spreading vaccination coverage levels among children worldwide. And little persuasion is needed about the vital importance of information and education in the behavioral changes needed to stop the spread of HIV/AIDS, given what is known of its means of transmission, nor the role that both these forces for action play in road accident prevention campaigns. The antispeeding campaign recently launched in France, for example, marks a turning point in the use of driver education and acceptance of responsibility. Instead of hiding the new radar speed traps to catch offenders, advance warning is given of their presence, and the campaign relies on this openness to persuade drivers to watch their own speed, using this system as an aid to self-control rather than trying to find ways of beating a new form of clampdown.

e. Health Funding

Detailed consideration of the complex relationships between health economics and health policies is clearly beyond the scope of this chapter. But this review cannot be brought to an end without a brief look at the varied ways in which financial resources are allocated to health in the different countries of the world. The data for 2001 collected by the WHO (2004) are used here, but limited to countries with populations over 15 million for comprehensible comparisons that nevertheless encompass most of the range of situations covering the overwhelming majority of the world population. These figures obviously must be approached with caution, as the data collected are those supplied by each country to the WHO, and consistency issues may arise between the estimates produced. Nevertheless, the WHO provides two sets of estimates, one presented in U.S. dollars at official exchange rates, the other in dollars converted at purchasing power parities or "international dollars." The data analyzed here are based on the estimates in international dollars. Arguably, however, the broad traits they reveal are sufficiently pronounced to reflect an undoubted measure of the reality.

To Each According to Its Own Means

The most striking thing about the WHO figures is that when it comes to translating the rhetoric and political spin about reducing world inequalities in health into money terms, the limits are soon reached. As Figure 107–3 shows, there is an almost perfect fit between per capita health expenditure and gross domestic product (GDP). Among the 60 countries with populations over 15 million for which data are available, the income scale range from under 400 (Democratic Republic of Congo and Ethiopia) to over 35,000 (United States) international dollars per capita, and the self-same countries are found at the extremes of the total per capita health expenditure distribution, which itself ranges from under 15 in Congo and Ethiopia to nearly 5,000 in the United States. There is a very high correlation coefficient of 0.960, although the best fit of the clusters would clearly be more a slightly concave curve than a straight line. All the wealthy countries naturally fall into the upper right quadrant of the scattergram, with the poor countries in the lower left quadrant, while a handful of emerging countries, already well on the way to economic development (South Korea, Argentina, Chile, Saudi Arabia) occupy the hesitant middle ground between the two extreme clusters. Wherever "health for all" would dictate that the distribution of health funding should maximally reflect needs, the stark fact appears to be that they are allocated by reference to each country's own means.

The trend is even more pronounced than shown on Figure 107–3, because not only do the rich countries spend more on health than poor countries, but they also allocate a bigger share of GDP to it (Figure 107–4). At $35,000 per capita, the United States spends nearly 14% of its national income on health, whereas Congo and Ethiopia spend under $400 per capita, or 3.5%, on health.

Differential Health Efforts

Notwithstanding the very strong linkage of health budgets to national wealth, there is a clear differential to be seen in the effort made by each country for its population's health. With a correlation coefficient of 0.80, the Figure 107–4 clusters are significantly less well aligned than those of Figure 107–3. For the same level of per capita GDP, the health expenditure share of GDP can vary widely, especially in poor countries. For the same very low per capita income (under $1,500), Madagascar and Myanmar spend just 2% on health, compared to Kenya's 8%—four times more. A gap, admittedly less significant, is also seen between very rich countries on broadly the same level: with a

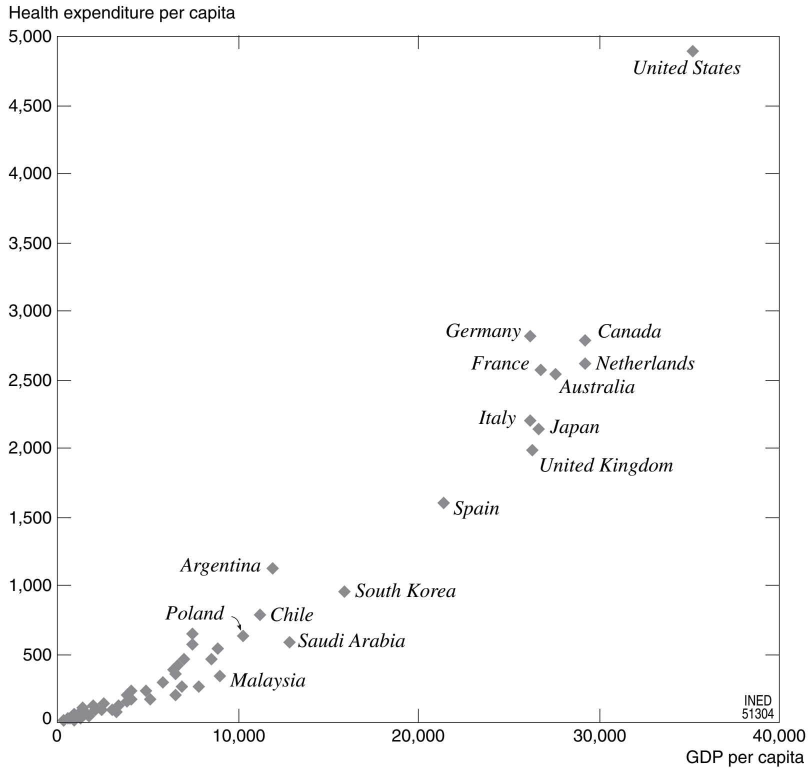


FIGURE 107–3 Total per capita health expenditure relative to per capita gross domestic product in 2001, in countries with populations over 15 million. (Source: WHO, 2004.)

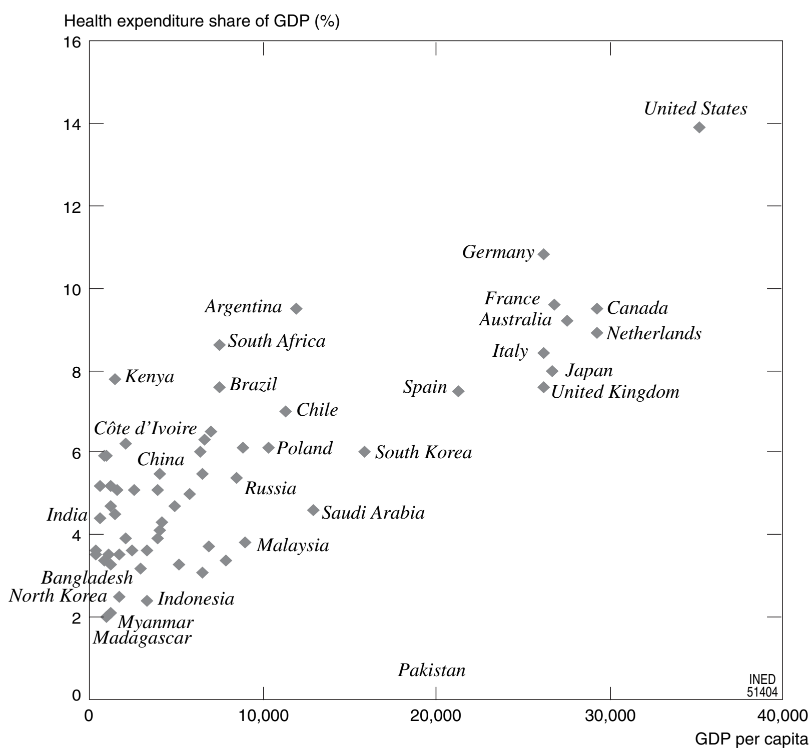


FIGURE 107–4 Total health expenditure share of GDP relative to per capita GDP in 2001, in countries with populations over 15 million. (Source: WHO, 2004.)

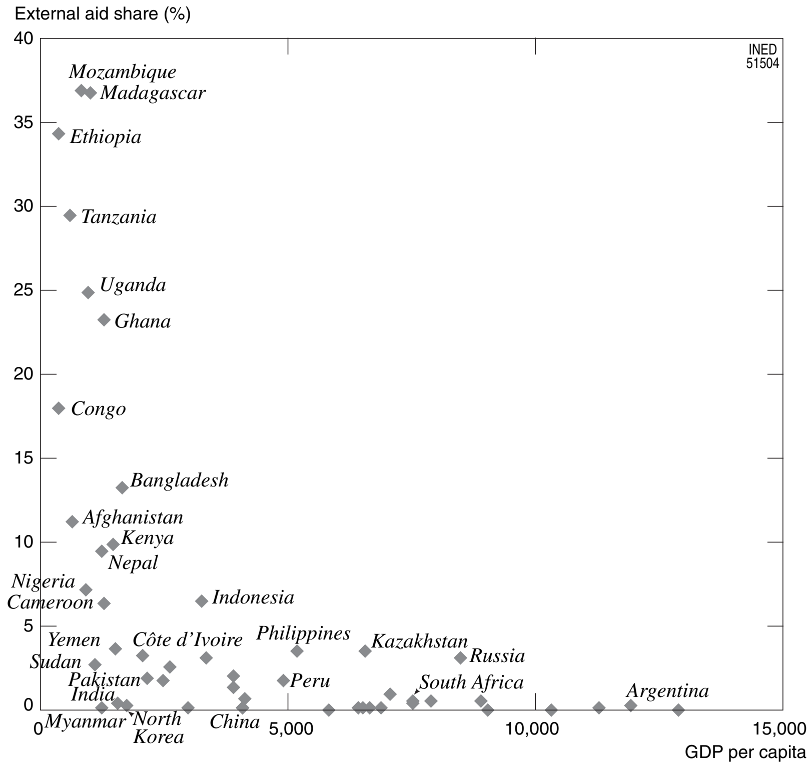


FIGURE 107–5 External aid share of total health expenditure relative to per capita GDP in 2001, in countries with populations over 15 million. (Source: WHO, 2004.)

per capita income of $26,200, Germany spends 10.8% of its GDP on health, and the United Kingdom just 7.6%. Conversely, markedly lower-income countries like Argentina ($12,000 per capita) and even South Africa ($7,500) spend as big a share of their GDP on health as does France or Italy.

The Confusing Role of External Aid

It might be thought for the poorer developing countries at least, that external aid for health (included in the WHO computation of total health expenditure) lies behind this diversity. Figure 107–5 shows that this share is highest in the poorest developing countries. But it also shows the extreme inequality in treatment of countries on the same level, especially the poorest ones. For a per capita income under $2,000, the external aid share of health expenditure varies from over 35% (Mozambique, Madagascar) to close to 0 (Myanmar, India, North Korea). Nor is there any evident link with the previous observation; quite the contrary since, among these countries, Kenya's 8% of GDP spent on health is less than 10% funded for the purpose by external aid, while the maximum level of external funding is received by the two countries which spend only a total of 2% of their GDP on health. In other words, it is not the poorest countries making the most effort that get most assistance.

Figure 107–5 also clearly shows how middle-income countries are affected by unequal treatment from international aid. So, albeit hard hit by AIDS, South Africa's health expenditure is less than 0.5% funded from external aid, while higher per capita GDP Russia's health spending is over 3% covered by external aid—more than 6 times more.

The Publicly Funded Share

On a different note, a comparison of the share of public funding from all sources (state, local authorities, compulsory insurance schemes) in health expenditure reveals two completely different worlds (Figure 107–6). In poor countries, there is almost no connection between total health expenditure and the publicly funded share of spending. The bracket of under $50 per capita spent on health includes, for example, North Korea (nearly 75% publicly funded) and Myanmar and Sudan (under 20% publicly funded). In this bracket—between $100 and $200 per capita on health—exactly the same gap between public contributions recurs, widest in Uzbekistan and Algeria, narrowest in India and Côte d'Ivoire. Ostensibly, indeed, the extent of the

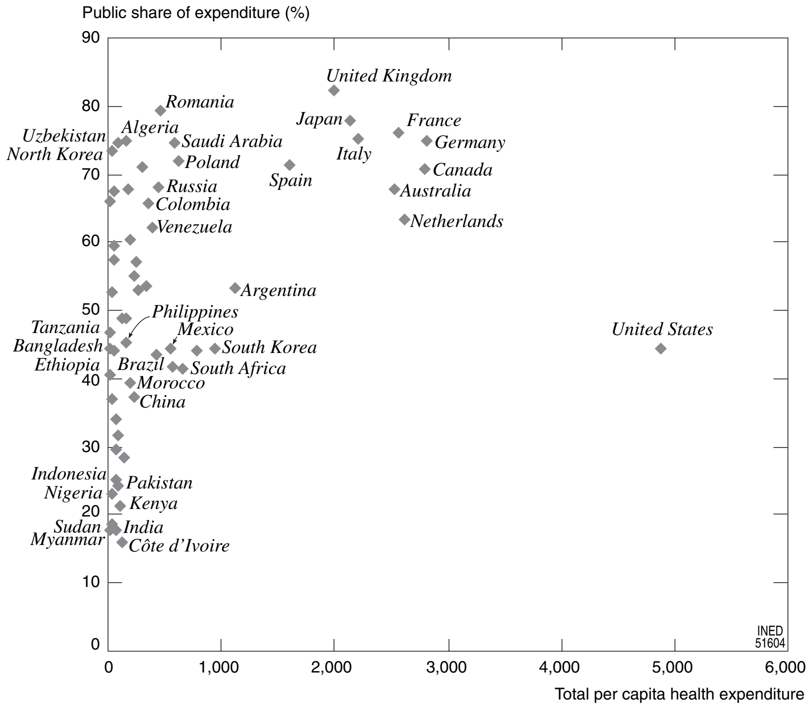


FIGURE 107–6 Public share of health expenditure relative to total per capita health expenditure in 2001, in countries with populations over 15 million. (Source: WHO, 2004.)

public effort in poor countries, which still falls far short of the needs, has little effect on total expenditure and varies more with the political regime than any real health agenda.

The situation is very different in wealthy countries, where there is a close correlation between total health spending and the public share of expenditure. It is highest in the United Kingdom (over 80%) where total per capita expenditure is lowest of all the rich countries included in Figure 107–6 (only Spain is lower, but it is also appreciably less wealthy than the United Kingdom). Conversely, the public share is lowest in the United States, where total health expenditure is well above the rest.

When looking at Figure 107–6, it is difficult to understand why collective health care funding systems have been so hard hit in the developed countries in recent decades. The free-market ideology, which has held the world economy in almost unchallenged sway since the fall of the Berlin Wall, operates in health matters on the twofold principle that the public management of economic affairs is nonsense because the squandering of resources can only be avoided by private enterprise operating in response to market constraints, and that health is a good like any other and no exception to the rule. The conclusion is then obvious: to halt the recent staggering increases in health expenditure, no time must be lost in returning more of health provision to private hands. The very opposite is what we are seeing here: the higher the public share of health spending, the lower total expenditure. The case of the United States is extreme to the point of distortion: while the public share of total spending is barely over 40%, not only does it have by far the highest per capita spending in the world (approaching $5,000), but also its spending accounts for the biggest share of national income (nearly 15%). For all that, it cannot be said that it has a higher life expectancy or quality-adjusted life years than anywhere else. By contrast, it is reasonable to wonder whether inequalities in health and mortality may not be higher than might be expected from such a level of expenditure. But that would be to beg the question of one of the issues in the following chapter, which analyzes health policy outcomes.

Acknowledgment

This chapter was translated from the French by Glenn D. Robertson.

References

ANNAN Koffi, 2000. *We the peoples. The role of the United Nations in the 21st century*. New York, United Nations, 80 p.

ANTONIOTTI Stéphanie, PELLISSIER Véronique, SIMÉONI Marie-Claude, and MANUEL Catherine, 2002. Déclaration obligatoire des maladies infectieuses. Des maladies "pestilentielles" aux maladies "émergentes", *Santé publique*, vol. 14(2), p. 165–178.

Bachmann Christian and Coppel Anne, 1989. *Le dragon domestique. Deux siècles de relations étranges entre l'Occident et la drogue*. Paris, Albin Michel, 564 p.

Beach Marilyn V., 2001. "Blood heads" and AIDS haunt China's countryside, *The Lancet*, vol. 357, p. 49.

Beghin Ivan and Vanderveken Marc, 1985. Nutritional programmes, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 81–100. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Beghin Ivan and Vandeveken Marc, 1985. Les programmes nutritionnels, *in*: Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 77–98. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Béresniak Ariel and Duru Gérard, 1997. *Économie de la santé*, 4th ed. Paris, Masson, 182 p.

Berger Guy, Mauffret Marie, Rousseau-Giral Anne-Chantal, and Zaidman Catherine, 2000. *La loi relative à la lutte contre le tabagisme et l'alcoolisme. Rapport d'évaluation.* Paris, La documentation française, 555 p.

Beveridge William H., 1943. *Social insurance and allied services.* London, Allen and Unwin, 300 p.

Bonnici Bernard, 2003. *La politique de santé en France*. Paris, PUF, 128 p.

Bordry Monique, Fayard Françoise, and Musée Curie, 1992. Les rayons X sortent de l'anonymat, *Comprendre et agir. Journal de l'Institut Curie*, vol. 21(4). (Available at http://www.curie.u-psud.fr/histoire/4/Histoire4.html).

Breslow Lester, 1985. Le cas des maladies cardio-vasculaires, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 195–216. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Breslow Lester, 1985. The case of cardiovascular diseases, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 197–216. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Brunet-Jailly Joseph (ed.), 1997. *Innover dans les systèmes de santé. Expériences d'Afrique de l'Ouest*. Paris, Karthala, 435 p.

Caldwell John C., 1979. Education as a factor in mortality decline: an examination of Nigerian data, *in:* OMS (ed.), *Délibération de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité*, p. 172–192. Geneva, New York, OMS/WHO, 622 p.

Caldwell John C., 1986. Routes to low mortality in poor countries, *Population and Development Review*, vol. 12(2,) p. 171–220.

Caldwell John C., 1999. Peut-on modifier les comportements pour préserver la santé? *Revue internationale des sciences sociales*, no. 161, p. 371–377.

Caldwell Pat, 1995. Prostitution and the risk of STDs and AIDS in Nigeria and Thailand, *Health Transition Review*, vol. 5(supp.), "The Third World AIDS epidemic" (edited by I. O. Oruboloye, John C. Caldwell, Pat Caldwell, and Shail Jain), p. 167–172.

Card David, Dobkin Carlos, and Maestas Nicole, 2004. *The impact of nearly universal insurance coverage on health care utilization and health: evidence from Medicare.* Cambridge, National Bureau of Economic Research, 73 p. (Working Paper 10365; http://www.nber.org/papers/w10365).

Ceccaldi Dominique, 1997. *Les institutions sanitaires et sociales.* Paris, Foucher, 405 p.

Cérisola Anne-Sophie and Mistral Jacques, 2004. *L'obésité aux États-Unis. Enjeux économiques et défis politiques.* Washington, French Embassy, 27 p. (Document de travail 2004–01).

Chesnais Jean-Claude, 1976. *Les morts violentes en France*. Paris, INED et PUF, 346 p. (Travaux et documents, Cahier no. 75).

Chesnais Jean-Claude, 1981. *Histoire de la violence en Occident de 1800 à nos jours.* Paris, Robert Laffont, 497 p. (Pluriel).

Chesnais Jean-Claude, 1985. L'efficacité de la prévention des morts violentes, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 257–274. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Chesnais Jean-Claude, 1985. The prevention of death from violence: lessons from experience, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 261–280. Liège, Ordina Édition, 558 p.

Chesnais Jean-Claude, 2003. Les morts violentes dans le monde, *Population et sociétés*, no. 395, novembre, p. 1–4.

Chesnais Jean-Claude and Vallin Jacques, 1977. Évolution récente de la mortalité et de la morbidité dues aux accidents de la route, *Population*, vol. 32(6), p. 1239–1265.

Collovald Annie, 2000. Le verre de lait de Pierre Mendès France ou l'oubli des origines politiques, *in:* Alain Carrigou (ed.), *La santé dans tous ses états*, p. 35–54. Biarritz (France), Atlantica, 362 p.

Dalmaz Patrick, 1998. *Histoire des sapeurs-pompiers français.* Paris, PUF, 128 p. (Que sais-je?, no. 3116).

Delbos Victor, 1971. Introduction. Notice sur la vie et les oeuvres de Kant, *in: Emmanuel Kant*, *Fondements de la métaphysique des mœurs*, p. 3–20. Paris, Librairie Delagrave, 210 p.

Desgrées du Loû Annabel, 1996. *Sauver les enfants: le rôle des vaccination.* Paris, CEPED, 261 p. (Les Études du CEPED, no. 12).

Désesquelles Aline, 1997. *Consommation de stupéfiants en France: expertise d'un système d'information et mesure du phénomène*. Bordeaux, Université de Bordeaux, 400 p. (Thèse de démographie).

Dumont Jean-Pierre, 1998. *Les systèmes de protection sociale en Europe.* Paris, Economica, 313 p.

Faure Olivier, 1994. *Histoire sociale de la médecine (XVIII^e^-XX^e^ siècle).* Paris, Anthropos, 272 p.

Fillaut Thierry, 1999. Pouvoirs publics et antialcoolisme en France sous la Troisième République, *in:* Thierry Fillaut, Véronique Nahoum-Grappe, and Myriam Tsikounas (eds.), *Histoire et alcool*, p. 127–182. Paris, L'Harmattan, 220 p. (Logiques sociales).

Fingerhut Lois A. and Kleinman Joel C., 1990. International and interstate comparisons of homicide among young males, *Journal of the Medical Association of America*, vol. 263(24), p. 3292–3295.

Flamant Robert, 1976. Risques carcinogènes et prévention des cancers, *Revue d'épidémiologie et de santé publique*, vol. 24(3–4), p. 345–353.

Frederiksen Harald, 1960. Malaria control and population pressure in Ceylon, *Public Health Reports*, vol. 75(10), p. 865–868.

Frydman Nathalie and Martineau Hélène, 1998. *La drogue: où en sommes-nous? Bilan des connaissances en France en matière de drogues et de toxicomanies*. Paris, La documentation française, 417 p.

Girard Alain, Henry Louis, and Nistri Roland, 1960. *Facteurs sociaux et culturels de la mortalité infantile.* Paris, INED, PUF, 210 p. (Travaux et documents, Cahier no. 36).

Goldman Noreen and Takahashi Shigesato, 1996. Old age mortality in Japan: demographic and epidemiological perspective, *in:* Graziella Caselli and Alan Lopez (eds.), *Health and mortality among elderly populations*, p. 157–181. Oxford, Clarendon Press, 360 p. (International Studies in Demography).

Hannoun Claude, 1999. *La vaccination.* Paris, Presses universitaires de France, 128 p. (Que sais-je?).

Hatton Françoise, Flamant Robert, and Bouvier-Colle Marie-Hélène, 1985. La lutte contre la mortalité cancéreuse, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*.

p. 217–238. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).
Hatton Françoise, Flamant Robert, Bouvier-Colle Marie-Hélène, and Maujol Léone, 1985. The fight against cancer, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.) *Health policy, social policy and mortality prospects*, p. 217–240. Liège, Ordina édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).
Henry Louis, 1948. La situation démographique, *Population*, vol. 4(3), p. 523–538.
INSERM, 2001. *Alcool: effets sur la santé. Synthèse et recommandations.* Paris, Institut national de la santé et de la recherche médicale, 49 p. (Expertise collective).
IRD, 2000. *Paludisme:* Anopheles funestus, *un vecteur complexe.* Paris, Institut de recherche pour le développement, 3 p. (Fiche d'actualité scientifique no. 110).
Jauffret-Roustide Marie (ed.), 2004. *Les drogues, approches sociologiques, économiques et politiques.* Paris, La documentation française, 156 p. (Les études de la Documentation française).
Jauffret-Roustide Marie and Simmat-Durand Laurence, 2004. État de la consommation: produits, modalités d'usages, usagers, *in:* Marie Jauffret-Roustide (ed.), *Les drogues, approches sociologiques, économiques et politiques*, p. 11–46. Paris, La documentation française, 156 p. (Les études de la Documentation française).
Jones Stephen, Waldman Ronald J., and Foege William, 1985. Le rôle des programmes de vaccination, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 41–52. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).
Jones Stephen, Waldman Ronald J, and Foege William, 1985. The role of immunization programmes, in: Jacques Vallin, Alan Lopez et Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 44–55. Liège, Ordina Édition, 558 p.
Kielmann Arnfried and Taylor Carl E., 1969. *Child and maternal health services in rural India: the Narangwal experiment*, Vol. 1. Baltimore, Johns Hopkins University Press, 263 p.
Kitagawa M. and Hisamichi S., 1970. Follow up study of gastric mass survey in Miyagi prefecture, *Gastroenterologia Japonica*, vol. 5(4–3), p. 182.
La Documentation Française, 1993. Procès et débats autour du sang contaminé, *Regards sur l'actualité*, no. 194–195, spécial "Sida, sexe et société," p. 57–76.
Letourmy Alain, 2000. Les politiques de santé en Europe, vue d'ensemble, *Sociologie du travail*, vol. 42(1), p. 13–30.
Lévy Claude, 1980. La mortalité par accident des enfants et des adolescents dans huit pays développés, *Population*, vol. 35(2), p. 291–320.
Meslé France, 1990. Symposium on the Contribution of Epidemiology to Cancer Prevention and Therapy, Vienna, 7–9 March 1990, *Population*, vol. 45(3), p. 667–669.
Meslé France, Shkolnikov Vladimir, and Vallin Jacques, 1994. Brusque montée des morts violentes en Russie, *Population*, vol. 49(3), p. 780–790.
Meslé France and Vallin Jacques, 1981. La population des établissements psychiatriques: évolution de la morbidité ou changement de stratégie médicale?, *Population*, vol. 36(6), p. 1035–1068.
Meslé France and Vallin Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in:* UIESP/IUSSP (ed.), *Congrès international de la population. Montréal 1993*, vol. 2, p. 365–382. Liège, UIESP, 493 p.
Meslé France and Vallin Jacques, 2003. *Mortalité et causes de décès en Ukraine au XX*[e] *siècle.* Paris, INED, 396 p. (Les cahiers de l'INED, cahier no. 152, with contributions from Vladimir Shkolnikov, Serhii Pyrozhkov and Serguei Adamets).
Meslé France, Vallin Jacques, and Shkolnikov Vladimir, 1998. Reversal of mortality decline: the case of contemporary Russia, *World Health Statistics Quarterly/Rapport trimestriel de statistiques sanitaires*, vol. 51(2–4), p. 191–206. (Historical epidemiology: mortality decline, and old and new transitions in health, special issue edited by Odile Frank).
Molineaux Louis, 1985. The impact of parasitic diseases and their control, with special emphasis on malaria in Africa, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 13–44. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).
Molineaux Louis, 1985. La lutte contre les maladies parasitaires: le problème du paludisme, notamment en Afrique, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 11–50. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).
Mosley W. Henry, 1985. Les soins de santé primaires peuvent-ils réduire la mortalité infantile? Bilan critique de quelques programmes africains et asiatiques, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 101–136. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).
Mosley W. Henry, 1985. Will primary health care reduce infant and child mortality? A critique of some current strategies, with special reference to Africa and Asia, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (ed.), *Health policy, social policy and mortality prospects*, p. 103–138. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).
Moulin Anne-Marie (ed.), 1996. *L'aventure de la vaccination.* Paris, Fayard, 498 p. (Penser la médecine).
Nadelmann Ethan, 1994. Pour un droit d'usage contrôlé, *Futuribles*, no. 185, p. 105–109.
Nations unies/United Nations, 1968a. *Convention sur la circulation routière. Fait à Vienne le 8 novembre 1968.* Vienna, Nations unies/United Nations, 72 p.
Nations unies/United Nations, 1968b. *Convention sur la signalisation routière. Fait à Vienne le 8 novembre 1968.* Vienna, Nations unies/United Nations, 61 p.
Nations unies/United Nations, 1998. *Programme d'action adopté à la conférence internationale sur la population et le développement, Le Caire, 5–13 septembre 1994.* New York, Nations unies/United Nations, Département de l'information économique et sociale et de l'analyse des politiques, 204 p.
Nayar P. K. B., 1985. The case of Kerala, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 371–382. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).
Nayar P. K. B., 1985. Le Kerala, Inde, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 357–368. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).
Newman Peter, 1965. *Malaria eradication and population growth: with special reference to Ceylon and Guiana.* Ann Arbor, The University of Michigan, School of Public Health, 259 p. (Bureau of Public Health Economics Research Series, no. 10).
Nor Malika, 2001. *La prostitution.* Paris, Le cavalier bleu, 128 p.
Norvez Alain, 1990. *De la naissance à l'école. Santé, modes de garde et préscolarité dans la France contemporaine.* Paris, INED, PUF, 464 p. (Travaux et documents, Cahier no. 126).
OMS/WHO, [1946] 1985. *Constitution de l'Organisation mondiale de la santé.* Geneva, Organisation mondiale de la santé/World Health Organisation, 18 p.

OMS/WHO, 1978. *Déclaration d'Alma Ata. Rapport de la Conférence.* Geneva, Organisation mondiale de la santé/World Health Organisation, 60 p. (Available at http://policy.who.int/cgi-bin/om_isapi.dll).

OMS/WHO, 1981. *Stratégie mondiale de la santé pour tous d'ici l'an 2000.* Geneva, Organisation mondiale de la santé/World Health Organisation, 90 p. ("Santé pour tous" series, no. 3; doc. hfa-fr2).

OMS/WHO, 1985. *Recueil des résolutions et décisions de l'Assemblée mondiale de la santé et du Conseil exécutif. Vol. 2, 1973–1984.* Geneva, Organisation mondiale de la santé/World Health Organisation, 424 p.

OMS/WHO, 1994. *Neuvième programme général de travail pour la période 1996–2001.* Geneva, Organisation mondiale de la santé/World Health Organisation, 60 p. ("Santé pour tous"series, no. 11).

OMS/WHO, 1999a. Révision du règlement sanitaire international, *Relevé épidémiologique hebdomadaire de l'OMS*, vol. 74(4), p. 25–27.

OMS/WHO, 1999b. Comparaison des rôles de trois organisations internationales, *Relevé épidémiologique hebdomadaire de l'OMS*, vol. 74(25), p. 193–201.

OMS/WHO, 2001. *Programme de travail 2002–2005.* Genève, Organisation mondiale de la santé, 10 p. (Available at http://www.who.int/gb/ebwha/pdf_files/WHA54/fgpw0205.pdf).

OMS/WHO, 2004. *Rapport sur la santé dans le monde 2004. Changer le cours de l'histoire*. Geneva, Organisation mondiale de la santé/World Health Organisation, 188 p.

OMS/WHO and Banque mondiale/World Bank, 2004. *Rapport mondial sur la prévention des traumatismes dus aux accidents de la circulation.* Geneva and Washington, Organisation mondiale de la santé/World Health Organisation and Banque mondiale/World Bank, 235 p. (Prevention).

ONUSIDA/UNAIDS, 2004. *Rapport sur l'épidémie mondiale de sida. Résumé d'orientation*. Geneva, ONUSIDA/UNAIDS, 18 p.

Oshima Akira, Hanai Aya, and Fujimoto Isaburo, 1979. Evaluation of mass screening program for stomach cancer, *National Cancer Institute Monograph*, no. 53, p. 181–186.

Puska Pekka, Tuomilehto Jaakko, Salonen Jukka T., Nissinen Aulikki, Virtamo Jarmo, Björkqvist Stig-Göran, Koskela Kaj, Neittaanmäki L., Takalo K., Kottke Thomas, Mäki J., Sipilä P., and Varvikko P., 1981. *Community control of cardiovascular diseases the North Karelia Project: Evaluation of a comprehensive community program for control of cardiovascular diseases in 1972–1977 in North Karelia.* Copenhagen, World Health Organisation, Regional Office for Europe, 351 p.

Puzin Daniel (ed.), 1996. *La juste place des hôpitaux universitaires dans les systèmes de santé.* Paris, Flammarion, 102 p.

Ravenholt Reimert T., 1990. Tobacco's global death march, *Population and Developement Review*, vol. 16(2), p. 213–240.

Robin François and Nativi François, 1987. *Enquête sur l'erreur médicale.* Paris, La Découverte, 355 p.

Roques Bernard, 1999. *La dangerosité des drogues: rapport au secrétariat d'état à la Santé*. Preface by Bernard Kouchner. Paris, Odile Jacob. La Documentation française, 316 p.

Rosero-Bixby Luis, 1985. The case of Costa Rica, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 341–370. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Rosero-Bixby Luis, 1985. Le Costa Rica, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort*, p. 331–356. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Royal College of Physicians, 1962. *Smoking and health: summary and report of the Royal College of Physicians of London on smoking in relation to cancer of the lung and other diseases.* New York, Pitman Publishing Co., 70 p.

Ruiz Magda and Rincón Manuel, 1996. Mortality from accidents and violence in Colombia, *in:* Ian Timaeus, Juan Chackiel, and Lado Ruzicka (eds.), *Adult mortality in Latin America*, p. 337–358. Oxford, Clarendon Press, 367 p. (International Studies in Demography).

Selden Mark, 1972. *Health and revolution. The Chinese medical system, essay on contemporary China.* New York, Peck, Schurman & Riskin.

Shahpar Cyrus and Li Guohua, 1999. Homicide mortality in the United States, 1935–1994: age, period, and cohort effects, *American Journal of Epidemiology*, vol. 150(11), p. 1213–1222.

Shan Hua, Wang Jing-Xing, Ren Fu-Rong, Zhang Yuan-Zhi, Zao Hai-Yan, Gao Guo-Jing, Ji Yang, and Ness Paul M., 2002. Blood banking in China, *The Lancet*, vol. 360, p. 1770–1775.

Shkolnikov Vladimir and Nemtsov Alexander, 1997. The anti-alcohol campaign and variations in Russian mortality, *in:* Jose Luis Bobadilla, Christine A. Costello, and Faith Mitchell (eds.), *Premature death in the New Independent States*, p. 239–261. Washington DC, National Academy Press, 504 p.

Soucat Agnès, Bangoura Ousmane, and Knippenberg Rudolf, 1997. L'initiative de Bamako au Bénin et en Guinée, *in:* Joseph Brunet-Jailly (ed.), *Innover dans les systèmes de santé. Expériences d'Afrique de l'Ouest*, p. 193–208. Paris, Karthala, 435 p.

Sournia Jean-Charles, 1986. *Histoire de l'alcoolisme.* Paris, Flammarion, 322 p.

Taylor Carl E., Sarma R. S. S., Parker Robert, Reinke William A., and Farugee Rashid, [1968] 1983. *Child and maternal health services in rural India: the Narangwal experiment. Vol. 2: Integrated family planning and health care*. Baltimore, Johns Hopkins University Press, 252 p.

United Nations, 1995. *Population and development. Vol. 1, Programme of action adopted at the International conference on population and development, Cairo, 5–13 September 1994.* New York, United Nations, Department for Economic and Social Information and Policy analysis, Population Division, 100 p.

United Nations, 2000. *United Nations Millennium Declaration.* New York, United Nations, 9 p. (Resolution 55/2 adopted by the General Assembly of the United Nations).

Vallin Jacques, 1979a. Facteurs socio-économiques de la mortalité dans les pays développés, *in:* OMS/WHO (ed.), *Délibération de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité. El Colegio de Mexico, Mexico, 19–25 June 1979*, p. 266–302. Geneva, New York, OMS/WHO, PAHO, 622 p.

Vallin Jacques, 1979b. Morbidité dans la région parisienne: apport et intérêt des statistiques de la Caisse primaire centrale d'assurance maladie, *Population*, vol. 34(1), p. 109–140.

Vallin Jacques and Chesnais Jean-Claude, 1975. Les accidents de la route en France. Mortalité et morbidité depuis 1953, *Population*, vol. 30(3), p. 443–478.

Vallin Jacques and Meslé France, 1999. *Le rôle des vaccinations dans la baisse de la mortalité.* Paris, INED, 16 p. (Dossiers et recherches, no. 74).

Vallin Jacques and Meslé France, 2001. Évolution de la mortalité en Europe depuis 1950: mortalité par âge, sexe et causes de décès, *in*: Conseil de l'Europe/Council of Europe (ed.), *Tendances en matière de mortalité et mortalité différentielle*, p. 33–188. Strasbourg, éditions du Conseil de l'Europe, 344 p. (Études démographiques no. 36).

Victora Cesar G., Bryce Jennifer, Fontaine Olivier, and Monasch Roeland, 2000. Reducing deaths from diarrhoea through oral

rehydration therapy, *Bulletin of the World Health Organization*, vol. 78(10), p. 1246–1255.
WHO, [1946] 1989. *Constitution of the World Health Organization*. Geneva, World Health Organization, 18 p.
WHO, 1983. *International Health Regulations (1969)*, 3rd ed., annotated. Geneva, World Health Organisation, 79 p.
WHO, 1985. *Handbook of resolutions and decisions of the World Heath Assembly and of the Executive Board. Volume 2, 1973–1984.* Geneva, World Health Organization, 440 p.
WHO, 2001. *General programme of work 2002–2005.* Geneva, World Health Organization, 10 p. (Available at http://www.who.int/gb/ebwha/pdf_files/WHA54/egpw0205.pdf).
WHO, 2003. *Alert, verification and public health management of SARS in the post-outbreak period.* Geneva, WHO. (Retrieved July 19, 2004, from http://www.who.int/csr/sars/postoutbreak/en/).
Yach Derek, 2000. *The tobacco epidemic: a global overview.* Geneva, World Health Organization, 5 p. (Presented at "Global Dialogue for Health," EXPO 2000 Hannover, Germany, August 29–31, 2000).

APPENDIX Acronyms

English Names		French Names	
[AFM]		AFM	Association française contre les myopathies
AIDS	Acquired Immunodeficiency Syndrome	SIDA	Syndrome d'immunodéficience acquise
[AME]		AME	Aide médicale d'État
[AMG]		AMG	Aide médicale gratuite
[ATD-QM]	ATD Fourth World	ATD-QM	Aide à toutes détresses—Quart-Monde
[BCG]		BCG	Bacille de Calmette et Guérin
[CEPED]		CEPED	Centre français sur la population et le développement
[CHU]		CHU	Centre hospitalier universitaire
[CMU]		CMU	Couverture médicale universelle
[CNAM]		CNAM	Caisse nationale d'assurance maladie
[DDASS]		DDASS	Direction départementale de l'action sanitaire et sociale
DDT	Dichloro-diphenyl-trichloroethane	DDT	Di (para-chloro-phenyl)-trichloroethane
EPI	Expanded Program on Immunization	PEV	Programme élargi de vaccinations
FAO	Food and Agriculture Organization	[FAO]	Organisation pour l'agriculture et l'alimentation (OAA est inusité)
GDP	Gross Domestic Product	PIB	Produit intérieur brut
HBV	Hepatitis B virus	VHB	Virus de l'hépatite B
HIV	Human immunodeficiency virus	VIH	Virus de l'immunodéficience humaine
HRW	Human Rights Watch	[HRW]	
IAEA	International Atomic Energy Agency	AIEA	Agence internationale pour l'énergie atomique
IBRD	International Bank for Reconstruction and Development	BIRD	Banque internationale pour la reconstruction et le développement
ICPD	International Conference on Population and Development	CIPD	Conférence internationale sur la population et le développement
ICSID	International Centre for Settlement of Investment Disputes	CIRDI	Centre international de réglement des différends internationaux
IDA	International Development Association	AID	Association internationale de développement
IFC	International Finance Corporation	SFI	Société financière internationale
IHR	International Health Regulations	RSI	Règlement sanitaire international
ILO	International Labour Office	BIT	Bureau international du travail
ILO	International Labour Organization	OIT	Organisation internationale du travail
IMF	International Monetary Fund	FMI	Fonds monétaire international
[INED]		INED	Institut national d'études démographiques
[INSERM]		INSERM	Institut national de la santé et de la recherche médicale
[IRD]		IRD	Institut de recherche pour le développement
IUSSP	International Union for the Scientific Study of Population	UIESP	Union internationale pour l'étude scientifique de la population
[IVS]		IVS	Institut de veille sanitaire
JAMA	Journal of the Medical Association of America	[JAMA]	
LN	League of Nations	SDN	Société des nations
MDG	Millennium Development Goal	[MDG]	Objectif de développement pour le millénaire
MGA	Muscular Dysmorphy Association	[MGA]	
MIGA	Multilateral Investment Guarantee Agency	AMGI	Agence multilatérale de garantie des investissements
MRI	Magnetic Resonance Imaging	IRM	Imagerie par résonance magnétique
NGO	Nongovernmental Organization	ONG	Organisation non gouvernementale

(continues)

APPENDIX (*continued*)

English Names		French Names	
NHS	National Health System	[NHS]	
NIH	National Institute for Health	[NIH]	
OAU	Organization of African Unity	OUA	Organisation de l'unité africaine
OECD	Organization for Economic Cooperation and Development	OCDE	Organisation pour la coopération et le développement économique
OEEC	Organization for European Economic Cooperation	OECE	Organisation européenne pour la coopération économique
[OIHP]		OIHP	Office International d'Hygiène Publique
OXFAM	Oxford Committee for Famine Relief	[OXFAM]	
PAHO	Pan American Health Organization	[PAHO]	
[PMI]		PMI	Protection maternelle et infantile
PSA	Prostate Specific Antigen	[PSA]	
[PUF]		PUF	Presses universitaires de France
[SAMU]		SAMU	Service d'aide médicale d'urgence
SARS	Severe Acute Respiratory Syndrome	SRAS	Syndrome respiratoire aigu sévère
[SCPRI]		SCPRI	Service central de protection contre les rayonnements ionisants
[SEITA]		SEITA	Service d'exploitation industrielle des tabacs et allumettes
UNAIDS	Joint United Nations program on HIV/AIDS	ONUSIDA	Programme commun des Nations unies sur le VIH/SIDA
UNDP	United Nations Development Program	PNUD	Programme des Nations unies pour le développement
UNESCO	United Nations Educational Scientific and Cultural Organization	[UNESCO]	Organisation des Nations unies pour l'éducation la science et la culture
UNFPA	United Nations Funds for Population Activities	FNUAP	Fonds des Nations unies pour les activités en matière de population
UNHCR	United Nations High Commissioner for Refugees	HCR	Haut commissariat des Nations unies pour les réfugiés
UNICEF	United Nations International Children's Emergency Fund	[UNICEF]	Fonds international des Nations unies pour la sauvegarde de l'enfance
UNIDO	United Nations Industrial Development Organization	ONUDI	Organisation des Nations unies pour le développement industriel
UNO	United Nations Organisation	ONU	Organisation des Nations unies
WFP	World Food Program	PAM	Programme alimentaire mondial
WHO	World Health Organization	OMS	Organisation mondiale de la santé
WHOSIS	WHO Statistical Information System	[WHOSIS]	
WTO	World Trade Organization	OMC	Organisation mondiale du commerce

CHAPTER

108

Health Policies: Can the Results Be Evaluated?

JACQUES VALLIN AND FRANCE MESLÉ

Institut national d'études démographiques (INED), Paris, France

Can the results of health policies be evaluated? There is no simple answer to this question. First of all, it is possible to accept several notions concerning health policies according to whether one considers a health policy, in the strict sense, as being a policy defined by a health authority with the precise objective of improving health, or whether one refers, on the contrary, to all public measures by whatever public body concerned which render it possible, knowingly or unknowingly, to improve health. Moreover, whatever limits are set, the evaluation of the global effect of the policy implemented must be distinguished from that of the specific effects obtained from a given measure taken to attain a given objective. Between the two, one might focus on either the action of a given collective body and attempt to evaluate the global result of the measures it has taken or, in another register, consider a given objective and measure the results obtained in this precise area thanks to a raft of measures taken by the different organizations concerned. One might also vary this almost infinitely by multiplying the different possible combinations of actors/objectives/measures taken. We will of course not be able to review all the possible subjects of evaluation, but we will attempt to give a few examples at different possible levels of evaluation.

However, the complexity of the subject is far from being the only obstacle to obtaining an answer. We will only refer to two obstacles, which are sometimes insurmountable given the current state of knowledge. On the one hand, if one may truly speak of a result, can any such result really be attributed to a public measure? Or may it not be linked to other factors? The question is already a very tricky one when it comes to measuring the effect of a specific measure. It is true that we have in this case a certain number of theoretical methods, but often the data from observations are missing in order to apply such methods unless a research-action program is designed beforehand with the measuring of the result as an inherent aspect. However, the situation becomes practically insoluble when looking at the global results of policies. In most cases, it will be necessary to resort to explanations that are practically unverifiable, unless one day, observation systems should become available, such as those that will be evoked in the following chapter by Viviana Egidi and Vittoria Buratta, possibly blazing the trail for a complete monitoring of operations.

On the other hand, measures taken to improve health may require varying lengths of time before producing an effect. Therefore, there are even fewer cases in which perfectly adequate data for measuring results are available. Moreover, the effect of a measure taken may be spread over time. A sudden reduction in alcohol consumption may, for example, have immediate effects on the mortality of alcoholics whose death was only awaiting the absorption of the last, lethal dose, but it will also have effects in the longer term on the evolution of different ailments linked to alcohol. There are even cases where one measure taken may have positive effects in the short term and negative effects in the long term. For example, a campaign for the eradication of malaria based on the destruction of mosquitoes using an insecticide may result in a rapid decline in the endemic disease in the short term, but if this also leads to the development of the resistance of mosquitoes to insecticides, the long-term consequence may be a strengthening of the disease.

Therefore, we cannot assure the reader that after reading this chapter he or she will have complete and final answers to the question. However, we do hope to provide a few ideas as to how to approach the problem by preparing the terrain on three levels: the appraisal of the global results of a health policy, the measuring of the specific effects obtained by a program targeting a specific pathology, and, finally, the evaluation of the impact of specific measures taken in favor of target groups. On the other hand, in each case, we will also try to do more than answer the question, "Is it possible to measure?" since, whether the answer is positive or negative, we will at least try to discuss, each time, the most plausible hypotheses concerning the reality of the effects obtained.

I. GLOBAL RESULTS: HEALTH POLICIES AND INCREASED LIFE EXPECTANCY?

There are two potential pitfalls in trying to globally assess the results of a health policy. The first, already mentioned, is that of not being able to decide between a stricter or wider definition of the health policy. Should it be limited exclusively to the actions of organizations specifically in charge of the question, as for example, a ministry of health? Certainly not, if we wish to remain consistent with the position adopted in the previous chapter. But should all collective actions aimed at improving health, even if not acknowledged as such, be included? Here again, we would adopt a wider view. After all, has not the World Health Organization (WHO) itself explicitly included in its strategy of primary health care the development of basic agricultural production? And it is a known fact that education has an impact on mortality decline. Therefore, all policies having an effect on health must be evaluated. Furthermore, a health policy is decided not only at the government level, but many other collective bodies, whether infranational or supranational, public or private, play a part, including the populations concerned, whether through a juxtaposition of converging individual behaviors or through associations. However, is there not another pitfall, that is, the risk of confusing the results of the policies (in the wider sense of the term) that have an effect on health and life expectancy and those of the simple development "of the progress of the human mind" in the way Marie-Jean-Antoine de Condorcet ([1795]) understood it? There are two pitfalls therefore, but also two ways of considering things globally, which may combine, since if one considers that man is a "political animal" according to Aristotle[1] and that all human progress is fundamentally political in nature, why not extend our ambition so far as to evaluate the effect of this human progress on human health and life expectancy? But what progress and in what field? In the medical domain? Or in the economic and social domains? We have come full circle. Finally, are not all attempts at global evaluation in vain? Let us at least briefly evoke what such an evaluation might be before admitting that the context must undoubtedly be taken into account, simply for reasoning in terms of comparable economic development or better maybe, at a comparable stage in the health transition.

1. Can a Global Approach Be Attempted?

Can it be said, that in a given country, life expectancy is higher because, globally, all of the policies that have an influence on it are better? In wanting to encompass everything, is there not a risk of being reduced to expressing such banalities as "The best health policy is that of the country that has the highest life expectancy"? In that case, it would obviously be easy to claim without even having to check the figures that the policy of the United States is better for health than that of Nigeria! Would it be more credible to base the comparison no longer on the level of life expectancy attained but on its progression observed over a given period? This is far from certain. Can it be said, for example, that policies implemented in the second half of the 20th century were better for health in Tunisia than in France, since from 1946–1951 to 1997 life expectancy in the former country improved by almost 40 years, increasing from 33.3 years to 71.9 years (Waltisperger *et al*., 2001), whereas it only increased by 12 years in France, from 65.8 years in 1948 to 78.5 years in 1997 (Vallin and Meslé, 2001a)? Obviously the comparison is not justified, as the effort concerning health (in the wider sense) to be deployed to gain years of life expectancy starting from the age of 33

[1] In his "Politika" Aristotle (384–322 BC) insists on the high degree of social organization of man as crowned by the organization of the state (Aristotle, [1874]. As Marcel Prélot wrote (1961, p. 6): "Ancient man, as defined by Aristotle thus appears as a "civic animal." The scope of the definition is greatly reduced when "zoon politikon" is translated as "social animal." The animal is also social, but only man is political. Instead of living in herds or hordes, his specific characteristic is to live integrated within the social organism constituted by the Polis, the City; and this is for him, both a natural and ideal necessity."

is not at all the same as that which would enable an improvement of the same magnitude over the age of 65. Would it be more justifiable to say that Tunisia has nonetheless implemented a better health policy than France because France took 160 years to go from a life expectancy of 34.9 years in 1806 to 71.5 years in 1967 (Vallin and Meslé, 2001a)? Even less so, it is the historical context that is not comparable, the medical means (to mention only the most obvious) having changed radically in the meantime. A better question would undoubtedly be to ask whether, in a given country, a change in policy produced a result in terms of health. But what kind of change? It is possible to imagine what a study of a specific change of direction concerning health practices, food production, or education might be: we will deal with the evaluation of the results of such changes in Section 2. But does the question lend itself to a global approach? This is doubtful. It would be necessary to refer to a strong policy change, globally revolutionizing policies followed in all areas. Let us explore two possibilities: comparing the evolution in the life expectancy of a country before and after a revolution or comparing two countries according to whether or not they have experienced the revolution in question.

a. Before and After a Revolution?

There have been plenty of revolutions during the last three centuries; what are lacking, however, are the data to explore this first approach. With the instruments for historical reconstructions available to us today (see Chapter 126), it would be conceivable to launch a vast research project on this question, but for now we will have to be content with answering it with just a few rudimentary elements.

According to the answers of the major historical demographic survey conducted in France by Louis Henry (1980; see also Fleury and Henry, 1958; Séguy *et al.*, 2001), the French Revolution seems to have brought about new policies that resulted in a leap forward in life expectancy, especially through a decline in infant mortality. According to this survey, life expectancy stagnated between ages 27.7 and 28.9 during the three decades of the 1760s, 1770s, and 1780s (Blayo, 1975) but increased by 10 years in the four following decades, reaching 38.8 years in 1820–1829.[2] It is a fact that important economic and social reforms were in preparation under the Ancien Régime and they might have also resulted in a marked improvement in life expectancy, but it is still possible that the Revolution, in spite of the negative effects of the violence, which undoubtedly would not have occurred without it, amplified the increase in life expectancy.

Can the same be said of the October 1917 Revolution in Russia? There is unfortunately too few data to risk such an assumption. But what can be said, for instance, of the advent of the Communist regime in a country such as Czechoslovakia? The fact that it occurred among the ruins caused by World War II obviously poses a problem. However, comparing only two periods of peace from 1921 to 1937, life expectancy in Czechoslovakia increased for men by 8.5 years in 16 years, progressing from 46.4 years to 54.9 years (OFS, 1974), whereas in the same lapse of time (1950–1966), it only gained 6.3 years under the Communist regime, from 60.9 to 67.3 years (OFS, 1980). The value of such a comparison may seriously be called into question. On the one hand, increasing life expectancy beyond the age of 61 years in the 1950s was perhaps more difficult than increasing it beyond the age of 46 in the 1920s. But above all, due to a lack of reliable data between 1937 and 1950, to what should the 6-year increase be attributed? To the last years of the Czechoslovakian democracy? To the German occupation? Or to the first years of the Communist regime? Unless it was simply due to the effect of the arrival of antibiotics, an effect independent of the political regime in place.

It would be interesting to discuss the case of China which, according to United Nations estimates (United Nations, 2003), had an increase in life expectancy of close to 18 years in the 15 years after Mao Tse-tung took power, increasing from 42.9 years in 1950–1955 to 60.6 in 1965–1970. Although we lack any solid point of comparison for the period before that, it is hard to deny the fact that this period marked a break in the evolution of the health status of the Chinese population (in the third part of this chapter we will come back to this issue). However, in Cuba, a country often cited as an example for its improvements in terms of health compared with other poor countries, when Fidel Castro came to power in 1959–1960, there was no break in the evolution of life expectancy estimated by the United Nations. Another interesting point to discuss would be that of the effects of decolonization, namely in Africa. In the case of Tunisia, for example, it may be assumed that the policy implemented during the first decade of Independence was better for improving the health of the population than the policy implemented by France during the last decade of the colonial era

[2] The survey conducted by Louis Henry does not provide results on all of the intermediary decades, but for females it gives the following progression: 28.1 in 1780–89, 32.1 in 1790–99, 34.9 in 1800–09, 37.5 in 1810–19 and 39.3 in 1820–29 (Blayo, 1975). According to our own reconstruction for the 19th century it seems that life expectancy for the two sexes reached close to 35 years from 1806 and 40 years in 1823 (Vallin and Meslé, 2001a).

because from 1956–1961 to 1968–1969 life expectancy leapt forward by 10.6 years, whereas from 1946–1951 to 1956–1961 it had only increased by 6.7 years (Waltisperger *et al*., 2001). But in a country like Niger, United Nations estimates tend to indicate the opposite, with an increase of only 4.0 years in life expectancy in the 15 years following Independence compared with 4.5 years in the previous 15 years.

Another possible field of study for this approach is the fall of Communism in Europe. The case is all the more interesting because it concerns countries that had experienced, since the middle of 1960s, a long period of stagnation, of regression even, in life expectancy before abandoning the Communist regime. But here again, the same event, this time occurring at exactly the same moment and in countries with highly comparable standards of living and health statuses, shows rather different results. Whereas countries such as Russia or the Ukraine plunged deeper into crisis (Meslé and Vallin, 2003b), others, such as Poland, the Czech Republic, and Hungary, started to experience an increase in life expectancy just after the fall of the Berlin Wall (Meslé, 2004b).

This global historical approach is not yet conclusive. It is true that major political changes rarely have a neutral impact on changes in health, but, assuming that the data exist to conduct a highly global analysis, such an analysis would be of little interest if it were not completed by an in-depth study of the aspect of the change that had a real effect on the health situation.

b. With or without a Revolution?

It would be possible to enhance the analysis by comparing, for a given event, two countries in other respects rather similar, one of which would have experienced the event and the other not. For example, a comparison may be attempted between the French Revolution and the case of England or Sweden. We saw that France had gained 11 years in life expectancy between 1780–1789 and 1820–1829. According to the estimates of Wrigley and Schofield (1981), between 1791 and 1826 England only gained 2.6 years (passing from 37.3 years to 39.9) and, according to estimates of Ingvar Holmberg (1970), Sweden only gained 1.6 years (from 37.9 years to 39.5). However, we hesitate to attribute this increase in life expectancy in France to the Revolution alone. One may well be intrigued by the fact that in all three cases, the point of arrival is the same (approximately 40 years life expectancy around 1826), which means that France in 1789 was very far behind Sweden and England. The success of revolutionary policies did no more than enable France to catch up. It seems that it was not so much due to the fact that the policies were better than those implemented in England and Sweden but that those of the French Ancien Régime were worse. There again, an in-depth analysis of the effects of the different changes brought about by the Revolution would be indispensable in order to reach a conclusion.[3]

Closer to home, one of the most interesting comparisons is the one of the changes observed between the two Germanys after the creation of the Federal Republic (FRG) and the Democratic Republic (GDR), in 1949. Starting with very similar levels (even though very precise data are missing), life expectancy in the two Germanys evolved rather differently after their separation (Figure 108–1). Initially, after a few years of Communist rule, the life expectancy of East Germans caught up with, and even overtook, that of West Germans, the progression of which started to flag at the beginning of the 1960s. Therefore, in the period between 1970 and 1974, life expectancy in the GDR maintained a half-year advance on that of the FRG. Milton Roemer (1985) attributed this phenomenon to the fact that while West Germany had just kept the health insurance system implemented by the Bismarkian reforms of the 19th century, East Germany had developed a health system entirely run by the State and which focused on prevention and universal access to comprehensive health services free of charge. This explanation was all the more plausible as the difference in life expectancy, to the advantage of the GDR, was accompanied with an even greater difference in the standard of living to the benefit of the FRG. However, a short time afterward, whereas the differences between policies remained, life expectancy in the FRG resumed a rapid increase while that of the GDR stopped growing. On the eve of reunification, the difference was much greater, but this time it was to the advantage of the FRG: 2.5 years in 1989 and even 3.1 in 1990.

This reversal of the situation can only be explained if one places oneself in the more general context of the health crisis that hit all of the European countries in the Communist block from the mid-1960s. In all the industrialized countries, when the decline in infectious mortality was such that to obtain new substantial increases in life expectancy it was necessary to counter the growing negative effects of industrialization and to fight mortality in new areas, Western countries managed to control manmade diseases and found

[3] Whilst bearing in mind the fact that the differences in life expectancy observed for the 18th century may have been partly due to differences in the methods of estimation employed (see Chapter 68).

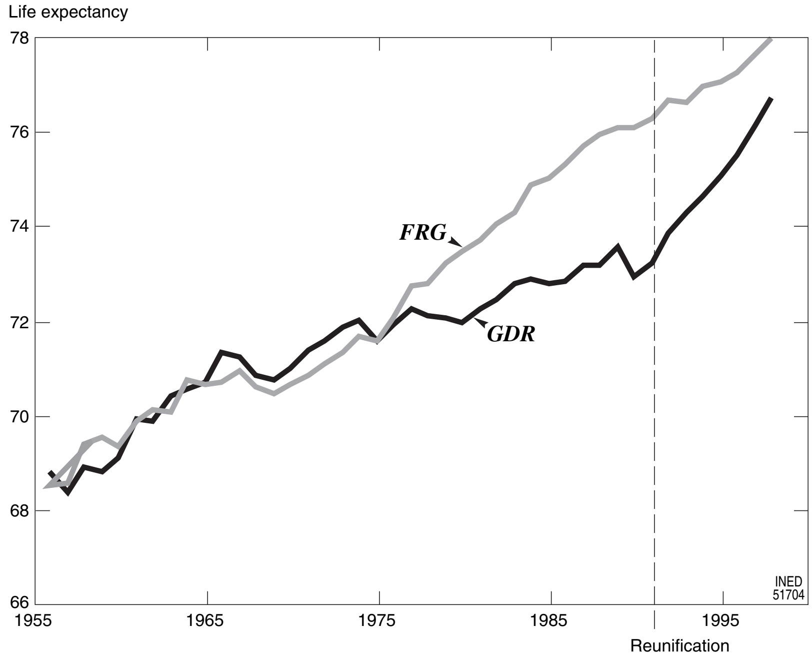


FIGURE 108–1 Compared changes in life expectancy at birth in East Germany and West Germany (both sexes). (Source: Human Mortality Database: http://www.humanmortality.org; Scholz and Maier, 2003.)

means of rapidly reducing cardiovascular mortality that Eastern block countries were unable to adopt, while the burden of manmade diseases was developing (Meslé and Vallin, 2002). However, in a third phase, with the reunification and the extension within the GDR of the West-German political system, the increase in life expectancy of East Germans accelerated, making up for a great deal of the difference in a few years. This difference was of only one year in 1999.

Two major conclusions stem from these examples and especially from the latter. On the one hand, if one wishes to build a remotely intelligible reasoning based on international comparisons, one must absolutely be able to distinguish between the main factors behind improved health, and if one wishes to confine oneself to a global comparison the minimum requirement is undoubtedly to make sure that standards of living are relatively similar from the outset. Moreover and, maybe, above all, it should be noted that the same global political system may be advantageous for health at a given stage of the health transition and a disadvantage at another stage. It is therefore necessary to reason in terms of equal stages in the health transition.

2. At an Equal Level of Economic Development

In the study he carried out in 1985, Milton Roemer sought to compare, two by two, developed countries (with the exception of the two Germanys) having not only comparable per-capita incomes but also, and more generally, major sociocultural similarities. He thus compared the United States and Canada, Belgium and the Netherlands, Austria and the United Kingdom, and New Zealand and Australia.

As he observes, in spite of a slightly lower rate of income per capita, a clearly less hospitable climate, and a substantially smaller share of national income dedicated to health expenditure, Canada benefited from a life expectancy greater—by one year—than that of the United States.[4] He attributed this difference to the fact that the United States is one of the few Western industrial nations that has not adopted a collective system of health care coverage, whereas since 1962 Canada has had a provincial-federal program for medical and hospital insurance covering practically the whole population (Roemer and Roemer, 1981). Even better, according to Milton Roemer, other countries, where income per capita is substantially lower than in the United States, such as Japan, Sweden, and the Netherlands, also have higher life expectancies, thanks to their collective coverage of health care expenditure. Roemer goes on to say that the argument

[4] Milton Roemer quoted the latest result available at the time which were life expectancies for 1977: 74.3 in Canada and 73.3 in the United States. In 2000, the difference, in the same direction, was of 2.2 years for women (with 81.9 years compared with 79.7) and 2.4 for men (76.7 compared with 74.3).

often put forward to challenge the assumption of a greater heterogeneity of the American population, which includes large African-American and Latin-American minorities whose health is apparently weaker, is not valid because even when only the white American population is considered, life expectancy in the United States remains lower than that of the overall population of Canada and of those of the other countries quoted.

In the same way, even between countries with advanced social welfare systems, Milton Roemer showed that the result in terms of life expectancy is better when the social welfare coverage is complete. He thus attributed the difference in life expectancy observed between Belgium and the Netherlands at the end of the 1970s to the fact that, in the Belgian system, health insurance was much more imbued with liberalism (imposing on patients a large financial participation, which penalized the poor) than the Dutch system in which almost everyone could consult doctors without paying anything. And it is for the same reason that the life expectancy of Austrians at the time was lower than that of the British who apparently had a lower standard of living or of that of the Australians which was no higher than life expectancy among New Zealanders, although their income per capita was almost twice as high.

In the developing world there are also many examples of countries that are culturally similar and which had, at the outset, comparable living standards but where life expectancy has evolved differently. Figure 108–2 presents several cases based on life expectancy estimates provided by the United Nations for the period 1950–2000 (United Nations, 2003). In East Africa, for example, it is possible to contrast, at least in a first instance, the high increase in life expectancy observed in Kenya and Zambia to the slow progress of Uganda and Malawi. It probably is due to the substantial efforts made by Kenya and Zambia in favor of economic and social development in general, but also of primary health care. Subsequently, however, AIDS has hit these countries more violently, eating away at a large proportion of the progress made. Is this to say that their mode of development has made them more fragile in the face of the AIDS epidemic? It may also be that Uganda, which has resisted the epidemic better than the other countries in spite of being hit first, has adopted a better strategy in the fight against AIDS.

In Western Africa, before the advent of AIDS, Côte d'Ivoire (Ivory Coast) or Benin apparently implemented policies for development and for health that were much more efficient than those implemented in Niger. It may of course be said that Côte d'Ivoire was able to bolster such policies thanks to much more favorable economic prospects, but this was not so true for Benin. On the other hand, as in the case of Benin, Côte d'Ivoire was harshly hit by AIDS while Niger was spared, which reduced the difference in life expectancy to nothing. In West Africa again, two English-speaking countries belonging to the same subregion, Sierra Leone and Gambia, had life expectancies that diverged greatly. Whereas Gambia, a small country quietly implementing a development policy that also addressed the issue of health, had a much faster-than-average increase in life expectancy for sub-

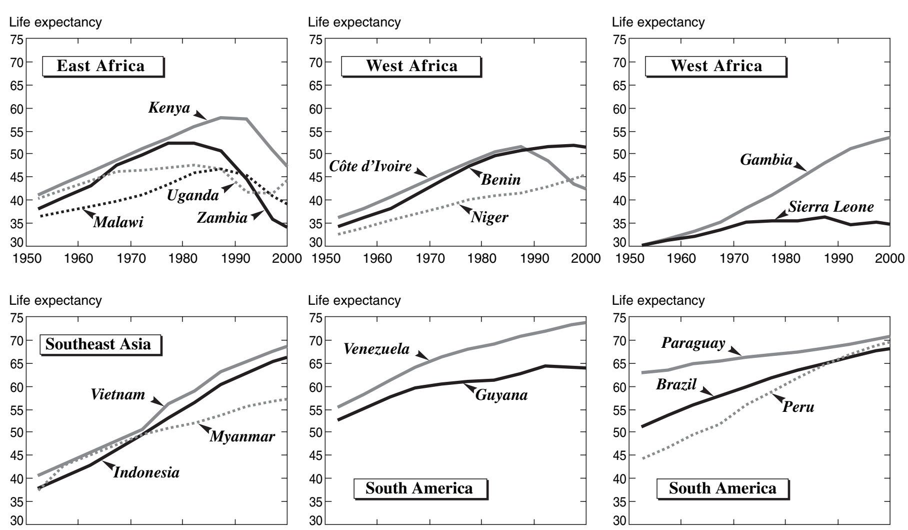


FIGURE 108–2 Compared changes in life expectancy in several countries belonging to the same cultural area and with comparable baselines for standards of living. (Source: United Nations, 2003.)

Saharan Africa, life expectancy stagnated in Sierra Leone without it being possible to attribute this to the recent political-military events only, however dramatic these may have been.

In Southeast Asia, in Indonesia and to an even greater extent in Vietnam, in spite of 30 years of war out of the 50 years observed here, life expectancies increased much faster than in Myanmar, a country which, in spite of its potential wealth, is far from being a star in terms of economic and social-sanitary development.

Figure 108–2 also illustrates two types of contrasts in South America. Starting with comparable levels of life expectancy in 1950, two countries of the Guyana Plateau, Venezuela and Guyana, have diverged a great deal. Due in large part to its exploitation of oil resources, Venezuela has benefited from a strong increase in life expectancy. In Guyana, on the contrary, life expectancy has only increased slightly, especially since Independence, which did not produce policies equal to the hopes it had fostered.

Inversely, Peru, Brazil and Paraguay, starting with very different levels of life expectancy in 1950 have converged greatly. Peru, an Andean country with difficult living conditions, has made spectacular progress nonetheless, whereas Paraguay has progressed only very slowly. Here again, differences in the approach to economic, social, and health development undoubtedly have something to do with this.

3. At a Given Stage of the Health Transition

There are two different angles from which it is possible to try to evaluate the effects of policies on changes in life expectancy, according to whether one considers countries with similar conditions at the outset and for which a divergence in mortality changes are observed or whether, on the contrary, one compares countries with relatively low levels of life expectancy with more advanced countries that the former are catching up with. In a recent study (Vallin and Meslé, 2004), we showed that, at the level of the general history of health improvements, these phenomena of divergence and convergence could be related to an explanatory model that should help provide a better understanding of the relations between these movements of life expectancy and the overall effectiveness of policies. The idea is that when a major health innovation emerges, such as vaccines and antibiotics in the past or, more recently, new strategies in the fight against cardiovascular diseases, a certain number of pioneering countries are able to develop policies to ensure that the entire population benefits from these innovations. In these countries life expectancy increases rapidly, whereas in the others it stagnates (unless they have not yet fully exploited all the resources of the previous innovation and continue to do so). Subsequently, however, countries lagging behind re-orient their policies to acquire the means to benefit from the new technologies, and they do so all the more easily when these technologies become increasingly accessible, whereas the pioneers tend to reach their upper limit because they have almost exhausted their resources. The countries lagging behind are then able to catch up with the pioneers, thanks to a more rapid increase in their life expectancy.

If one reasons in terms of a group of countries studied at a comparable stage of economic development, for example, industrialized countries, and if following an innovation in health care some of these countries progress rapidly whereas others stagnate, it would be normal to attribute the advantage gained by the leaders to a better overall policy. On the contrary if situations converge, it would not be possible to claim that the countries where life expectancy is increasing fastest, to the point of catching up with those that were in advance, practice a better policy than the pioneers because the latter find themselves blocked by the absence of another innovation. These countries are not at the same stage of the health transition. However, it is possible to compare countries that are lagging behind according to their capacity to catch up.

We have shown that in two or three centuries of health transition, there have been at least two major divergence-convergence movements of this type and that a third may have started. The first is related to the discovery of modern means of fighting against infection. It initially caused profound divergences between several pioneering industrialized countries and the others, and then the divergences developed between developed countries and developing countries until there emerged, especially right after World War II, a global movement of convergence. The second, linked to the implementation of policies able to check man-made diseases and of efficient means to fight against cardiovascular diseases, concerns countries which in the 1960s had reached a stage at which possible gains from fighting against infectious diseases had become negligible, basically, the industrialized countries. The third may have already begun due to the major successes of some countries, which are leaders in the fight against aging and in mortality decline at advanced ages.

The most tangible example of divergence-convergence and of the global role of policies in the explanation is certainly the one that has, in the last 40 years, opposed industrial countries in the East and the West (in the very political sense of these two cardinal

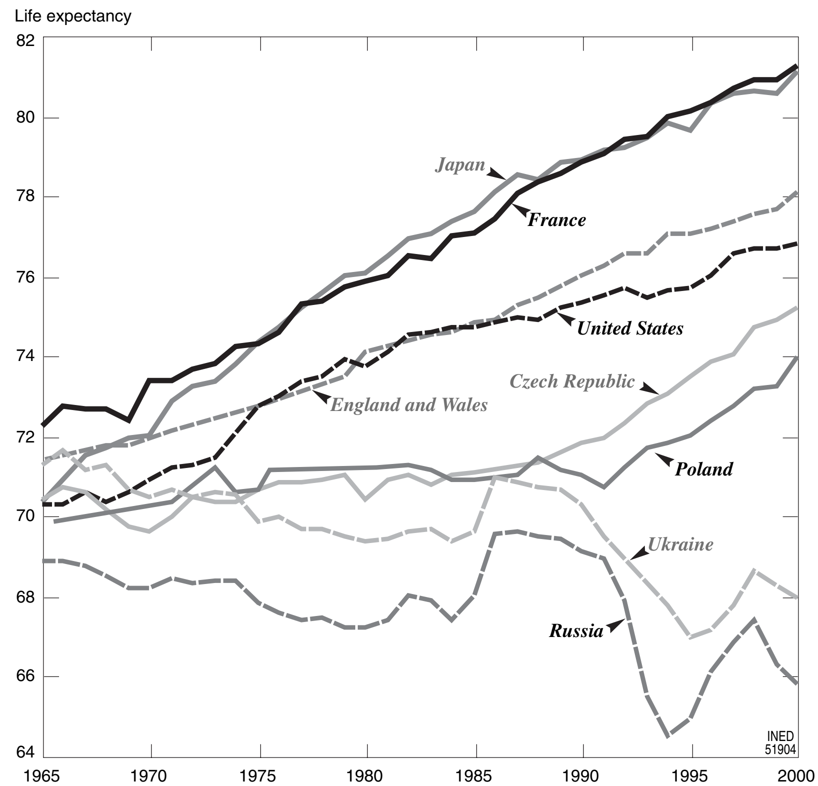


FIGURE 108–3 Compared changes in life expectancy in four Western countries and four Eastern countries since 1965 (sexes combined). (Source: Vallin and Meslé, 2004.)

points). Figure 108–3 illustrates the evolution in life expectancy among the most significant countries, four in the West and four in the East (Vallin and Meslé, 2004).

At the time these eight countries, all major industrial countries in the developed world in the beginning of the 1960s, had very similar life expectancies (between 69 and 71 years) and had reached levels of economic development, which even though dissimilar (it is impossible to deny the fact that Eastern countries were at a disadvantage) were very much in advance of the Third World. Everywhere, infectious mortality was reduced to such a degree that subsequent improvements could no longer have a substantial impact on life expectancy, whereas increases in life expectancy were slowed down or even stopped because it was impossible to make any substantial gains on other fronts and because of the growing threat of man-made diseases (traffic accidents, smoking, alcoholism, etc.). From the second half of the 1960s, however, Western countries simultaneously managed to curb man-made diseases and reduce cardiovascular diseases, whereas Eastern countries were unable to do so. In the West, increasingly well-targeted policies concerning the consumption of tobacco, road safety, work accidents, alcoholism, and so forth bore fruit while new means were used in the fight against cardiovascular diseases: technological means, of course, but also changes in behaviors, namely in eating habits. In Eastern countries, on the contrary, the mode of industrial development made the fight against pollution and accidents more difficult, whereas the decline in the value of individual initiatives created mentalities that were not very conducive to adapting new behaviors in the prevention of cardiovascular diseases (which require the participation of each individual) in countries where people were used to everything being done by the State. Hyper-centralized policy strategies that had proved very efficient in making available all the simple means for fighting against infectious diseases turned out to be inoperative for the much more subtle and complex fight against cardiovascular diseases. Of course, to these reasons must be added the fact that the arms race and space exploration led the Soviet regimes to economic and budgetary choices that had negative on the development of other sectors, namely that of health, depriving these countries of the means of deploying new technologies in the fight against cardiovascular diseases for the benefit of the greater number. The divergence of life expectancies observed from 1965 to 1980 also directly results from the global political divergence between the two groups of countries.

From the end of the 1980s, however, this situation evolved even before the fall of the Berlin Wall. In the USSR (in Russia and Ukraine; see Figure 108–3), in 1985, the campaign against alcoholism conducted by Gorbachev resulted in a leap forward in life expectancy. This operation has since fizzled out, and the tendency is toward deterioration in spite of substantial fluctuations. On the contrary, in the Czech Republic, where the policy started to change even before the country abandoned Communism, growth in life expectancy resumed from 1988–1989, and Poland followed suit two or three years later. These two countries are now on the verge of catching up with the countries of the West, at least concerning cardiovascular diseases, since in the West, Japan, France, and a few other countries may have already begun a new stage (Vallin and Meslé, 2004), leaving behind other countries such as the United States.

However, even making comparisons at comparable stages of health development, the benefit of such a highly global approach to the effectiveness of policies soon reaches its limit. It is not possible to determine what, within a general policy orientation, is truly responsible for success or failure in terms of improved health. It is not known, except if one takes information from other sources, which implemented measures actually have had which effects.

It is therefore necessary to examine the different measures or groups of measures (programs of action) implemented to attain the objective. This is what we will attempt to do now from two different angles: that of (a) programs aimed at reducing and even eradicating a disease or a cause of death and that of (b) policies aimed at globally improving health or the survival of a specific group within the population.

II. REACHING TARGETS IN PATHOLOGICAL TERMS

Concerning programs of action targeting specific pathologies, it is necessary to distinguish between several types of evaluations and to identify what we are truly interested in.

There are two very different fields of investigation in this area (Khlat, 1997b): (a) clinical trials, which make it possible to test a procedure from the point of view of its medical effectiveness as well as of its feasibility (acceptability, practicability, etc.); and (b) large scale programs of action, within the framework of a health policy for the benefit of the entire population. The first level is obviously fundamental; it is on the basis of its results that the political authorities may decide, in full possession of the facts, on such or such a health policy orientation aimed at making best use of scientific knowledge. For example, thanks to clinical trials it is possible to know whether one vaccine is more effective than another, whether a given antibiotic is more adapted in certain circumstances or if a procedure for the early screening of cancer may help in improving the survival of patients. In addition, many clinical trials are run, they are conducted according to strict scientific rules, and they provide proof of the effectiveness or ineffectiveness of a given procedure.

But it is obviously not this type of evaluation that we are interested in here. We have to judge, not the theoretical results of a procedure for the fight against a disease, but the actual consequences on the population of a program of action aimed at reducing, or even eliminating, a disease or a cause of death. However, there may be a great difference between theoretical efficacy of a procedure and its practical effectiveness when it is employed on a large scale in the field. For example, within the framework of a clinical trial, the protocol for the prevention or the treatment is administered under the direct and permanent control of specialized medical staff, whereas in the case of a large-scale program the application of the protocol depends largely on the willingness to participate and attention on the part of the public or the patients. Inversely, especially concerning contagious diseases, action on a massive scale may change epidemiological conditions and benefit the health of those who are outside the scope of the program. In other words, the same protocol may have more or less effectiveness in the field than efficacy in a clinical trial.

Between the two, there are what might be called epidemiological tests in which an attempt is made to test the effect in the field by applying the protocol in the same conditions as in a global policy but on a small sample of the population of which the changes in morbidity and mortality are followed as the program is implemented. At the same time the changes in the same indicators in another control sample not subjected to the program are followed (Khlat, 1997a). This type of test is not in itself quite what we are aiming for here since whatever is done there generally remains a substantial difference between the test and the real-life program itself. What is more, it is often difficult to make a comparison with the control sample since, apart from the ethical problems the method poses, the latter may indirectly benefit from spin-offs of the action conducted in the test area, as in the case of several studies in Bangladesh for example (Mostafa and Rahman, 1997; van Ginneken *et al.*, 1998). What we would like to discuss here, inasmuch as possible, is the real demographic effect of a large-scale program of action aimed at fighting a given disease or cause of death. It is in this area unfortunately that it is most difficult to obtain proof and where useful data are least abundant. We will reason here in terms of mortality given that morbidity data are even more incomplete.

A last preliminary precaution concerning language, given that a large-scale intervention is involved, has to be mentioned. As Myriam Khlat (1997b) points out, at this level, there are two types of evaluation that must be differentiated: that of the program and that of the program results. It is the second type—the program results—that interests us more.

1. Infectious Diseases

It was essentially the widespread use of vaccines and antibiotics that made the spectacular decline of infectious mortality possible in all of the developed countries initially and subsequently now in most developing countries, even though in the old Europe their decline had already begun thanks to improved diets and hygiene. Although in many developing countries, infection diseases remain major public health issues, this is also greatly due to the fact that the policies implemented have not yet made it possible for all populations to systematically benefit from them. Beyond this obvious fact we will, however, try to take stock of the progress of some of the major programs of action.

a. The Eradication of Smallpox

In Chapter 106 we evoked the long history of the fight against smallpox from the ancient Chinese practice of variolization until its near disappearance in developed countries thanks to the widespread use of the vaccine developed by William Jenner at the end of the 18th century and finally its total eradication all over the world. Thanks to political action, the large-scale use of Jenner's technique was made possible in Europe and North America, and it was political action orchestrated on a global scale by WHO that made its total elimination possible (see Chapter 107). This is the greatest historical example of a concerted and sustained fight against a disease, and it is the only case to date of the global eradication of a human disease through the implementation of policies voluntarily directed against it.

b. The Case of Measles

The case of measles is rather different. This disease had disappeared in developed countries as the main cause of infant deaths well before the first effective vaccine was developed in 1969 (Vallin, 1992; Vallin and Meslé, 1999). However, it continued, and still continues to be a serious disease and a major cause of infant and child mortality in a number of developing countries, namely in sub-Saharan Africa. Is the vaccine the tool on which the health policies of these countries should be based? Measles is undoubtedly the most interesting example for a discussion on the role of vaccination in the reduction of mortality. Since the end of the 1980s, there have been two opposing theses on this subject.

The first is based on the idea that measles is not a serious disease in itself but only becomes so through its possible complications, which only manifest themselves and turn to disaster if the child is badly fed (Morley, 1973; Walsh, 1983). Therefore, any policy based on seeking to provide good vaccination coverage against measles is pointless because the only children who risk dying from the disease are malnourished children who would die in the short term anyway even if not from measles (Mosley, 1985b). At the end of 1980s, however, the Danish anthropologist Peter Aaby (1988, 1989, 1992) conceived of a completely different explanation for the high lethality of measles in Africa, thus reaffirming the importance of anti-measles vaccination. According to him, the lethality of measles is not linked to a characteristic of the host (malnutrition) but to the mode of transmission, as children who receive a strong dose of the virus die more frequently, whatever their nutritional state, than those who receive only a small dose.

Until the middle of the 1990s, there were few studies that allowed valid comparisons between the all-cause death rate of children vaccinated against measles with that of non-vaccinated children, even though a number of limited studies supported Aaby's thesis (Hartfield and Morley, 1963; Kasongo, 1981; Aaby *et al.*, 1984), and, in a rather general manner in sub-Saharan Africa, the results of vaccination campaigns were rarely as effective as hoped. In 1996, however, Annabel Desgrées du Loû wrote a book on the subject based on the experience of Senegal. Her main conclusion was that, although the results obtained during the vaccination campaigns often seem disappointing, it is probably because they are limited to spectacular operations involving the massive distribution of vaccines at a given moment and without a permanent structure able to maintain the vaccination coverage thus obtained. On the contrary, in the rarer cases when such a follow-up is properly provided, measles disappears, which in turn results in a substantial decline in total mortality. But Annabel Desgrées du Loû also shows that, even in the case of a vaccination campaign that is not followed up, it is possible to observe a marked decline in total mortality, as long as the "crackdown" operation in question is sufficiently massive. The difference is that once the immediate benefit of the vaccination has passed measles returns and mortality increases again because the new cohorts of children are not vaccinated. It should also be noted, that for quite a long time afterward, the elder children are protected and the new measles epidemics are less violent.

More generally, recent studies prove that the progress made by the Expanded Program on Immunization (EPI) makes it possible to obtain substantial declines in infant mortality in Africa (see Chapter 107). Such is the case, for example, of a study in Sierra Leone (Amin, 1996). However, in Africa, at the national level, mortality is only measured approximately and occasionally, within the context of household surveys. No African country is able to measure annual changes in mortality and to link them, at this level, with that of vaccination coverage. It is, of course, even less realistic to hope to make a comparison at a national level between total mortality and mortality from measles, as the large-scale gathering of causes of death poses a huge problem.

However, there is much hope today that measles will soon be eradicated as was smallpox. In 1994, the 24th pan-American Health Conference set the objective of eradicating measles on both North and South America. The results have been remarkable due to systematic vaccination. Measles has ceased to be endemic in the Americas, most of the countries having proved that the transmission has been interrupted. The last case of indigenous transmission, detected in Venezuela, dates back to 2002 (de Quadros, 2004). It is true, that as long as eradication is not worldwide, vaccination remains necessary to maintain this situation in the Americas, but if a single concerted policy could be conducted with as much firmness on a global scale, total eradication could be possible: "A world free of measles by 2015 is not a dream" (de Quadros, 2004; p. 138).

c. Influenza

Influenza, which 40 years ago was still very deadly, is another story. After smallpox, influenza is perhaps the only major cause of mortality for which vaccination has played a decisive role in Europe. In France, for example (Figure 108–4), until toward the end of the 1960s, this very infectious disease was responsible for a mortality that varied between 5 and 50 per 100,000. The development of the first vaccine (1945) had no visible effect. For each flu epidemic, it was necessary to have a vaccine capable of anticipating the mutations of the virus. As soon as the polyvalent vaccine was put into service and introduced in France in 1970, mortality caused by flu collapsed. The disease is still an epidemic but it kills fewer and fewer people as the vaccination coverage progresses, namely among elderly people. From this point of view, the availability (free of charge in France) of this vaccine from the age of 65 and the campaigns to promote it have played a very important role in the decline in mortality at advanced ages. Flu epidemics were accompanied by high excess mortality due to associated diseases (pneumonia, bronchial pneumonia, etc.). At the same time their decline eliminated in France the last cause for annual fluctuations in total mortality and substantially contributed to the increases in life expectancy obtained during the 1970s and 1980s.

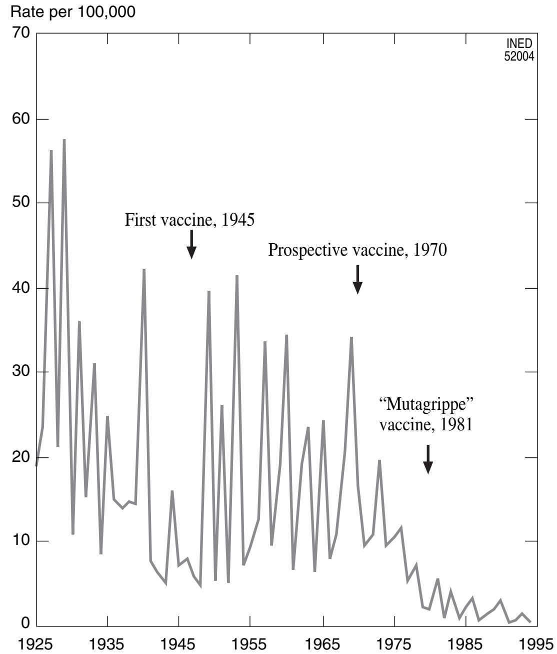


FIGURE 108–4 Changes in standardized death rates per flu epidemic, France, 1925–1994. (Source: Vallin and Meslé, 1999.)

d. Malaria

Malaria, caused by a parasite (*Plasmodium*) transmitted to humans by a mosquito (the anopheles), is still responsible for high mortality in many tropical countries, namely in Africa. Although it is one of the first tropical diseases that modern medicine has tackled, namely for the benefit of the colonizers whom it hit hardest, there still is no vaccine against it, in spite of periodic announcements. Two types of actions are implemented: (a) action against the parasite, through the administration of quinine, known about since the 18th century, or synthesized products developed after World War II, both preventively and curatively; and (b) action against the vector, either through destroying its favorable environment (draining of wetlands), through the setting-up of a barrier again the host vector (mosquito nets, protective clothing, etc. impregnated with repulsive products) or by the destruction of the mosquitoes themselves, through the pulverization of insecticides, the most famous being DTT, which was developed during World War II.

It is against this highly endemic disease in many developing countries that the World Health Organization directed one of its first major vertical health programs.This strategy was very successful in Sri Lanka at the end of the 1940s. Following a massive campaign in 1945–1946 to destroy the mosquitoes by systematically spraying DDT all over the island (70 000 km^2), malaria disappeared (only to return two decades later) and life expectancy made a spectacular leap forward from 46 years in 1945–1947 to 60 years in 1954. The relation between these two latter phenomena was indeed questioned (e.g., by Frederiksen, 1960, 1961, 1966) on the grounds that the mortality decline had already begun before and that the causes of the decline following the eradication campaign had no more to do with the results of this campaign than before. However, several meticulous studies have since confirmed the major role played by the eradication of malaria in the mortality decline (Newman, 1965, 1977; Gray, 1974; Molineaux, 1985). By analyzing mortality change by district, it is possible to compare the results obtained according to the highly variable degree of malaria infection in each district. Districts never having been infected can serve as control areas, as in a real-life test-case study (Figure 108–5). It is also possible to establish a model according to the degree of malaria infection of the districts. Assuming that without eradication of malaria, mortality decline would be the same in absolute value in all the districts, Paul Newman demonstrated that 43% of the decline of the crude death rate observed on the scale of the island between 1946 and 1960 (representing a reduction of 20% of the rate of 22 per 1,000 observed in 1946) was due to the eradication of malaria. However, Ronald Gray, believing this hypothesis underestimated the reduction of possible mortality without the eradication of malaria, considered it more appropriate to assume the decline in the absence of eradication to be proportional to the initial death rate. He thus showed that eradication had nonetheless caused a 27% decline in total mortality. Therefore neither the (temporary) success of the eradication campaign nor its major impact on total mortality may be denied.

Unfortunately eradication was not definitive, as malaria reappeared in Sri Lanka in the 1970s. As long as eradication is not obtained at a global level, the disease can reinfest a country and become endemic once again. Sri Lanka is the only example of the successful eradication of malaria at the level of an entire country. Malaria was indeed eradicated in Mediterranean areas, but the endemia had never been as strong and as generalized as in tropical regions. There is of course the experience of Guyana where malaria was also eradicated from the most populated part of the country, at the same time and with the same

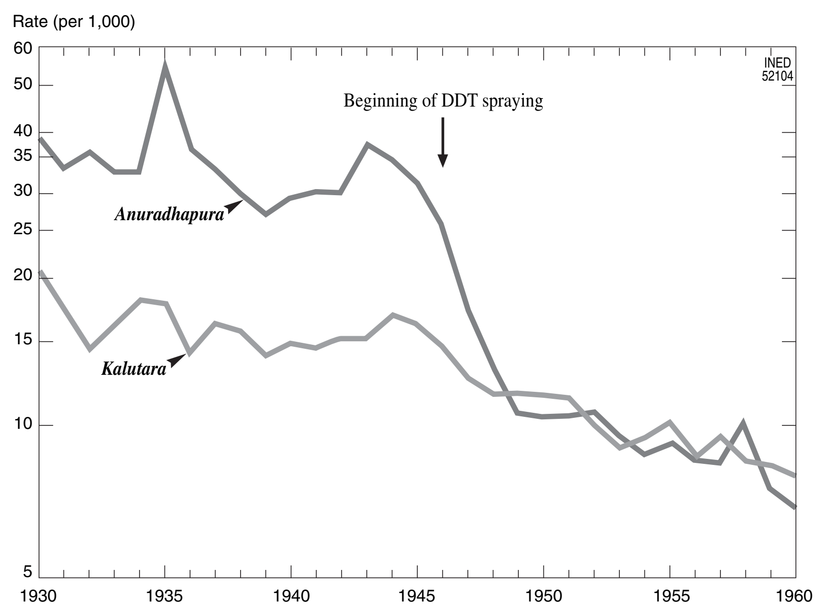


FIGURE 108–5 Sri Lanka: Evolution (before and after eradication of malaria) in the crude death rate in the district most infected with malaria (Anuradhapura) and in the district least infected with malaria (Kalutara). (Source: Newman, 1965.)

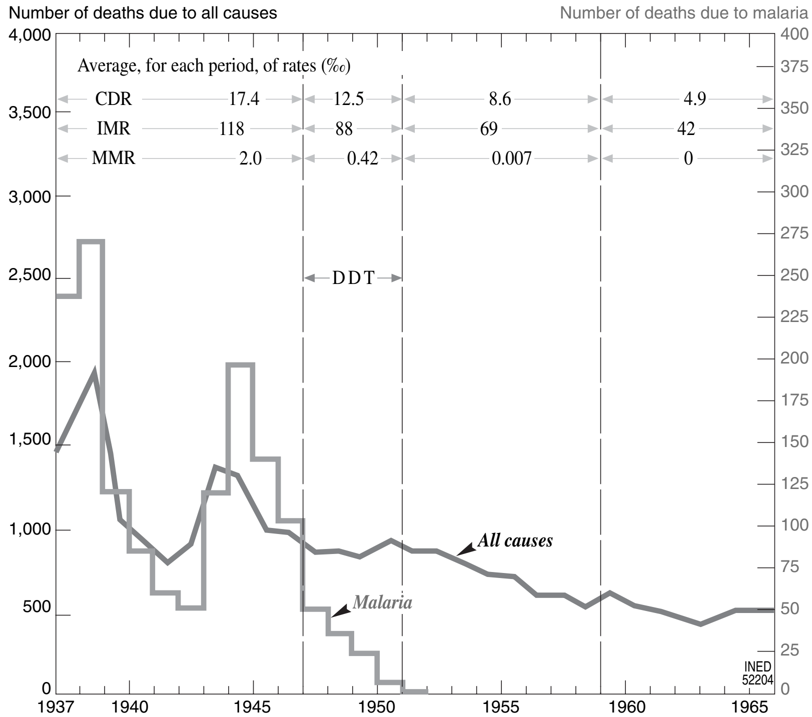


FIGURE 108–6 Sugar plantations in Guyana: Evolution in all-cause and malaria death numbers (the average crude death (CDR), infant mortality (IMR), and malaria mortality (MMR) rates are noted for each period above the graph). (Source: Giglioli *et al.*, 1976.)

methods used in Sri Lanka (Molineaux, 1985), with comparable results (Newman, 1965; Giglioli *et al.*, 1976) (Figure 108–6). But this did not concern a whole country. The operation was conducted only along the narrow coastal stretch where the sugar cane plantations are situated (and 90% of the population), and it was not very difficult to systematically spray the DDT. For the rest, only very localized operations were done, more for experimental reasons than with the aim of eliminating the disease from the country, as for example in Kisumu in Kenya (Payne *et al.*, 1976) or in Garki in Nigeria (Molineaux and Gramiccia, 1980; Cohen, 1988). Even though these trials generally produced rather convincing results, the means needed to extend them over the whole country were never made available. It should be said that to do in a country like Nigeria or the Democratic Republic of Congo what was done on the Island of Ceylon or in the coastal region of Guyana would have required investing considerable financial and human resources to which priority has never really been given. Lastly, it should be added that when one considers the experience of Sri Lanka, which also proves that even on an island malaria can return after 20 years, without strong international concerted action, reluctance to do anything at the level of a single country is understandable. Rightly or wrongly, there has never been, in the case of malaria, anything done on the same scale as the vast global operations for the eradication of smallpox or as the EPI. However, in the meantime the mosquitoes have acquired a certain resistance to insecticides whereas the parasites have mutated toward forms that are increasingly resistant to quinine and its derivatives. New products are constantly being developed, but the disease always ends up adapting itself.

e. AIDS

AIDS, a disease that appeared in the early 1980s, occupies an exceptional place in medical history and represents a major challenge to present-day health policies. It has even cast some doubt over the celebrated triumph of modern medicine over infectious diseases. In fact, it has not totally called into question the first phase of the health transition, that is, the phase in which infectious diseases are eradicated or largely controlled (Caselli *et al.*, 2002; Vallin and Meslé, 2004), which is nothing more than what Abdel Omran (1971) referred to as epidemiologic transition (see also Chapter 57). Rather, it has only delayed, sometimes very seriously, its completion in a certain number of countries with very vulnerable economic and sociocultural contexts. It is particularly interesting to examine here, at different levels, some of the aspects of the health policies implemented in response to the epidemic.

Even if it is almost certain that its origin is African (Grmek, 1989), the epidemic was first observed in developed countries, namely in North America where it severely hit homosexual populations. In San Francisco, for example, which was at the epicenter of the beginning of the epidemic, the homosexual community was rapidly decimated (Hessol *et al*., 1992). As soon as the syndrome was described, the cases multiplied very quickly and all the patients seemed fated to die. At the beginning of the 1980s in the United States, the number of AIDS cases doubled every six months (Allen, 1984, Grmek, 1989). From 200 cases identified at the end of 1981, the number of affected people increased to 450 in mid-1982, to 750 at the end of 1982, to 1,800 in mid-1983, to 3,000 at the end of 1983 (Grmek, 1989, p. 72). This rapid growth provoked a great deal of emotion within the scientific community and led very early on to nightmare scenarios since no treatment seemed to have an effect. The fact that this sudden eruption first appeared to threaten developed countries had important consequences for the orientation of the policies implemented, some of which were very positive, others much less so.

On the positive side it should be stressed that a considerable effort was immediately deployed in the area of research. In the field of biology, but also in epidemiology, demography, and sociology, impressive and unprecedented means were mobilized for research on this infectious disease, inciting lively competition between researchers on each side of the Atlantic. Very soon, as early as 1983, the pathogenic agent was identified by the team of Luc Montagnier at the Institut Pasteur (Barré-Sinoussi *et al*., 1983), opening the way to more targeted research toward finding potential ways of fighting the disease. But the less positive side, residing in the fact that the alert was first triggered off in developed countries and in a specific context of the mode of transmission of the disease is that, very quickly too, the race to find an effective and lucrative treatment prevailed over research into finding a vaccine. A first step toward treatment was made in 1984 with the discovery of the anti-retroviral properties of AZT, which made it possible to increase patients' life expectancy, which had hitherto been very short (Fischl *et al*., 1987). Then in 1996, much more effective tri-therapies were developed, even though they presented considerable disadvantages and do not yet provide a total cure (Amoroso *et al*., 2002). A vaccine still has not been found.

The way the epidemic has evolved in the developing countries and especially in sub-Saharan Africa shows to what extent such a vaccine is needed. In the developed countries, the epidemic never really spread beyond three specific categories of population at risk: homosexuals, drug users, and persons having received blood transfusions. In particular, heterosexual transmission, the means for the virus to access the population as a whole, has always remained very limited in the North.[5]

For each of these three populations (homosexuals, drug users, and receivers of blood transfusions), specific policies have been conducted, based on the mode of transmission. Concerning persons receiving blood transfusions, the epidemic was almost immediately stopped thanks to a reorganization of blood collection procedures, making it possible to exclude contaminated blood. Concerning drug users, the channel of infection is the shared syringe. Policy targeting drug users consisted of convincing the people concerned not to use the same syringe twice, either through information campaigns or through needle exchange programs that provided for the exchange of syringes free of charge (Stine, 1998, Chap. 9). The results were radical in the countries of the West. Since the end of the 1990s, drug users have only represented a small proportion of the new cases of contamination (see, e.g., for France: Lert *et al*., 2004). Concerning homosexuals, the most urgent messages were prevention and avoiding unprotected sex. It was also just as important to understand why in the countries of the North sexual transmission was essentially homosexual. When it became clear that the main factor for the difference was due to the greater frequency in changes of partners or to the fact of having multiple partners (Brouard, 1994), greater insistence was also made on the importance of changes in behaviors and on the absolute necessity of practicing safe sex in unstable relationships. Targeted information campaigns also played an important role. Lastly, even though heterosexual transmission has not developed much, it is nevertheless important to avoid it happening and, here again, since prevention depends essentially on the use of condoms in casual sexual relations, general campaigns for the promotion of new behaviors have been directed toward the young, for example, whereas easier access to condoms has been provided.

As for those who have HIV/AIDS, the number of which has never been huge compared to the other causes of morbidity, in addition to looking for treatments the main effort has been making such treatments available to them as much as possible as soon as they are developed. This has been especially important because these treatments are very expensive.

[5] It should be noted, however, that one factor contributing to the continued prevalence of AIDS in certain European countries is the arrival of immigrants carrying the virus, especially from sub-Saharan Africa. In France, for example, this is the main component in female prevalence (Lert *et al*., 2004).

However, given that the number of infected persons has been contained, it has been relatively easy to ensure the collective funding of these treatments in rich countries.[6]

As a result in the developed countries, the initial dizzying increase in the number of new AIDS cases was very quickly reduced. In Europe (except for the countries of the former USSR), progression of HIV prevalence seemed to stop at the end of the 1980s, and if since the beginning of the 1990s it has remained at an almost constant level of 50 per 1 million inhabitants (EuroHIV, 2003, p. 40), it is essentially because of improved screening. But above all, the arrival of tri-therapies made it possible, in the second half of the 1990s, to reduce the prevalence of the AIDS disease faster than it increased. From over 60 per 1 million inhabitants in 1994 it has declined to 25 in 1998 and 20 in 2002, whereas mortality from AIDS dropped to under 10 per million after having culminated at close to 50 in 1995 (EuroHIV, 2003, p. 40).

The situation has been almost the same in North America. In the United States, after having reached a peak of close to 80,000 in 1993, the annual figure of new AIDS cases dropped to under 50,000 in 1997 and has stabilized at approximately 40,000 since then. The number of deaths due to AIDS peaked at 51,000 in 1994 before declining quickly and leveling off at 20,000 in 2000–02 (CDC, 1995, 2002).

Although the disease has not been eradicated, it has ceased to be epidemic. On the negative side is the situation in the former USSR (especially Russia) where HIV prevalence has recently soared, burdening these countries with what the OMS refers to as "the seeds of potential massive epidemics" (WHO-BRE, 2002, p. 22).

The situation is completely different in developing countries and especially in sub-Saharan Africa. In this region, by far the most affected in the world, the transmission of AIDS, which is essentially heterosexual, presents a grave threat to the whole population. This is due to the same reasons as the high degree of homosexual transmission in Europe and North America: frequent changing of partners and multiple partners (Brouard, 1994). Even if these phenomena are less developed as they were among homosexuals,[7] they have led to the whole of the population being affected by the epidemic which in certain countries, especially in Southern Africa, has reached frightening proportions, the prevalence of HIV infection reaching as much as 15% or 20% of the total adult population. In some countries, life expectancy has declined by between 15 and 20 years or even more. According to the United Nations, in 15 years, South Africa has lost 15 years in life expectancy, Zambia 20 years, Botswana 25 years, and Zimbabwe almost 30 years (United Nations, 2003).

These countries are almost completely powerless in the face of this disaster. Those worst hit are not, however, those lagging furthest behind in Africa in terms of health development. The four countries mentioned earlier were even those where life expectancy, in sub-Saharan Africa, had progressed most since World War II. Yet, their health systems were completely overwhelmed by the scale of the disaster and their economic and social structures badly affected by a sudden mortality crisis in the age groups of the economically active. And the result was just as disastrous in countries slightly less affected but much less provided for before the epidemic. Of course, no African economy was able to cover the cost of caring for AIDS victims and this remains largely the case, even with access to cheaper generic drugs. Patients continue to die in a short time. The only policy that is realistically possible is prevention through information campaigns and the promotion of changes in behaviors in favor of more stable unions and the use of condoms in casual sexual relations. Substantial efforts have been made in this direction by the countries themselves and by the international community. But here again the task is huge since not only well-defined categories of the population need to be contacted, as in northern countries, but the population as a whole, in the most remote rural areas as well as in the big cities. Until now, in spite of the already very high prevalence rates reached, the epidemic is far from having been brought under control in Africa and is becoming an even greater threat in other regions of the South.

In Asia, Thailand was the first country to be hit by AIDS and has suffered hardest with approximately one million persons infected since the beginning of the epidemic for a total population of 61 million inhabitants. Government response to AIDS in Thailand was rapid with the implementation from 1987 of a multisector program to fight the epidemic. This program, which associated all the ministries, nongovernmental organizations, and different private donors and companies, set five priorities: informing and educating the public (prevention); human rights and social assistance; promoting research; providing patients with treatment; and covering its cost. These efforts soon

[6] It should be noted, however, that although AIDS treatment benefits from collective coverage, either totally or for the most part, in most European countries and in Canada, this is much less the case in the United States where Medicare and Medicaid now cover only 20% of patients.

[7] According to Nicolas Brouard (1994), whereas the rate of partner change (average number of new partners per year) was 7 in the cohort of homosexuals in San Francisco, among heterosexuals it is approximately 3 in sub-Saharan Africa compared with hardly more than 1 in Europe or in North America.

bore fruit as shown by the remarkable decline in HIV prevalence among young army recruits, which dropped from 12% to 8% between 1991 and 1995 due to changes in their behaviors (reduced frequentation of prostitutes, increased use of condoms) (Nelson *et al.*, 1996). It is striking to note in this instance the importance of experience acquired previously in the area of family planning. The fact that Senator Meechai Viravaidhaya, often nicknamed "Mr Condom" who heads the Association for Population and Community Development, was the national leader for the family planning campaign before being in charge of the fight against AIDS certainly played a major role.

In Uganda, the prevalence rate peaked at 12% in the mid-1980s and even declined slightly after that (Mulder *et al.*, 1994), much less than the level reached in South Africa, where the epidemic only really took off in the mid-1990s. But even if the remarkable prevention policy implemented in Uganda may be evoked, the reasons for this difference remain uncertain (Caraël and Holmes, 2001).

For the African countries most affected, as well as for other developing countries where the epidemic remains a serious threat, only a vaccine could provide tangible hope to gain real control over the disease. But the development of such a vaccine has turned out to be a long and costly process and for the pharmaceutical laboratories, the drug market is financially much more interesting. The African AIDS epidemic is one of the most serious international health policy failures in recent decades.

2. The Fight against Man-made Diseases

It is a known fact that during the 1960s growth in life expectancy was curbed, even blocked for a time, in most industrial countries (apart from the notable but almost unique exception of Japan). Apart from the fact that infectious mortality had become so low that subsequent successes in this area had no more effect on life expectancy, the latter increasingly came under the attack of what Abdel Omran (1971) called "man made diseases." How have health policies contributed to reversing this trend? We will evoke three cases: alcoholism, smoking, and road accidents.

a. Alcoholism

Even if the (very moderate) consumption of alcohol is becoming increasingly acknowledged as having some positive effect on health (Poikolainen, 1995; Holman *et al.*, 1996), it is obvious that an abusive consumption of alcohol is deadly. It is deadly either directly and brutally (acute alcoholism, sudden aggravation of cerebrovascular diseases) or more slowly (cirrhosis of the liver, cancer of the oesophagus, alcoholic psychosis, etc.), or indirectly (accidents at work or road accidents, homicide, suicide). In certain countries, such as France, Russia, and other Eastern European countries, mortality due to alcoholism, which grew rapidly at the end of the 19th century, became exceptionally high after World War II. In the mid-1960s there was a reversal of this trend in France, whereas it increased in Russia.

In the case of alcohol-related mortality a simple policy, not even necessarily intended principally to that end, can have a considerable impact. For example in France the 1957 *Loi d'assainissement économique et financier* (Law for Economic and Financial Stabilization) resulted in a sudden increase of one year in life expectancy, essentially due to the immediate decline in mortality caused by acute alcoholism and alcoholic psychosis, but also by alcoholic cirrhosis of the liver (Figure 108–7). Not only did this law considerably restrain the capacity for consumption of the French, but it was also accompanied by an increase in taxes per liter of red wine which raised the price from 0.92 francs to 1.40 francs per liter (Meslé and Vallin, 1993). This type of effect, however, is short-lived because in the absence of a sustained campaign against alcohol abuse, consumption and mortality return as soon as the standard of living rises again. In fact, in France, this measure was only a sort of accident in a panorama that included many other, more precisely targeted measures,[8] the cumulated effects of which finally resulted in a lasting change in the trend of alcohol-related mortality after 1965 (Vallin and Meslé, 1988, Nizard and Munoz-Perez, 1993).

[8] Measures reducing access (e.g., the gradual closing down of drinking establishments, the number of which continually declined between 1954 and 1991; regulation of sales outlets, especially at the workplace; and the gradual disappearance of the privilege of home distillers in 1960), development of preventive measures against road accidents (creation of the offense of driving under the influence of alcohol in 1958) and in schools (screening of early alcoholism in schools, and from 1955, education programs on alcoholism), organization of anti-alcohol advertising campaigns (the slogans of some of which have remained very popular, such as the one of 1956, "les parents boivent les enfants trinquent" [The parents drink and the children suffer] or "boire ou conduire, il faut choisir" [Drink or drive, you have to choose], "boire un petit coup, casse cou" [To have a drop is reckless], "un verre ça va, trois verres bonjour les dégâts" [One drink's fine, three are trouble]) etc., the regulation of advertising in 1960 (forbidding the association of alcohol with sports or with automobiles, or forbidding claims of stimulating or aphrodisiac properties, etc.), control of the quality of wine (with the obligation, in 1964, of distilling or converting into vinegar low-quality wines and the creation, in 1965, of a VDQS label) (Berger *et al.*, 2000).

FIGURE 108–7 First decline in comparative rates of male alcohol-related mortality in France with the Law for economic and financial stabilization (*Loi d'assainissement économique et financier*) of 1957 and the sustained reversal of the trend induced by more long-term measures. (Source: Nizard and Munoz-Perez, 1993.)

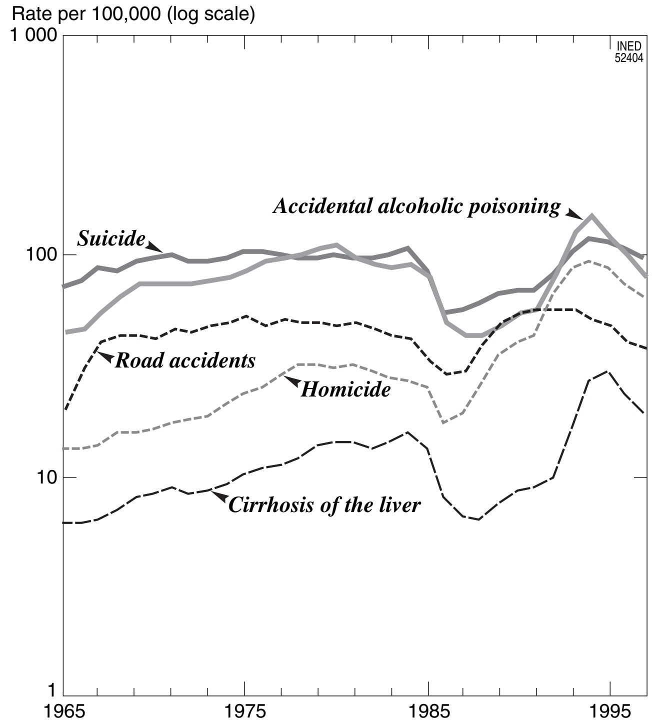


FIGURE 108–8 Changes in age-standardized death rates according to different alcohol-related causes of death in Russia. (Source: Meslé *et al.*, 1998.)

The anti-alcohol campaign conducted by Mikhail Gorbachev in 1985 in the USSR under Perestroika was different and involved a reduction in production, reduction in sales (controlled by the State), increase in prices, and restrictive measures against home distilling (Shkolnikov and Nemtsov, 1997). The effect on life expectancy was immediate: whereas it had been declining for 20 years, from 1984 to 1987, it suddenly increased by 3.2 years for men and 1.4 years for women (Meslé *et al.*, 1994; Shkolnikov *et al.*, 1995). It has been demonstrated that this sudden improvement was entirely due to a sudden decline in alcohol-related mortality, either directly under the form of accidental alcohol poisoning (under which heading acute alcoholism and alcoholic psychoses were included) or of cirrhosis, or indirectly through a decline in the number of homicides, suicides, and accidents in the workplace and on the roads (Figure 108–8) (Meslé *et al.*, 1998).

The decline in violent deaths alone explains 53% of the increased life expectancy among males and 33% among females (Meslé and Vallin, 2003a), and the rest is in all likelihood due to a slight reduction in deaths from other causes (namely cardiovascular diseases) greatly influenced by alcohol. Lastly, even if these improvements had begun slightly before the official launch of Gorbachev's anti-alcohol campaign, it was also clearly demonstrated that most of the decline in alcohol-related mortality and the increases in life expectancy were directly linked to a reduction in the consumption of alcohol following the campaign itself (Figure 108–9) (Meslé *et al.*, 1994; Shkolnikov and Nemtsov, 1997). In that sense, it can indeed be said that this energetic anti-alcohol campaign had a major impact. However, the effect was short-lived. The Gorbachev government soon had other urgent matters to deal with, and the anti-alcohol restrictions soon fell into abeyance. From 1989 life expectancy started to decline once again and this decline accelerated further in 1993–1994, after the collapse of the USSR because of the serious economic and social crisis set off in Russia by the sudden and ill-controlled transition to a market economy. Before this crisis, the relapse was entirely attributable to a resumption in the consumption of alcohol; with the crisis, of course, many other causes came into the equation, but alcohol continued to play an important role (Meslé *et al.*, 1998).

However effective at the time, a crackdown such as the one in 1985 in the USSR can only have very temporary effects. The fight against alcoholism requires a real cultural revolution and only a long-term policy focusing on information and education rather than on repression can have a sustained impact. In the countries most affected, strong belief in the virtues of alcohol is the main obstacle that needs to be overcome.

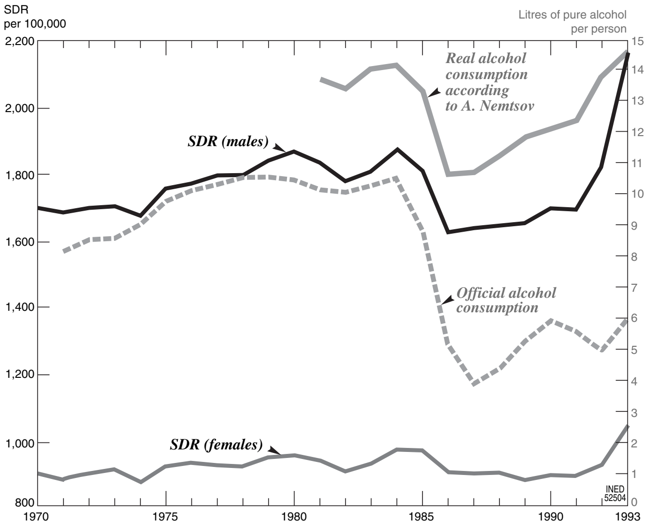


FIGURE 108–9 Compared changes in age-standardized male death rates and alcohol consumption in Russia. (Source: Shkolnikov and Nemtsov, 1997.)

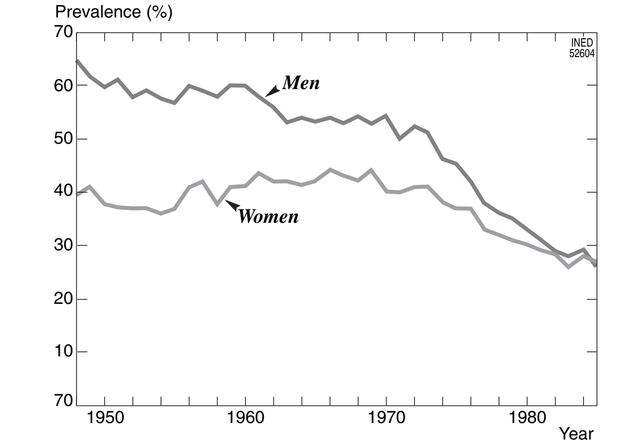


FIGURE 108–10 Evolution of the proportions of smokers among British men and women, 1948–1997. (Source: RCP, 2000; 1948–1971, Tobacco Advisory Council; 1972–1996, General Household Survey; 1997, Omnibus Survey.)

b. Smoking

Tobacco is harmful to health in several ways. It especially affects the aerodigestive tracts (oral cavity, pharynx) and the rest of the respiratory system (larynx, trachea, lungs), by provoking cancers of these tissues but also by causing other respiratory diseases (emphysema, chronic bronchitis, etc.). Moreover, some of its components, passing through the blood, are also aggravating factors for a number of diseases, of which the following are at the top of the list: ischemic heart diseases, aortic aneurism, and cancers of the bladder and the pancreas. A number of studies have revealed the impressive evidence of excess mortality provoked by smoking. Cuyler Hammond (1966) demonstrated that out of a sample of one million Americans, for example, men aged 45–54 smoking 40 cigarettes a day or more had an all-cause death rate nearly three times higher than that of nonsmokers, whereas, for France, Michel Blanc (1985) calculated that the elimination of smoking as it prevailed in 1968 would enable males to increase their life expectancy by 3 years.

As in the case of alcoholism, the situation cannot be improved simply by waving a magic wand. It requires a sustained policy in which health information and education of the masses play as much of a role as regulating the production and distribution of tobacco as well as protecting living environments. In the case of tobacco the problems are even greater than with alcohol because its addictive power operates even faster.

It is in the United Kingdom that the first benefits were reaped from a policy of information and prevention. However, paradoxically, actual government action was relatively limited (an information campaign against the dangers of smoking and the setting-up of health centers to help people who wish to stop smoking, and then, in 1964, the banning of advertising of tobacco on television), but it was greatly backed up by private initiatives and strong public awareness, namely thanks to follow-up work by doctors (RCP and ASH, 2002; Doll *et al*., 2004).

A study by The Royal College of Physicians (RCP, 2000) shows that smoking, which had taken unprecedented proportions after World War II (as much as 60% of smokers among men and more than 40% among women), had declined slightly at the beginning of the 1960s, at the time when a major survey among British doctors was being conducted, but it took until the second half of the 1970s for a continued decline to occur, first among men, then very soon after, among women (Figure 108–10). In 1997, the last year for which data were available for the drafting of the report, the proportion of smokers had dropped to an equal 25% level among both sexes.

The example of the United Kingdom was soon followed by other English-speaking countries (namely the United States) and Northern European countries, and later by the other European countries.

In terms of mortality, it is easy to identify the consequences of these changes on the evolution of mortality caused by lung cancer, presented in Figure 108–11 for six European countries and by sex. The effect of the decline in smoking was obviously not direct. One sees, for example, that in the United Kingdom the first decline in the 1960s only slowed down the increase in male mortality by respiratory

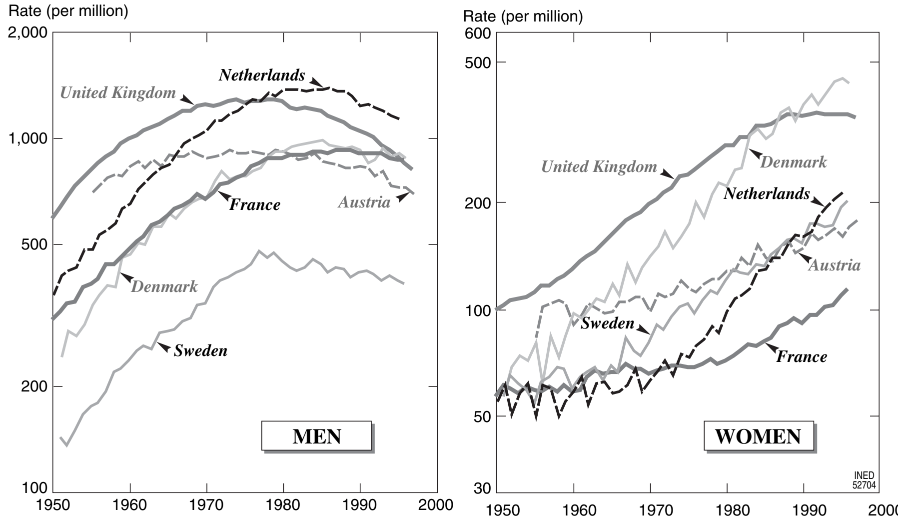


FIGURE 108–11 Evolution since 1950 of age-standardized death rates for cancer of the lungs, the bronchus and larynx in various European countries (logarithmic scale). (Source: Vallin and Meslé, 2001b.)

cancer and the reason why the strong decline begun in 1976–1977 coincides with a complete halt in the increase of this mortality is especially because male smoking ceased to grow from the end of the 1940s. Mortality among women, whose consumption of tobacco only really ceased increasing at the beginning of the 1970s, continued to get worse. However, it is the decline in smoking among both sexes from the end of the 1970s which also finally resulted in a decline in male lung cancer mortality and halted the increase in female lung cancer mortality. At the end of the 1990s, the decline obtained among men was substantial (close to 40%), and it seemed that the decline was beginning among women.

Concerning men, rather similar results were obtained at about the same time in Sweden, a country where respiratory cancer mortality has never been as high as in the United Kingdom. But this was also the case, a little later, in the Netherlands, where male lung cancer mortality reached a level at the beginning of the 1990s that was even higher than the maximum ever observed in the United Kingdom. On the other hand, in France, although the increase in male lung cancer mortality was curbed from the end of the 1970s, the real turnaround was still hardly noticeable at the end of the 1990s. This is because the anti-smoking campaigns there were started later and have only been raised to a higher pitch in recent years.

Concerning women, the situation is even more problematic. Indeed, everywhere traditionally women have always smoked much less than men, but this difference has greatly reduced until it has sometimes completely disappeared and, except in the United Kingdom, female consumption has constantly increased in each of the countries represented in Figure 108–11, which explains the reason why the increase in lung cancer mortality continues there. Why do anti-smoking policies have less influence on women than men? Women, notably young women, undoubtedly have more reasons than men to resist the new social pressure from anti-smoking campaigns. For many of them smoking was and still is a symbol of emancipation to which they attach more importance than to the need to stop smoking in order to prevent hypothetical lung cancer.

c. Road Accidents

Road accidents did not start with the advent of the automobile but the latter increased their impact very much. The development of road traffic with the invention of mechanical horsepower made road accidents one of the leading causes of violent death. At the end of the 1960s, more than 15,000 people died each year on the roads in France each year, representing a crude death rate of more than 30 per 100,000 inhabitants, comparable to figures in the United States (26) and Germany (30). In fact, the number of road deaths in France exceeded 15,000 in 1967 and reached its maximum of 18,630 in 1972 (Vallin and Chesnais, 1975). Policies were relatively quickly implemented to deal with this problem at different levels (see Chapter 107), but it was only starting in the 1970s that this modern-day problem was curbed in most European countries (Figure 108–12). It was also at this time that road accident mortality declined noticeably in Japan,

FIGURE 108–12 Evolution since 1950 in age-standardized road accident male death rates in European countries (logarithmic scale). (Source: Vallin and Meslé, 2001b.)

North America, and Australia. However, it is continuing to grow in many developing countries.

One of the major ways of preventing road accidents has been and still is to improve the road network and the vehicles. However, the results of these improvements are ambiguous to say the least. Let us take, for example, the construction of motorways. There is no doubt that motorways are much less dangerous than ordinary roads. In France, for example, the number of deaths per kilometer traveled is 3.5 times less on motorways than on primary or secondary roads (DSCR, 2004). The increasing importance of motorways in road transport has therefore saved many human lives. However, the development of motorways has enabled a much greater number of vehicles to get onto the roads and has subjected a growing proportion of the population to the risk of accidents. It is true that the development of transport is an essential factor for economic development, but, at this level, it is transport policy as a whole that is called into question. A considerable investment into road transport has obviously been made to the detriment of other much less dangerous means of transport, such as air travel or railway. The development of roads and automobiles has also created the conditions of a cultural preference among the public for the most dangerous means of transport, whereas the lack of investment into railways or fluvial navigation has diverted goods traffic to the roads, making them even more dangerous. This choice concerning the society we live in was made in all the industrial countries at the turn of the 1950s–1960s and is something that is difficult to reverse today.

Nevertheless, road safety policies have been reinforced everywhere (see Chapter 107), and in almost all the industrial countries, a significant change occurred in the mid-1970s. The movement was particularly synchronous in Europe, as shown in Figure 108–12. It was the consequences of two factors combined. The accumulation of preventive measures (improvement of the road network and of the safety of the vehicles, speed limits, wearing of safety belts, etc.) finally yielded its dividends, but at the same time the shock of the oil crisis in 1973 provoked among drivers a change in attitude toward speed, a great consumer of fuel.

Thus, thanks to an accumulation of complementary measures, road safety policy has made it possible to reverse the trends in most developed countries. Nevertheless, something that should not be ignored is the importance of the influence of temporary phenomena which may act either positively or negatively, nor, above all, should it be forgotten that this question remains far from resolved in developing countries where road accidents are increasingly becoming a major public health issue.

3. The Fight against Cardiovascular Diseases

Beyond curbing man-made diseases, the success won over cardiovascular diseases was a crucial factor in the return of a sustained increase in life expectancy in industrialized countries from the beginning of the 1970s. This event even occupies such an important place in changes in health in the last three decades of the 20th century that we have already substantially covered it in our first section on the global evaluation of health policies. We have, in particular, already stressed the profound difference between East and West engendered by the opposition between the success of the policies conducted in the West in this area and the incapacity of Communist countries to implement adequate strategies.

We shall therefore simply complete it here with a few points of detail based on the example of France, the country for which we have reconstructed death series by constant cause definition since 1925 (Vallin and Meslé, 1988, 1998; Meslé and Vallin, 1996), an absolute prerequisite for following over time the dif-

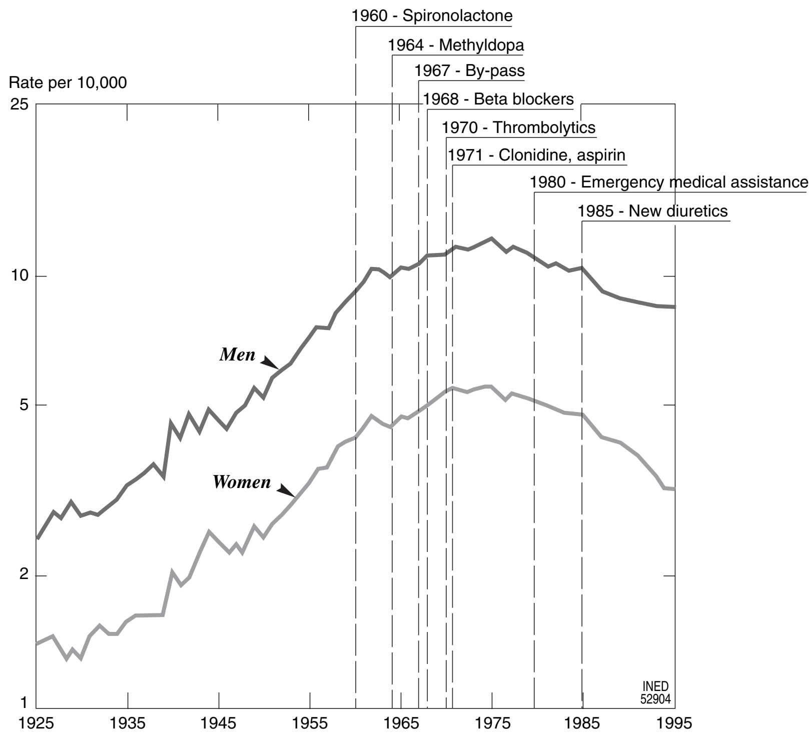


FIGURE 108–13 Evolution of age-standardized myocardial infarction death rates in France since 1925 and dates of introduction of the main medical innovations. (Source: Caselli *et al.*, 1999.)

ferent components of cardiovascular mortality. In fact, in France, total cardiovascular mortality had started to decline long before the 1970s, but this slow global decline was the result of contrary movements of causes toward decline (heart failure, pulmonary heart diseases), stagnation (cerebrovascular diseases), or increase (acute myocardial infarction), and it was when cerebrovascular mortality started to decline and, above all, when the trend concerning myocardial infarction was reversed, that the global decline in cardiovascular mortality took off decisively.

Two major factors have played a part in ensuring the success of the policies conducted in this area: the dissemination of new technologies and changes in behavior. Figure 108–13 clearly shows how the accumulated positive effects of new technologies have weighed in the balance toward a reversal of the unfavorable trend of myocardial infarction mortality:[9] monitoring and control of hypertension (with the distribution of spironolactone from 1960, of methyldopa from 1964, of beta blockers from 1968, and of clodinine since 1971), prevention of angina pectoris (beta blockers), treatments against thrombosis (a second generation of thrombolytics effective against atheromatus coronaries and brought out from 1970), coronary by-passes (performed for the first time in 1967), the organization of emergency procedures that became very effective from the 1980s with the setting-up of *Service d'assistance médicale d'urgence* (Emergency Medical Assistance Service), and so forth (Biraben, 1988).

But the second factor behind the reversal in the trend is no less important. The increase in mortality was due to that of the risk factors. We have already mentioned the role played by smoking and alcoholism. The period of the 1950s and 1960s was also marked by a change in diet under the effect of a higher standard of living (with the increase in the consumption of animal fats), the transformation toward a sedentary lifestyle (with work evolving, less manual and more office work), not to mention pollution of the environment, all of which contribute to high cholesterol and hypertension. The reversal of the trend toward an increase in infarction mortality observed at the turn of the 1970s reflects not only the distribution of new medical technologies enabling early diagnoses and treatment, but also a change in behaviors toward risk factors.

[9] However, one should be cautious about the extent of the increase in myocardial infarction mortality as about the decline due to certain ill-defined causes such as, namely heart failures and, especially, ill-defined heart diseases. Even after reconstruction of death series by constant cause definitions, the real content of the different sections remains subject to changes due to improved diagnostic techniques. It is almost certain that the increase in infarction mortality during the 1950s and 1960s is lower than Figure 108–12 suggests. This bias does not, however, call into question the reasoning.

In both cases the precise role played by health policies, in the strictest sense of the term, is rather difficult to determine. It is obvious that in a country like France where costs are essentially covered collectively, this greatly facilitated quick access for the majority of the population to often very costly new technologies. In the same way efforts to provide information and education concerning a healthy diet and attention to early symptoms and screening for cholesterol and hypertension also have played an important role. But more generally, success has come in the form of awareness on the part of the public and the patients themselves of the fact that improved health in general and in the area of cardiovascular diseases in particular requires that everyone should play their part.

Figure 108–13 shows rather clearly that, at least until 1996, the year at which the graph stops, the decline in myocardial infarction mortality was more rapid among women than men in spite of the latter having higher rates. The reason is that women, because of the greater attention they pay to their bodies, were more willing than men to adopt the necessary changes in behavior and to benefit from technological progress (see Chapter 53). Since then, however, we know that men have made up for most of their disadvantage by adopting behaviors more favorable to their health (Meslé, 2004a). Success in this case results therefore more from a global social change than from the implementation of precise health measures.

III. FULFILLING OBJECTIVES TARGETED IN TERMS OF GROUPS

A third way of evaluating health policies consists of examining whether or not and to what extent specific programs aimed at reducing the morbidity or mortality of certain target groups are crowned with success. At a given stage in socio-sanitary development it is indeed in the collective interest for certain groups, considered as disadvantaged or exposed to specific risks, to be the object of specific measures. Here again examples abound. Let us examine two types of concerns among others: protection at certain stages in life and the reduction of inequalities.

1. The Protection of Certain Stages in Life

At certain times in life the human being is exposed to specific health risks that most societies have more or less tried to contain. In modern societies, specific policies are developed to protect the newborn, children, schoolchildren, workers, pregnant women or women just out of childbirth, and the elderly. Let us take three examples: infancy, maternity, and old age.

a. Infancy

Developed Countries

In developed countries, after a brilliant victory over infectious infant mortality, greatly due to traditional policies of maternal and child protection, the main objective of recent decades has been to obtain a decline in neonatal mortality. Many countries have set up special programs in this area. Such was the case in France with the launch in 1971 of an incentive program aimed at prevention and care and benefiting from major budget contributions for several years.[10] This program (MSPSS, 1971; Chapalain, 1975) included seven measures: training and advanced training of health care workers, vaccination against German measles, reinforcement of the medical supervision of pregnancies, improvement in the supervision of deliveries, the resuscitation of newborn babies in the delivery room, and the equipment of obstetric wards and neonatal intensive care units. A series of three surveys (1972, 1976, and 1981) was organized to monitor the development of the program and its results (Rumeau-Rouquette, 1979).

The results were not long in coming. It is possible to follow the program with a simple but significant indicator: the proportion of women in childbirth having had 4 prenatal consultations or more, which was only 49% according to the survey of 1972, rose to 60% in 1976 and 80% in 1980 (Rumeau-Rouquette and Blondel, 1985). Moreover, the state of health of newborn babies considerably improved, with reductions in the proportion of both premature births and late births.[11] But neonatal mortality especially fell sharply. Indeed, like all the other components of child mortality, neonatal mortality had been declining since World War II, but at a much slower rate than post-neonatal mortality and, since the latter had dropped to a very low level, the resistance to the decline in neonatal mortality started to curb the decline in total infant mortality (Figure 108–14). With the launch of the peri-

[10] With the finance act of 1971, a budget of 9.5 million francs had been earmarked for new measures and for the duration of the sixth four-year plan this program was a part of, a total of 257 million francs (in constant francs) was provided (Rumeau-Rouquette and Blondel, 1985).

[11] The proportion of births at under 37 weeks of pregnancy fell from 8.2% in 1972 to 5.6 in 1981, and the proportion of births after 43 weeks of pregnancy or more fell from 4.0% in 1972 to 3.4 in 1981 (Rumeau-Rouquette and Blondel, 1985).

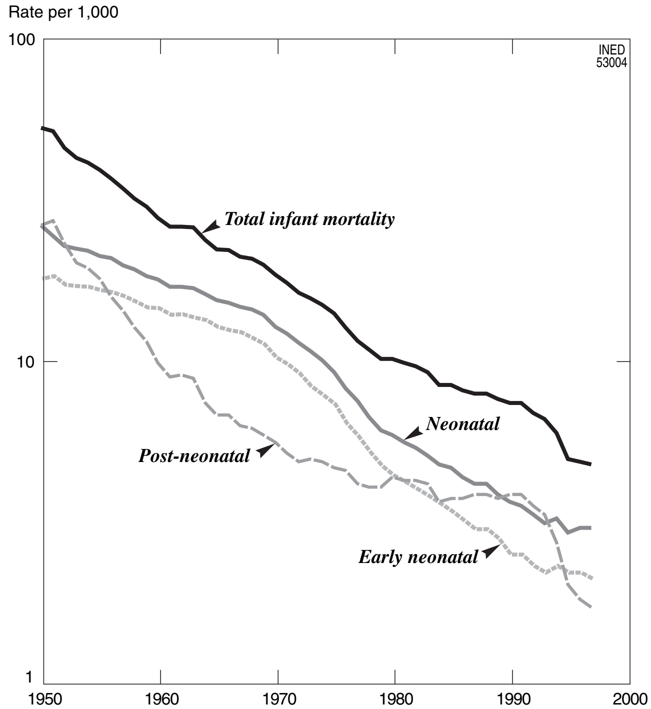


FIGURE 108–14 Evolution between 1950 and 1997 of the different components of infant mortality in France. (Source: INSEE, statistical yearbooks.)

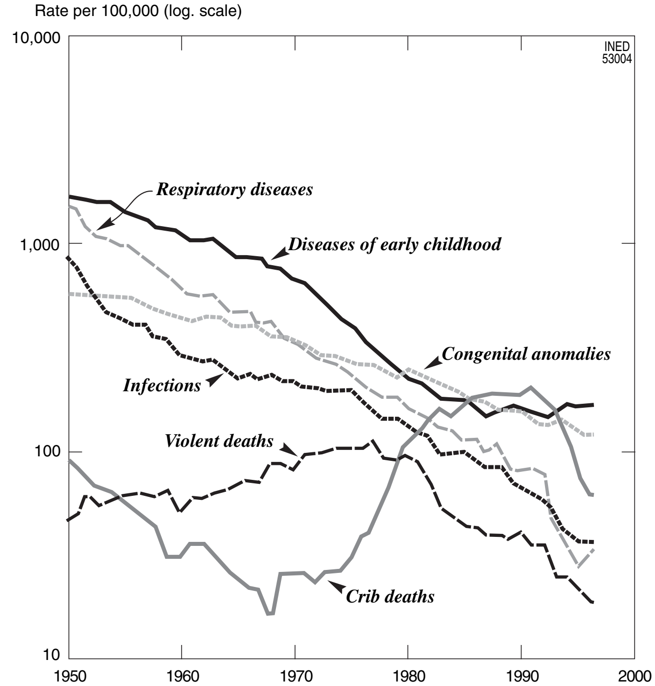


FIGURE 108–15 Evolution of crib death mortality among other causes of infant death. (Source: Meslé, 2000.)

natology program of 1971, these trends suddenly reversed. Whereas the decline in post-neonatal mortality slowed down because of the very low level it had reached, neonatal mortality collapsed. And this success concerned both total neonatal mortality (0–27 days) and early neonatal mortality (0–6 days).

From the 1980s, infant mortality continued to decline thanks to the continued and very rapid decline in neonatal mortality. However, post-neonatal mortality stopped declining and at the end of the 1980s, it rose once again above neonatal mortality, calling into question the decline in infant mortality. In fact, this unfavorable evolution of post-neonatal mortality was to a great extent due to an error in medical policy which was the cause of a sudden increase in crib deaths. The extraordinary increase of this cause of death between 1975 and 1985 is due to the generalized placing of babies on their stomachs in their cribs. It is true that this practice initially saved lives of children by substantially reducing mortality due to anoxia and hypoxemia (Meslé, 1995), but it later became apparent that it caused a surprising increase in crib deaths (Sénécal *et al*., 1998). It was enough for the medical establishment to agree and to change policy by imposing on maternity wards that babies should be placed on their side and by advising mothers to do the same for the increase in crib deaths to stop, then suddenly to collapse when the new practice was generalized (Figure 108–15).

Developing Countries

In the case of developing countries everything said above concerning vaccines constitutes in itself an effort almost entirely focused on improving the health of children and reducing their mortality. Nonetheless, it is not the only effort made in this direction. We believe it is particularly important to complete the picture here by evoking two other major health policy actions in favor of children: the fight against diarrhea with oral rehydration procedures and the fight against malnutrition with food supplement programs.

i. Programs against Diarrhea Toward 1980, diarrhea was the first cause of death among children in developing countries (and in the whole world, given the very low infant-child mortality of developed countries). Efforts made since then to counter this cause of death have essentially relied on oral rehydration therapy (ORT), a technique developed in Bangladesh at the end of the 1960s, the dissemination of which was greatly encouraged by the WHO from 1979, following the Alma Ata Conference.[12]

[12] In fact, this expression often covers complementary actions such as administration of oral rehydration salts, rehydration using domestic fluids, overhydration, and sometimes even medical treatment.

Although many National Control of Diarrhea Diseases Programs (NCDDP) have been established, few have been evaluated at a national level. In an exception to this, Cesar Victora *et al*. (2000) list four cases in which global evaluation elements are available: in Brazil, the Philippines, Egypt, and Mexico. In Brazil, the proportion of hospital admissions for diarrhea out of the total number of admissions of children (in the major pediatric hospitals in the nine northern states) fell from 57% in 1980 to 30% in 1990. At the same time, whereas the campaign for the use of rehydration salts had been launched in 1980, a survey showed that in 1991 mothers had used the therapy in 35% of all the diarrhea cases and in 62% of the cases they judged to be severe. Lastly, the study of a simulation showed that this therapy accounted for three-quarters of the decline in diarrhea disease mortality. In the Philippines, with the development of NCDDP, diarrhea disease mortality declined by 5% per year during 20 years, whereas infant-child mortality for other causes declined much more slowly. In Egypt, decline in diarrhea mortality was even faster, with reductions of 4.2% per year in 1970–1977, 7.8% in 1978–1983, and 15.9% in 1984–1990, greatly accelerated by the introduction of ORT. In Mexico, lastly, the proportion of infant-child diarrhea mortality (0–5 years) fell from 26% in 1983 to 11% in 1993.

On a global scale, it is estimated that the number of deaths caused by diarrhea fell by 4.6 million in 1980 to 3.3 million in 1990 and to 1.5 million in 1999. The NCDDPs and especially the oral rehydration programs seemed to have been the reason for this massive decline (Victora *et al*., 2000).

ii. Food Supplement Programs Nutritional deficiencies among children are frequent in poor countries. Different studies have shown that in the 1970s, two-thirds of children's deaths could be attributed to nutritional deficiencies (Puffer and Serrano, 1973; Kielman and McCord, 1978; Chen *et al*., 1980). Several approaches to the infant nutritional problem have been the object of programs, with varying degrees of success. The simplest in its concept is obviously the one that advises maintaining or reinforcing prolonged breastfeeding traditions. With modernization, breastfeeding practices have declined without mothers having access to healthy, balanced artificial milk for their children. This problem was revealed as early as the 18th century in Europe and already at the time was the object of campaigns promoting breastfeeding. The promotion of breastfeeding has also been, in the last 20 years, one of the five mainstays of the GOBI-FFF (growth monitoring, oral rehydration, breastfeeding, immunization–food, female education, and family planning) launched by UNICEF in the 1980s. According to UNICEF (Chen *et al*., 1997), 57 out of 102 countries launched campaigns promoting breastfeeding, but have not, unfortunately, provided evaluations of the results.

Other programs focus on vitamin A supplementation. Vitamin A deficiency is associated with high child mortality because of the different diseases it provokes, for example, xerophtalmia. An Indonesian study on 26,000 children of preschool age showed that the group having received vitamin A tablets twice, at six-month intervals, had a death rate at 12 to 71 months, 33% lower than that of the control group in the year that followed the vitamin supplementation (Sommer *et al*., 1986). Similar results were obtained in southern India (Rahmathullah *et al*., 1990) and in Nepal (West *et al*., 1991) but neither in Sudan (Herrera *et al*., 1992) nor in Ghana (Ross *et al*., 1995) where the vitamin A supplementation had practically no effect on mortality. Out of the 67 countries involved in this action through the GOBI-FFF program, 35 "will come close to eliminating the problem" (Chen *et al*., 1997).

More generally, food supplement programs have been implemented, the first of which, in the 1950s, were intended to fight against malnutrition in young children. In centers for nutritional rehabilitation, undernourished children received a balanced diet made up of food available on site and prepared according to local recipes, with the participation of the mother in the choice and preparation of the ingredients. The dual objective was for the children to recuperate and for the mothers to become informed. These programs are effective in the short term: Beghin and Viteri (1973) found that approximately 70% to 80% of malnourished children recuperated during their stay if the center was well run. Although this type of action was in vogue in the 1960s, it was soon abandoned. Resources needed to ensure coverage for all malnourished children mobilized were never mobilized, and the real impact of these centers has always been marginal (Beghin and Vanderveken, 1985). Much less costly per day than hospital care, these programs nevertheless stretched public health budgets of developing countries.

Programs for the noncommercial distribution of food to selected children with the aim of improving or protecting their nutritional state took over in the 1970s and 1980s and constituted, in the mid-1980s, the most widespread form of intervention in this area. Although substantial funds were dedicated to them (Beghin and Vanderveken, 1985), the results were disappointing (Beaton and Ghamessi, 1979). There were several reasons for this. On the one hand, the management of these programs turned out to be complex in that it proved to be extremely difficult to avoid both high dropout rates and diversion of food. In addition, the cost of extending the programs to all children at risk is very high.

Except for periods of serious crisis, complementary food programs have proved to be costly and difficult to implement. Two approaches seemed, for a time, more promising: supplementary diets for pregnant women, which make it possible to reduce the proportion of newborn children with low weight (Mora, 1979; Cole *et al*., 1995), and the nutritional parts of programs integrating health care and nutrition. In both cases, due to the fact that food programs are associated with other actions, their cost is diminished. On the other hand, it is more difficult to measure actual performances. In a few cases, it has been proved that nutritional supplements had a direct impact on the survival of children, as, for example, in the Narangwal study (Taylor *et al*., [1968] 1983). However, here again, pilot studies should not be confused with actions on a national scale. Such programs have never been generalized anywhere. Some authors have even suggested that this cannot be done efficiently (Gwatkin *et al*., 1980).

b. Maternal Risk

Decline to an almost imperceptible level of maternal mortality in the most developed countries has been one of the major successes of health policies. Specific policies were implemented in this area from the second half of the 19th century, and in most cases, in fact, these were associated with those aimed at the health of the children (namely in the framework of maternal and infant protection). It is true, maternal mortality greatly declined—in part due to improved general hygiene and diet conditions but also, in countries like France and Sweden, due to the early decline in fertility—before these policies were explicitly aimed at making medical progress available to the majority of the population; however, they did play a major role in the final result.

It is possible to measure maternal mortality by relating the number of deaths of pregnant women or women in childbirth either to the number of births (maternal mortality ratio) or to the number of women of reproductive age (maternal mortality rate) (see Chapter 45). In this case we are using what is most common: the maternal mortality ratio. In most cases, maternal mortality is due to five main causes: infection, hemorrhaging, eclampsia, obstetrical accident, and abortion.

The first major breakthrough was the discovery, well before Pasteur, of the contagious nature of puerperal fever, perhaps as early as 1790 by the Scotsman Alexander Gordon, but in any case in 1840, simultaneously in the United States by Olivier Webdell Holmes and in Austria by Ignaz Semmelweis. This discovery met with resistance on the part of some obstetricians until the 1930s (Loudon, 1992), but it was rapidly adopted by many others. In Sweden, for example, the generalization of precautions to prevent contagion greatly contributed to reducing the incidence of puerperal infection from 5,300 per 100,000 in 1881 to 220 in 1900 (Högberg and Broström, 1985).

At the same time, surgery progressed, encouraged by the introduction of the first methods of anesthesia, which made it possible to reduce the number of obstetrical accidents. Later, of course, came sulphonamides, then antibiotics, for the efficient treatment of infections, new progresses in surgery and anesthesia facilitating caesarean births, modern methods of transfusion, and other means of fighting against hemorrhages. Recently, it has been possible to follow the impact of all these technological advances on the evolution of maternal mortality ratio, as Judith Fortney does in Chapter 45 (see Figure 45–1). However, it is thanks to maternal health policies that it has been possible for these advances to be adopted for the benefit of all mothers. These policies were fully implemented in 1950s with the generalization of childbirth in hospital and more recently by increased medical supervision of pregnant women. From a maternal mortality ratio of the order of 1,500 to 2,000 per 100,000 in the 18th century, ratios fell to between 500 and 800 at the beginning of the 20th century and have since fallen to ratios of less than 10 today (sometimes even less than 5) (Chapter 45).

Unfortunately, the best proof that technical progress is not enough and that the final victory in this area as in many others concerning health depends, above all, on the policies implemented is the diversity of situations in developing countries today. The financial cost of the basic elements for a very strong reduction in maternal mortality (prevention, antibiotics, conditions of delivery) is very low. Yet, in 2000, the maternal mortality ratio still varied, in developing countries, from 92 per 100,000 (Sri Lanka) or even 43 (Costa Rica) to more than 1,000 (Kenya, Burkina Faso) or even 1,500 (Tanzania) or close to 2000 (Afghanistan, Sierra Leone, Malawai) (WHO *et al*., 2004). Not only have countries like Costa Rica or Sri Lanka done a great deal to enable the majority of the population to benefit from the elementary means available today in the fight against maternal mortality, but they have done so in context of general policies for the promotion of women's rights placed at the very heart of their economic and social development policies. From the 1980s and 1990s and, even more so, since the last Conference on Population and Development (held in Cairo in 1994), WHO and UNICEF have coordinated their efforts to eliminate most of maternal mortality, an objective which should be easily attained, and even to go as far beyond this as possible by attempting to have all countries adopt the concept of "safe motherhood" (see www.safemother hood.org).

c. Old Age

Many developed countries have conceived health policies directed especially toward elderly persons. In countries where general coverage of health costs is extensive, these programs only play a subsidiary role, that is, dealing with problems not sufficiently covered by the general health plan. An interesting case is that of the United States, where the collective coverage of health costs is not the general rule but where a special program for elderly people was set up about 40 years ago.

Medicare, created in 1965, is a government program that covers all Americans aged 65 and over (see Chapters 107 and 117 on the United States). The law was adopted as a reaction to the fact that medical costs were increasing much faster than was the cost of living. Toward 1960 it became evident that private insurance could no longer provide sufficient coverage for elderly citizens, whether or not they benefited from a retirement pension. It was estimated at the time that only one out of four Americans aged 65 and over benefited from insurance covering 75% of a hospital bill.

Given that Medicare is, as Carl Haub writes in Chapter 117, "the largest government health programme in history [of the United States]," it is important to know if it has had any positive effects in an area where some are to be expected: life expectancy at age 65. Among men, the pace of increase in life expectancy at 65 was faster from 1965 than between 1950 and 1965. Whereas in 15 years, Americans aged 65 had only gained 0.1 years of life expectancy between 1950 and 1965, they gained 1.3 years between 1965 and 1980, more than 10 times more. And they gained another 2.3 years from 1980 to 2001 (Figure 108–16). However, it is not certain that this leap forward is really attributable to the Medicare program (Card *et al*., 2004). First the situation is much poorer for women. Whereas they

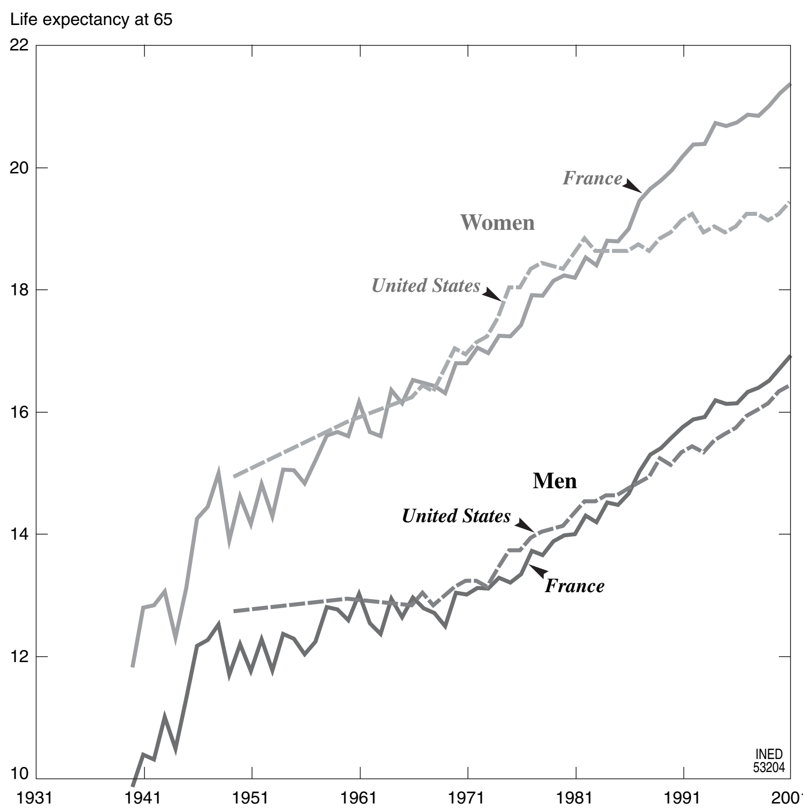


FIGURE 108–16 Evolution of life expectancy at 65 according to sex in the United States and in France, since 1940. (Sources: United States: life tables published annually in *Vital Statistics of the United States, Volume II*; France: Vallin and Meslé 2001a.)

had gained 1.3 years in the 15 years preceding Medicare and had gained another 2.1 years during the 15 following years, they have only gained another 1.1 years in the last 21 years.

But above all, the creation of Medicare coincided with the beginning of the cardiovascular revolution which has opened in all the Western industrial countries a new area for an increase in life expectancy, especially at advanced ages. If one compares the evolution in American life expectancy at 65 to that of the French, for example, we see that the acceleration of the 1970s probably has little to do with Medicare, except if one assumes that without Medicare life expectancy would have continued to stagnate among Americans and that the slowdown observed among American women from the 1980s would have occurred earlier.

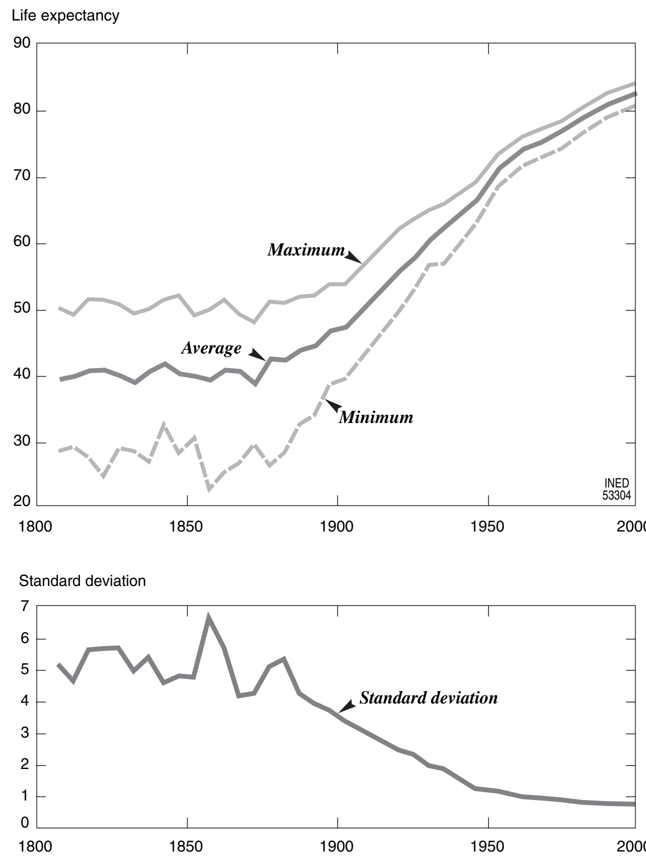


FIGURE 108–17 Evolution in two centuries of geographical variations in female life expectancy in France by department. (Source: Chap. 55, Fig. 14.)

2. The Reduction in Inequalities

Inequality in health is multiform: inequalities between sexes, differences between urban and rural areas, geographical variations, differences in social or ethnic groups, differences between developed and developing countries, and so on. It is also multifactorial: biological differences, selection factors, environmental constraints and nuisances (domestic, professional, residential, etc.), cultural and behavioral differences, and influences due to social (positive or negative) and economic policies. It has also in a diverse number of cases been the object of specific policies aimed at reducing these inequalities. Since it is not possible to deal with all of these in this chapter we shall limit ourselves to three types of disparities: geographical disparities, social inequality in relation to death, and finally North-South inequalities.

a. Geographical Disparities

The role of health policies in reducing geographical variations in mortality was discussed in Chapter 107. In this area, the policy objective is rarely specifically aimed at health but more generally at addressing inequalities in living conditions throughout the whole territory, the idea being that all the regions of a given country should have access to the same development opportunities and that measures implemented by the State should correct the disparities that may be caused by natural conditions in a free market context. France is a good example of the extraordinary change from a health point of view through the somewhat equal distribution throughout the whole of the country of the benefits of medical progress (Figure 108–17).

Whereas during the first half of the 19th century female life expectancy at birth was approximately 40 years nationally, it varied according to the *département* from less than 30 years to more than 50 years. In other words, the difference between the extremes was more than 20 years, more than half the average life expectancy. The standard deviation itself was about 5 years, approximately 13% of the average. This huge gap started to close during the second half of the 19th century, and the phenomenon accelerated from the 1890s, after the coming of the Third Republic, the discoveries made by Pasteur, and the creation of obligatory school attendance and free public, nonreligious schools. From the beginning of the 1950s, for a mean age of 71 years, the deviation between the extremes was no more than 5 years (7.5%) and the standard deviation of 1.1 (1.5%). Since then, geographical differences have lessened but more slowly. At the turn of the century, with a life expectancy of 82 years, the maximum deviation was only 3.3 years (4%) and the standard deviation was 0.7 (0.8%). It would be interesting to know to what extent this massive long-term reduction in geographic inequalities observed in France occurred elsewhere and if it differs according to the degree of national solidarity deployed as a result of the economic, social, and/or health policies. Unfortunately, we are far from having the required historical data.[13]

b. Social Inequalities

Reducing social inequalities in death is at the heart of health policies, namely in industrial countries which just after World War II developed collective systems for the coverage of health expenditure. These systems attempted to combine the philosophy of insurance (the healthy individual pays contributions for the sick person he may one day become) with that of social solidarity (the rich pay for the poor). If one judges by recent changes in mortality according to social category, the least one can say is that the result is not very good. In all European countries where it has been possible to measure them, social differences in mortality have tended to worsen in recent decades.

This evolution is clearly discernible in a country such as Finland, where it has been possible to follow it thanks to a systematic coupling of census data and civil registry information (Valkonen and Martelin, 1999). It is true that the trend for life expectancy at 35 (Figure 108–18) is toward an increase for all social categories, but the difference between workers and managers has greatly increased in 30 years, going from 4.5 years in 1972–1974 to 6.1 years in 1993–1995, among men. Among women, the phenomenon is less acute but just as real, the difference increasing from 2.3 to 3.0.

In Britain also, differences have increased (Figure 108–19). For men, the life expectancy differential at birth between category V (unskilled blue collar workers) and category I (managers) has gone from 5.5 years in 1972–1976 to 9.5 years in 1992–1996.

In France too, an unfavorable evolution in social mortality has been observed for the years from 1950 to 1980 (Desplanques, 1984, 1985), but more recent data do not give any clear indication as to this trend continuing.

Anton Kunst (1997) compared the evolution of social mortality in six European countries from the 1970s until the 1980s. Everywhere, the difference between manual and nonmanual occupations (Table 108–1) increased. And given what is revealed by more

[13] An attempt made by Tapani Valkonen (2001) at summarizing regional deviations in life expectancy in Europe shows to what extent this remains elusive. On the one hand, this study describes changes since the end of the 1960s and therefore begins at a time when the phenomenon was practically over in France. On the other hand, Valkonen was only able to gather data for fewer than half a dozen countries.

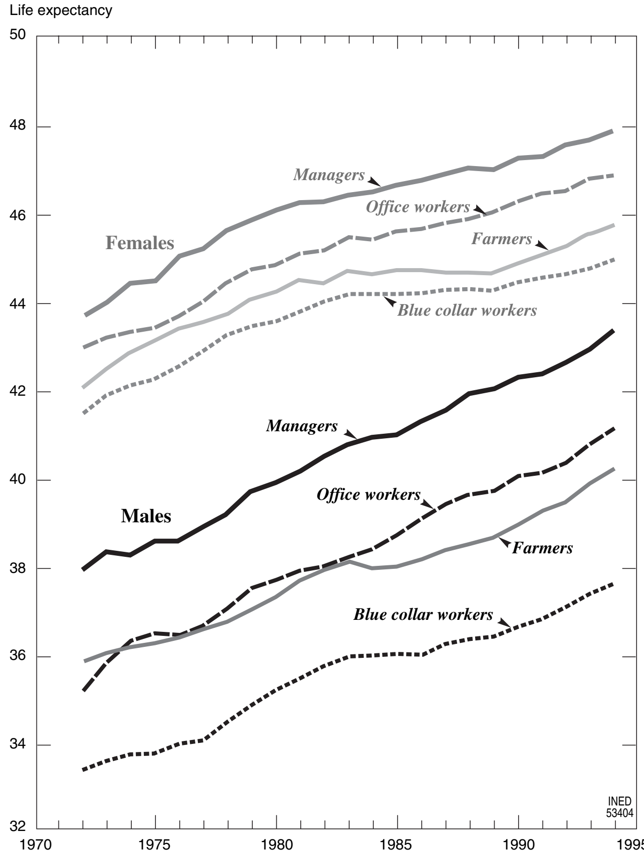


FIGURE 108–18 Evolution in life expectancy at 35 in Finland, by social category and sex, since 1970. (Source: Valkonen and Martelin, 1999; Valkonen, 2001.)

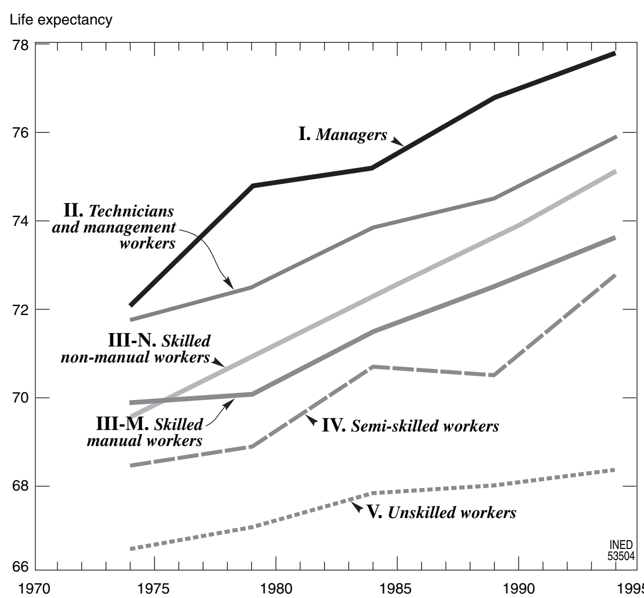


FIGURE 108–19 Evolution in male life expectancy at birth in England and Wales by social group since 1970. (Source: Hattersley, 1997; Valkonen, 2001.)

TABLE 108–1 Excess Mortality among Manual vs. Nonmanual Workers in Six European countries, in the 1970s and in the 1980s

Country	Corrected ratio[a]	
	1970s	1980s
Norway	1.18	1.34
Denmark	1.22	1.33
Sweden	1.26	1.41
England and Wales	1.25	1.44
Finland	1.40	1.53
France	1.61	1.71

[a] Ratio corrected to eliminate bias caused by not accounting for the nonworking population. Source: Kunst, 1997.

recent data for Finland and Britain, it is clearly apparent that differences will continue to widen.

This widening of social mortality differentials is obviously a failure in relation to the objective defined in 1985 by the WHO Regional Bureau for Europe: "By the year 2000, the actual differences in health status between and within countries should be reduced by at least 25%" (WHO, 1985).

This failure, however, is not necessarily that of the collective health coverage systems prevailing in Europe. It is not possible to know what the situation would have been in the absence of such systems which, it is assumed have made it possible to contain the adverse effects of strong factors. Featuring among these factors are indeed, for certain countries, increasing social inequalities and income differentials (Carrol and Davey Smith, 1997). But this explanation is not adequate since although it certainly applies for countries such as the United Kingdom, it has not played a role in countries, such as Finland, where differences in income have tended to decrease. The main factor in the recent aggravation in social mortality comes from the decline in cardiovascular mortality. The most privileged social groups benefited faster and earlier from this decline. This follows the same principle as the one evoked in the first part of this chapter concerning global evaluation: each major breakthrough in health introduces a phase of divergence because the countries with the greatest capacity to benefit from the progress do so before the others. The same thing applies, within a country, to social groups, the most privileged having a greater capacity to immediately benefit from the progress (Vallin and Meslé, 2004). An eloquent example is that of the role played in Britain by the decline in smoking. The change in attitude which contributed to reversing the trend of increasing lung cancer mortality first occurred among white-collar workers, greatly contributing to the increase in life expectancy differential between social groups (Harding *et al*., 1997).

It would be interesting to compare the European results with those of the United States where there is much less welfare coverage. Unfortunately, we do not have data for the United States that are comparable to the data concerning Europe because of the lack of a strict match between census and civil registry data. On the other hand, the United States has produced a great deal of statistics concerning what is referred to as "race."[14] Because black Americans have on average a socioeconomic status inferior to that of white Americans, this provides another angle from which to view social differentials in mortality.

We already mentioned in the previous section the rather disappointing results produced by the Medicare program on the evolution of life expectancy at 65. If the evolution of life expectancy among black Americans[15] is compared with that of white Americans, one notes, among other things, that not only has the differential not declined, but it has increased considerably (Figure 108–20).

Whereas life expectancy at 65 of whites increased by close to 2 years from 1959–1961 to 1985–1990 (13.0 to 14.9 years), that of blacks increased by hardly more than 1 years (11.9 to 13.1), and the absolute mean deviation almost doubled from 1.0 to 1.8 years. For women, the deviation—which was already greater than that for men at the beginning of the 1960s (1.3 years)—also increased in the same proportions, reaching 2 years in 1985–1990. Yet, the creation of Medicare was accompanied by that of Medicaid, intended to facilitate access to health care for the poor (see Chapter 117), and the elderly black population should therefore have largely benefited from these two programs. Here again, failure may not necessarily be attributable to the programs themselves, which probably lessened what would have occurred otherwise, but to the dual influence of the widening social gap in the United States and the cardiovascular revolution white Americans benefited from earlier.

[14] The notion of race is ambiguous because of its biological connotation. But what Americans refer to as "race" is more of an ethnic and cultural characteristic associated with a skin color with very ill-defined biological criteria (for more details, see Chapter 95).

[15] Mortality among black Americans is greatly underestimated at advanced ages by official statistics because of the exaggeration of the ages declared. However, in a recent article, Irma Elo and Samuel Preston (2004) published life expectancies at 65 since the 1940s, reestimated to correct this bias.

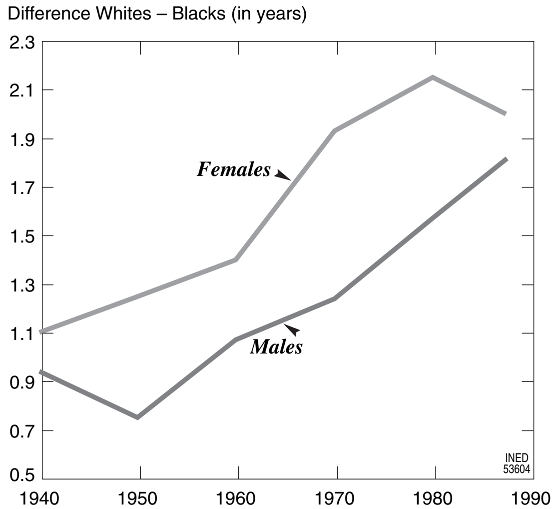


FIGURE 108–20 Difference in life expectancy at 65 between whites and blacks in the United States, by sex, 1940–1990. (Data for blacks from Elo and Preston, 2004; data for whites from life tables published annually in *Vital Statistics of the United States, Volume II*.)

c. North-South Inequalities

One of the major challenges in the reduction of differences in health statuses between developed and developing countries, to which WHO has been committed since its creation, has been to provide rural populations, which constitute the great majority of the population of developing countries, access to medical advances that are already available to most urban populations in the Third World. Such was the objective promoted by the primary health care strategy. And it is a fact that developing countries that have applied this strategy the most effectively have also been the most successful in rapidly reducing excess mortality in rural areas and in generally getting closer to the life expectancy of developed countries. However, this is on condition that it is done within the context of a development policy in which social issues weigh as much, if not more, than economic issues.

China, which inspired this strategy, has been one of the best examples of success, at least up until a certain stage, in spite of all the reservations that one might have concerning the political regime of this country. Unfortunately, there are very few data on rural mortality rates in China in the 1950s. Judith Banister (1987) quotes, with much precaution, a few results of recorded infant mortality, according to which the latter apparently declined from 73 to 20 per 1,000 between 1957 and 1974 in rural areas and from 35 to 10 in urban areas. Even though these figures are certainly underestimated, notably at the beginning of the period and in rural areas, the reduction in the difference between the city and the country is beyond doubt, indicating

TABLE 108–2 Evolution in Life Expectancy at Birth and Infant Mortality per 1,000 since the 1950s in China

	1950–1955	1960–1965	1970–1975	1981	1989–1990	1995
e_0	40.8	49.5	63.2	67.7	69.6	69.6
M_0	138.4	85.6	51.5	37.7	33.4	34.0*

Source: Attané, Chapter 112 (this volume).
*In 1997.

that the key to global success resides in a wide dissemination of improved health standards, the main assets of which undoubtedly have been the administrative organization of the territory, the mobilization of the masses, health education, and, above all, a definite will to satisfy the elementary needs of the great majority of the population (Mosley, 1985a, 1985b; see also Chapter 112). However, after the death of Mao Tse-tung, the liberalization of the economy benefited the urban minorities on the coast and considerably slowed down national health improvements. Thus, Chinese life expectancy, which had increased by close to 30 years from the beginning of the 1950s to the end of the 1970s—closing most of the gap with developed countries—reached a ceiling in the 1980s at just under 70 years (Table 108–2). Likewise, infant mortality, after having rapidly fallen from 140 to 38 per 1,000, has not declined any further in the last two decades.

In fact, since the beginning of the 1980s, it was clear that the strategy of primary health care could not succeed independently from the general context defined by the economic and social mode of development. Henry Mosley (1985b) cited, for example, the ambitious health operation of the province of Lampang in northern Thailand, where the recruiting and training of 6,000 primary health care assistants (for 600,000 inhabitants) and substantial financial support were not efficient in that the system did not focus enough on the basic needs of the population (food, infant health, drinking water, and public health) but on the contrary tended to insist on medical care, which had less of an effect on the health status of the population.

It is also at this time that in India, a vast politically heterogeneous federation, the success of the health policy of a state such as Kerala (Nayar, 1985) contrasted with the stagnant situation in Western Bengal (Nag, 1982). If, according to classic indicators, these two states had comparable medical resources, the different use made of them resulted in a wide difference in life expectancy. Having focused more on social development (social services, training, health services, public transport, etc.) and having encouraged greater awareness of the need and the right to use these services, Kerala made it possible to make the most of what was already—ahead of its time—a primary health strategy. In contrast to Kerala, Western Bengal, concerned about giving the priority to economic development, failed. Succeeding in considerably reducing social inequalities in death and in reducing the gap between urban and rural populations, Kerala attained one of the highest life expectancies in the developing world, whereas Western Bengal remained far behind.

Also very early on is the case of Costa Rica, which in 1970 implemented a model policy in primary health care for poor social groups in rural areas. Whereas in 1960–1964, the national infant death rate was more than twice that of the capital, San José, in 1978–1980 the relative deviation was no more than 5% (Rosero-Bixby, 1985). Thanks to this improvement, with a national rate of 20 per 1,000, the health status of Costa Rica caught up with that of industrial countries while in terms of economic performance it remained well below these countries.

But beyond examples of marked differences of this type, that one could list almost endlessly, what is the situation regarding the global reduction in North-South inequalities? To take into account the numbers of people concerned, Figure 108–21 represents the distribution of accumulated populations from different countries in the world according to life expectancy at three successive periods (1950–1955, 1975–1980, and 1995–2000). Developed countries are assembled under the continuous blue line and developing countries under the continuous pink line. At the beginning of the 1950s, life expectancies ranged between 30 and 75 years. The Third World, with life expectancies ranging for the most part between 30 and 55 years (only a few countries with low population rates did better), differed greatly from industrial countries whose life expectancy rates varied essentially between 60 and 70 years. This considerable dichotomy was characteristic of the huge difference that separated those whom Frantz Fanon (1961) called "the damned of the Earth" from the rich.

At the end of the 1970s, the global spectrum of life expectancies slightly narrowed to come closer to the maximum observed in 1950: there are no more countries with a life expectancy below 35 years and the 35–40 group is only composed of a few rare countries with low population rates, whereas no country has gone beyond the maximum bracket of 70–74 already present in 1950–1955. What is more, although the great majority of developed countries have got closer to this peak of the 1950s, a good many developing countries, including China, have closed most of the gap that separated them from developed countries whereas most, including India, had exceeded the threshold of 50 years life expectancy. At this time, however, developed and developing countries still remained far apart, the 65–69 years life expectancy bracket being only very weakly represented.

In 1995–2000, the panorama was upset once again, but under the effect of two different phenomena. On the one hand, the larger populations of developing countries continued to catch up with the top of their distribution, even exceeding the absolute maximum observed in 1975–1980, as certain countries in the group exceeded 75 years of life expectancy. On the other hand, because of phenomena of divergence, within both developed countries and developing countries, a rear guard emerged. As a consequence, the overall range widened, now going from 35 to 85 years. On the one hand indeed, sub-Saharan African countries (under the darker, dashed line) were practically

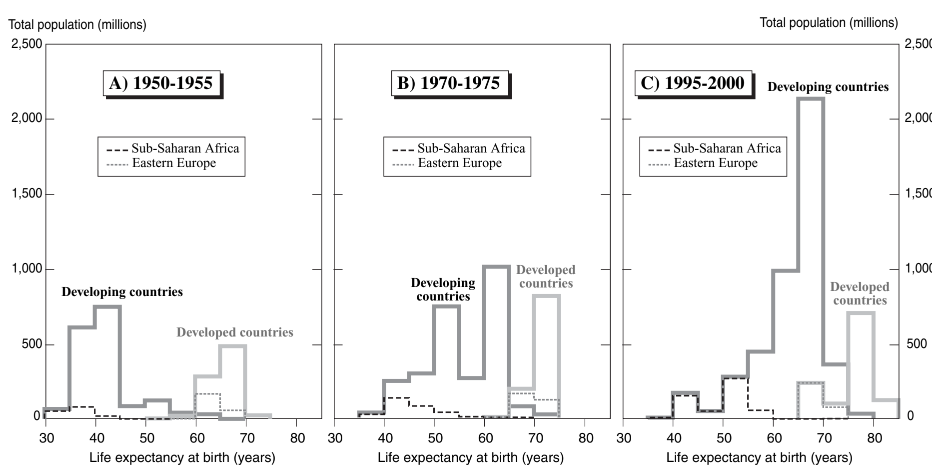


FIGURE 108–21 Total populations according to life expectancy level (Source: Meslé and Vallin, 2003a.) (Figure 108–21 also reproduced in color plate section.)

alone in having life expectancies below 55 years and some of them, namely those hardest hit by AIDS, significantly increased the number in the 40–44 life expectancy bracket. On the other hand, the cardiovascular revolution made it possible for Western industrial countries to take another leap forward, and they are clearly separate from East European countries (represented under the dotted blue line) which now find themselves at the same level as the larger populations of former Third World countries.

In other words, overall, the policies conducted, whether they concern health directly or more global actions and whether they have been conducted locally or are part of international collaborations, have for the majority of them definitely been part the driving force behind a decline in the North-South gap concerning the larger populations. But this result can only leave a bitter taste in the mouths of populations in the North and the South, which have for different reasons and in different ways been abandoned by the wayside.

CONCLUSION

Evaluating the results of health policies, especially when this expression is used to cover all collective measures directly or indirectly aimed at improving health, is usually a great challenge. However, the difficulty resides in the difference in degree according to the angle of observation chosen. It is at a maximum when one seeks to evaluate the results of a political system globally. It is less difficult and sometimes relatively easy to measure the success or the failure of a specific program aimed at obtaining a clearly defined objective. We have seen this, for example, in the case of the global eradication of smallpox, of the anti-malaria campaign of the end of the 1940s in Sri Lanka, of anti-alcohol campaigns, and with certain actions aimed at reducing infant mortality. It is then possible to evaluate the success or failure as well as the scale and sustainability of the results. Concerning policies aimed at reducing inequalities, there again, situations vary, but in most cases problems of interpretation of the same order of complexity are to be found as those concerning global evaluations. This is especially true concerning social inequalities or recent evolutions that appear to present a failure in systems for the collective coverage of health expenditures, but where in fact the aggravation of social inequalities is undoubtedly less than it would have been in the absence of such systems, because of the powerful divergence factors affecting the societies concerned.

More than in any other area, health policies are confronted here with a reality that is becoming increasingly obvious: it would be pointless to assume a health policy can reduce social inequality in death independently from a reduction in inequalities in living conditions in all domains, from the economic to the cultural. More generally, it seems that the main conclusion of this chapter is that it is impossible to dissociate the global results of a health policy from all of the effects produced by economic, social, and cultural policies with which it forms a whole, even though this whole is made up of contradictions. In addition, our analytical instruments are for the time being far too crude and inaccurate to be able to evaluate global results with precision. We may be better equipped to do so if one day we are able to have statistical series allowing long-term comparisons based on more precise indicators such as those suggested in Chapter 109 by Viviana Egidi and Vittoria Buratta.

Acknowledgment

This chapter was translated from the French by Paul Belle.

References

Aaby Peter, 1988. *Le surpeuplement, un facteur déterminant de la mortalitépar rougeole en Afrique.* Paris, INED, 53 p. (Dossiers et recherches, no. 15).

Aaby Peter, 1989. La promiscuité, un facteur déterminant de la mortalité par rougeole en Afrique, *in:* Gilles Pison, Étienne van de Walle, and Mpembele Sala Diakanda (eds.), *Mortalité et société en Afrique au sud du Sahara,* p. 295–324. Paris, INED, PUF, 446 p. (Travaux et documents, Cahier no. 124).

Aaby Peter, 1992. Overcrowding and intensive exposure: major determinants of variations in measles mortality in Africa, *in:* Étienne van de Walle, Gilles Pison, and Mpembele Sala Diakanda (eds.), *Mortality and society in sub-Saharan Africa,* p. 319–348. Oxford, Clarendon Press, 450 p. (International Studies in Demography).

Aaby Peter, Bukh Jette, Lisse Ida Maria M., and Smits Arjon J., 1984. Measles vaccination in child mortality: a community study from Guinea-Bissau, *Journal of Infection*, vol. 8(1), p. 13–21.

Allen James R., 1984. Epidemiology of the acquired immunodeficiency syndrome (AIDS) in the United States, *International Seminars in Surgical Oncology,* vol. 11(1), p. 4–11.

Amin Ruhul, 1996. Immunization coverage and child mortality in two rural districts of Sierra Leone, *Social Science and Medicine*, vol. 42(11), p. 599–604.

Amoroso Anthony, Davis Charles E., and Redfield Robert R., 2002. Antiretroviral therapy in resource-limited settings, *in:* Max Essex, Souleymane Mbup, Phyllis J. Kanki, Richard G. Marlink, and Sheila D. Tlou (eds.), *AIDS in Africa,* 2nd ed., p. 322–344. New York, Kluwer Academic and Plenum Publisher, 724 p.

Aristotle, [1874]. *La politique.* Translated by Jules-Barthélemy Saint Hilaire. Paris, Ladrange, 545 p.

Banister Judith, 1987. *China's changing population.* Stanford, Stanford University Press, 488 p.

Barré-Sinoussi Françoise, Chermann Jean-Claude, Rey Françoise, Nugeyre Marie-Thérèse, Chamaret Sophie, Gruest Jacqueline, Dauget Claudine, Axler-Blin Charles, Vézinet-Brun

Françoise, Rouzioux Christine, Rozenbaum Willy, and Montagnier Luc, 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science*, vol. 220(4599), p. 868–871.

Beaton George H. and Ghassemi Hossein, 1979. *Supplementary feeding programmes for young children in developing countries*. Unpublished report. New York and Toronto, UNICEF and University of Toronto.

Beghin Ivan and Vanderveken Marc, 1985. Nutritional programmes, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 81–100. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Beghin Ivan and Vanderveken Marc, 1985. Les programmes nutritionnels, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 77–98. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Beghin Ivan D. and Viteri Fernando E., 1973. Nutritional rehabilitation centers: an evaluation of their performances, *Journal of Tropical Pediatrics and Environmental Child Health*, vol. 19(4), p. 404–416.

Berger Guy, Mauffret Marie, Rousseau-Giral Anne-Chantal, and Zaidman Catherine, 2000. *La loi relative à la lutte contre le tabagisme et l'alcoolisme. Rapport d'évaluation.* Paris, La documentation française, 555 p.

Biraben Jean-Noël, 1988. Les progrès de la médecine, *in:* Jacques Dupâquier (ed.), *Histoire de la population française*, vol. 4, p. 263–277. Paris, PUF, 590 p.

Blanc Michel, 1985. Les effets à long terme des programmes d'intervention contre le tabagisme: application à la France, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort. Influence des politiques sociales et des politiques de santé sur l'évolution de la mortalité,* p. 238–256. Paris, INED/PUF, 542 p. (Travaux et documents, Cahier no. 108).

Blanc Michel, 1985. The long-term effect of anti-smoking programmes, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 241–260. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Blayo Yves, 1975. La mortalité en France de 1740–1829, *Population*, vol. 50(special issue), p. 123–142.

Brouard Nicolas, 1994. Aspects démographiques et conséquences de l'épidémie de sida en Afrique, *in:* Jacques Vallin (ed.), *Populations africaines et sida,* p. 120–178. Paris, La Découverte, CEPED, 224 p. (Recherche).

Caraël Michel and Holmes King, 2001. The multicentre study of factors determining the different prevalences of HIV in Sub Saharan Africa, *AIDS*, vol. 15(suppl. 4), p. 1–132.

Card David, Dobkin Carlos, and Maestas Nicole, 2004. *The impact of nearly universal insurance coverage on health care utilization and health: evidence from Medicare.* Cambridge, National Bureau of Economic Research, 73 p. (Working Paper 10365; http://www.nber.org/papers/w10365).

Carrol Doug and Davey Smith George, 1997. Health and socioeconomic position. A commentary, *Journal of Health Psychology*, vol. 2, p. 275–282.

Caselli Graziella, Meslé France, and Vallin Jacques, 1999. Le triomphe de la médecine, *in:* Jean-Pierre Bardet and Jacques Dupâquier (eds.), *Histoire des populations de l'Europe. III. Les temps incertains, 1914–1998,* p. 125–181. Paris, Librairie Arthème Fayard, 792 p.

Caselli Graziella, Meslé France, and Vallin Jacques, 2002. Epidemiologic transition theory exceptions, *Genus*, vol. 58(1), p. 9–52.

CDC (Centers for Disease Control), 1995. AIDS cases, case-fatality rates, and deaths by half-year and age group, through December 1995, United States, *HIV/AIDS Surveillance Report*, vol. 7(2), p. 18, Table 13.

CDC (Centers for Disease Control), 2002. AIDS cases and deaths by year and age group, through December 2001, United States, *HIV/AIDS Surveillance Report*, vol. 13(2), p. 30, Table 21.

Chapalain Marie-Thérèse, 1975. Présentation de la rationalisation des choix budgétaires au Ministère de la santé, *Cahiers de sociologie et de démographie médicale*, vol. 15(2), p. 43–45.

Chen Lincoln C., Chowdhury A. K. M. Alauddin, and Huffman Sandra L., 1980. Anthropometric assessment of energy-protein malnutrition and subsequent risk of mortality and use of health services surveys, *International Journal of Epidemiology*, vol. 33(8), p. 1836–1845.

Chen Lincoln C., Singh Sagari, and Victora Cesar G., 1997. *Sustainability of the World Summit for Children Goals: concepts and strategies.* New York, UNICEF, 45 p. (UNICEF Staff Working Papers, no. EVL-97–004).

Cohen Joel E., 1988. Estimating the effects of successful malaria control programs on mortality, *Population Bulletin of the United Nations*, no. 25, p. 6–26.

Cole Timothy J., Foord Frances A., Watkinson M., Lamb W. H., and Whitehead Roger G., 1995. The Keneba pregnancy supplementation study, *Bulletin of the WHO*, vol. 73(suppl.), p. 72–74.

Condorcet Marie-Jean Antoine Nicolas, Cariat Marquis De, [1795] 1933. *Esquisse d'un tableau historique des progrès de l'esprit humain.* Edited and revised by O. H. Prieur. Paris, Boivin et Cie, 239 p.

De Quadros Ciro A., 2004. Can measles be eradicated globally? *Public Health Reviews*, vol. 82(2), p. 134–138.

Desgrées Du Loû Annabel, 1996. *Sauver les enfants: le rôle des vaccinations*. Paris, CEPED, 261 p. (Les Études du CEPED, no. 12).

Desplanques Guy, 1984. L'inégalité sociale devant la mort, *Économie et statistique*, no. 162, p. 29–50.

Desplanques Guy, 1985. *La mortalité des adultes. Résultats de deux études longitudinales (période 1955–1980).* Paris, INSEE, 212 p. (Les Collections de l'INSEE, série D, no. 102).

Doll Richard, Peto Richard, Jillian Boreham, and Sutherland Isabelle, 2004. Mortality in relation to smoking: 50 years' observations on male British doctors, *British Medical Journal*, vol. 328(7455), p. 1519–1533.

DSCR, 2004. *Les grandes données de l'accidentologie. Caractéristiques et causes des accidents de la route.* Paris, Direction de la sécurité et de la circulation routières, 6 p.

Elo Irma T. and Preston Samuel H., 2004. The African American population, 1930 to 1990, *in:* Elijah Anderson and Douglas S. Massey (eds.), *Problem of the Century,* p. 168–223. New York, Russell and Sage Foundation, 470 p.

Euro HIV (WHO and UNAIDS Collaborating Centre on AIDS), 2003. Mid-year report 2003, *HIV/AIDS Surveillance in Europe*, no. 69, p. 1–54.

Fanon Frantz, 1961. *Les damnés de la terre.* Preface by Jean-Paul Sartre. Paris, François Maspéro, 244 p.

Fischl Margaret A., Richman Douglas D., Grieco Michael H., Gottlieb Michael S., Volberding Paul A., Laskin Oscar L., Leedom John M., Groopman Jerome E., Mildvan Donna, Schooley Richard T. *et al.*, 1987. The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex, *New England Journal of Medicine*, vol. 317(4), p. 185–191.

Fleury Michel and Henry Louis, 1958. Pour connaître la population de la France depuis Louis XIV: plan de travaux par sondage, *Population*, vol. 13(4), p. 663–686.

Frederiksen Harald, 1960. Malaria control and population pressure in Ceylon, *Public Health Reports*, vol. 75(10), p. 865–868.

Frederiksen Harald, 1961. Determinants and consequences of mortality trends in Ceylon, *Public Health Reports*, vol. 76, p. 659–663.

Frederiksen Harald, 1966. Determinants and consequences of mortality and fertility trends, *Public Health Reports*, vol. 81, p. 715–727.

Giglioli George, Ch'en Wan-I, Marchant, D. E., and Howell P., 1976. Malaria eradication in Guyana, *Tropical Doctor*, vol. 6(3), p. 126–132.

Gray Ronald H., 1974. The decline of mortality in Ceylon and the demographic effects of malaria control, *Population Studies*, vol. 28(2), p. 205–229.

Grmek Mirko D., 1989. *Histoire du sida.* Paris, Éditions Payot, 392 p.

Gwatkin Davidson R., Wilcox Janet R., and Wray Joe D., 1980. *Can health and nutrition intervention make a difference?* Washington (DC), Overseas Development Council, 76 p. (Overseas Development Council Monograph, no. 13).

Hammond E. Cuyler, 1966. Smoking in relation to the death rates of one million men and women, *in:* William Haenszel (ed.), *Epidemiological approaches to the study of cancer and other chronic diseases,* p. 127–204. Bethesda (Maryland), National Cancer Institute, 495 p. (National Cancer Institute Monograph, no. 19).

Harding Seeromanie, Bethune Ann, Maxwell Roy, and Brown Joanna, 1997. Mortality trends using the Longitudinal Study, *in:* Frances Drever and Margaret Whitehead (eds.), *Health Inequalities. Decennial Supplement,* p. 143–155. London, The Stationery Office, 257 p. (Series DS no. 15).

Hartfield J. and Morley David, 1963. Efficacy of measles vaccine, *Journal of Hygiene (Cambridge)*, vol. 61(March), p. 143–147.

Hattersley Lin, 1997. Expectation of life by social class, *in:* Frances Drever and Margaret Whitehead (eds.), *Health Inequalities. Decennial Supplement,* p. 73–82. London, The Stationery Office, 257 p. (Series DS no. 15).

Henry Louis, 1980. *Techniques d'analyse en démographie historique.* Paris, Éditions de l'Institut national d'études démographiques, 177 p.

Herrera M. Guillermo, Nestel Penelope, El Amin Alawia, Fawzy Wafaie, Mohamed Kamal A., and Weld Leisa, 1992. Vitamin A supplementation and child survival, *The Lancet*, vol. 340(8814), p. 267–271.

Hessol Nancy A., Buchbinder Susan P., Colbert David, Scheer Susan, Underwood Ronald, Barnhart J. Lowell, O'Malley Paul M., Doll Lynda S., and Lifson Alan R., 1992. Impact of HIV infection on mortality and accuracy of AIDS reporting on death certificates, *American Journal of Public Health*, vol. 82(4), p. 561–564.

Högberg Ul and Broström Göran, 1985. The demography of maternal mortality. Seven Swedish parishes in the 19th century, *International Journal of Gynecology and Obstetrics*, vol. 23(6), p. 489–497.

Holman C. D'Arcy J., English Dallas R., Milne Elizabeth, and Winter Michael G., 1996. Meta-analysis of alcohol and all-cause mortality: a validation of NHMRC recommendations, *Medical Journal of Australia*, vol. 164(3), p. 141–145.

Holmberg Ingvar, 1970. A study of mortality among cohorts born in the 18th and 19th century, *in:* Statistiska centralbyran (ed.), *Cohort mortality of Sweden. Three studies describing past, present, and future trends in mortality,* p. 71–86. Stockholm, Statistiska centralbyran/National Central Bureau of Statistics, 86 p. (Statistiska Meddellanden/Statistical reports 1970).

Kasongo project team (The), 1981. Influence of measles vaccination on survival pattern of 7–35-month-old children in Kasongo, *The Lancet*, vol. 317(8223), p. 764–767.

Khlat Myriam (ed), 1997a. *Demographic evaluation of health programmes.* Paris, CICRED, UNFPA and Ministère de la coopération, 226 p.

Khlat Myriam, 1997b. Program evaluation in health field, *in:* Myriam Khlat (ed.), *Demographic evaluation of health programmes,* p. 5–13. Paris, CICRED, UNFPA and Ministère de la coopération, 226 p.

Kielman Arnfried A. and McCord Colin, 1978. Weight for age as an index of death in children, *The Lancet*, vol. 311(8076), p. 1247–1250.

Kunst Anton E., 1997. *Cross-national comparisons of socio-economic differences in mortality.* Doctoral thesis. Rotterdam, Erasmus University, 264 p.

Lert France, Obadia Yolande, and L'équipe Vespa, 2004. Comment vit-on en France avec le VIH/sida? *Population et Société*, no. 406, p. 1–4.

Loudon Irvin, 1992. *Death in childbirth: an international study of maternal care and maternal mortality 1800–1950.* Oxford, Clarendon Press, 622 p.

Meslé France, 1995. La mortalité en France: le recul se poursuit, *Population*, vol. 50(3), p. 743–776.

Meslé France, 2000. *Mortalité infantile: une chute spectaculaire.* Paris, INED, 2 p. (Fiches d'actualité scientifique, no. 3).

Meslé France, 2004a. Espérance de vie: un avantage féminin menacé, *Population et sociétés*, no. 402, p. 1–4.

Meslé France, 2004b. Mortality in Central and Eastern Europe: long-term trends and recent upturns, *Demographic Research*, vol. Special Collection, no. 2: "Determinants of diverging trends in mortality," p. 46–70.

Meslé France, Shkolnikov Vladimir, and Vallin Jacques, 1994. Brusque montée des morts violentes en Russie, *Population*, vol. 49(3), p. 780–790.

Meslé France and Vallin Jacques, 1993. Développement économique et espérance de vie: la transition sanitaire au tournant des années soixante, *in:* IUSSP/UIESP (ed.), *Congrès international de la population. Montréal 1993*, vol. 2, p. 365–382. Liège, IUSSP/UIESP, 493 p.

Meslé France and Vallin Jacques, 1996. Reconstructing long-term series of causes of death, *Historical Methods*, vol. 29(2), p. 72–87.

Meslé France and Vallin Jacques, 2002. Mortalité en Europe: la divergence Est-Ouest, *Population-F*, vol. 57(1), p. 171–212.

Meslé France and Vallin Jacques, 2002. Mortality in Europe: the divergence between East and West, *Population-E*, vol. 57(1), p. 157–198.

Meslé France and Vallin Jacques, 2003a. La mortalité dans le monde: tendances et perspectives, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (ed.), *La population du monde: géants démographiques et défis internationaux,* p. 527–546. Paris, INED, PUF, 768 p. (Les Cahiers de l'INED, no. 149, second edition).

Meslé France and Vallin Jacques, 2003b. *Mortalité et causes de décès en Ukraine au XX^{e} siècle.* Paris, INED, 396 p. (Les cahiers de l'INED, PUF, cahier no. 152).

Meslé France, Vallin Jacques, and Shkolnikov Vladimir, 1998. Reversal of mortality decline: the case of contemporary Russia, *World Health Statistics Quarterly/Rapport trimestriel de statistiques sanitaires*, vol. 51(2–4), p. 191–206. ("Historical epidemiology: mortality decline, and old and new transitions in health," special issue edited by Odile Frank).

Molineaux Louis, 1985. The impact of parasitic diseases and their control, with special emphasis on malaria in Africa, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (ed.), *Health policy, social policy and mortality prospects,* p. 13–44. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

MOLINEAUX Louis, 1985. La lutte contre les maladies parasitaires: le problème du paludisme, notamment en Afrique, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 11–50. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

MOLINEAUX Louis and GRAMICCIA Gabriele, 1980. *The Garki project. Research on the epidemiology and control of malaria in the Sudan savana of West Africa.* Geneva, WHO, 311 p.

MORA J. O., 1979. Nutrition supplementation and the outcome of pregnancy. I: Birth weight, *American Journal of Clinical Nutrition*, vol. 32(2), p. 455–462.

MORLEY David C., 1973. *Pediatric priorities in the developing world.* London, Butterworths, 470 p.

MOSLEY W. Henry, 1985a. Biological and socioeconomic determinants of child survival. A proximate determinants framework integrating fertility and mortality variables, *in:* IUSSP (ed.), *International Population Conference, Florence 1985*, vol. 2, p. 189–208. Liège, UIESP/IUSSP, 456 p.

MOSLEY W. Henry, 1985b. Les soins de santé primaires peuvent-ils réduire la mortalité infantile? Bilan critique de quelques programmes africains et asiatiques, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 101–136. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

MOSLEY W. Henry, 1985b. Will primary health care reduce infant and child mortality? A critique of some current strategies, with special reference to Africa and Asia, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 103–138. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

MOSTAFA Golam and RAHMAN Mizamur, 1997. Impact of birth care pratices on neonatal mortality in Matlab, Bangladesh, *in:* Myriam Khlat (ed.), *Demographic evaluation of health programmes,* p. 143–164. Paris, CICRED, UNFPA and Ministère de la coopération, 226 p.

MSPSS (Ministère de la santé publique et de la sécurité sociale), 1971. *Pour une politique de la santé.* Paris, La documentation française, 14 fasc.

MULDER Daan W., NUNN Andrew J., KAMALI Anatoli, NAKIYINGI Jessica, WAGNER Hans-Ulrik and KENGEYA-KAYONDON Jane F., 1994. Two-year HIV-1 associated mortality in Uganda rural population, *The Lancet*, vol. 343(April 23), p. 1021–1023.

NAG Moni, 1982. *Impact of social development and economic development on mortality: a comparative study of Kerala and West Bengal.* New York, Population Council, 51 p.

NAYAR P. K. B., 1985. The case of Kerala, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 371–382. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

NAYAR P. K. B., 1985. Le Kerala, Inde, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 357–368. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

NELSON Kenrad E., CELENTANO David D., EIUMTRAKUL Sakol, HOOVER Donald R., BEYRER Chris, SUPRASERT Somboon, KUNTOLBUTRA Surinda and KHAMBOONRUANG Chirasak, 1996. Changes in sexual behavior and a decline in HIV infection among young men in Thailand, *New England Journal of Medicine*, vol. 335(5), p. 297–303.

NEWMAN Peter, 1965. *Malaria eradication and population growth: with special reference to Ceylon and Guiana.* Ann Arbor, The University of Michigan, School of Public Health, 259 p. (Bureau of Public Health Economics Research Series, no. 10).

NEWMAN Peter, 1977. Malaria and mortality, *Journal of the American Statistical Association*, vol. 72(358), p. 257–263.

NIZARD Alfred and MUNOZ-PEREZ Francisco, 1993. Alcool, tabac et mortalité en France depuis 1950. Incidence de la consommation d'alcool et de tabac sur la mortalité, *Population*, vol. 47(4), p. 975–1014.

OFS (Federal Statistical Office), 1974. Statistikaj Supervidoj, *Demosta*, vol. 7(3–4).

OFS (Federal Statistical Office), 1980. Statistikaj Supervidoj, *Demosta*, vol. 13(3).

OMRAN Abdel R., 1971. The epidemiologic transition: a theory of the epidemiology of population change, *The Milbank Memorial Fund Quarterly*, vol. 49(4), p. 509–538.

OMS/WHO-ROE, 2002. *Rapport sur la santé en Europe 2002.* Copenhagen, World Health Organisation, Regional Office for Europe, 156 p. (WHO, Regional publications, European series, no. 97).

PAYNE D., GRAB B., FONTAINE R. E., and HEMPEL J. H., 1976. Impact of control measures on malaria transmission and general mortality, *Bulletin of the World Health Organisation*, vol. 54, p. 445–452.

POIKOLAINEN Kari, 1995. Alcohol and mortality: a review, *Journal of Clinical Epidemiology*, vol. 48(4), p. 455–465.

PRÉLOT Marcel, 1961. *La science politique.* Paris, PUF, 128 p. ("Que sais-je?," no. 909).

PUFFER Ruth Rice and SERRANO Carlos V., 1973. *Patterns of mortality in childhood. Report of the inter-American investigation of mortality in childhood.* Washington, Pan American Health Organisation (PAHO), 470 p. (PAHO Scientific Publication no. 262).

RAHMATHULLAH Lakshmi, UNDERWOOD Barbara A., THULASIRAJ Ramilla D., MILTON Roy C., RMASWAMY Kala, RAHMATHULLAH Rahem and BABU Ganeesh, 1990. Reduced mortality among children in southern India receiving a small weekly dose of vitamin A, *New England Journal of Medicine*, vol. 323(14), p. 929–987.

RCP (Royal College of Physicians), 2000. *Nicotine addiction in Britain.* London, Royal College of Physicians, 190 p.

RCP and ASH, 2002. *Forty fatal years.* London, Royal College of Physicians and Action on Smoking and Health, 30 p.

ROEMER Milton, 1985. Politiques sociales et systèmes de santé: leurs effets sur la mortalité et la morbidité dans les pays développés, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort. Influence des politiques sociales et des politiques de santé sur l'évolution de la mortalité,* p. 523–535. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

ROEMER Milton, 1985. Social policies and health care systems: their effects on mortality and morbidity in developed countries, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 541–552. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

ROEMER Milton and ROEMER Ruth, 1981. *Health care systems and comparative manpower policies.* New York, Marcel Dekker Co, 411 p. (Health Policy).

ROSERO-BIXBY Luis, 1985. The case of Costa Rica, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects,* p. 341–370. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

ROSERO-BIXBY Luis, 1985. Le Costa Rica, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *La lutte contre la mort,* p. 331–356. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Ross David A., Kirkwood Betty R., Binka Fred N., Arthur Paul, Dollimore Nicola, Morris Saul S., Shier Rosaleen P., Guapong John O., and Smith Peter G., 1995. Child morbidity and mortality following vitamin A supplementation in Ghana: time since dosing, number of doses, and time of year, *American Journal of Public Health*, vol. 85(9), p. 1246–1251.

Rumeau-Rouquette Claude, 1979. *Naître en France.* Paris, INSERM, 162 p. (Santé publique).

Rumeau-Rouquette Claude and Blondel Béatrice, 1985. The French perinatal programme, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 299–312. Liège, Ordina Edition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Rumeau-Rouquette Claude and Blondel Béatrice, 1985. Un exemple: le programme français de périnatalogie, *in:* Jacques Vallin, Alan Lopez, and Hugo Behm (ed.), *La lutte contre la mort,* p. 293–304. Paris, INED, PUF, 542 p. (Travaux et documents, Cahier no. 108).

Scholz Rembrandt and Maier Heiner, 2003. *German unification and the plasticity of mortality at older ages.* Rostock, Max Planck Institute for Demographic Research, 22 p. (Working paper no. wp-2003–031).

Séguy Isabelle, Caleçon Hélène, Méric Corinne, and Le Sager Fabienne, 2001. *La population de la France de 1670–1829. L'enquête de Louis Henry et ses données.* Paris, INED, 208 p. (Études et enquêtes historiques).

Sénécal Jean, Roussey Michel, Bouvier-Colle Marie-Hélène, Hatton Françoise, and Paclot Catherine, 1998. À propos de la mort subite du nourrisson, *Population*, vol. 53(4), p. 841–846.

Shkolnikov Vladimir, Meslé France and Vallin Jacques, 1995. La crise sanitaire en Russie. I. Tendances récentes de l'espérance de vie et des causes de décès de 1970–1993, II. Évolution des causes de décès: comparaison avec la France et l'Angleterre (1970–1993), *Population*, vol. 50(4–5), p. 907–982.

Shkolnikov Vladimir, Meslé France, and Vallin Jacques, 1995. Health crisis in Russia. I. Recent trends in life expectancy and causes of death from 1970 to 1993, II. Changes in causes of death: a comparison with France and England and Wales (1970 to 1993), *Population, An English Selection*, vol. 8, p. 123–190.

Shkolnikov Vladimir and Nemtsov Alexander, 1997. The anti-alcohol campaign and variations in Russian mortality, *in:* Jose Luis Bobadilla, Christine A. Costello and Faith Mitchell (ed.), *Premature death in the New Independent States,* p. 239–261. Washington DC, National Academy Press, 504 p.

Sommer Alfred, Tarwotjo Ignatius, Djunaedi E., West Keith P., Loeden A. A., and Tilden Robert, 1986. Impact of vitamin A supplementation on childhood mortality. A randomized controlled community trial, *The Lancet*, vol. 8491(1), p. 1169–1173.

Stine Gerald J., 1998. *Acquired immune deficiency syndrome. Biological, medical, social, and legal issues.* Upper Saddle River (New Jersey), Prentice Hall, 610 p.

Taylor Carl E., Sarma R. S. S., Parker Robert, Reinke William A., and Farugee Rashid, [1968] 1983. *Child and maternal health services in rural India: the Narangwal experiment. Volume 2: Integrated family planning and health care*. Baltimore, Johns Hopkins University Press, 252 p.

United Nations, 2003. *World population prospects. The 2002 revision.* New York, United Nations, Department for Economic and Social Affairs, Population Division, vol. 1: 781 p., vol. 2: 939 p. and CD-Rom (ST/ESA/Ser.A/225).

Valkonen Tapani, 2001. Tendances en matière de mortalité différentielle en Europe, *in:* Conseil de l'Europe, *Tendances en matière de mortalité et mortalité différentielle,* p. 189–331.– Strasbourg, éditions du Conseil de l'Europe, 344 p. (Études démographiques, no. 36).

Valkonen Tapani, 2001. Trends in differential mortality in European countries, *in:* Council of Europe, *Trends in mortality and differential mortality,* p. 185–321. Strasbourg, Council of Europe Publishing, 332 p. (Population Studies, no. 36).

Valkonen Tapani and Martelin Tuija, 1999. Social inequality in the face of death, *in:* Juha Alho (ed.), *Statistics, registries and science. Experiences from Finland,* p. 211–224. Keuruu, Statistics Finland and Otava Book Printing, 256 p.

Vallin Jacques, 1992. Impact démographique de la vaccination dans les pays en développement, *in:* Jean-Paul Moatti and Claude Mawas (ed.), *Évaluation des innovations technologiques et décisions en santé publique,* p. 235–247. Paris, INSERM, 278 p. (Analyses et prospective).

Vallin Jacques and Chesnais Jean-Claude, 1975. Les accidents de la route en France. Mortalité et morbidité depuis 1953, *Population*, vol. 30(3), p. 443–478.

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925–21978.* Paris, INED, PUF, 608 p. (Travaux et Documents, Cahier no. 115. With 7 annexed vols).

Vallin Jacques and Meslé France, 1998. Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925–1993, *in:* Gérard Pavillon (ed.), *Enjeux des classifications internationales en santé,* p. 113–156. Paris, Éditions INSERM, 220 p. (Questions en santé publique).

Vallin Jacques and Meslé France, 1999. *Le rôle des vaccinations dans la baisse de la mortalité.* Paris, INED, 16 p. (Dossiers et recherches no. 74).

Vallin Jacques and Meslé France, 2001a. *Tables de mortalité françaises pour les XIX^e et XX^e siècles et projections pour le XXI^e siècle.* Paris, INED, 102 p. + CD-rom (Données statistiques, no.4-2001).

Vallin Jacques and Meslé France, 2001b. Évolution de la mortalité en Europe depuis 1950: mortalité par âge, sexe et causes de décès, *in:* Conseil de l'Europe, *Tendances en matière de mortalité et mortalité différentielle,* p. 33–188. Strasbourg, Éditions du Conseil de l'Europe, 344 p. (Études démographiques no. 36).

Vallin Jacques and Meslé France, 2001b. Trends in mortality in Europe since 1950: age-, sex- and cause-specific mortality, *in:* Council of Europe, *Trends in mortality and differential mortality,* p. 31–186. Strasbourg, Council of Europe Publishing, 334 p. (Population Studies no. 36).

Vallin Jacques and Meslé France, 2004. Convergences and divergences in mortality. A new approach of health transition, *Demographic Research*, vol. Special Collections no. 2 "Determinants of diverging trends in mortality," p. 1–23.

van Ginneken Jeroen, Bairagi Radheshyan, De Francisco Andres, Sarder Abdul Mazid, and Vaughan Patrick, 1998. *Health and demographic surveillance in Matlab: past, present and future.* Dacca, International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), 44 p.

Victora Cesar G., Bryce Jennifer, Fontaine Olivier and Monasch Roeland, 2000. Reducing deaths from diarrhoea through oral rehydration therapy, *Bulletin of the World Health Organization*, vol. 78(10), p. 1246–1255.

Walsh Julia A., 1983. Selective primary health care: strategies for control of diseases in developing world. IV measles, *Review of Infectious Diseases*, vol. 5(2), p. 330–340.

Waltisperger Dominique, Vallin Jacques, and Ben Mrad Abdelhamid, 2001. La dynamique naturelle de la population depuis l'Indépendance, *in:* Jacques Vallin and Thérèse Locoh (eds.), *Pop-*

ulation et développement en Tunisie: la métamorphose, p. 53–88. Tunis, Cérès Éditions, 801 p.

West Keith P. Jr., Pokhrel Ram P., Katz J. Joanne, Leclerq Steven T., Khatry Subarna K., Shrestha Sharada Ram, Pradhan Elizabeth Kimbrough, Tielsch James M., Pandey M. R. and Sommer Alfred, 1991. Efficacy of vitamin A in reducing preschool child mortality in Nepal, *The Lancet*, vol. 338(2), p. 67–71.

WHO, 1985. *Targets for health for all. Targets in support for the European strategy for health for all*. Copenhagen, World Health Organization, Regional Office for Europe, 201 p.

WHO, UNICEF and UNFPA, 2004. *Maternal mortality in 2000.* Geneva, World Health Organization, 30 p.

Wrigley Edward A. and Schofield Roger S., 1981. *The population history of England 1541–1871.* London, Edward Arnold, 781 p.

CHAPTER

109

Information Systems for Health Policies

VIVIANA EGIDI AND VITTORIA BURATTA

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Instituto Nationale di Statstica (ISTAT), Rome, Italy

INTRODUCTION

Over the recent decades, numerous countries have reorganized their health systems considerably in order to increase effectiveness, improve the quality of service, improve the ability to respond positively to the population's needs, and assess and improve the system's fairness. This revision had become all the more urgent in the need to curb the generalized trend of rapid increase in the cost of health services brought about by both technological progress and the aging of the population (Moise and Jacobzone, 2003). This trend was all the more worrying as it coincided with a stagnating economy and increasingly limited resources.

The solutions adopted, whether at an institutional level or at a management level, are as varied (Docteur and Oxley, 2003) as the historical and cultural contexts of the different countries, but they have all led to health policies becoming increasingly dependent on empirical information which enables them to be evidence based. Whether it is a question of defining a public health policy or of assessing its results, information has become a major tool of this strategy. In a field as complex as health, where economic objectives of efficiency and profitability must obviously be reconciled with objectives of fairness and social justice in access to health care, the availability of information that makes it possible to describe the numerous aspects of the system, and to understand the overall dynamic as well as the underlying interrelations, is an increasingly decisive factor in the success of actions undertaken and in the continuity of the system itself. Political leaders at various levels (government, health workers, citizens) continually require information to enable them to make decisions, establish priorities, evaluate the costs and the benefits, as well as the effects and results of their actions. At the same time, a high demand for information comes from the public which increasingly demands to be able to check that requirements for rationalizing the system and managing costs do not go against individual and collective rights to health or to the fairness of the system.

As attentive observers have often pointed out (Zurkowski, 1974; Shenk, 1997; Stanley and Cripsham, 1997; Hill, 1999), in our modern societies largely based on statistical information, the anarchistic multiplication of all types of information, sometimes apparently or truly contradictory, can lead to a lack of effectiveness in information tools, at the very time when they have become essential; public opinion can therefore be diverted and the decision makers deprived of an essential tool. The extraordinary progress made over recent years in information technology has reinforced this risk by substantiating the false idea that intensive use of advanced technological tools necessarily improves information and makes its use by public authorities, institutions, health workers, and citizens more effective. The thaumaturgic capacity too frequently attributed to new information technologies, which provide access to all types of data, and which are supposed to make it possible to extract the best information at a lesser cost, is confirmed by the recent proliferation of survey results and administrative statistics made available to the general public. This is undoubtedly a case of "rich data, poor information" (Chen and Frolick, 2000).

In reality, in order to obtain real information from basic quantitative or qualitative data, it is necessary to go through a process of selection, which can only be carried out in the context of an initial conceptual framework, which alone can give meaning and informative power to the data available. The best database, even with the most modern and effective analysis, can never alone satisfy the need for interpretation hypotheses. It is the hypotheses themselves which, throughout the process from basic data collection to the development of useful information, determine the quality and the explanatory capacity of this information.

There is an additional crucial aspect to these problems when the user of statistical information is required to stimulate or monitor policies at the supranational level. Beyond the usual requirements of relevance, exactitude, and opportunity, statistical information should also satisfy the overriding necessity of comparability, despite social, cultural, or normative differences which separate countries and which frequently make it very difficult to make a simple comparison of phenomena which are apparently of the same nature.

Thus, in recent decades, new requirements for rationalization have brought new information needs to the forefront of the health policy scene. In the next seven sections we will briefly discuss the necessity of the following actions:

1. Creating information systems
2. Identifying the objectives
3. Identifying the stakeholders
4. Defining a conceptual framework
5. Building information bases
6. Taking into account a complex reality
7. Ensuring comparability

An eighth section will illustrate these subjects with some examples of information systems.

I. CREATING INFORMATION SYSTEMS FOR HEALTH POLICIES

To implement a public intervention within the context of a health policy, or, more generally, a social policy, requires at least four basic stages to be carried out:

1. The precise identification of the *objectives* to be reached

2. The development of a *conceptual framework* highlighting the aspects of the phenomenon that need to be explored or that should be focused on as they condition the achievement of the objective, as well as the relationships that may exist between these different aspects

3. The creation of an *information base* on these different aspects and the organization of its update from administrative sources and surveys

4. The definition and implementation of a *system of statistical indicators*, making it possible to monitor the evolution of the targeted phenomenon and its differentials, to regularly check that the direction followed is the one that will lead to the objective, and to evaluate the effect of the results obtained

The explicit or implicit reference to a conceptual framework is the key criterion to any element of statistical information, which otherwise would only be simple quantitative data of little interest beyond that of the process that generated them. Furthermore, in order to convert statistical information into indicators, it is necessary to be very sure of its adequacy to the targeted objective, in knowledge or in action, as well as the basic hypotheses on which the indicators were built or chosen within the wide panoply of existing statistical information. The traditional chain of knowledge (data-information-knowledge-wisdom) is thus enriched by an additional link, the indicator, where the selection is made, among the available information, of the elements necessary to identify the needs, to orient the choices in the different policy actions possible, to evaluate the costs and benefits, to evaluate the effects of the actions taken and their relevance with regard to the targeted objectives (Laaser, 1996). Recognizing this additional link should make it possible to break out of the vicious circle which seems to have seized information from modern societies and resulted in the availability of an enormous amount of data, notably statistical, and which runs the risk of obscuring rather than facilitating real knowledge of these phenomena.

But what is an indicator? There are numerous definitions of this concept, which are frequently inconsistent with one another in the literature. According to a classic World Health Organization (WHO) definition, indicators are "pieces of crucial, selected information that help to measure changes in relation to priority needs and allow for monitoring of specific aspects of health policy or of factors relevant to the determination of health care and related policies" (WHO 1981c). Other definitions place the emphasis on the function of indirect measurement of the indicators in relation to the studied phenomenon. In any case, the indicator never has a value in itself, apart from that which it could have with regard to the frame of reference that makes it possible to provide an interpretation. The majority of the definitions are consistent on this point.

The debate regarding the requirement in terms of social indicators and their use is certainly not new. Before the 1960s what is known as the "social indicator movement" (UNSO, 1975; Bauer, 1996) attracted the public's attention on the necessity of equipping society with measurement tools that would make it possible to take into account not only economic progress but also social and human development. This enthusiastic phase, which lasted up to the end of the 1970s, was followed by a period of disenchantment and loss of interest in social indicators, the idea being that too many expectations had been placed on the tools themselves, which could obviously never be a substitute for the political process of decision making or the definition of priorities by decision makers and managers (OECD, 1996).

Nevertheless, the indicators have continued to play a limited but important role in some social areas such as education and health. For example, the strategy of the "Health for All by 2000" program launched by WHO at the beginning of the 1980s, made substantial use of a set of indicators, aiming at the measurement of the process followed by the different countries in Europe to achieve the 38 targets laid down in the regional strategy for the continent (WHO ROE, 1993). It can be said that in the field of health, reference to sets of indicators has never been missing from national (van de Water and van de Herten, 1996) or international organizations, even though their greatest advantage has always been on the provision of service rather than on demand, and, even less, on the evaluation of quality of life or the level of user satisfaction. It is only very recently that these aspects have entered on an equal footing into the field of analysis and have led to the development of new indicators.

II. IDENTIFYING THE OBJECTIVES OF HEALTH POLICY

To build an information system, a first fundamental distinction must be made between the objectives of knowledge and decision making and the objectives of policy action and evaluation. These two types of objectives do not necessarily alternate, as the first type is frequently a premise to identifying specific actions from which the aims of the second type could be defined. In practice, they frequently result from distinct subsystems within the same information system. This distinction makes it possible to clarify the nature of the objective and to know how to build the indicators. The objective of the first type is general, such as the evaluation of a population or a group's good or bad health, the evaluation of differences in health in order to promote greater equity in the matter, the identification of groups or geographical zones suffering from specific problems, monitoring specific illnesses that have a high social or health impact, and so forth. In all these cases, the system of indicators is built a priori and can lead, or not, to identifying particular objectives for health policy initiatives, according to the use that is made of them.

The case of evaluating policy actions is different. Here, the system of indicators is closely linked to the targeted objectives and is generally built as an integral part of the policy activity whose effectiveness must be evaluated. Thus, for example, an initiative aiming to improve the effectiveness of a health structure must indicate in advance the evaluation tools that should be used for measuring the manner in which the program attains its objectives. It is a well-known fact that certain evaluation criteria can enter into conflict. More intense use of the structures, measured by a higher rate of use does not necessarily correspond to better effectiveness but can result in unduly prolonged stays in hospital.

Evaluating the overall performance of a health system, which has retained the attention of WHO and national governments in the past few years (Murray *et al.*, 2003), is a different objective again, which requires elements that are relevant to both the first and the second types of objective. Indeed, it requires, at the same time as a general and multidimensional vision essential to evidence-based decision making, care in the evaluation of a comparable nature to monitoring interventions. The debate on the sets of indicators to implement and regularly update is currently ongoing and several initiatives have been undertaken to create a consensual conceptual framework.

Regarding the characteristics of these types of tools, we must emphasize the problems caused by the frequent confusion between the objectives of strategic evaluation and the objectives of monitoring specific actions and how these can skew the evaluation. Indeed, particularly at a national level, the recent debate has focused on short-term objectives (moderation of costs, contribution of individuals and families to health expenditure, patient satisfaction, etc.), often losing sight of the main objective of any health system, which should be to improve the population's health while combating inequality (Murray and Evans, 2003). It is absolutely necessary to revert to a more correct use of the various available tools and to a more long-term view of evaluation. A set of indicators that only fulfill sectoral objectives, and are frequently limited to certain specific functions of the health system when they are not limited to one particular process, is only significant with regard to these objectives. It is not rare that a short-term success in achieving these objectives coexists with a mid-term and long-term failure in the general improvement of a population's health and in the fairness of the health system.

III. IDENTIFYING THE STAKEHOLDERS OF HEALTH POLICY

The objectives of health policies can also vary according to the stakeholders: from local, to national, to supranational, each player is situated at a different level of general knowledge, has different abilities, and needs different anchorage points for his or her activities and decisions. Thus, for example, the role of the European Union with regard to the health policies explicitly specified by the Maastricht Treaty is to formulate a common policy at the European level for general problems of major social importance (cancer, drugs, AIDS, health promotion and education, accidents, pathologies related to pollution). This obviously requires the availability of comparable statistical information in each member state. Regarding the objectives targeted by the national and local governments, these can vary considerably from one country to another according to the type of internal organization and the system of delegation that exists in the field of health policies and health management.

Moreover, stakeholders can distinguish themselves according to how the health system is defined, for example, according to what is understood in the strict sense, that is, all the structures and resources intended to intervene directly on health, or in the broader sense, that is, all the institutions and resources that carry out activities and make decisions which, directly or indirectly, affect public health (quality of the environment, food production, control of risk factors associated with behavior, safety in the workplace, etc.). In this latter case, which is certainly more effective for reaching the ultimate objective of improving the health of the population, management of responsibilities, as well as the information medium necessary to direct the various actions, can become extremely complex. This vision, despite its complexity, has the advantage of highlighting the articulated system of interrelations that exists between health and the other aspects of economic and social life, and favors improvement in the coherence of policy actions in the various fields.

Finally, identifying the objectives targeted through public health intervention pre-supposes that choices have been made on the basis of a list of priorities. The objectives should therefore be chosen on the basis of a cost-benefit analysis and should be measurable. From this point of view, a circular process is set up that encompasses, on the one hand, the requirements in statistical information to direct the definition of the policy goals and to establish priorities, and on the other hand, the need for objectives to direct the collection and analysis of statistical material. The solution to this contradiction can only be found through an iterative procedure of progressive reciprocal adjustments.

IV. DEFINING A CONCEPTUAL FRAMEWORK

Creating a conceptual framework from which the system of indicators will obtain its justification and its direction presupposes that it either refers to a health theory or, at least, to a system of relations between individual and collective health and different factors and phenomena that could influence it. Without an explicit explanatory framework, it is difficult for researchers and policymakers to reach a consensus. In our pluralist societies, researchers and policymakers should be involved in making policy decisions and in choosing objectives and therefore also in collecting information on which they can rely.

There is currently no general framework for describing and analyzing the health of individuals and the population, or the relations that link health to different living conditions. However, there is already a general consensus on the numerous elements that should be taken into consideration in such a plan. A brief description of the logic followed by the creation of a system of relations concerned with health objectives can help in understanding the choices made. Individuals, with their own biodemographic characteristics (sex, age, genetic inheritance, and biological heritage) are presumed to live their lives exposed to the influences of the different environments that surround them (Caselli *et al*., 1990): the *physical environment*, made up of a system of natural resources and external manmade conditions, the *vital environment*, made up of elements indispensable to life (water, air, food); the *social environment*, that is, the system of social relations to which individuals belong: family, network of relatives and friends, school, work, and the cultural system and dominant values in the society and in individuals' reference groups; the *medical environment*, from which individuals obtain their information on behavior that is beneficial or harmful to health and to which they turn when they have a health problem or they become aware of worrisome symptoms, to obtain a diagnosis on their health and the necessary treatment to reestablish their psychophysical balance or to restore their well-being. Numerous tools with which individuals are equipped during their lives (starting with their level of education) influence their ability to construct positive relations with these different environments by understanding the possible negative effects and by establishing more effective relations with them to satisfy their needs. Naturally, the relationship established between the different environments, and between these and an individual's medical history is not a one-way street. On the contrary, in the case of some nonformalized environments, such as the family or the network of relatives and friends, individuals and their history considerably condition the type and quality of the relationships that they establish with these external contexts and the sequences of action-reaction put in place.

Individual medical history is developed in close connection with an individual's life history of which it is part. This is marked by various important morbid events: the occurrence of an illness, the shock of a trauma, or an anomaly in the psychological, physiological, or anatomical structure or functions, according to the type of health process that is under analysis. In particular, in the case of illness, such facts may not be immediately noticed by the individual, thereby often considerably delaying diagnosis and the onset of treatment. The diagnosis of the problem or the transition of a latent condition, structural or functional, physical or mental (a deficiency), the intermediate result (a disability), and the final result of the process, positive (cure, rehabilitation), or negative (chronic illness, permanent disability, death) are all affected. Individuals live and interpret their life history, including their medical history, through the prism of their own perception (Bowling, 1991). This prism substantially conditions their ability to interact with their different environments and has important consequences on the development and the results of their medical history.

From a demographic point of view, collective phenomena and processes correspond to each of these individual events and phases. These phenomena and processes are the targets of a health policy's interventions. In this plan, the traditional morbidity indicators such as incidence, prevalence, lethality, mortality, and the duration of the illness, or the health indicators, such as the different measurements of disability, are all at the same level as the indicators based on the perception that individuals have of their own health.

The various contexts considered as favorable to health can also result in specific sub-systems of indicators. Different typologies of indicators are used in association with the characteristics of the sector and the type of presumed relation to health: morbidity indicators, service indicators, resource indicators, indicators of the demand for health care, and so on.

Other subsystems are then developed to complete the framework necessary for managing the system, especially at the macro level (i.e., funding and expenditure).

V. THE INFORMATION BASE: A COMPROMISE BETWEEN CONTENT AND TECHNOLOGY

As we mentioned earlier, a number of illusions have been generated by the extraordinary technological progress of the last few years. It is almost as if we hoped that the speed and effectiveness of calculation of computers could make up for theoretical weakness and lack of relevance in the interpretation hypotheses.

Today, while it is no doubt essential to maintain elementary information bases that are useable for developing solid and reliable measurements of the phenomena to be studied, it is also true that direct use of elementary statistics (micro-data) is only required for a small number of specific objectives, such as those for research into the relation between the various behaviors of the same individual, and the effect of some of these characteristics and behaviors on his or her health.

Information systems based on indicators are all the more useful, as a support for evaluating decisions and policies, when they are created in such a way that they can be continually updated and gradually adapted to the knowledge requirements that may become apparent over time. Their usefulness is even greater if they have been created in such a way that they can answer questions at different levels, in other words, when they make it possible to reconcile the information needs of decision makers who plan health interventions, each at his or her level of responsibility (local, national, supra-national) in the context of compatible objectives.

VI. CREATING SYSTEMS OF INDICATORS REPRESENTATIVE OF A COMPLEX REALITY

No unidimensional indicator can alone represent the diversity of phenomena or the multiplicity of the dimensions that should be taken into consideration in order to evaluate a population's health or to implement policies aimed at all of the health problems in the complex world in which we live today. It is therefore necessary to develop systems of indicators which, as a whole, can take into account the different dimensions that are useful in the description and analysis of the health of populations, and which are thus capable of providing an integrated view of the different contexts that, directly or indirectly, influence the health of populations.

The definition of the objectives provides a guide for selecting the indicators by making it possible to monitor the variables whose evolution must be evaluated or which are targeted by the health policies (e.g., the incidence of AIDS, the mortality rate by cancer, changes in lifestyles that are dangerous to health, etc.). The available statistics or those which can be obtained through surveys rarely satisfactorily explain the real determinants of a population's health or the risky etiological factors, even when a precise definition of the risk has been provided. What is generally available is information on variables that, at best, can be considered as proxies for the real determinants or risk factors. Most frequently the information available only enables the population to be listed according to the various characteristics linked to health, without the possibility of interpreting these relations as cause-effect relations. We can therefore only identify groups of individuals that form potential targets for various specific interventions.

A set of indicators developed to take into account different contexts that influence health or different aspects of its condition does not necessarily create a system. Collecting and analyzing basic information, even guided by an explicit conceptual framework, needs to be improved through the evaluation of the characteristics of each element of the system (the indicators) and of their interrelations.

The relevance, credibility, exactitude, topicality, and sensitivity of each indicator must be evaluated. Particular attention must be paid to the purity of the indicator, in other words, to its capacity to quantify a phenomenon in the absence of spurious effects of other phenomena (for demand indicators, the population's age and sex structure, for example). Furthermore, the indicators should be comparable, especially when the objectives pre-suppose the need to compare different situations within the same country or between different countries. From this point of view, it is necessary to be aware that researching comparability can reduce the quality of the indicators in terms of relevance and reliability. In some cases, taking the quest for comparability too far can result in empiricism such that the indicators not only lose all interest for the public and for decision makers, but can even cause errors. For international comparison, for example, indicators of results are preferable to data on implementation methods insofar as health care systems can vary considerably from one country to another, and it is therefore practically impossible to obtain measurements of comparable methods. Without knowledge of specific legislation regulating the provision of health care and its funding in each country, the international comparison of method indicators can lead to an erroneous vision.

Once the elements that can be used to create a system of indicators have been identified, the selection then needs to be carried out followed by the clarification of the linkages between the various elements selected. The system should not be redundant, in the sense that it should not only contain relevant indicators or only the most important, but should also contain elements concerning the availability of medical resources and elements relating to needs. Furthermore, the system should be flexible enough to guarantee its usefulness for different types of users, the possibility of reproducing it, and that of regularly updating it.

In fact, when a system of indicators is set up, it is important to consider the different categories of potential users. As stated previously, from the implementation phase the different levels of decision making in force in the area of health policies must be considered. The system of indicators should provide these different levels of decision making with elements of direction, decision, allocation of resources, and the monitoring and evaluation of the actions undertaken. However, it should not be forgotten that important users of the system are the public in their dual role of decision maker for their own families' health strategies, and as censors of the government's activities. Certain organizations of civil society devote their activities to the evaluation and control of specific aspects of the relation between public demand and the service provision of health structures. These organizations are interested in monitoring the results of health actions and are therefore potential users of the information systems; their specific needs also should be taken into consideration.

Another specific type of users are health workers who, in their daily routine, need elements of information that enable them to evaluate their work and apply it to other realities. These requirements assume a very specific level of detail, which could be difficult to ensure with an information system directed toward strategic decisions. In the context of this purpose, particularly in the clinical field, systems for communicating and sharing information on medical practices have been developed, which have contributed significantly to the spread of improved approaches and reduced the considerable differences that existed previously (OECD, 2003a). Evidence-based medical practice set up on this basis and with these tools has affirmed itself over recent decades and has opened the way to considerable progress in the treatment of the most important pathologies. The problem now is to create connections between these specific information systems and the higher level information systems, which due to satellite systems, for example, make it possible to supply the higher level with composite indicators without losing the specificity required by the daily activity.

It is not necessary to expand further on the various stages for a successful implementation of a perfect system of indicators that includes all these characteristics and properties to understand that such a system is more an aspiration toward an ideal situation than feasible instructions. It is improbable that real conditions will make it possible to conform to this theoretical framework in the near future. In practice, numerous compromises are necessary.

VII. ENSURING COMPARABILITY THROUGH HARMONIZATION

In Europe, comparability became a crucial requirement during the 1980s and 1990s, following the publication of WHO's Regional Office for Europe of the *Targets for Health for All by 2000* (WHO, 1985), which established a vast panoply of indicators for measuring progress toward the objective of "Health for All by 2000." Indeed, the strategy of "Health for All by 2000," launched by the World Assembly of WHO in 1977, made the countries increasingly aware of the need to have adequate and comparable information for the formulation, monitoring, and evaluation of policies. Indeed, this strategy stresses the need for quantitative objectives, which are translatable in the form of statistical indicators, which facilitate the monitoring, evaluation, and transparency of health policy decisions (WHO, 1981b). WHO Regional Office for Europe, in adapting the world strategy to the European context, has adopted a set of objectives (38 in total), which include the need to implement a health information system based on a core of internationally comparable indicators in each country (WHO ROE, 1980, 1993). Consequently, the need to develop common methods and tools to carry out comparative analyses of the population's health has increased. The main problem was that the majority of the indicators selected posed enormous problems of comparability between countries. Despite the progress made regarding the definitions to adopt and the measuring tools to employ, these issues are far from being completely resolved.

Fifteen years have gone by since the first *Consultation to develop common methods and instruments for health interview surveys*, held under the aegis of the WHO Regional Office for Europe, and the Central Bureau of Statistics of the Netherlands in 1989. Two other consultations followed in 1990 and in 1992, which were also devoted to household health surveys. Both meetings made considerable progress toward improved comparability between countries, especially in the field of measuring disability. However, much remains to be done before we can conclude that enough headway has been made in harmonizing health indicators (a process which, in fact, has only just begun), or even sufficiently explored the perspectives. Different measurements based on household health surveys must be analyzed, as must other sources (registers, hospital statistics, etc.). When the problems that impede the comparability of data from these sources are resolved, they will become information sources of great value.

The first cycle of consultations started a process which has since then resulted in numerous confrontations. As an example, let us quote two significant cases. The first is that of the Eurohis project (*Common methods and instruments for health interview surveys in Europe*) directed by WHO Regional Office for Europe.

The overall objectives of the project were as follows:

- The development of Health Interview Survey (HIS) tools
- Common methods and tools (a guide) for use in household health surveys, as an integral part of national health information systems
- The promotion of their use in national surveys on household health
- The adaptation of national data for international comparison
- The development of analytical methods to improve the comparability of health survey data already available
- The adaptation of national HIS for international comparison, to demonstrate, as far as possible, their potential with regard to international comparisons

The project aimed to establish indicators in the following fields:

Chronic physical illnesses
Mental illnesses
Alcoholism
Physical exercise
Medical consumption: recourse to health services
Medical consumption: recourse to medication
Preventative action: vaccination, check-ups, etc.
Quality of health

The Eurohis project came to an end in 2002 and resulted in a set of recommendations, which appeared in a WHO publication (Nosikov and Gudex, 2003).

The second example concerns a project for harmonizing data pertaining to disability. It was generated by the Washington City Group on Disability Statistics whose main objective is "to guide the development of a small set(s) of general disability measures, suitable for use in censuses, sample-based national surveys, or other statistical formats, which will provide basic necessary information on disability throughout the world" (http://unstats.un.org/unsd/methods/citygroup/washington.htm). The project is still in the development phase and should conclude in 2005.

The aim is now to follow this path and widen the field to include other tools and indicators, to analyze their characteristics and effectiveness under different conditions, as well as to eliminate the factors that prevent or impede access to a satisfactory level of international comparison, beginning with numerous differences in the quality of statistical data on methods. The identification and eradication (or at least control) of these problems would improve the comparability of the indicators without sacrificing the specificity of the information. This is obviously a difficult task (as can be seen in the scale of efforts made that have not always met with success), but it is a task which must be carried out if we wish to build a system that allows for international comparison of the population's health, as well as a basis for formulating recommendations for the health and social policies of each country and the monitoring of the actions undertaken by each country to achieve "Health for All."

VIII. SOME EXAMPLES OF INFORMATION SYSTEMS

At the international level there have been numerous attempts to build systems of indicators, for the purposes of description and perspective, as well as for policy purposes. Among the most visible is the OECD (Organization for Economic Cooperation and Development) Health Data system, the most recent edition (annual) of which dates back to 2003, and WHO's "Health for All for the 21st Century" system of indicators.

OECD's system is one of the historical systems of reference in the area of health. It is a typical cognitive system, that is, a very large collection of indicators, updated annually, which relates to multiple fields: from the health of the population to the operation of health systems and to the environmental and social frame of reference (Table 109–1). The data are presented at a national level for the various countries and in historical series (since 1960). Since 1991, the system has been available in electronic form as a CD-Rom containing the whole consultable database, supplemented by meta-data.

As is recalled in the brief history on WHO's website,

> Health for All was adopted in 1977 and launched at the Alma Ata Conference in 1978 to underline the fact that despite the ambitious proclamations enshrined in the WHO Constitution,[1] large numbers of people and even whole countries, were not enjoying an acceptable standard of health. By the late 1970s nearly 1 billion people were living in poverty. The HFA renewal process was launched in 1995 in response to accelerated global change and to ensure that individuals, countries and organizations are prepared to meet the challenges of the 21st Century.
> (http://www.who.int/archives/hfa/history.htm)

This strategy was accompanied very early by the implementation of a system of indicators intended to ensure monitoring, in order to periodically verify its progress and to evaluate the national initiatives of adaptation to the strategy. In fact, while on the one hand, this strategy is aimed at all WHO's member states in the same way, its adaptation to national realities is left to each country in the context of each country's own health policies. The year 2020 is the target date for completion. It can be viewed online at http://hfadb.who.dk/hfa/.

The available indicators relate to many aspects of the health of populations and its determinants. Definitions are provided for each indicator, as well as a classification. The principal fields selected are as follows:

Demography and socioeconomic situation
Mortality
Morbidity, disability, and hospital discharges
Lifestyles
Environment
Health care resources
Health care utilization and expenditure
Maternal and child health

FINAL REMARKS AND CONCLUSION

As we emphasized at the beginning, in recent years a number of the world's large countries have been faced with problems caused by increasing constraints

[1] In its preamble, WHO's Constitution proclaims, inter alia, "The enjoyment of the highest attainable standard of health is one of the fundamental human rights of every human being without distinction for race, religion, political belief, economic or social condition." The Constitution also stipulates that "the health of all peoples is fundamental in the attainment of peace and security and is dependent upon the fullest cooperation of individuals and states."

TABLE 109–1 List of the Variables Included in the OECD's 2003 Health Database

Part 1. Health status
Mortality
- *Life expectancy*
- *Causes of mortality*
- *Maternal and infant mortality*
- *Potential years of life lost*

Morbidity
- *Perceived health status*
- *Healthy life expectancy/Disability-free life expectancy*
- *Infant health*
- *Congenital anomalies*
- *Dental health*
- *Communicable diseases (HIV/AIDS)*
- *Cancer*
- *Injuries*
- *Absence from work due to illness*

Part 2. Health care resources
Health employment
In-patient beds
Employment-to-beds ratio
Medical technology
Education in health and welfare

Part 3. Health care utilization
Prevention (immunization)
Consultations
In-patient utilization
Average length of stay
- *Average length of stay for in-patients and acute care*
- *Average length of stay by diagnostic categories*

Discharges
- *Discharge rates by diagnostic categories*

Surgical procedures
- *Total surgical procedures*
- *Surgical procedures by ICD-CM*
- *Transplantation and dialyses*

Part 4. Expenditure on health
National expenditure on health
- *Total expenditure on health*
- *Expenditure on personal health care*
- *Expenditure on collective health care*
- *Prevention and public health*
- *Expenditure on health administration and insurance*
- *Expenditure on health-related functions*
- *Expenditure on medical services*

Total expenditure on medical services by functions
- *Expenditure on in-patient care*
- *Expenditure on out-patient care*
- *Expenditure on home care*
- *Expenditure on ancillary services*

Medical goods, dispensed to out-patients
- *Total expenditure on medical goods*
- *Pharmaceutical and other medical non-durables*
- *Therapeutic appliances and other medical durables*

Part 5. Financing and remuneration
Health expenditure by source of funds

Part 6. Social protection
Social expenditure
Health care coverage

Part 7. Pharmaceutical market
Pharmaceutical industry activity
Pharmaceutical consumption
Pharmaceutical sales

Part 8. Nonmedical determinants of health
Life styles and behavior
- *Food consumption*
- *Alcohol consumption*
- *Tobacco consumption*
- *Body weight and composition*

Environment: air quality

Part 9. Demographic references
General demographic
Population age structure
Labor force
Education and training

Part 10. Economic references
Macroeconomic references
Monetary conversion rates

Source: Organization for Economic Cooperation and Development (OECD).

on health resources while an increase in demand has been reinforced by technological progress and population aging. The case of the United States, the United Kingdom, Germany, and the Netherlands are emblematic. Even though they involve national systems that are completely different from one another (from the private model of the United States to the universal public system of the United Kingdom, from the social insurance model of Germany to the combined system of the Netherlands), since the end of the 1980s, they have all experienced a period of crisis and reorganization (Docteur and Oxley, 2003). The curbing of costs has, as noted by some, become the "war cry" of decision makers, of managers, and, sometimes even, of practitioners (Patrick and Guttmacher, 1983).

In this general context of the desire to rationalize health systems, the role played by indicators has become a fundamental one, as we have already mentioned. Just as fundamental is creating a link between the system of indicators and the decision making system. However, too frequently when discussing indicators, we run the risk of becoming immersed in the swirl of an operationalization logic, by imagining that indicators should only be used for specific objectives. Vice versa, the current situation obviously requires systems of overall cognitive indicators as well as systems of indicators directed toward specific problems.

As is the case for the system that Eurostat is currently setting up in cooperation with all the countries of Europe, and as is the case for national systems (Italian, German, British, and Dutch, e.g.), the first type of system should supply a framework of information constituting the base from which institutions and governments should steer, to decide on which phenomena to take into consideration, the actions to carry out, and the methods of health surveillance (Egidi, 2003). At certain stages of activity and programming, compartmentalized reasoning is impossible; an overall vision is absolutely necessary in order to identify the effects, desired or undesirable, of the interrelations between health and the health system and other economic and social sectors. The availability of systems of this type should make it possible to combat the tendency to promote actions in the absence of a wide view and perspective. Furthermore, these systems are also systems of democratic information accessible to all, that enable the public as well as experts to become aware of the stakes of a situation, which concerns the general public as much as it does public and private decision makers.

The objective of the second type of system is to provide information on which operators and decision makers can base their decisions or commitments. A small number of general indicators are no longer satisfactory. Information at the level of specific needs is required. The increasing complexity of problems consequently leads to an equivalent complexity in the required information systems, which must respond to personalized choices through their diversification.

In short, health information systems should become everyday tools in both the programming and evaluation process as well as in access to understanding public health issues. However, the measures taken should be integrated in a conceptual framework that reflects the interpretative plan of the phenomena and processes concerned. A model that should be based on the debate involves not only the decision makers but also researchers and experts because information alone is never neutral and should be explicitly positioned in an explanatory framework.

Let us emphasize two aspects. First, the need to persevere in this sometimes very difficult path toward the harmonization of indicators and systems of indicators for analyzing the health of the population and the effectiveness of health systems, as well as for monitoring health policies and actions recommended at the international level. In these fields, supranational organizations play a crucial role; they should promote systems that are coherent with one other. Unfortunately, there is no shortcut to this goal. The illusion of the effectiveness of standardized tools that are incapable of taking into account the specific health systems and the social organization of each country, requires that problems of interpreting differences in level or evolution be overcome, and that we become aware of the continuing impossibility of carrying out direct comparisons between countries, even with data that have apparently been collected on the basis of identical definitions and methods (Egidi and Buratta, 1997).

Secondly, the differences and the evolution of the mechanisms that are the basis of demographic phenomena must be evaluated. It is perhaps superfluous to draw attention to the fact that, over time, natural changes occur in the population and the groups comprising it, caused by the aging of individuals and the passing of successive cohorts. Too frequently a cross-sectional approach to analyses and to monitoring progress toward the chosen goals does not permit the correct evaluation of real developments. Thus, for example, the variation in the prevalence of a given disease at a given age between two successive periods can occur without any specific public health initiative. This could be a simple consequence of certain individuals being progressively substituted for others whose life histories cause a change in the incidence of the same disease. The diffusion of attitudes with regard to prevention can be favored by the arrival of better educated cohorts of individuals. The reduction of disability among elderly people can be a result of the aging of individuals who have been better protected during their working lives against risks specific to certain particularly dangerous jobs. It is rare to have the information necessary for the adequate development of a longitudinal approach which makes it possible to evaluate the relations between the life histories of different cohorts. Nor is it common to have access to information that makes a longitudinal evaluation of consequences in terms of health possible, or the calculation of indicators capable of measuring the effect. More often we must be satisfied with cross-sectional or period data and it is necessary to rely on the good judgment of users to at least indirectly take into consideration the effects of such mechanisms.

Acknowledgment

This chapter was translated from the French by Accenta Ltd.

References

Bauer Raymond Augustine, 1966. *Social indicators.* Cambridge (Mass.), MIT Press, 357 p.

Bowling Ann, 1991. *Measuring health: a review of quality of life measurement scales.* Buckingham, Open University Press, 199 p.

Caselli Graziella, Duchêne Josianne, Egidi Viviana, Santini Antonio, and Wunsch Guillaume, 1990. *A matter of life and death: methodologies for the life history analysis of adult mortality.* Louvain-la-Neuve, Université catholique de Louvain, Institut de démographie and Académia, 39 p. (Working Paper no. 151).

Chen Lei-da and Frolick Mark N., 2000. Web-based data warehousing, *Information Systems Management*, vol.17(2), p. 80–87.

Docteur Elizabeth and Oxley Howard, 2003. *Health care systems: lessons from the reform experience.* Paris, OECD, 85 p. (Health Working Paper).

Egidi Viviana, 2003. *Promoting a better use of statistical information: an integrated information system on disability.* New York, United Nations, 10 p. (UN Expert Group Meeting on Setting the Scope of Social Statistics, New York, 6–9 May, Doc. no. 30).

Egidi Viviana and Buratta Vittoria, 1997. Comparability of definitions and data on morbidity, *in:* Guillaume Wunsch and Attila Hancioglu (eds.), *Morbidity and mortality data: problems of comparability. Proceedings of the European Association for Population Studies and the Hacettepe Institute of Population Studies workshop, Ürgüp (Turkey), 18–20 October 1995,* p. 11–29. Ankara (Turkey), Hacettepe Institute of Population Studies, 236 p.

Hill Michael W., 1999. *The impact of information on society: an examination of its nature, value and usage.* London, Bowker-Saur, 292 p.

Laaser Ulrich, 1996. *Perspectives for a European public health information system.* Düsseldorf (Germany), NRW, 32 p.

Moise Pierre and Jacobzone Stéphane, 2003. Population ageing, health expenditure and treatment: an ARD perspective, *in:* OECD (ed.), *A disease-based comparison of health systems. What is best and at what cost?*, p. 163–179. Paris, OECD, 363 p.

Murray Christopher J. L. and Evans David B. (eds.), 2003. *Health systems performance assessment: debates, methods and empiricism.* Geneva, WHO, 917 p.

Murray Christopher J. L. and Evans David B., 2003. Health systems performance assessment: goals, framework and overview, *in:* Christopher J. L. Murray and David B. Evans (eds.), *Health systems performance assessment: debates, methods and empiricism,* p. 3–23. Geneva, WHO, 917 p.

Nosikov Anatoliy and Gudex Claire (eds.), 2003. *EUROHIS: developing common instruments for health surveys*. Amsterdam, IOS Press, 208 p. (Biomedical and Health Research, no. 57).

OECD (Organization for Economic Cooperation and Development), 1996. *The German health information system. A possible means to improve the international comparison of health care data?* Paris, 18 p. (Document for the Working Party on Social Policy.)

OECD, 2003a. *A disease-based comparison of health systems. What is best and at what cost?* Paris, OECD, 363 p.

OECD, 2003b. *OECD health data.* Paris, OECD, CD-Rom.

Patrick Donald L. and Guttmacher Sally, 1983. Socio-political issues in the use of health indicators, *in:* Anthony J. Culyer (ed.), *Health indicators: an international study for the European Science Foundation*, p. 165–173. Oxford, Martin Robertson, 223 p.

RIVM, 1995. *Public health status and forecasts.* The Hague, Research Institute for Men and Environment, 44 p.

Ruwaard Dirk, Kramers Pieter G. N., van den Berg Jeths Anneke, and Achterberg Pieter W. (eds.), 1994. *Public health status and forecasts. The health status of the Dutch population over the period 1950–2010.* Bilthoven, Rijksinstituut voor Volksgezondheid en Milieuhygiëne, 239 p.

Shenk David, 1997. Data smog: surviving the information glut, *Technology Review*, vol.1(4) May/June, p. 18–26.

Smith Alwyn, Jacobson Bobbie (eds.), 1990. *The nation's health. A strategy for the 1990s. A report from an independent multidisciplinary committee, chaired by Professor Alwyn Smith.* London, King Edward's Hospital Fund for London, 331 p.

Stanley Andrew J. and Clipsham Philip S., 1997. Information overload: myth or reality?, *in:* IEE (ed.), *Colloquium on IT strategies for information overload*, p. 1–4. Montreal, Canada, IEE, 520 p. (Digest no. 1997/340).

UNSO, 1975. *Toward a system of social and demographic statistics.* New York, United Nations, Statistical Office, 45 p. (Studies in Methods, Series F, no. 18).

van de Water Harry P. A. and van de Herten Lois M., 1996. *Bull's eye or Achilles' heel: WHO's European health for all targets evaluates in the Netherlands.* The Hague, TNO Prevention and Health, 229 p.

WHO, 1981a. *Development of indicators for monitoring progress towards health for all by the year 2000.* Geneva, World Health Organisation, 91 p. (Health for All Series, no. 4).

WHO, 1981b. *Global strategy for health for all by the year 2000.* Geneva, World Health Organisation, 90 p. (Health for All Series, no. 3).

WHO, 1981c. *The use of health indicators. How to make statistics talk.* Copenhagen, World Health Organisation, 68 p.

WHO-Regional Office for Europe 1985. *Target for health for all*, Copenhagen, European Health for All Series, No. 1, 35 p.

WHO, 1993. *Implementation of global strategy for health for all by the year 2000.* Geneva, WHO, 178 p. (8th Report on the World Health Situation).

WHO, 1994. *Concern for Europe's tomorrow.* Copenhagen, WHO, 88 p. (European Series, no. 53).

WHO-ECEH (European Centre for Environment and Health), 1995. *Concern for Europe's tomorrow. Health and the environment in the World Health Organization European Region.* Stuttgart, Wissenschaftliche Verslagsgesellschaft, 573 p.

WHO ROE, 1980. *Regional strategy for attaining health for all by the year 2000.* Copenhagen, WHO Regional Office for Europe, 30 p. (Document EUR/RC308, Rev. 2).

WHO ROE, 1989. *Consultation to develop common methods and instruments for health interview surveys.* Voorburg, 21–23 June 1988. Voorburg (Netherlands), WHO Regional Office for Europe, 85 p. (Document EUR/RC308, Rev. 2).

WHO ROE, 1993. *Health for all targets. The health policy for Europe.* Copenhagen, WHO Regional Office for Europe, 42 p. (European Health for All Series, no. 4).

Zurkowski Paul G., 1974. *The information service environment. relationship and priorities.* Washington, DC, U.S. National Commission on libraries and Information Science, Government Printing Office, 5–6 p.

P A R T

IV

MIGRATION POLICIES AND MANAGEMENT OF THE POPULATION PYRAMID

GRAZIELLA CASELLI, JACQUES VALLIN AND GUILLAUME WUNSCH

Along with fertility and health, two other major spheres of population policy action relate to the third big factor of population dynamics—external migration—and changes in population structure. Strictly, this latter sphere is another way of coming at the first three, as whatever influences fertility, mortality or migration may feed into the population structure. Combining migration policies and those for the management of the population pyramid in the same part when a whole part has been given to each of the first two factors of dynamics—fertility and health—therefore, smacks of the arbitrary. We take full responsibility for that. It is partly due to the fact that this Treatise has from the outset focused consistently on the natural forces of population development, because we are mainly concerned with large-scale populations (countries, groups of countries, world), where fertility and mortality are much bigger shaping factors than migration. A single chapter on migration policies—chiefly international migration—was thought to suffice here, therefore, given that internal migration has little effect on total population dynamics, and chiefly affects composition. As to the latter, it was also decided not to attempt exhaustive coverage. Many aspects of population structure are of government concern and may be the subject of policies. Our chosen aspect—age structure—is far from the only one. Many countries have strict policies on geographical distribution, where internal migration is a powerful factor, whether in respect of urbanization or broader area planning and development. However, a separate chapter on this was not thought to be required; instead the authors of the country case studies were asked to cover it if necessary, as the analysis of national policies is where an issue so tied to the geographical context of each country can best be understood. Other key aspects of a population structure may also be policy concerns: the social structure, for example, is highly susceptible over the long-term to economic and social policy options;

the sex structure, a growing focus of concern to the governments of some countries where large-scale sex preselection is beginning to be practiced; etc. But limits must be drawn, and so again, the choice has been to focus on the aspect of population structure that has most influence on its development: age structure, which is entwined with cohort replacement.

This fourth part is therefore divided into two chapters: Chapter 110, by Stefano Baldi and Raimondo Cagiano, on migration policies, and Chapter 111, by Michel Loriaux, on the economic and social management of the population pyramid.

C H A P T E R

110

Migration Policies

STEFANO BALDI AND RAIMONDO CAGIANO DE AZAVEDO

Facoltà d'Economia, Università degli Studi di Roma "La Sapienza," Rome, Italy

Broadly speaking, most political decisions have an impact on migrants and can therefore be considered as including, whether in an explicit way or not, some migration policy elements. For example, everything that directly or indirectly affects the social sector necessarily influences migration behaviors. However, for reasons of simplicity, the following discussion will only deal, except occasionally, with those policies that are explicitly related to migration; it will also be about labor migration only, without taking into account such movements as those of refugees, displaced persons or asylum seekers (concerning those points, see Chapter 62).

The major reason why workers emigrate to a foreign country is that they are seeking for better living standards, and even for better chances of survival sometimes. The present economic situation in the world, particularly in the poorer countries, is likely to intensify international migration. The growing demographic imbalance between the different countries or regions in the world, particularly in those places where it is combined with wide disparities in terms of wealth distribution, is bound to increase the pressure for migration.

As a matter of fact, international labor flows are a source of immense potentialities, both for the countries of departure and those of reception; however, they have also necessarily become a major political problem.

According to the Program of Action of the International Conference on Population and Development (ICPD) organized by the United Nations in Cairo in 1994 (United Nations, 1995b, item 10.15), "it is the right of every nation State to decide who can enter and stay in its territory and under what condition. Such right, however, should be exercised taking care to avoid racist or xenophobic actions and policies." In fact, governments and bureaucracies have paid little heed to the problems caused by immigration, which has often resulted in negative attitudes developing among local populations toward immigrants, making it easier for xenophobic rhetoric to emerge, as well as opinion campaigns aimed at limiting the admission of foreign workers.

The analysis of migration policies can be divided into four points:

1. The migration policies of the receiving countries
2. The migration policies of the countries of emigration
3. International organizations and migration policy agreements
4. New elements in migration policies

I. THE POLICIES OF THE RECEIVING COUNTRIES

Although any classification of social phenomena is necessarily arbitrary and questionable, still the various areas for migration policies intervention must be distinguished so as to make the discussion as clear as possible. Only, it must be emphasized that such distinctions should not be understood as drawing impassable borders. There are necessarily blurred areas and discrepancies between the different spheres, whatever classifying system is adopted.

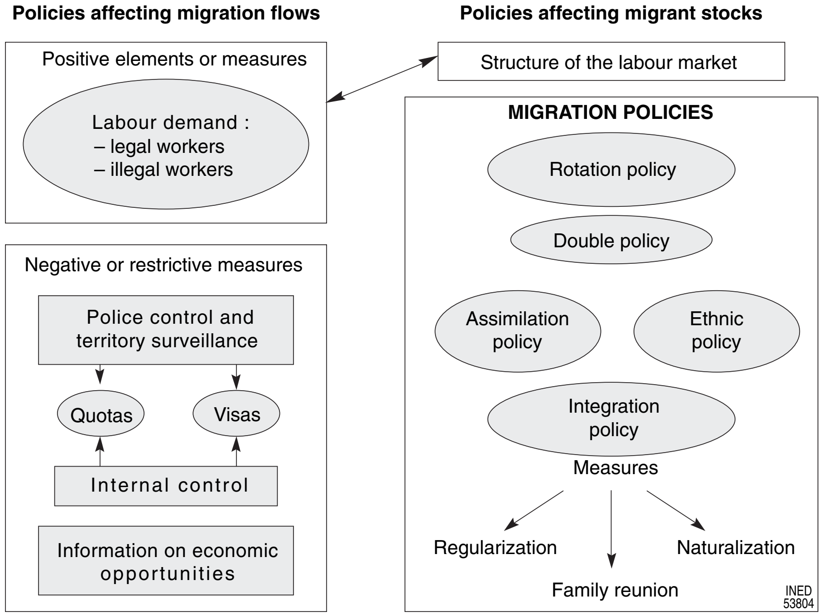


FIGURE 110–1 A schematic comparison between the logistics and methods of flow policies and stock policies as determined by labor demand in the countries of reception.

The simplest, and probably most efficient, approach consists in distinguishing between policies that aim at affecting flows and those that influence stocks. We will proceed along such distinction, first concerning the policies of the countries of reception, and again in Section 2 of this chapter, concerning the countries of departure.

Policies concerning flows and those affecting stocks are obviously closely interdependent, but the methods used by those two types of policies are different, and they each have their own characteristics, as summarized in Figure 110–1.

1. Policies Affecting Immigration Flow in Receiving Countries

Those policies that aim at influencing migration flow may clearly work several ways. Measures can be adopted to increase (positive measures), reduce (negative measures), or keep stable the net number of admissions.

The most significant positive element that affects migration flows is demand for employment. The volume, composition, and schedule of migration flows are, more often than expected, determined by employment demand rather than by supply.

Moreover, the labor market is segmented in such a way that in certain developed countries, substantial unemployment (among the natives) coexists with quite sizeable immigration. This is because immigrants usually accept work in sectors that imply risks or even danger (e.g., in mining or industry), or in illegal activities (such as prostitution, drug trafficking), particularly where the natives are not inclined to do so.

On the other hand, the richer countries, where the populations are aging, may also need to use immigration to compensate for the lack of native labor in some particular sectors, including services (domestic services, catering), the building industry, and public works.

Several reasons account for immigrant workers accepting low-skilled, or unskilled, underpaid jobs from the start. Firstly, most immigrants from the South come from rural areas, have low educational levels or are even illiterate, and have no experience working in industry. They also have language challenges. But above all, because of their status—due to precarious work permits, and sometimes to their being illegal immigrants—they are in a weak position when it comes to negotiating salaries, all the more so because to them survival is a matter of emergency.

As a matter of fact, many immigrant workers are in breach of immigration regulations. Such a situation has been termed *illegal* or *clandestine immigration*, and the migrants themselves *illegal* or *undocumented immigrants*. As the presence of an illegal worker implies the existence of an illegal employer, the size of the phenomenon is closely related to tolerance—or need—of illegal labor.

Migration flows may have substantial consequences over the structure of the population, in the country of reception as well as that of departure. Employment-related migration usually concerns groups of male young people of working ages (25–45 years old). Migration policies that aim at family entry and settlement may correct the imbalance characteristic of populations of migrant workers. Differences are often to be observed in the fertility—and sometimes the mortality—between the migrants and the natives. Fertility rates are usually higher among the former than the latter. The situation is not quite so clear as far as mortality is concerned, because of the effects of selection as the migrants enter a country or leave another, but the fact is that immigrants are usually provided with health facilities that are not quite as good as those provided to the natives. However, those differences, which have a direct effect over growth rates, tend to fade as time goes by and the migrants integrate or assimilate.

Volume II, Section II in this book is on the determinants of migration; those clearly have direct and indirect implications for migration policies, which vary in time as well as from one country to another. Those policies may be either positive, in favor of immigration, or negative, aimed at restricting it or even at encouraging emigration.

Positive measures aimed at encouraging immigration may result from different objectives and make more or less intensive use of various instruments with a view to reaching the implicit or explicit targets of the government in the country of reception.

It is usually a question of meeting internal needs, mostly economic ones. It may be a matter of general need for reconstruction or economic growth, such as experienced by many European countries after each of the two world wars or during the Thirty Glorious Years (1945–1975), but there may also be more specific needs in certain sectors of the economic system that suffer from lack of one type of labor or another. For example, immigration was encouraged in the last century in Europe for the specific needs of the mining sector and the steel industry. In more recent times, the sector of personal services (domestic, medical, and assistance services) has been making much use of immigrants from Central and Eastern Europe, and even from Asia. Apart from Europe, the Gulf countries, and before them North America, Australia, and New Zealand have always used immigration to serve their economic development.

In the latter cases, what was at stake in those policies was even, in a more general way, to get colonists to settle on uninhabited lands, as is still the case nowadays as Israel actively implements its policy of drawing the Jewish diaspora together. Likewise it is mainly for political and cultural reasons that West Germany has always claimed willingness to take in any person of German descent who would like to come and settle there; the ultimate point of such a policy was reunification.

Specific migration flows were also encouraged, or even planned, with a view to settling conflicts (e.g., in the aftermath of colonial relations) or to organizing cooperation for development. There are many recent agreements in which extraordinary quotas for agreed immigration were set between the countries of destination and those of departure, just as such cooperation agreements include all kinds of interventions for development—loans at favorable rates, investment flows, letting the partners off their debts, and other supporting measures for underdeveloped economies. Such agreements have been concluded between Germany and Turkey, Italy and Albania, Tunisia and Morocco, and others.

Besides such economic and political objectives, there are also humanitarian ones contained in the measures concerning the admission of refugees, whether as part of international agreements on the right of asylum or of specific policies such as those on territorial asylum in France, which makes it possible for persons who are threatened by war or persecution to get asylum in a receiving country. In the last few decades, the restrictive policies enforced in the richer countries with a view to slowing immigration down have badly impaired the existing system (particularly since the Geneva agreements). However, there are a few remnants of it, which could be reactivated. An example is provided by the recent history of the Balkans.

In more recent times a specific need for support has emerged in the countries of the North of the World, where very low fertility rates, combined with high levels of life expectancy, have caused unprecedented demographic aging of the population. True, no country has explicitly adopted an immigration policy aimed at meeting that need yet; still, it is not impossible that such types of measures be taken in the near future. In fact, a recent UN report strongly recommends that the countries concerned begin to do so (United Nations, 2000).

In any case, in all those different areas, the trend is in favor of negotiated and international agreements. Most of the time such agreements are developed and implemented due to the pressure of growing needs. As a consequence migration flows are taking extremely varied forms, and can be defined in very different ways depending on which questions are asked—Who? How many? How? Where from? Where to? Under what legal conditions? In what political contexts? (Golini, 2003).

Negative measures, still more than positive ones, are also intended to meet national concerns in the countries of destination. The latter are clearly bent on controlling the flows, and when immigration seems undesirable to them, they enforce drastic policies so as to limit the flows, or even to stop them altogether, or again to turn the immigrants out of the country.

Such policies may be related to chauvinism, as was the case under the fascist regime in Italy, or to a situation of conflict, such as prevails between Israel and the occupied Palestinian territories. More often yet they express a more general concern with regulating immigration flows for economic or socio-cultural reasons, which often merge in a more or less confused, or even demagogical way, particularly in times of elections.

In several European countries, including those where immigration is recent, such as Spain, Italy, or Greece, measures setting a limit on admission flows are combined with regularization action through which the situation of immigrants who previously entered the country in a disorderly, or even illegal way, can be resolved.

In the countries where there has been long-standing immigration (France, Germany, the United Kingdom, Canada, the United States, Australia), regularization and control measures are also usually combined with integration policies in which the provisions for naturalization or access to citizenship play an important part, with wide differences from one country to another. For example, in Europe, France, where the right to citizenship by virtue of birth in the country prevails, can be considered favorable to naturalization, as opposed to Germany, where the right to citizenship by virtue of kinship prevails, making naturalization more difficult to get; on the other hand, American and Canadian legislations are far more favorable to naturalization than the French legislation.

In any case the economic situation, particularly that of the labor market, plays an extremely important part; however, internal and international rules in favor of basic human rights, of which the different countries are more or less observant, are also important. It is in that view that family entry and settlement is one of the most widely accepted exceptions to restrictive policies.

In spite of controls, a number of migrants manage to get across most borders. The United States is the country with the largest number of illegal migrants (7 million of them in 2000, by Immigration and Naturalization Service estimates) because they have such a long border with Mexico. According to the Organization for Migration (IOM, 2003), illegal entries into the European Union may well be as many as 500,000 a year.

Controlling immigration may eventually lead to militarization of borders, which may have an impact on the political system in the country of destination.

> There is a fear that militarizing external borders may eventually lead to the militarization also of society. [. . .] A complete control is nearly impossible in the Western democracies today, considering the political costs of militarizing/policing borders and internal structures. Neither does it seem possible due to the fact that the labor market in itself represents a "pull" factor in many of the countries, indicating that (sometimes powerful) groups are having vested interests in a continued flow. Besides, it seems difficult to stem immigration without discouraging tourism, as a significant number of illegal immigrants today legally entry as tourists in the first place. (Brochmann, 1993)

The *quota* system, in which the number of immigrants is set based on their initial nationality or other such criteria (IUSSP, 1982), is used with a view to limiting the number of newcomers and stabilizing immigration. Switzerland provides an example of countries that offer different quotas to immigrants depending on the type of work permit (on a year, season, or short visit basis) that they apply for.

Visas and border controls can be more efficient if they are buttressed with computerized databases on undesirable persons (as stipulated in the *Schengen* agreement, e.g.). The criteria used by host countries for work visas (and hence for immigrant selection) are widely varied—family entry and settlement, skills, random methods (such as lotteries), and so on. The efficiency of visas as a means to control entries is closely related to the internal controls to which foreigners are subject.

The authorities in charge of immigration policies can establish several types of internal control. These range from internal surveillance to access to a resident's permit or an identity card, access to lodging and social benefits, measures taken against employers, etc.

Theoretically, potential migrants will ground their decision to migrate on assessment of several elements. Salaries in real terms in the country of destination rank among the most important elements. The level of employment, the possibilities of getting accommodation, and last but not least, the financial and moral or psychological cost of migration, are also important. Last, specific information, focused on the economic opportunities offered by the host countries, may help potential migrants to make correct assessments.

2. The Stock Policies of the Host Countries

A stock policy is any measure or regulation that either encourages migrants to settle in the host country for long periods of time, or even for their entire lives,

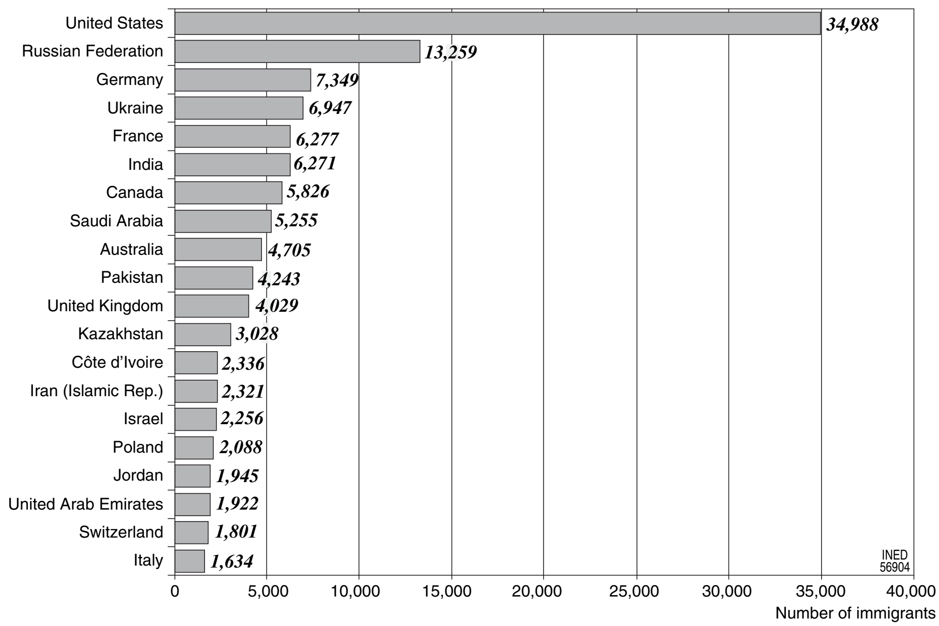


FIGURE 110–2 Numbers of immigrants (in thousands) in those countries where stocks were largest in 2000. (Source: United Nations, 2002.)

or to stay there for a very limited period only. Considering equal flows, such policies will determine the size of the stock, but also the rotation pace and hence the characteristics of the stock.

Figure 110–2 is an illustration of the comparison between the sizes of the largest immigrant stocks in the world.

Immigrants from the same country or region who have been living in a country for a long time eventually form a community. The latter may be more or less attached to their origins, depending on both the policies enforced by the host country and the immigrants' attitude.

Before studying the various degrees of immigrant integration into host societies, we must mention the part played by labor market structures as a determining factor in the formation and composition of migrant communities.

Immigrants, as mentioned earlier, usually find jobs among the lowest segments of the labor market. Supply of cheap labor from immigrants makes it easier for the host country to keep up its social system, particularly in times of economic recession. Besides, as far as illegal workers are concerned, the cost of labor is further brought down through nonpayment of welfare costs (health insurance, pension funds, etc.).

Despite the economic benefits derived from it, immigration is often described as a threat to the host country's economy. It is considered one of the factors in rising unemployment (particularly for unskilled workers) and is sometimes perceived as a danger to social stability. Such a biased image fails to take into account the fact that the labor markets, in most developed countries, are characterized by segmentation. It also ignores the fact that certain chores are economically and socially unattractive in the eyes of native workers. It thus appears that immigration has no effect either on the nationals' income or their employment (OECD, 1993).

One of the most obvious effects of immigration for the host country is that it reduces the cost of labor. In the short and the middle term, immigrant workers are underpaid and overexploited, as compared with native workers, because of their weak, marginal position in the labor market; they are also usually employed in particularly difficult sectors (the steel industry, mining, fisheries, etc.), in services that do not pay well (as domestic servants, in the food industry, etc.), or in seasonal work (agriculture) that nationals, even when they are unemployed, find unattractive. Small and medium-sized enterprises are particularly in need of cheap foreign labor, which makes it possible for them to remain flexible and keep their production costs down. Immigration thus generates immediate profit for many different groups (heads of companies, middle-class households, etc.), while the cost for society (in terms of welfare, educational, and health services) is deferred.

In most cases, foreign immigration has proved not to be a temporary phenomenon. Most immigrant workers who entered the host country in periods of economic growth subsequently decided to stay longer than might have been expected.

However, there are exceptions. Switzerland provides a famous example of workers' rotating immigration. The system is characterized by short-term work permits and seasonal authorizations. Workers then have to make frequent moves between their native country and the host country. To make sure that temporary workers actually remain so, rotation must be fast. Naturalization and family entry and settlement are very rare. Egyptians migrating to the oil-producing Arab countries where there is regular rotation of foreign labor is another example of temporary migration.

There is still a lot of seasonal work in many European countries, such as Germany (particularly from East European countries) or Italy (from North Africa).

The various degrees of integration in the host society can be classified as (a) assimilation, (b) the ethnic approach, or (c) integration. *Assimilation* can be considered as full acceptance of immigrants by the host society, until they are completely absorbed into it. Immigrants are scattered within the national community, with which they identify, and from which they eventually stop thinking of themselves as different. The much-used "melting pot" image gives an idea of ethnic identities mixing into a new homogeneous identity that emerges from the mixing. Any form of assimilation makes integration easier, because it accelerates the learning process and the use of public and private services in the host country.

Conversely there are cases (as Asians do in Britain) when immigrants do not really wish to get assimilated into the host society. In such cases they tend to establish close collaboration within the immigrant community that they form and to keep up close links with their native country.

The *ethnic approach* has negative aspects, such as the emergence of rivalry between the different groups, but also positive ones, such as cultural enhancement. The United Kingdom provides a typical example of that form of approach. The government guarantees all individuals the right to participate in the country's social, economic, and public life, while fully preserving their freedom to practice their own faith and keep their cultural identity. "Ethnic minorities" are thus recognized and classified. Even in censuses, the population is classified in different ethnic categories—White, Caribbean Black, African Black, other Black, Indian, Pakistani, Bangladeshi, Chinese, other. One side effect of such an approach is that it may heighten the risk of discrimination in terms of housing, schooling, and so forth.

Multiculturalism is a slightly different approach, in which ethnic and other differences are not only tolerated but even encouraged sometimes. In that way, identifiable groups coexist and interact so as to create a heterogeneous, stable society. In the near future, a new ethnic model may possibly emerge and develop in the East European countries where immigration is now a serious problem.

The *integration* of immigrants can be considered from two points of view—the integration of immigrants into the native society (which is the concept most in use), but also the integration of the natives into the immigrant community (in terms of acceptance of their traditions, religions, etc.). In both cases, integration is made complicated by interaction between the policies and societies of the host countries and those of departure.

Integration is not always the result of a deliberate decision; it sometimes happens through necessity rather than by choice. The legislation or the practices of the administration on the labor market may prove to be means to force migrants to seek to be integrated into the host country.

Within the European Union, the aim of integration is to gradually eliminate legal, cultural, linguistic and other obstacles, so that immigrants can have the same way of life as the nationals of the host country. Along that line, the European Council stated at the Tampere meeting (October 15–16, 1999) that

> The legal status of third-country nationals should be approximated to that of Member States' nationals. A person, who has resided legally in a Member State for a period of time to be determined and who holds a long-term residence permit, should be granted in that Member State a set of uniform rights which are as near as possible to those enjoyed by EU citizens: e.g. the right to reside, receive education, and work as an employee or self-employed person, as well as the principle of non-discrimination vis-à-vis the citizens of the State of residence. The European Council endorses the objective that long-term legally resident third-country nationals be offered the opportunity to obtain the nationality of the Member State in which they are resident. (European Union, Conclusions of the Presidency, para. 21)

Assessing integration possibilities requires taking immigrants' attitude into account. In many cases, particularly soon after they have arrived, immigrants will clearly express a wish that they be recognized in their minority groups, an attitude that could be interpreted as a wish not to get integrated into the host society and remain an ethnic minority. High concentration of immigrant populations may delay the integration process, because it encourages keeping up the traditions of the native country, traditional ways of life, and preferential marriages.

However, keeping up the original culture does not necessarily mean a loss in terms of integration. People can well be integrated through their job and daily life, while continuing to practice their religion and preserve their beliefs and customs. The United States and Canada provide good examples of multiethnic integration: the Greeks, the Spaniards, the Italians, the Irish, and the Poles live together, and are integrated into American society in many ways, but to a large degree they continue to preserve their individual ethnic identities and spend their leisure time with their own ethnic communities.

Integration is a psychosocial, long-term process. It takes both time and willingness (or alternatively, necessity). Temporary immigrants are seldom interested in becoming integrated simply because they think they will be staying a short time. Later they may try to get a little more integrated, just enough for them to live quietly.

Integration policies are very slow in producing effects. Also, they are not really compatible with political situations in which governments are expected to find rapid solutions to the social pressure problems that are derived from immigration. Not only the central authorities, but also the local ones, play a crucial part in implementing integration policies that usually target minorities whose demands are very heterogeneous.

The problems related to immigrant integration concern housing, health, education, employment, as well as the social and cultural dimensions. Each of those issues is important and can provide indicators as to the degree of integration among immigrant communities. However, the problems of housing, welfare, employment, and education do not concern only immigrants, and hence cannot be considered as solely related to such or such an integration policy. Local administrations are faced with those problems at a general level.

Immigrants only put in an additional dimension, which the authorities must take into account as they seek for such solutions as may promote local development within society. This applies at a local as well as a national level. In other words, certain policies aim at overall improvement of living conditions (housing, health, employment, services, etc.) with a view to securing social stability, economic growth, and welfare; other policies are more specifically intended to promote the integration of the foreign part of the population; but all have an influence over the degree to which immigrants are integrated.

Accommodation is usually the very first problem that newcomers have to face. Immigrants are subject to accommodation conditions that are inferior to those of the nationals. They often suffer discrimination on the part of landlords, and eventually go and live in areas where they are tolerated because such areas are spurned by the natives. As a result, they are concentrated in specific residential areas, where the quality of accommodation is bad and living conditions are far below standards.

"Education is of vital importance to the integration of immigrants' children. Both the level of schooling achieved and the diplomas obtained are determining factors in their success in the labor market and their integration into the society of the host country" (OECD, 1995). The concentration of immigrants in specific areas has consequences for education, because the normal implementation of school curricula may be disturbed by the fact that so many children speak different languages and have such different cultural backgrounds. Some countries, such as the United Kingdom, have been trying to distribute immigrants' children among different schools, so that they are not concentrated in some schools only. As education is the main means for young people to make their way into society, immigrant parents usually wish their children not to lose their cultural and religious heritage. Integration policies at school must then guarantee both that cultural identities are preserved and that there is no discrepancy between the education provided to immigrant and to native students.

The issue of immigrants' access to health insurance and old-age pension is closely linked to that of illegal employment. Clearly there is discrimination in the labor market, though it is so complex that it would be difficult to assess. Employers are usually less interested in their employees' integration than in the quality and quantity of the work they provide, whatever their nationality. Immigrants' position on the labor market is one of the basic indicators of their integration.

The level and quality of the education of immigrants' children play a crucial role in their access to vocational training. In this respect, there are two possible approaches: supporting them through the development of specialized training, or implementing antidiscrimination policies at recruitment level.

The idea of the cultural and social dimension of integration includes all the activities of volunteer associations that work toward the benefit of immigrants. Among such associations are those created and run by natives (religious associations, trade unions) and those set up and managed on an ethnic or interethnic basis. The latter are particularly involved in political, social, and cultural activities, either within or outside their communities. Some associations are created with a view to securing well-being and support for their members; if, as sometimes happen, the necessity for such associations is felt to decline, then they either die out or change their activities. For example, some regional ethnic associations are now no more than meeting places for elderly immigrants.

Integration should not necessarily be considered in a positive sense. Immigrants also may be integrated into corrupt groups, with criminal or antidemocratic activities. Paradoxically a high rate of crime among immigrants may be an obvious sign of their capacity to get integrated in the host country. One significant example is the trafficking of women from developing countries, or in more recent times, from Central and Eastern Europe to Western Europe.

> Trafficking in women occurs when a woman in a country other then her own is exploited by another against her will and for financial gain. The trafficking element may—cumulatively or separately—consist of: arranging legal or illegal migration from the country of origin to the country of destination; deceiving victims into prostitution once in the country of destination; or enforcing victims' exploitation through violence, threat of violence or other forms of coercion. (IOM, 1995b)

Sweden and Germany provide two interesting cases for an understanding of the various degrees and methods of integration.

Sweden has been gradually reinforcing its integration policy since the 1970s. Now, every foreign-born resident who has decided to live in the country must demonstrate a degree of integration. The Swedish immigration policy is based on a tolerant kind of multiculturalism. Immigrants are regrouped in parallel organizations by nationality. In that integration process, Sweden, along with the other Scandinavian countries, forms the most advanced region, as it guarantees the right to vote in local elections to those immigrants who have permanently settled on Scandinavian territory.

Germany is totally different, in that the country does not officially consider itself an immigration country. In the past, Germany firmly supported rotation policies and was harshly critical of integration policies. Now, they have changed their approach, and the integration of foreigners and their families who are legal residents on their territory (particularly the "invited workers," who immigrated in the 1960s and 1970s) is one of the principles of the government's "foreigner policy". In Germany's explicitly expressed position, migration has always been reduced to the needs of the country's economic development. As a consequence, even the very limited forms of integration offered to the invited workers who had been living for many years in Germany were never intended to make the immigrants more German-like.

The success of an integration policy depends, to a degree, on the measures intended to limit future immigration. The most significant legislative instruments to control and stabilize immigrant communities are regularization, naturalization, and family entry and settlement. Policies in this respect are extremely varied, even within the European Union (Table 110–1).

TABLE 110–1 The Degree of Integration and the Migration Policies Adopted in Some European Union Countries (as of 1994)

Easy naturalization	**Difficult naturalization**
Sweden, France	Germany
Recognition of "ethnic minorities"	**"Immigrants" or "Foreigners"**
United Kingdom, the Netherlands	All other countries
Formal "multiculturalism"	**Assimilation or departure**
Sweden, the Netherlands	Germany
No residence restriction	**Local residence restrictions**
United Kingdom	Belgium

Source: Macura and Coleman, 1994, p. 60.

Regularization is a means to reduce the number of illegal immigrants. That type of measure is usually used before tightening immigration restriction measures. It has not been proved that regularization measures affect future immigration flows. The regularization of illegal migrants is always combined with sterner measures of immigration limitation. It does ensure that immigrants are better protected, but past experience has shown that it does not prevent new illegal immigrants from coming.

Naturalization can be used as a means to accelerate the integration process. It is easier to secure for those immigrants (or their offspring) who grew up in the host country.

The host countries have widely different policies as far as naturalization is concerned: some countries, such as Canada or Sweden, encourage immigrants to get naturalized as soon as possible so as to make their integration easier. On the contrary, other countries, such as Switzerland or Germany, are very restrictive and consider naturalization to be a reward for immigrants. The issue has also been discussed in world conferences such as the World Summit for Social Development that was held in Copenhagen in 1995, when the final declaration mentioned that "governments of receiving countries are urged to consider giving to documented migrants having the right to long-term residence, civil and political rights and responsibilities, as appropriate, and facilitating their naturalization" (United Nations, 1995b, point 77).

Family entry and settlement is one of the most important means used by the host countries to promote or reduce the integration of legal immigrants. According to the Programme of Action of the Cairo Conference:

"All Governments, particularly those of receiving countries, must recognize the vital importance of family reunification and promote its integration into their national legislation in order to ensure the protection of the unity of the families of documented migrants" (United Nations, 1995e, point 10.12). It is important to note that the Conference, after lengthy discussions, could not come to an agreement on recognizing reunification as a basic right. The right to immigrate may be granted to spouses who are separated, or extended to other family members of those immigrants who have a long-term resident's permit. Family reunification is particularly important for migrant women who are admitted into a country mostly on the grounds that they have family ties with nationals of the host country or with other immigrants. Therefore, they are dependent on the migration and status of other persons. The women who are admitted as workers are usually confined to traditionally feminine jobs (as domestic workers, in services, care, etc.). In fact family reunification policies have made family flows possible. Due to drastic reduction of workers' flows, the latter are now the very first source of residence permits in some countries, such as Italy.

Mixed marriages (between nationals and immigrants) may be considered an indicator of integration. However, it must be kept in mind that marriage is sometimes used as a means to get a resident's permit or the nationality of the host country. A few recent surveys (Tribalat, 1996) have shown that mixed marriages are comparatively unstable, which suggests that they do not necessarily entail reinforcement of the assimilation or integration of immigrants into the host society.

Many different attempts have been made to measure the integration of immigrants (Cagiano de Azevedo *et al.*, 1992; also see Chapter 84). The most useful sources in developing integration indicators include basic demographic data, position within the labor market, household and family structures, the length of the stay, economic activity, education, income, accommodation conditions, naturalization, personal attitudes, and legal status. Two basic aspects must be taken into consideration in measuring migration. On the one hand, there are problems related to the diversity of sources, as a result of data on international movements being probably incomplete; and on the other hand, there are conceptual problems:

> While births and deaths can be defined biologically and marriages and divorces legally, migration is more difficult to define. Migration entails not only the crossing of international borders but also entails the intention of staying in the country or area of arrival. Although international migration may well in many cases entail legal procedures, it remains the most difficult of demographic phenomena to define correctly. (United Nations, 1991, p. 95)

II. THE POLICIES IN THE COUNTRIES OF EMIGRATION

The distinction previously adopted for immigration countries between those policies that impact on flows and those that affect stocks also applies to the countries of origin (Figure 110–3).

1. The Flow Policies of Emigration Countries

Before we study the political measures that affect the flows in the countries of emigration, we will mention the main basic factors that determine those flows: the demographic situation, the labor market, the migrants' networks, and educational level and profile. Next we will discuss how policies can be implemented through border controls, administrative structure, and information on receiving countries.

The demographic situation in the emigration countries is obviously a key factor in migration. The most important variables include growth pace, fertility, birth and mortality rates, and population distribution by age and sex. When fertility is high and mortality is low, the number of people of working age who enter the labor market grows very fast, and the phenomenon continues for some time after fertility has begun to decline. The pace of job creation then can hardly keep up with that of the growth of manpower. In many developing countries, the emigration of part of the population of working age has been a means to solve this demographic problem, at least in part; simultaneously it has also been a source of additional income for development, through foreign currency transfers.

In Western countries population growth is under 0.5% a year, as opposed to over 2% in the developing countries. When the resources of a region prove not to be enough to meet the basic needs of its population, migration appears to be a safety valve. The current demographic situation in the North African countries, for example, despite the very fast decline in fertility that they have experienced in recent times, still implies substantial growth of the population of working age in the next 20 years. Those countries will thus remain a source of migration pressure for another two decades, particularly on the countries in the South of the European Union, unless their economic development is fast enough to absorb that potential excess of labor.

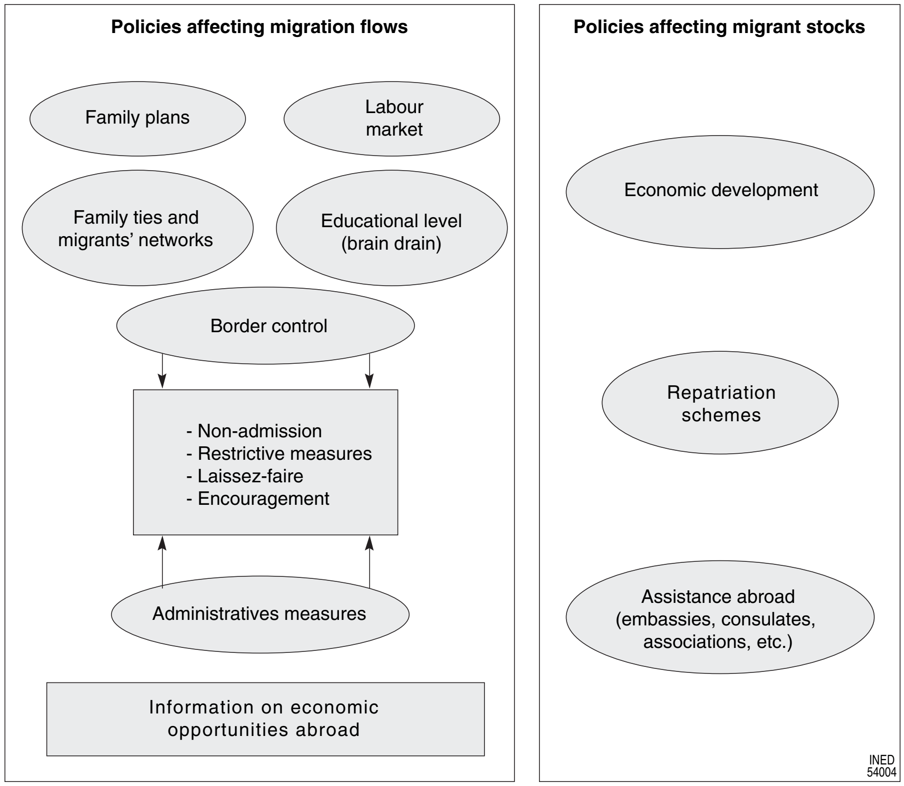


FIGURE 110–3 A schematic comparison between the logistics and means used in flow policies and stock policies in the emigration countries.

> Development and demographic balance are no longer considered the sole concern of the individual states involved but rather as problems facing a wider human community comprising all of the countries affected by these issues, which, in accordance with the subsidiarity principle, may be dealt with as appropriate by local, regional, national or international institutions. (Council of Europe, 1995)

A distinction must be drawn if the effects of emigration on the labor market of the countries of departure are to be assessed. In the case of unskilled or low-skilled workers, emigration can be considered a cheap, fast means of alleviating unemployment. In the cases of highly skilled persons, emigration (which is then referred to as "brain drain") may have long-term negative consequences over the development of the country of departure. Brain drain can be viewed as resulting from an understandable wish among highly skilled workers to make maximum use of their investment in education. In the available literature, brain drain is often considered as a loss for the emigration country, though it is not always the case. Nowadays, the mobility of skilled persons is usually related to the training process (e.g., doctorates) or to progress in their careers (in international research laboratories). The negative aspect of brain drain is in the lack of skilled labor in the countries of origin (e.g., the lack of doctors in developing countries). Their departure is a problem for the countries of emigration. However, if the inactivity or unemployment rates are high among the skilled workers of the country of origin, the emigration of skilled workers may have positive effects on employment for the skilled persons who stay in the country, and in the long term, the consequences are not necessarily negative for the population as a whole, because foreign currency transfers may then play a major part in the development of the country of emigration. Besides, cultural (scientific or vocational) transfers, in the long term, are important stimuli for an improvement of the internal situation. Some countries are in favor of the emigration of their skilled nationals for that very reason.

Migration networks are generally a means to disseminate information on employment opportunities and administrative procedures in the receiving countries. The existence and efficiency of such networks have been proved through many surveys conducted among migrants. The rapid development of Turkish emigration to Germany can largely be accounted for through the efficient network established in the 1960s. The same applies to Filipinos who emigrate to the

United States. On the one hand, such networks may help emigrants at the time of departure, and later protect them in their new environment; on the other hand, they may slow down the integration process.

The brain drain issue is related not just to the distortions of the labor market, but also to the educational system. The lack of demand for skilled workers or lecturers and researchers encourages such highly skilled persons to emigrate, because they must have a more favorable environment if they are to progress in their careers. Emigration may also result from lack of infrastructure or junior technical staff, as the organization of their research or specialized work is dependent on those. Such emigrants are unlikely to return to their country of origin.

A wide variety of border controls and other such administrative measures, ranging from prohibition to encouragement, can be enforced by the governments of the countries of emigration with a view to adjusting the flows (Appleyard, 1989). Stark prohibition of emigration is seldom efficient (the case of Algeria in 1973, though, is a success story). Better results can be reached through selective control in which limits are set to the emigration of persons with certain qualifications. In the case of laissez-faire policies, governments try to remain neutral, only establishing minimal or no control. On the contrary, some countries actively encourage workers to emigrate and do their best to increase the number of their nationals abroad (as was the case for the Philippines in the 1980s). It must be kept in mind that migration policies also reflect the internal political regimes. In totalitarian countries, emigration, far from being an objective of the government, is usually strictly forbidden.

Other measures have also been taken to regulate emigration flow as deemed desirable—instituting authorizations prior to emigration, defining the standards required of private recruiting agents, setting maximum recruiting fees to be paid by workers, requiring minimum guarantees in contracts of employment, and so forth.

It is very important for the country of emigration to provide potential emigrants with data and information on the economy and labor market of the possible receiving countries. Potential emigrants have very poor and superficial knowledge on the laws in those countries. In that respect, the media have great influence, though they may mislead potential emigrants by disseminating false information, particularly by giving the public an overly positive image of the economic situation and employment opportunities in those countries that are perceived as potential receiving countries. This is probably what happened to many Albanians who emigrated to Italy in the early 1990s, grounding their decision on the attractive image of Italy disseminated by Italian radio and television stations.

Information on the receiving countries would empower migrants in several respectsby explaining how to select recruiting sources, examine and accept a contract of employment, organize money transfers, and set up adequate saving and investment plans (Gunatilleke, 1992). In fact, all of these activities require full collaboration between the countries of departure and those of destination.

2. The Stock Policies of the Countries of Emigration

The economic situation and its prospects in the country of emigration are very important in the decision to emigrate. As long as the country remains in an underdeveloped situation, with no signs of economic take-off or will to achieve development, and gaps widen between salaries in the countries of the North and those of the South, not only will the number of people who wish to emigrate increase, but those who wish to come back to their native country are also deterred from doing so.

The economic effects of migration over the country of emigration are also to be felt through transfers of foreign currency and goods. Foreign currency transfers may have several kinds of positive consequences on the economy of the emigration country. They provide a source of foreign currency and are a factor of national income growth by means of the additional yields of savings and investment they generate. The level of those profits depends not only on the number of emigrants but also on their average incomes (salaries and other earnings), their propensity for saving, and the portion of their savings that they decide to transfer and invest in the country.

Here again, the case of Albania is interesting. According to an IOM study (IOM 1995a), 43% of Albanians have relatives who emigrated between 1992 and 1995. Most left the country illegally; the emigrants usually had good education and qualifications. In return, the transfers received by Albanians who stayed behind involve large sums, which often are their main means of subsistence.

For most countries of emigration, transfers are a major source of foreign currency, and this is reflected, in a way, in the set of policies devised to encourage emigrants to transfer their savings. For example, in some countries preferential conditions are granted in terms of exchange rates for such transfers and interest rates for their investment. Transfers are supposed to be individual and family actions, and consequently cannot be directly oriented by governmental policies. Theoretically when emigrants send money to their family, the country of origin should not interfere in the way the funds are used. In practice, however, the government can enforce financial and investment policies through the banking system; in developing countries, however, banking systems are usually unsatisfactory and sometimes corrupt. As a consequence, emigrants had rather send their money through informal channels (through relatives or friends, etc.) or use such safe, fast channels as the famed Western Union. Often, their funds are not sent home and are instead used to pay debts or contribute to the ethnic mutual assistance networks in the receiving country.

The debate on whether international migration promotes or delays development prospects, in both the countries of emigration and of immigration, will continue. The question is related to such various issues as poverty alleviation, improvement of living standards, and job seeking or creating, conflict prevention, reinforcement of democracy, and observance of human rights. Bilateral agreements between developed and developing countries increasingly tend to include a significant migration component, just like those on economic cooperation, external debt settlement, reduction of obstacles to commercial exchange, and other initiatives intended to promote the economic development of the labor-exporting countries. Personal mobility (emigration and repatriation) has thus come to play a crucial part in international relations.

The countries of emigration may develop repatriation plans so as to provide emigrated workers with economic opportunities in their native country and facilitate their reinsertion into their original society. Due to continued economic constraints in developing countries, the promotion of return migration has proved difficult hitherto. The countries of emigration usually contend that the issue of return migration must be considered within the context of economic development.

To a certain extent, the countries of emigration may provide assistance to their emigrated nationals. The existence of an embassy or consulate in the receiving country may play a major part in implementing migration policies. In fact, not only can they be used as a point of reference for emigrants, but they can also help maintain links with the native country. At least, consular issues (bilateral agreements, repatriation, money transfers, administrative affairs, etc.) can be solved more easily thanks to the presence of an official representation. However, many countries of emigration, as they are developing countries, cannot afford to have permanent diplomatic missions in every country.

International Organizations and International Agreements in the Area of Migration

Many different international organizations deal with issues related to migration policies at various levels. The main ones are the Population Division of the United Nations in New York, the United Nations Office of the High Commissioner for Refugees (UNHCR) in Geneva, the United Nations Population Fund (UNFPA) in New York, the International Labour Organization (ILO) in Geneva, the International Organization for Migration (IOM) in Geneva, the Organization for Economic Cooperation and Development (OECD) in Paris, the Council of Europe in Strasbourg, the European Union (EU) in Brussels. As mentioned in the *Action Plan* of the Cairo Conference:

> The role of international organizations with mandates in the area of migration should be strengthened so that they can deliver adequate technical support to developing countries, advice in the management of international migration flows and promote intergovernmental cooperation through, inter alia, bilateral and multilateral negotiations. (United Nations, 1995e, point 10.8)

The world conferences that have dealt with the issue have played, and may further play, a very important role in devising policies for the future. In recent times, the three most important ones were the International Conference on Population and Development (ICPD, Cairo, 1994), which issued a *Programme of Action*, the World Summit on Social Development (Copenhagen, 1995), which issued a *Declaration* and a *Programme of Action*, and the 1999 meeting of the General Assembly of the United Nations, when the efforts made to reach the ICPD objectives were reviewed.

Bilateral and multilateral agreements may also play a very significant part in the evolution of migration policies. As mentioned in the Cairo *Programme of Action*, the governments of the countries of destination are invited to

> consider the use of certain forms of temporary migration, such as short-term migration or migration related to a specific project, as a means of improving the skills of the nationals in their countries of origin [. . .]. To that end they should consider, as appropriate, entering into bilateral or multilateral agreements. (United Nations, 1995c, point 10.5)

Two major recent examples of regional integration agreements related to migration policies should be mentioned at this point—the North American Free Trade Agreement (NAFTA) and the Schengen Treaty, the latter of which was concluded between some of the member countries within the European Union. Both agreements also were intended to limit illegal immigration.

NAFTA, the most extensive area for international free trade in the world, is mainly intended to support Mexican economy with a view to reducing illegal immigration to the United States. However,

> because of the differences between the Mexican and the US productive structures and the size of their respective economies, rapid convergence of their standards of living is inconceivable, and, irrespective of signs that the gap is beginning to close, migration flows over the next few years will remain high. A reduction in the propensity to emigrate as a result of greater economic integration can only be a long-term objective; to recognize it as such is to avoid making the mistake concerning the nature of migration and development processes that characterized European policies after the frontiers were closed in the 1970s. (Tapinos, 1994)

Border control is abolished among the countries that signed the Schengen Treaty. However, identity control remains possible, for the prevention of international crime. The agreement also deals with police and judicial cooperation as well as coordination of control on the external borders of the Schengen space, through a common approach to the visa and asylum procedure policies. Even though the unification process has begun in the area of immigration, there are still countries in the European Union that are unwilling to give up control on their own borders for non-Union nationals.

The European Union has taken a somewhat tortuous path before it could come to a common approach to migration. The initial objective of the common policy had been zero immigration, before it took a new turn with the Barcelona Declaration in 1995, which was both a point of achievement and a new start, as soon afterward the European Council, meeting in Tampere in 1999, announced for the first time that the objective was to develop a common immigration and asylum policy. The declaration of the European Commission (the *Community Immigration Policy*, 2000) followed immediately, putting an end to the zero immigration policy and recognizing the European Union as a zone for immigration. (Zero immigration was referred to as part of a policy in which borders are controlled, or even closed, effectively barring new immigrants out. Despite such a policy, no decrease was observed in the flows, due to the demand for immigrant labor from the market.) The Laeken European Council meeting (2001) then emphasized the necessity to define a migration flow policy as an essential component of any common external policy. The next year, however, in Seville (2002), the Council somewhat left aside its overall view of migration flows, paying more attention to short-term initiatives aimed at reinforcing control over illegal immigration and terrorism, thus pushing the principle of a common security and justice space adopted in Tampere into the background. Eventually, the draft Constitutional Treaty was signed at the European Council held in Salonica (2003). It included a specific section on security and justice and placed immigration issues among decisions to be made by a qualified majority. The draft treaty was thus giving a mandate to the European institutions for them to develop a common immigration policy such as could facilitate efficient management of migration flows. Henceforth it was impossible for member states to fail individually to enforce the relevant rules that had been decided on by a qualified majority. This gives full measure of how much progress has been achieved, if the Constitutional Treaty is ratified.

III. NEW ELEMENTS IN MIGRATION POLICIES

1. Cooperation in Migration Matters

However strict they could be (with actual, rigorous control at the time of admission as well as subsequently), the policies adopted by developed countries clearly never had any direct effect over the basic factors that cause emigration from developing countries. On the contrary, cooperation and development program do have an impact on the source of international migration movements; at the same time, they are a way of refusing to accept economic and political underdevelopment as the result of fate (OECD, 1994).

Immigration is becoming a major issue for the main developed countries, particularly European ones, who are trying to establish new laws and regulations so as to reduce the access of immigrant workers to the labor market. The very fact that their legislative activity is intensifying is a token of the political importance of the issue. The measures taken to fight illegal immigration are more or less the same everywhere (border control, territory surveillance, prison or expulsion sentences against workers who have no employment permit).

However, it is recognized that the

> governments of countries of origin and destination should seek to redress the causes of emigration in order to alleviate the massive and uncontrolled international migration flows. The redressing of these causes would require increased effort to achieve sustainable economic and social development, avoid international and internal conflicts, respect the rule of law, promote good governance, strengthen democracy, promote human rights, support education, nutrition, health and population-relevant programs, and ensure effective environmental protection. (UN-ECE, Council of Europe and UNFPA, 1994, *Recommendation 31*)

A new form of cooperation is necessary. Developing countries, with the assistance and cooperation of developed ones, must ensure that economic growth is ever concomitant with job creation. Concurrently they must establish such social and institutional structures as are likely to slow down their demographic growth.

As recently emphasized by the United Nations,

> international migration, with its intricate web of demographic, social, economic and political determinants and consequences, is a topic that has moved to the forefront of national and international agenda. In the last few years, immigration has become a major issue of concern in an increasing number of countries. More recently, in the aftermath of the events of 11 September 2001, some countries have further tightened their policies towards immigrants, refugees and asylum seekers. (United Nations, 2002)

Through international cooperation, it is probably possible to limit migration flows so as to adjust them to the receiving capacity of the host countries. Many governments in the receiving countries have realized that it is not possible to reduce immigrant flows without taking the economic and social situation in the countries of origin into consideration. In such a view, many options have been examined (and experimented in some cases), such as opening up international trade, increasing direct investment, rescheduling external debts, or targeting public development aid in the emigration countries.

Many receiving countries now have to find a compromise between investing in integrated development plans in the emigration countries and investing at home in social infrastructure and services for immigrants. In fact, short of militarizing societies and fuelling the kind of tension on which fundamentalist movements feed, migration flows cannot be controlled without achieving development, or at least without development actually beginning to be visible in the countries of origin. If no action is taken in that respect, the level of social tension between the nationals and the immigrants, or between the receiving countries and those of origin, can only go on deteriorating.

It is necessary to develop joint (multilateral or bilateral) agreements on migration flow control. Such agreements must ensure that development aid is efficiently used, particularly in the areas from which migration flows start, on the one hand, and take into account the needs of intensive labor sectors, on the other hand.

Political stability is essential to developing long-term economic and social reforms in the countries of emigration. Regional integration, with all the efforts made to achieve it, might be one means of reducing tension in some regions where governments have difficulties establishing stability by themselves.

Immigration measures have been viewed as part of those on international cooperation, particularly those on cooperation with the countries of origin. Such cooperation measures include not only financial aid, investment in the countries of origin, vocational training, and economic and technical cooperation, but also admission of a limited number of immigrants. The number must be discussed with the countries of origin so as to secure their cooperation, which is indispensable in controlling the migration flows that exceed the levels set in the cooperation programs. The very existence of economic cooperation provides good motivation for the governments of countries of origin to actively and efficiently exert preventive control over migration flows: otherwise, they would see little point in doing so (Council of Europe, 1995).

2. Possible Future Developments in Migration Policies

The measures to be taken in immigration must be planned in close coordination with those relating to international cooperation, particularly with the migrants' countries of origin. Agreements must be secured with emigration countries on the circumstances under which they would be willing to help in controlling the migration flows that are in excess of the levels agreed in cooperation programs.

Cooperation between receiving countries and those of origin, which is essential in controlling migration flows, also requires that great attention be paid to the actual use of aid in the regions and villages from which migrants come, where the decisions to emigrate are taken. Although such decisions are prompted by national and international considerations, they are eventually determined by local pressure emergency.

South–South cooperation still remains a largely theoretical plan. Developing countries have not really managed to increase collaboration between neighbors. Future effective South–South cooperation is likely to have positive effects over migration. In fact migration policies are one of the major potential areas for cooperation between developing countries. Operating collaboration agreements between economic systems in the Third World (or South) countries may well solve the development problem, at least partly. The North could find enough resources to support such cooperation, so as to decisively contribute to launching a development process. South to South relations do not necessarily have to involve all the countries simultaneously; on the contrary, it would be best to bring together those countries that belong to the same geographical area and have similar problems, or comparatively homogeneous characteristics (Arab countries, the Sahel region, Latin America). Moreover, those countries should decide by themselves which countries should associate with which others.

As Claudio Grua (1991) wrote, the fundamental functions of such regrouping could include:

Creating sound structures for cooperation between countries

- Liberalizing trade within the zone and defining a common trade policy (customs duty, tariffs, negotiation on the location of international headquarters)
- Defining some common development programs in major areas of common interest
- Creating supranational agencies for the management of strategic resources
- Establishing a regional union for payment purposes so as to reduce both needs in hard currency and debts with the international monetary system

Besides stimulating synergies in the zone, such a policy would offer the possibility of speaking with one voice in international institutions, at least on subjects on which general agreement can be reached—which would promote the emergence of a political group on an international scale.

In such a context, what the developed countries can do is quite clear: aid must not only be part of a program of organic transfer of resources, as mentioned earlier, but also be granted under specific conditions and organized so as to ensure that it is successfully used. Reference to the Marshall Plan could be a source for useful suggestions.

3. Co-development Processes and Decentralized Cooperation

Among host countries as well as those of origin, many belong to the same region, though located on different continents (e.g., the Mediterranean region). A new approach to the migration issue could be devised as a kind of co-development process, in which countries (or certain regions) of origin would be considered as internal underdeveloped regions within the host countries. Such zones would then receive all the aid and facilities that migrant-receiving countries provide to their own regions.

Usually, at the time when cooperation programs are developed, no study is conducted on the current demographic situation, nor on the possible effects of such programs over the situation; and no follow-up system has ever been set up so as to assess the expected or the final results from a demographic point of view. In fact, the criteria for selecting cooperation activities should be changed so that they are better suited to the political and economic context; priority sectors should be defined and appropriate strategies established.

Prospective demographic growth is expected to remain high in developing countries. This is an obstacle to absorption of the expanding manpower through local economic development alone (not to mention the needs in improved public services, particularly in the social sector); consequently, it is a strong incentive to emigration, particularly from Africa to Europe. Whereas in the past, emigration was a possible means of overcoming the obstacle, it is now drastically limited by the strict policies enforced in the host countries, which are also favorite destinations for migrants from Eastern Europe—hence the necessity to establish close cooperation between developing and industrialized countries, such as make it possible for every party to work complementarily to one another.

Developing countries will have to examine how to reorganize and expand their social sector and implement population policies in which the most isolated categories of the population are included. They must also campaign at all levels to improve the situation of women, whose social condition is still often precarious and who suffer from discrimination in education and employment. In many countries, the law and the customs even grant them rights that are inferior to those of men, and women find it difficult to assert their role and personality in societies that still force them to have large families so as to ensure that they get minimum support in old age.

There is a risk to using temporary emigration to promote vocational training and to profit from migrants' financial transfers: emigrants may be unwilling to return to their native countries at the end of the agreed period. This, however, is not sufficient reason for host countries to put an end to the experiments conducted here and there on controlled temporary immigration, in which promise of return to their native country is secured from the workers. Certain trials made in European regional or local institutions provide promising examples.

On the other hand, new forms of migration are emerging, such as transit migration, in which migrants go to one country with a view to trying to go to another later. This is usually a type of illegal migration that takes advantage of certain flaws in the system. A typical example is when people migrated by stages from East European countries to Western Europe, in the wake of the 1989–1991 political upheavals (OECD, 2000).

The issue of differences in the cultures and value systems of emigration societies and host ones must also be taken into consideration. It is a long way still to harmonious cohabitation. Cultural cooperation must be firmly planned and developed, giving high priority to improving mutual understanding. Some programs have been initiated along that line, but they are still too timid, with respect to their priorities as well as to the resources that are granted to them.

4. Migration as a Move Planned at the Micro Level

It would be profitable to examine and develop new concepts with a view to better understanding and dealing with migration phenomena. An interesting approach consists in considering migration as an individual plan. The new migration flows from the Third World cannot be considered as though they were the same as the traditional migration of workers. The latter is usually regulated by international agreements between the country of origin and the host one, so as to benefit both countries. A large proportion of recent flows could be better accounted for if the decision to migrate is taken into account at the micro level.

In that prospect, an individual decision to migrate can be considered part of a rational plan developed on the basis of an economic and social assessment of costs and profits. In fact the decision to emigrate is usually made at the family level. In the distribution of individual roles and tasks, certain members of the family are selected to stay at home and participate in farming, while others are entrusted with small trading or transport activities, and yet others must go and earn money elsewhere by migrating, particularly abroad, and transfer as much of it as possible to the family, so as to keep the latter supplied with consumer goods or capital.

In such circumstances migration tends to strengthen family ties within the extended family in the most traditional sense, in the migrant's native village and region. As a rational plan based on integrated economic decisions, individual migration then contributes to family development from an economic and other points of view. Decisions concerning marriage, the question of how many children to have and when, and sometimes the decision to return to the native country, thus become stages in a well-planned process. This partly contributes to high demographic growth continuing through the stages of demographic transition among some populations. Migrants' plans thus appear to be fully part of a process of family building that may influence the plans of their future children later. This type of migration planned at micro level usually is a condition for the survival of village and regional economies.

The male or female members of a family who emigrate as a result of such rational plans are bound to supply the family with additional resources as soon as possible, not only in financial terms (transfers) but also in terms of consumer and capital goods, such as cars and other means of transport, goods for small traders, and consumer durables.

Once the decision to emigrate has been made, the persons involved have to face the many difficulties of migration, and accept the ups and downs of it. The first shock for migrants is usually at the macro level: they have to face borders and a State, away from the possibility of assistance from their family, whom they cannot get in touch with. This is not in conformity with their initial plan, particularly with their ambition to start making financial transfers soon after their departure. As for host countries, they have to choose between forbidding long periods of stay and integrating immigrants. These are two approaches that used to cause division among politicians in some European countries in the 1960s and 1970s.

The macro effects of this type of migration remain to be studied, though some of the consequences have been clearly identified already. In Morocco, for example, according to the National Trade Agency, the financial transfers made by emigrants (exchanges in kind not included) amounted to 20 billion dirham in 1994. This in value is half the amount of the country's exports, 30% of imports, 150% of the profits from tourism, and 250% of the payments for the external debt. The savings from emigrated Moroccan workers not only make it possible for the country to pay the bills for its external debt, but they also have consequences for local economies in the migrants' native regions.

The levels of resources transferred in foreign currency or in kind by emigrants are such that they are enough to create an actual transnational economic zone. Such transfers, though, may be partly derived from illegal, or even criminal, activities. The zone thus created by migrants also requires new regulations and new forms of cooperation. The part played by migrant communities in cooperation is no longer just rhetorical or abstract; it is a basic element in the current economies and has many political and cultural implications.

Acknowledgment

This chapter was translated from the French by Mireille Rabenoro.

References

Appleyard Reginald T. (ed.), 1989. *The impact of international migration on developing countries.* Paris, OECD, Development Centre, 404 p.

Appleyard Reginald T. (ed.), 1991. *International migration: challenge for the nineties.* Geneva, International Organization for Migration, 84 p.

Appleyard Reginald T., 2001. International migration policies 1950–2000, *International Migration*, vol. 39(6), p. 7–20.

Arditis Solon (ed.), 1994. *The politics of East-West migration*. London and New York, The Macmillan Press and St. Martin's Press, 258 p.

Bogue Donald, Arriaga Eduardo and Anderton Douglas (eds.), 1993. *Readings in population research methodology. Vol. 4: Nuptiality, migration, household, and family research.* Chicago, UNFPA/Social Development Center.

Brochmann Grete, 1993. Migration policies of destination countries, *in:* Council of Europe *(ed.), Political and demographic aspects of migration flows to Europe*, p. 105–128. Strasbourg, Council of Europe Press (Population Studies, no. 25).

Cagiano De Azevedo Raimondo, 2000. *Le migrazioni internazionali. Il cammino di un dibattito.* Turin (Italy), G. Giappichelli Editore, 196 p.

Cagiano De Azevedo Raimondo, Bacchetta Paola and Di Santo Paola (eds.), 1993. *International migrations as a security and cooperation in Europe and in the Mediterranean region. European seminar proceedings.* Rome, Università degli studi di Roma "La Sapienza," 195 p.

Cagiano De Azevedo Raimondo, Cantore Alessandra, Di Prospero Rita and Sannino Barbara (eds.), 1994 . *Immigrants integration policies in seven European countries.* Rome, Università degli studi di Roma "La Sapienza," 236 p.

Cagiano De Azevedo Raimondo, Di Prospero Rita and Di Santo Paola (eds.), 1992. *Measuring migrants integration.* Rome, Università di Roma "La Sapienza," 341 p.

Collinson Sarah, 1993. *Europe and international migration.* London and New York, Pinter Publishers for Royal Institute of International Affairs, 189 p.

Council of Europe, 1995. *Migrations and international co-operation.* Strasbourg, Council of Europe Press, 133 p. (Population Studies, no. 28).

Council of Europe, 1997. *Measurement and indicators of integration.* Strasbourg, Council of Europe Publishing, 198 p.

Council of Europe, 2000. *Framework of integration policies.* Strasbourg, Council of Europe Publishing, 49 p.

Council of Europe, 2001. *Migration. Summaries and texts of treaties.* Strasbourg, Council of Europe Publishing, 267 p.

Council of Europe, 2002. *The demographic characteristics of the immigrant populations.* Strasbourg, Council of Europe Publishing, 598 p.

Domenach Hervé and Picouët Michel, 1995. *Les migrations.* Paris, Presses universitaires de France, 128 p. (*Que sais-je?* series).

Dumont Gérard-François, 1995. *Les migrations internationales, les nouvelles logiques migratoires.* Paris, CDU-SEDES, 224 p. (Mobilitésociale).

Entzinger Han B., 1990. L'emergenza delle politiche di integrazione per gli immigrati in Europa, *in:* Albert Bastenier, Felice Dassetto, and John Rex (eds.), *Italia, Europa e nuove immigrazioni*, p. 179–200. Turin (Italy), Edizioni della Fondazione Giovanni Agnelli, 260 p.

European Union, Conclusions of the Presidency, 1999. *EU Newsletter*, no. 10, p. 4.

Eurostat, 2002. *European Social Statistics. Migrations.* Luxembourg, European Commission, Eurostat, 75 p.

Falchi Nino, 1995. *International migration pressures. Challenges, policy response and operational measures: an outline of the main features.* Geneva, International Organization for Migration, 49 p.

Frejka Thomas (ed.), 1996. *International migration in Central and Eastern Europe and the Commonwealth of Independent States.* Geneva and New York, UN Economic Commission for Europe and CIS (Commonwealth of Independent States). (Economic Studies, no. 9).

Grua Claudio, 1991. Dalla politica di aiuti a quella di cooperazione. Turin (Italy), Piemont Europa, 34 p.

Gunatilleke Godfrey (ed.), 1992. *The impact of migration on households: a comparative study in seven Asian countries*. Tokyo, United Nations University Press, 314 p.

Hammar Tomas (ed.), 1985. *European immigration policy. A comparative study.* Cambridge and New York, Cambridge University Press, 319 p.

IBRD, 1995. *1995 World Development Report. Workers in an integrating world*. Washington (DC), International Bank for Reconstruction and Development, 276 p. (18th Report).

IOM, 1995a. *Profiles and motives of potential migrants from Albania.* Budapest, International Organization for Migration (IOM), 18 p.

IOM, 1995b. *Trafficking and prostitution: the growing exploitation of migrant women from Central and Eastern Europe.* Geneva, International Organization for Migration (IOM), 38 p.

IOM, 2003. *World migration 2003: managing migration. Challenges and responses for people on the move.* Geneva, International Organization for Migration (IOM), 396 p.

IUSSP, 1982. *Multilingual demographic dictionary.* Liège (Belgium), International Union for the Scientific Study of Population and Ordina Éditions, 161 p.

Kritz Mary M., Lin Lean Lim and Zlotnik Hania (eds.), 1992. *International migration systems: a global approach*. Oxford, Clarendon, 354 p. (International Studies in Demography).

Lapeyronnie Didier, 1992. Les politiques locales d'intégration des immigrés en Europe, *in*: Didier Lapeyronnie (dir.), *Immigrés en Europe*, p. 5–19. Paris, La Documentation Française, 194 p.

Macura Miroslav and Coleman David A. (eds.), 1994. *International migration: regional processes and responses.* New York, United Nations, 200 p. (Economic Studies, no. 7).

OECD, 1993. *The changing course of international migration*. Paris, OECD, 263 p.

OECD, 1994. *Migration and development. New partnerships for co-operation.* Paris, Softcover Publishing and OECD, 305 p.

OECD, 1995. *Trends in international migration: Continuous Reporting System on Migration (SOPEMI) 1995.* Paris, OECD, 244 p.

OECD, 2000. *Migration policies and EU enlargement. The case of Central and Eastern Europe.* Paris, OECD, 196 p.

Ross John A. (ed.), 1982. *International encyclopaedia of population*, 2 vol. New York and London, The Free Press and Collier Macmillan, 750 p.

Segal Aaron, 1993. *An atlas of international migration.* London, Hans Zell, 233 p.

SOPEMI (Système d'observation permanente des migrations), 1995. Trends in international migration. Continuous Reporting System on Migration. Annual Report 1994. Paris, OECD, 243 p.

SOPEMI (Système d'observation permanente des migrations), 2002. *Trends in international migration. Continuous reporting system on migration: Annual Report: 2002 edition.* Paris, Organisation for Economic Co-operation and Development (OECD), 372 p.

Stalker Peter, 1994. *The work of strangers: a survey of international labour migration.* Geneva, International Labor Office, 327 p.

Tapinos Georges, 1994. Regional economic integration and the effects on employment and migration, *in:* OECD, *Migration and development. New partnership for cooperation*, p. 213–228. Paris, OECD, 305 p.

Tribalat Michèle, 1996. *De l'immigration à l'assimilation. Enquête sur les populations étrangères en France.* Paris, La Découverte/INED, 302 p.

UN-ECE, Council of Europe and UNFPA, 1994. *European Population Conference. Proceedings and recommendations, Geneva, 23–26 March*,

vol. 2. Geneva and New York, UN Economic Commission for Europe, Council of Europe and United Nations Fund for Population Activities, 429 p.
United Nations, 1991. *Demographic Yearbook 1989/Annuaire démographique 1989.* New York, United Nations, 926 p. (Forty-first issue/quarante-et-unième édition, Special topic: International Migration Statistics/Sujet spécial: Statistiques des migrations internationales). (Doc. ST/ESA/STAT/SER.R./19).
United Nations, 1995a. *International migration policies and the status of female migrants. Proceedings of the United Nations Expert Group Meeting, San Miniato, Italy, 1990.* New York, United Nations, 300 p.
United Nations, 1995b. *Population and development. Vol. 1: Programme of action adopted at the International Conference on Population and Development. Cairo, 5–3 September 1994.* New York, United Nations, Department for Economic and Social Information and Policy analysis, Population Division, 100 p.
United Nations, 1995c. *Population consensus at Cairo, Mexico City and Bucharest—An analytical comparison.* New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 165 p.
United Nations, 1995d. *Report of the World Summit for Social Development, Copenhagen, 6–12 March 1995.* New York, United Nations, 154 p. (Doc. A/Conf.166/9).
United Nations, 1995e. *Report on the International Conference on Population and Development. Cairo, 5–13 September 1994.* New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 194 p. (Doc. A/Conf.171/13/Rev.1).
United Nations, 1999. *General Assembly. Key actions for the further implementation of the Programme of Action of the International Conference on Population and Development*. New York, United Nations, 25 p. (Doc. A/S-21/5/Add.1 of 1 July 1999).
United Nations, 2000. *Replacement migration: is it a solution to declining and ageing populations?* New York, United Nations, Department of economic and social affairs, Population division, 143 p. (Final version published in 2001).
United Nations, 2002a. International Migration Report 2002. New York, United Nations, Department of Economic and Social Affairs, 323 p. (Doc.ST/ESA/SER.A/220).
United Nations, 2000b. 1990 International Convention on the Protection of the Rights of all Migrant Workers and Members of their Families, p. 60 (Doc.A/RES/45/158).
World Bank, 1995. *World development report 1995. Workers in an integrating world.* New York, Oxford University Press, 268 p.

APPENDIX Definitions Adopted by the International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (United Nations General Assembly, 69th Plenary Session, 18 December 1990)

Migrant worker refers to a person who is to be engaged, is engaged or has been engaged in a remunerated activity in a State of which he or she is not a national.

Frontier worker refers to a migrant worker who retains his or her habitual residence in a neighboring State to which he or she normally returns every day or at least once a week.

Seasonal worker refers to a migrant worker whose work by its character is dependent on seasonal conditions and is performed only during part of the year.

Seafarer, which includes a fisherman, refers to a migrant worker employed on board a vessel registered in a State of which he or she is not a national.

Worker on an offshore installation refers to a migrant worker employed on an offshore installation that is under the jurisdiction of a State of which he or she is not a national.

Itinerant worker refers to a migrant worker who, having his or her habitual residence in one State, has to travel to another State or States for short periods, owing to the nature of his or her occupation.

Project-tied worker refers to a migrant worker admitted to a State of employment for a defined period to work solely on a specific project being carried out in that State by his or her employer.

Specified-employment worker refers to a migrant worker:
(i) Who has been sent by his or her employer for a restricted and defined period of time to a State of employment to undertake a specific assignment or duty; or
(ii) Who engages for a restricted and defined period of time in work that requires professional, commercial, technical or other highly specialized skill; or
(iii) Who, upon the request of his or her employer in the State of employment, engages for a restricted and defined period of time in work whose nature is transitory or brief; and who is required to depart from the State of employment either at the expiration of his or her authorized period of stay, or earlier if he or she no longer undertakes that specific assignment or duty or engages in that work.

Self-employed worker refers to a migrant worker who is engaged in a remunerated activity otherwise than under a contract of employment and who earns his or her living through this activity normally working alone or together with members of his or her family, and to any other migrant worker recognized as self-employed by applicable legislation of the State of employment or bilateral or multilateral agreements.

Source: 1990 International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families (United Nations, 2002b, p. 59. UN Document A/RES/45/158).

CHAPTER

111

Economic and Social Management of the Population Pyramid

MICHEL LORIAUX

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

I. POPULATION POLICIES OR MANAGEMENT OF POPULATION STRUCTURES?

Our era has become the era of management in all directions: business management, personnel management, management of inheritances, natural resource management, time management, and why not, in this context, management of the population and more specifically management of the population pyramid?

It is true that the idea isn't new; a desire to influence population size and growth may be found in diverse epochs and in very different civilizations. Demographic interventionism on the part of States has often been justified by the need to combat demographic decline, real or anticipated, in acting on childbearing, marriage or migration.

On the other hand, while population policies, whether they be pronatalist in the North or antinatalist in the South, have been put into effect almost everywhere in the world beginning several decades ago, it is more uncommon for them to be explicitly oriented toward managing the population pyramid, that is, the distribution of the population by age and sex.

One could, however, think, a priori, that this distribution is more important than the absolute size of the population, in that it directly conditions the relationships among citizens, whereas the size of populations seems to come into play in a more indirect manner, and often as a function of exogenously defined objectives (e.g., to support a comparison with regard to a neighboring country, or satisfy expansionist visions) or internally defined objectives, but without marked differentiation between the different members of the community (increase collective well-being, improve the educational or health level of residents, etc.).

Between the praise of numbers or the fear of depopulation on the one hand, and the dread of numbers or apprehension about overpopulation on the other hand, history provides us with a complete and varied catalogue of attitudes regarding the pace of demographic growth.

> Demographic causality is invoked both by those whose understanding of the slowing of demographic growth of developed countries leads them, for example, to worry about the consequences of population ageing on health expenditures, and by those who attribute to demographic growth all of the planet's ills, from poverty to environmental degradation. (Tapinos, 1996).[1]

However, as the majority of Western countries seem to have settled into a situation of quasi-stationarity (with more or less no growth) and practically all of the countries described as developing have initiated their

[1] "La causalité démographique est invoquée aussi bien par ceux qui appréhendent le ralentissement de la croissance démographique des pays développés et s'inquiètent par exemple de l'incidence du vieillissement sur les dépenses de santé, que par ceux qui imputent à la croissance démographique tous les maux de la planète, de la pauvreté à la dégradation de l'environnement."

demographic transition and are taking great strides toward levels of fertility close to the replacement level, the obsession with world overpopulation has regressed and a more sustained attention is being given to age groups and to the relationships among generations.

For example, since the end of the 1990s, UNFPA reports on the state of world population have above all emphasized the "new generations," even if, both for reasons of prudence and for continuity regarding demographic policies previously extolled quite vigorously, the UN organization still greatly emphasizes the fact that "the era of rapid population growth has not yet come to an end" (UNFPA, 1998) and that it is desirable to remain vigilant.

Of course, there have always been young, adult, and elderly generations in a population, but the new twist is that "more young people than ever are entering their childbearing and working years. [. . .] At the same time, the number and proportion of people over age 65 are increasing at an unprecedented rate" (UNFPA, 1998). At least that is the case in the South; the situation that has already prevailed in the North for a long time, without having the same intensity or exactly the same characteristics, likewise poses new problems in terms of relationships among different age groups.

In sum, we are discovering (or rediscovering) that a population is not a monolithic entity that can be fully described by its size and rate of growth and that behind the placid and anonymous appearance of an age and sex structure, there are extremely complex problems of relationships among generations, and beyond, relationships between men and women who interact not only in households and families, but also in schools, enterprises, institutions, and associations.

II. DEMOGRAPHIC STRUCTURES CHANGE BECAUSE EVERYTHING CHANGES, AND VICE VERSA

But if these demographic structures were stable or relatively invariant over time, they wouldn't be at the heart of the passionate debates about the future of our societies that they presently stimulate. And the modes of social, economic, and political organization that would have been deemed satisfactory in the past wouldn't be called into question for the present or the future.

If it is otherwise, it is, of course, because population dynamics have a direct influence on population structure and any action seeking to modify the number or alter the pace of growth ends up, in reality, acting simultaneously on the composition of the population (see Chapter 4, Volume I). This is obvious for demographers, who are familiar with the interrelationships between demographic movement and population structure, even if they do not always perceive well the important implications of these interrelationships outside of the limited field of their discipline. But failure to recognize this is a big mistake for those who are ignorant of demographic analysis.

In past centuries, this condition of invariance, without being fully satisfied, was reasonably approached due to maintenance of fertility and mortality at high thresholds, which were only periodically disturbed by the random events of history (wars, epidemics, famines, etc.). In contrast, from the time when the demographic transition leading from regimes with high fertility and mortality to new regimes with low values of these parameters was set in motion in a sustainable manner, its effects on the age composition began to be felt, following a well-known process, which began with a distinct reduction in the young age groups as a consequence of the fall in birthrates, and which was followed by an increase in the importance of adult and elderly age groups, both in absolute and relative numbers, and described as population aging (see Chapters 68, 69, and 77).

The advantage, in the European countries that were the first to be affected by this process, was that the process extended over a relatively long period on the order of 150 to 200 years, and that all of the institutions destined to favor the maintenance of high fertility necessary to compensate for the ravages of very high mortality had the time to adapt progressively to the demographic change. Marriage practices, family structures, moral and religious codes, laws, types of education, inheritance rules, forms of labor organization, and systems of health and welfare were revised, often fundamentally, so that technological, economic, social, cultural, and political structures would be in more or less substantial conformity with the demographic structures.

Did the demographic change initiate the other changes, or was it the reverse? The question is relatively subordinate because, rather than a strictly causal approach which quickly gets lost in the sinuous meanderings of causes and consequences, it is preferable here to resort to a global and systemic approach. In effect, the main point is above all to understand that most of the changes that took place were not contingent, but constituted heavy currents strongly integrated among themselves. For example, we know that the progress in life expectancy was not simply the consequence of more effective medicine, but also and especially of better hygiene and training of mothers in infant welfare and child care as well as the general life conditions which were less demanding for all workers, and were only made possible thanks to progress in productivity, in the reduction of work time, and in the increasing of salaries.

In a parallel fashion, the passage from the traditional family to the modern family[2] was possible only because mortality declined, especially infant mortality, because solidarity of proximity was less solicited, and because the State and collective social institutions were able to substitute for the family to ensure schooling of children and the care of the elderly. The increase in the level of schooling as well as the improvements in health and in the standard of living of the elderly thus considerably modified the social and health structures. This resulted in turn in new progress in life expectancy at all ages and contributed, in the process, to the new and continually growing category of retirees, who pose today the problem of financing social security in a context of budgetary restrictions and reductions in the size of the potential labor force. That these changes weren't often necessarily perfectly synchronized or even entirely coherent does not call into question the principle of the existence of an overall outline that has been progressively put into place via linking of phases of order and disorder, of disorganization and reorganization.

[2] The forms of the family are very diversified and most typologies proposed to account for this diversity are incomplete or unsatisfactory. The classification adopted in the 19th century by Frédéric Le Play (1871) distinguished three models for France: the nuclear model (where children marry "outside of the parental home with an egalitarian inheritance sharing"); the model of the founding family (where only one of the children is chosen as the privileged heir and "once married, remains in the home of the parents"); and a community or patriarchal model (where "all the children who marry establish themselves in the paternal home"). But for Alain Collomp (1992), the theory developed by Le Play to support his conservative plea, according to which "the nuclear family would be a recent decomposition of the 'stable' forms of the family group, founding family and patriarchal family," is incorrect and should have been contradicted by the most recent work in historical demography that shows that

> the predominance of the egalitarian nuclear model, this modern form of the family, has been solidly implanted in Great Britain, in the Netherlands and in northern and western France since at least the end of the Middle Ages. On the other hand, under the effect of economic and demographic difficulties, the family group was able to evolve from the simple, nuclear structure to more complex forms, founding family, "*frérêches*" (cohabitation of multiple couples of the same generation) or big communities. This phenomenon appeared in certain provinces in the centre and south of France in the 15th and 16th centuries. In Ireland, the founding family only developed from nuclear structures in the second half of the 19th century, following the subsistence crises and the massive emigration of the young to Great Britain and the United States. (p. 16)

However, these important observations on the periods prior to the onset of the industrial and demographic revolutions do not constitute, in themselves, a denial of the idea that the passage from a traditional society to a so-called modern society has greatly upset the majority of societal structures, including family structures, at the same time with respect to the type of regulation of the relationships among generations and regarding the size of families. What is important to note here is that from the time that the strong constraints of high mortality that weighed on the family were attenuated and economic growth took off, the conditions emerged to encourage the establishment of public welfare systems (solidarity) among generations as a substitute for (or complementary with) the private solidarity that characterized rural societies.

III. TO ACT ON THE AGE AND SEX COMPOSITION DESPITE THE INERTIA AND RESISTANCE?

To recall these elementary principles of the dynamics of population change is useful for seeing not only that the age and sex composition of the population are at the heart of societal change, but also that these population structures are part of a complex, interdependent network. It is difficult, even dangerous, to try to intervene in these processes, under the different pretexts of ethics, justice, economic efficiency, political manipulations, and so forth, because societal systems, governed by a principle of homeostasis,[3] strongly

[3] The concept of homeostasis was introduced in 1932 by the American physiologist Walter B. Cannon, and it constitutes one of the cornerstones of the systemic approach. It literally means "remain constant," and designates a property of complex systems (living cell, human body, business, army, nation, etc.) to maintain their structure and functions by a series of regulatory mechanisms that seek to assure a dynamic equilibrium. When one of these systems is subject to disturbing influences, it reacts by corrective actions that seek to maintain its general internal equilibrium. If it does not succeed in doing this, it is obliged to change and to adopt a new way of functioning, or worse, to disappear. "Homeostasis in this way appears as an essential condition of stability and thus of survival of complex systems" (de Rosnay, 1975). A biological example is that of the human body, which, when it is attacked by a microbe or a virus, initiates diverse defense mechanisms such as a fever or certain biochemical reactions. If the response is sufficient to counter the attacker, the result is healing; otherwise, it is the destruction of the system, death. In the social arena, when the existence of a revolution is invoked, it is in principle the indication that the society in question has not successfully overcome the internal and/or external aggressions nor has it been able to maintain its previous social equilibrium. The case of the French Revolution is illustrative of such a situation, where the monarchy and the regime of social classes that supported it did not succeed via their instruments of control and repression in containing the rise of new ideas about liberty and equality and in limiting popular discontentment. Many policies (economic, social, or demographic) can be analyzed in the same way; generally, they only have a serious chance of success if they are implemented in a favorable context, without clashing with any strong resistance from the population or from influence groups. A typical example of homeostatic resistance is that which opposes

(continued)

resist these stimulations or aggressions and tend to maintain their equilibria by activating all of the regulatory mechanisms available. But these mechanisms are all the more numerous and complex as the systems that they regulate are themselves complex. Anyone who seeks to act on populations or their structures should contemplate these principles of the systemic approach, because they will often discover the reasons for the numerous failures of actions taken in this domain. These failures are generally perceived a posteriori as not understandable or foreseeable, but this is probably the case less than is supposed, so as to justify them or absolve those responsible for the actions.

In reality, what is true about demographic structures is also true about other structures (social, economic, technological, etc.), but it is reasonable to think that the demographic structures have some specificities such that they are probably more resistant to voluntary manipulations. The inertia of demographic phenomena is certainly there for a reason, even if it is noticeably weaker in periods of rapid change like our own than was the case in preceding periods, when changes were much slower.

The fundamental modifications of demographic variables and structures are operative in the medium or long terms, whereas short-term economic variations, while sometimes brutal, do not generally entail lasting consequences. These fundamental changes are often only visible if the behaviors of two successive generations are compared, that is, at intervals of around 25 to 30 years. Of course, sometimes, via the effect of natural chances or policy actions, we see events or circumstances that can have immediate and profound impacts on the population. For example, this may occur with a massive deportation of residents, a sharp diffusion of a deadly epidemic, the implementation of a highly coercive antinatalist policy, or a forced international migration resulting from an ecological catastrophe. In these cases, not only may the total numbers and the spatial distribution of the population be strongly affected in the short term, but also the gender composition (case of China's one-child policy) and the age composition (migration of workers, ethnic extermination of women and children to prevent reproduction of a group, etc.). Even if it is the outcome of deliberate strategies, the elimination or purification of a class, an age group, an ethnic group, or race does not constitute, in a democratic context, an acceptable policy for managing the population and its distribution by sex and age. We thus limit ourselves here to means and instruments more classically respectful of the founding principles of democratic societies.

The idea of taking action to influence the age and sex composition of a population in order to transform them in a sense deemed to be desirable a priori can, nevertheless, seem seductive. But such action immediately runs into numerous difficulties regarding choices of objectives; of institutional, policy, financial and other means of achieving them; of assessing their effectiveness; and, finally, of evaluating their moral legitimacy and their acceptability in a democratic context.

Different questions must be addressed or evoked. Is it possible to implement policies that seek to manage the population pyramid? What may their objectives be? Can they be effective? Are they scientifically based? Are they morally ethical and legitimate? It is preferable to deal with such questions in as concrete a manner as possible, based on reflection concerning a number of societal aspects.

IV. AGE POLICY: WATCH OUT FOR UNEXPECTED EFFECTS!

The intention to influence the evolution of the age and sex composition of the population assumes that the composition has been deemed to be inadequate or that it may evolve in the long term in an undesirable way. This implies right from the start that what is adequate or desirable is known, and implicitly that reference is made to a sort of theoretical norm of an ideal population, in the same way that Adolphe Quetelet (1869) considered the average man as a divine ideal from which human beings would deviate only due to the capriciousness of nature or the random effects of reproduction. But the existence of such an ideal is eminently debatable, for populations as much, if not more, than for individuals.

It is not a hindrance that a lot of demographers or politicians and experts continue to refer to this notion, explicitly or implicitly. The ideal most often evoked is that of a stationary population or, sometimes, one that is growing slightly, with an age structure such that the

demographic policies, pro- or antinatalist, when they are contrary to dominant cultural values and norms, often reducing their effectiveness to very little, or even nothing. There have been violent illustrations of this in developing countries (India, notably) when family planning programs were forced on the population (Verrière, 1978). In the other direction, in the European countries where government authorities and international agencies generally consider fertility to be too low, most efforts to stimulate the fertility of couples have turned out to be unfruitful (birth premiums, family allowances, parental leaves, tax deductions, etc.) because they did not offer a sufficient counterpart to stem the propensity of parents to limit their childbearing to less than two children because of their high cost, and also the competition from other opportunities (leisure, women's work, self-fulfillment, etc.).

burden of persons described as inactive (the elderly and the young) does not weigh too heavily on those in the labor force and that the products of economic growth (the *surplus* or the *rent*) can be devoted to productive investments, rather than to demographic investments that just allow maintenance of the same standard of living to additional individuals or to the too many who are inactive. From this perspective, a shrinking or aging population is considered to be a catastrophe that must, if not be avoided at all costs, at least be limited as much as possible, by actions seeking to repopulate the "empty cradles"[4] or to replenish the age pyramid in its central portion by the massive contributions of foreign populations of labor force age that can fill the holes or the depressions brought about by previous changes.

Such policies had their moment of glory and their fervent defenders, and they have sometimes had a certain efficaciousness. This is especially clear regarding controlled immigration, where the massive transfers of foreign labor have on occasion permitted sustaining economic activity in strategic sectors such as coal mines or the iron and steel industry, deserted by native populations in search of better-paying or less difficult work. Unfortunately, what pays off during several years or decades is not necessarily going to be profitable in the long term. And the problems of integration of migrant populations, which most European countries are confronting, are often the delayed consequences of migration policies and procedures put into effect previously (for economic, demographic and sometimes political reasons, such as welcoming those from former colonies). Today, those who were seduced by the sirens of prosperity and who thought they would be able to avoid their miserable conditions sometimes reproach their countries of origin for having sold their health or their cultural tradition "for a sack of coal" (Franciosi *et al.*, 1996). And their descendants live their dual national affiliation with difficulty, feeling rejected at the same time by both their countries of origin and adoption; they constitute, according to the media, the hotbed of urban violence and of suburban insecurity.

In addition to these delayed perverse effects, it can even be asked if the solutions to the sectoral shortages of labor that were provided by this migration, with apparent success initially, didn't in reality contribute to delaying clean-up and restructuring measures for certain sectors that subsequently rather rapidly showed themselves to be in a deficit position or insufficiently competitive.

These questions are even more pressing with regard to the measures taken to counter the fall of fertility and aging, to the extent that they often have seemed adapted to the current demographic context and to the perspectives of emphasizing in the future observed past trends, but without taking account of the changes that were under way in other social sectors. In other words, too many experts and demographers,[5] obsessed by the risks of weakening of all kinds (of military might, of productive power, of the capacity to colonize, of evangelization, etc.) that the decline in fertility and the aging had previously prompted, continued to talk about the future, all other things being equal, in amplifying the consequences linked to the following or even the acceleration of these trends, without taking into account the fact, rather evident, that if the demographic structures were changing, it was probably because the other basic structures of the society (technological, economic, social, cultural) were also changing profoundly.[6]

[4] The expression "empty cradle" was frequently used during the 20th century to criticize a level of fertility deemed to be too low. There is even a theatrical play from Quebec written right after World War II with this title (Vekeman, 1945). Moreover, it is also in Quebec, in the 1960s, that the expression "revenge of the cradles" appeared, curiously, as Jacques Dupâquier (2000) writes: "At the same time when this was ending, the Francophone minority being abruptly less fertile than the Anglophone majority." In France, it is also common to refer to "the empty cradles of Marianne" (Dûpaquier and Biraben, 1981).

[5] The notoriety of Alfred Sauvy undoubtedly explains why he is frequently cited as one of the principal individuals contemptuous of population aging, but in fact he belongs to a diverse line of French scientists and politicians which began before him and which still continues presently, even if we sometimes see certain changes in who is setting forth the argument: Pierre Chaunu and Jean Legrand (1979), Jean-Claude Chesnais (1995), Evelyne Sullerot (2000), Gérard-François Dumont (2002), Michel Godet (2003). In this regard, see the works of Patrice Bourdelais (1993) and Paul-André Rosental (2002), who have written about the history of French demography and especially of pronatalist sentiments.

[6] The following observation is at the base of this theory: a complex system can only assure its durability, its stability over time, if its principal elements are themselves in relative harmony or complementarity. When an important element undergoes a significant change, the set of interdependencies among the elements means that the other elements must adapt in turn, under the penalty of creating tensions and compromising the good functioning of the system. In our Western societies it is often technological innovations that initiate the processes of transformation of the other structures. For example, the mechanization of agriculture was responsible for an increase in productivity and a reduction in prices that forced the closure of less profitable enterprises and resulted in rural out-migration, with a long-term consequence being a total remaking of the social landscape and of the societal system. Closer to us, when the religious fundamentalists took power in a country with a secular government, this resulted in a change in the status of women and in relationships between the sexes, and also of lifestyles and cultural behaviors (rejection of Westernization, etc.).

V. THE BABY BOOM: GOOD LUCK OR BAD LUCK OF HISTORY?

Without a broad convergence of the great societal currents, the risks of destabilization of societies due to the demographic changes evoked above would be real, but a greater fear is that the demographic conditions will not change, rather than the contrary, when revolutions are at work in other sectors of societal activity. What these demographers have not perceived or understood is the fact that Western demographic aging was no longer simply the consequence of the reduction in fertility, that this first cause was increasingly being supported by a second cause linked to the prolongation of life, and that classical aging, from the bottom of the population pyramid, was being accentuated by aging from the top. But they also didn't understand that the quantity of labor was no longer the decisive criterion for economic development and that its quality was much more important, due to technological progress and the complete revision of the production function combining capital and labor. To produce more children in this perspective results in being exposed later to a swelling of the reserve army of the unemployed. All things considered, it may even be said that couples were more realistic and farsighted, in reducing their fertility, than the scientists who were advising politicians, often moreover condemning the selfishness and the lack of civic sense of their fellow citizens.

One can even ask, with the passage of time, if the rise in fertility after World War II, this "baby boom" that was welcomed by so many observers as an unexpected opportunity to make up for the damages caused previously by the secular trend toward declining fertility, wasn't in reality an unfortunate accident of history that would have been better to try and check rather than to encourage.[7] Indeed, the large number of births that persisted for 20 years, until 1965, and which seemed to compensate for the births lost in the interwar period, was a destabilizing factor of demographic structures, because after the baby boom came a baby bust, and these fluctuations in the size of successive cohorts will have economic and social consequences until at least halfway through the 21st century. Current and prospective problems of financing social security are in large part linked to the fact that the baby boomers, who presently constitute a large share of the labor force, will reach retirement ages when the small cohorts born after 1965 will themselves be in the labor force, accentuating the dependency ratios of those not in the labor force relative to those who are, and risking creation of strong tensions between those working and retirees.

If we were to realize a renewed growth of childbearing, as some call for (Chesnais, 1995; Godet, 2003), we would then run the risk of repeating the cycle of successive peaks and troughs impacting different sectors, beginning with schooling and going all the way to retirement, and including along the way marriage and occupational activity. The presence of more numerous youths would first necessitate greater efforts in the areas of infant and child care and schooling, and this would be followed by growth in the supply of labor, which, however, will be satisfied only if there is a corresponding growth in demand. For a while, the dependency ratio would improve and the tension between taxpayers and beneficiaries of social security payments would be reduced, but the opposite would occur without fail afterward if fertility did not remain indefinitely at its new higher level.

A certain degree of cyclical behavior of the phenomena responsible for demographic change is undoubtedly difficult to avoid. This is, for example, suggested by Richard Easterlin (1975). But to encourage it by a policy supporting fertility that does not take into consideration the long-term consequences of the expected immediate results would be a very inconsistent position.

[7] This thesis contradicts the most current theory, alluded to by many authors (Godet and Monti, 2000), Alfred Sauvy first among them, who think that there is a positive relationship between demographic growth and economic growth. The proof for them is given most notably by the fact that the baby boom apparently stimulated household consumption as a consequence of the growth in the number of children, and it isn't simply by chance that this period corresponds for the most part to the *Trente glorieuses* (Thirty Glorious Years) of Jean Fourastié (1979), whereas the drop in childbearing and the demographic aging that have followed have been accompanied by a recessionary phase whose effects are still being felt. As Alain Minc (1987) has written, in a phrase as brutal as it is short, "Expansion and youth always go together, as do recession and ageing." But in reality, things are probably a lot more complex, and there may not be any mechanical relationship between population growth and economic growth. For Jean-Claude Chesnais (1987), who does not at all conceal his pronatalist leanings, to reduce the question of development to a competition between two growth rates constitutes an abusive simplification. And even if we acknowledge that, in certain contexts, demographic pressure can constitute a factor speeding up technological progress and changes in political and social structures (the thesis of Boserupian adaptation), nothing proves that it is a necessary and sufficient condition. In the absence of which, one would have to accept that all of humanity is condemned to live eternally in recession and in crisis, to the extent that it has entered a long-lasting period of slowing of its demographic growth and even, in the long term, of reduction in its absolute numbers. At present, nothing proves that technological progress is slowing down, even in those regions of the world where the demographic decline is already observable or potentially under way.

VI. A POPULATION ISN'T A BUSINESS

In speaking of the management of population structures, it is tempting to make the analogy with management of business personnel. In both cases there are populations, even if one is a macro type while the other is micro, for which they have their own dynamics depending on the flows of entries and exits, and to which the principles of demographic analysis can be applied.

However, the comparison runs up against some major differences.[8] In a business organization, the new entrants are based on hiring, which is closely linked to the firm's policy as a function of its business activity and prospective profits. Exits, apart from a few cases of death among the employees, are essentially dependent on retirement and termination. Those in charge have numerous possibilities available to them for influencing the volume of employment, and the decisions made are most often directly operational. The firm can also makes its work force more youthful, by hiring younger workers and encouraging older workers to take their anticipated retirement, which can effectively be a means of reducing costs or following technological change. On the contrary, the firm may wish to avoid too rapid a reduction in the age of the work force following a large expansion, so as not to lose the know-how and the institutional memory of the enterprise. In brief, targeted policies can be used to address clearly defined objectives, because there are well-identified means to put them into action.

In contrast, at the macro level, for the population of a country or a region, the objectives are hardly ever clearly defined, and the means are at once poorly identified, of uncertain effectiveness, and often ambiguous with regard to their congruence with fuzzy objectives. For example, to not want to act against mortality seems immoral and humanly intolerable, even if official policy was to slow population aging and it is well known that progress in life expectancy at older ages accelerates the aging process. In the same manner, encouraging foreign immigration so as to make the age structure younger can conflict with the imperatives of an employment policy during a period of economic crisis and high unemployment. Or again, a lot of pronatalist policies, not daring to describe themselves as such for fear of shocking public opinion or of giving the impression of limiting the range of choices offered to couples, have failed, dissolving into a confused mix of measures seeking to influence behaviors, measures that were neither coherent nor capable of modifying reproductive behaviors.[9]

It is clear that at the outset demographic change and family policy were strongly associated and that family policies adopted by countries such as France had explicit demographic objectives. But the opportunity of pursuing this coupling today raises great doubts.

Didier Blanchet (2000) immediately eliminates the pronatalist scenario, deemed to be too "offensive and unrealistic," considering that "ageing from above, due to the prolongation of life [. . .] will subsist, even with fertility maintained at the replacement threshold," and that only a quite unlikely return to fertility levels comparable to those reached during the baby boom, on the order of 2.7 to 3 children per couple, would allow a stabilization of the dependency ratio (out of the labor force/in the labor force) at a level close to that reached in 2000, which

> would imply permanent demographic growth on the order of 1 to 2%, representing a doubling of the population every 50 years. It is on one hand an objective that is almost surely beyond the reach of the instruments of family policy, and on the other hand even if it can be envisaged, it hardly seems desirable. At best it would be desirable temporarily: it is

[8] Despite the similarity, it would not be realistic to claim to be able to manage the characteristics of a global population as can be envisioned doing for a very specific population, such as that of a firm or an occupational group. The issue would not be limited to simply borrowing of demographic techniques in order to attain human resources management objectives, according to a practice that has become common in large organizations: multinational firms, armies, and so on, as Régine Monti (2003) shows in her doctoral thesis on *La gestion prospective des âges des populations larges.* The dreaded and denounced danger would be that the sorcerer's apprentices of population policies are inspired by these experiences to export them in a context of global populations, believing to be able to demonstrate high operational efficiency.

[9] In many European countries, the measures adopted to stimulate births were limited to financial incentives (diverse payments, tax advantages, etc.), which did not necessarily produce the desired effect, probably because they were insufficient to bring about a radical change in the fertility behaviors of couples. In some of these countries debates even took place concerning granting a salary to stay-at-home mothers, a measure which would have undoubtedly resulted in greater success but which has never been put into effect, to our knowledge, for lack of adequate means to make the project operational. In addition, the policy encountered fierce opposition from feminist groups, who saw it as a way to reinforce sex role differentiation and to keep women in domestic and reproductive functions. Conversely, the Scandinavian countries have often shown greater imagination, with emphasis on measures designed to more easily reconcile the occupational and family roles of women (day care centers, parental leaves, etc.) and to better share the distribution of domestic and educational tasks between the partners (Monnier and Prioux, 2000).

indeed necessary, one of these days, to adjust to the prolongation of life other than by a risky acceleration of demographic growth. (p. 892)[10]

VII. GOOD OR BAD STRUCTURES: SOME PROPOSITIONS

Can we, nonetheless, identify some general principles that can guide action relating to the management of age structures, and make some propositions?[11]

There are not, in and of themselves, good or bad population age and sex structures. There are only populations that are well or poorly adapted to the societal contexts with which they are associated, as well as to the social ends pursued by the community or the nation.

Inertia, stupidity, or the apparently unavoidable character of certain demographic trends should not lead us to deal separately with population phenomena, without taking into account other societal phenomena. These other phenomena are at least as important as demographic phenomena, if not more so, but their changes are much harder to predict, and these changes can upset the relationships of the demographic system with the other basic systems of society (economic, sociological, ecological, etc.).

To fix population policy objectives with reference to intrinsic thresholds (threshold of zero growth, replacement level, dependency ratio of inactives to actives, etc.) stems from a dangerous belief that can focus all worries around parameters whose social significance is only relative and whose interest for the future of the community is limited and conditional.

The idea, apparently tempting, to limit the adverse consequences of temporal variations of the principal demographic variables (fertility, mortality, migration) should be received with prudence. Imagined corrective measures, generally uncertain, cannot only fail but they may also cause a delay in relation to the phenomena to be corrected, or they may cause an excess in the degree of correction envisaged, and thus result in bigger problems than those they were meant to resolve.

To the extent that demographic phenomena have a certain degree of inertia and rather important reaction lags, a population policy (including management of the population pyramid) must be designed from a medium- or long-term perspective. But political decision makers often have only a short-term vision of the domains that they must manage, so it is difficult to imagine that they would be able to give priority to long-term societal projects when those projects are, in addition, in conflict with their immediate electoral interests.

Even if it cannot be denied that the age and sex composition of a population can have a considerable impact on major macro-economic equilibria, and beyond, on the way of life of citizens and on the entire functioning of a society, it is also quite evident that scientific knowledge concerning the relationships among these phenomena is still largely in an embryonic stage. Hence, this knowledge alone can only rarely legitimate interventions to these population structures and a fortiori serve as the specific basis for active policies seeking to manage the population pyramid. Thus, the orientations of policies and the choice of measures to adopt come back at least as much to normative and ideological options as to scientific knowledge, and it is imperative first of all to clarify things for citizens by clearly indicating the real reasons that underlie proposed actions. No more than other policies, demographic policies or strategies for managing the age pyramid should not undermine the principles of self-determination and freedom of choice, and it is useful to avoid their resulting in limiting the exercising of certain fundamental rights (to education, to work, to health, to family reunification, etc.) or in creating growing disparities between social classes or generations.[12]

[10] "le vieillissement par le haut, dû à l'allongement de la vie [. . .] subsistera, même avec une fécondité maintenue au seuil de remplacement des générations" et que seul un retour tout à fait improbable à des niveaux de fécondité comparables à ceux atteints durant le baby boom, de l'ordre de 2,7 à 3 enfants par couple, permettrait une stabilisation du rapport de dépendance inactifs/actifs à un niveau proche de celui atteint en 2000, ce qui "impliquerait une croissance démographique permanente de l'ordre de 1 à 2% soit un doublement de la population tous les cinquante ans. C'est d'une part un objectif dont on est quasiment sûr qu'il est hors d'atteinte des instruments de la politique familiale et qui, d'autre part, quand bien même il serait envisageable, n'apparaît guère souhaitable. Il le serait au plus à titre transitoire: il faudra bien, un jour ou l'autre, s'adapter à l'allongement de la durée de vie autrement que par la fuite en avant dans la croissance démographique."

[11] Some of these have been taken, in whole or in part, from a paper presented in Lille (France) at the Sixth National Demography Colloquium (Loriaux, 1986).

[12] The big international conferences that marked the 50th anniversary of the United Nations produced reports that are types of universal charters often proposed to member states as sources of inspiration for putting into place action programs or policies. This is the case for the programmed of the International Conference on Population and Development (ICPD) held in Cairo in 1994, which includes many noble declarations concerning individual liberties, but which is astonishingly silent when it comes to proposing concrete measures in certain sensitive areas, for example, with regard to the right to family reunification for (legal!) immigrants or assistance with transfers of medical technology sought by all the countries of the South but rejected by most of the Western nations (with confirmation at later meetings of the United Nations Economic and Social Commission at New York) (Lassonde, 1996).

VIII. A NECESSARY REVERSAL OF THE PROBLEM

In view of all this, must we give up all hope of interventionism with regard to management of the age and sex composition of the population, and instead prefer a laissez-faire approach? This would leave us at the mercy of natural evolutions, which would at least have the merit of integrating a complexity that our models, mental or mathematical, don't seem to be able to provide. This way would at least have the advantage of avoiding the uncontrolled changes linked to the questionable experiments of sorcerers' apprentices in the management of human resources. But human reason acknowledges, only with difficulty, that leaving these evolutions to chance (or to something that resembles chance) may be an optimal solution. Rather than laissez-faire, wouldn't it be better to reverse the problem and acknowledge that it is not the population structures that one must try to change temporarily to align them with other societal structures, but rather these other structures and our types of organization that it would be better to put into harmony with the population structures?

Given that the age and sex compositions resist limited stimuli rather easily, while their profound changes are practically impossible to counter, it is therefore probably much simpler to transform our institutions, our laws, our rules, our moral codes, and even our cultural models to avoid their coming into conflict with our demographic structures. At least we have the means to do this via the diversity of instruments at the disposition of governments and public authorities. It would suffice if they would show the political will to use them and show that they have the imagination and innovativeness to find solutions adapted to new situations.

The population aging of industrial countries offers us some rather remarkable examples of this. In the absence of being able to prevent the elderly from living longer and longer or to find the magic wand that will make the numbers of young workers increase, that is, to find how to make the collective aging recede on all fronts, can't we find modes of organization that reduce the tensions between generations or that attenuate the adverse consequences of disparities between the sexes? Such modes of organization can be imagined; however, acknowledging them often runs into strong political, economic, or cultural resistance, so that it always seems preferable to try, often in vain, to modify at the margin the former modes of organization that previously demonstrated their worth but which, generally, are no longer at all adapted to the current societal contexts, rather than reforming them radically. It is possible that the democratic mode of functioning makes radical reforms difficult, if not impossible, due to the absence of a sufficiently broad political consensus. If this is the case, however, we will have to wait until our societies become virtually unmanageable in the face of the numerous challenges that will soon be imposed on them by the increasingly profound and rapid changes in the societal environment.

IX. THE FINANCING OF SOCIAL SECURITY: A FALSE DEBATE?

Many demographers (Guillemard, 1986; Euzéby and Euzéby, 1996; Bonnici, 1997), acknowledge that the introduction of social security systems around the second half of the 20th century constituted a great social advance, in allowing society to cover, as never before, two of the greatest risks to which workers were exposed: illness (or accident) and old age. It is, to a great degree, the combination of strong demographic growth (natural increase or via migration) and strong economic growth that created the opportunity without which the great social reforms after World War II would hardly have been conceivable. It is also, in large part, this combination that oriented, in many countries, the choices regarding retirement toward a pay-as-you-go (redistributive) system rather than a system with individual private accounts (system of capitalization). The large number of workers compared to the number of elderly people out of the labor force allowed pensions to be provided to those who had contributed no or very few payments, without compromising the equilibrium of the system. However, several decades later conditions had changed substantially: risks have grown sharply, principally the risk of old age, which has become a near certainty, while receipts have tended to regress in the face of continually expanding expenditures. Is this to say, as is often repeated in the media, that social security systems have become explosive and even, according to some (Minc, 1987), that the fuse of the bomb has already been lit? This is only really the case if nothing else is envisaged to help the systems evolve than to reduce payments, ration health expenditures, put a ceiling on pension amounts, or increase taxes on salaries.

One of the remedies most often recommended, including by international organizations like OECD (1998), is, for example, to increase the legal age of retirement, "which would have as a double consequence to increase the number of those in the labour force while at the same time reducing the number of retirees, thereby contributing to alleviating the financial pressure on the retirement system" (Tapinos, 1996).[13] Increasing the retirement age is certainly not an inconceivable scenario, in view of the lengthening of the average life, but the arguments supporting such measures are nearly always exclusively financial, without any consideration of the counterarguments, by nature more sociological or political (Gaullier, 1988).[14] In addition, the defenders of such policies often lose sight of the fact that such actions, even if they can be easily adopted by administrative order, risk encountering fierce opposition, not only from workers and their unions, but also from employers. Moreover, from a technical point of view, our societies, which are undermined by unemployment, seem to suffer more today from an excess of labor than a shortage, and to claim to take care of the problem by an artificial extension of salaried employment reflects more a mechanistic vision of social relationships than a systemic approach to the complexities of the labor market.

This type of proposition seems more inspired by a projection into the future of past situations and by a recourse to old recipes than by an expanded analysis of current and future trends of demographic structures. Such an analysis could inspire bold prospective scenarios regarding the economic and social roles of the elderly and concerning new relationships between generations, beyond the questions, most certainly important but nonetheless fairly small, of financing of pension funds. The day when the economy will cease to be an end in itself to which the society seems enslaved in all of its other dimensions, to become again what it should have never stopped being, namely, an instrument at the service of the realization of the collective objectives and of the satisfaction of the essential needs of the citizens, all of these questions, today apparently without answers, will perhaps find solutions for which we will ask why they didn't appear sooner.[15]

X. INTERGENERATIONAL EQUITY: A CLEVER DIVERSION FOR OTHER LESS AVOWABLE ENDS?

I am sometimes astonished to observe to what degree currents of thought, considered by many as innovative, seek to call into question earlier social advances in the name of economic efficiency, of political realism, and even of justice or fairness. In particular, the concept of equity among generations has become practically unavoidable, and it is hardly possible anymore to address the question of social policies without evoking it. A priori the idea can seem seductive, because it's about trying to ensure an equitable sharing of resources among generations, so that none of them takes any substantial advantages to the detriment of others. But demographic aging has given this theory a remarkable field of application, in that the number of elderly is increasing while the number of youth is diminishing, at least relatively, and the elderly are often considered not only as a burden to the society that devotes growing resources to them that could be used by other age groups, but also as an obstacle to development, as expenditures in their favor are ordinarily counted as consumption expenditures rather than investment.

Although most authors (Wisensale, 1997; Masson, 1999, 2000) acknowledge that the very notion of equity is difficult to define and even more difficult to render operational, it seems necessary to refer to it to be politically correct, even if the most immediately visible consequence is to exacerbate tensions among generations and to make more difficult the management of multigenerational societies. Preserving justice among generations thus risks creating the conditions for war among generations, whereas certain lucid authors haven't failed to notice that the discussion of intergenerational equity had been for many "a clever diversion" (Wisensale, 1997) to limit the intervention of governmental authorities in social programs and favor reductions in taxes, in particular for the benefit of the rich, in modifying the public image of elderly persons.

From there, in fact, it is no longer unreasonable nor unjust to ask if health care shouldn't be rationed as a function of age, dismantling the systems of protection and privatizing social security, and even, based on principles inspired by the works of a philosopher like John Rawls (1971, 1993), to restrain or suppress the

[13] "ce qui aurait pour double conséquence d'accroître le nombre des actifs tout en réduisant celui des retraités et contribuerait ainsi à relâcher la pression financière sur le système de retraites"

[14] It is relatively easy to acknowledge that if workers remain in their jobs longer there will be additional social security tax payments made and a reduction in the amount of pension payments (more receipts and fewer expenses). On the other hand, frequently neglected, for example, is the fact that if such measures extending the working life are taken (and accepted), they will also have as an indirect consequence delaying when retirees can get involved in diverse benevolent activities or second careers (Gaullier, 1988) and be able to offer their services on the market of the social economy.

[15] Without entering here into a discussion that goes beyond the focus of this chapter, it should be acknowledged that numerous ideas and propositions for restructuring our modes of functioning and of collective organization remain confined to small intellectual circles and have difficulty penetrating political milieus where they could be the subject of expanded democratic discussions. This is, for example, the case for alternative modes of financing social protection, such as universal allocation (van Parijs, 1990; Ferry, 1995) or taxation of capital and/or value added, but also for boldly innovative means of struggling against the early exclusion of older workers from the labor market thanks to the introduction of "multiple activity" contracts (Gaullier, 1999).

right to vote of the elderly (van Parijs, 1999). Have we thus arrived at this incredible paradox where the industrial societies, which are the only ones to have succeeded in mass-producing elderly people, the same as cars, televisions, or cell phones, are now trying to find means to get rid of them surreptitiously, as they would with toxic or nuclear waste? In addition, in the name of equity, an old humanist notion has been transformed for the occasion into a redoubtable ideological arm for confrontation of different age groups, most notably children, parents, and even grandparents.

The debate has, moreover, for several years taken a supplementary dimension with the rise of ecological movements, because the threats to the environment and the risks of running out of natural resources have expanded the search for equity from current generations to future generations, which of course does not simplify the problem. In reality, when it is defined in such a summary fashion, this notion of equity, which is only based on very incomplete criteria, leads us to treat a complex problem in a simplistic manner, with the risk of accentuating the injustices that we claim to be combating in referring to it.

Without a doubt, the efforts made to put in place systems of intergenerational accounting in order to better discern the flows of resources among generations should be encouraged, in particular to demonstrate the bilateral nature of the transfers. But this should be done on the condition that hasty conclusions not be drawn that go beyond the potential of such tools and that would leave obscure some important issues that are extremely difficult to resolve: What value should be given to economic and social investments made by earlier generations? Should we consider the expenditures made for elderly persons as an income from their investment for subsequent generations? Should we consider a part of the sums spent on children as an investment in the future productivity of society? How can we decide on the distribution of investments in research, preservation or protection of the environment and defense among the different age groups? (Wisensale, 1997).[16]

XI. A MORE POSITIVE APPROACH: SOLIDARITY OF THE AGES OF LIFE

With equity being a vague and suspect concept, more divisive than unifying, it seems more interesting to promote another concept, even if it is also relatively fuzzy and carries a heightened ideological charge: *solidarity*. This has the enormous advantage of favoring a positive vision of the social structure and searching for integrating formulas taking better account of the new stakes created by demographic and economic change.[17]

It is, indeed, necessary to acknowledge that aging societies present certain fragile aspects, certainly linked in part to the existence of elderly dependent subpopulations, but also and especially to the fact that the transformations of age and sex structures take place parallel to unprecedented changes in the global economy and make existing modes of organization obsolete. However, to take advantage of these fragile aspects to dismantle the edifice would be suicidal (or criminal) if the new principles founding the reconstruction don't take account of the requirements stemming from the new age structure. Because aging has strengthened the need for linkages among the generations, which were, paradoxically, less important[18] in the former demographic regimes (when generations followed one another practically in a chain, with limited periods of cohabitation), but which have become all the more necessary currently, as the progress of longevity has resulted in a strengthened coexistence in the same times and places of multiple

[16] For more on questions raised by generational accounting and its principles, see Laurence Kotlikoff (1992).

[17] This point of view is fully shared by one of the authors who has greatly contributed to the debate about equity and generational solidarity in France, André Masson (2000):

> [the concept of generational equity] suggests too much that each generation, working only for itself, must "evaluate its accounts" (assets and liabilities) with respect to the other generations if it is possible to do so. We thus lose sight of the unavoidable and fertile character of the generational link assured by the family and by the society, as "systems of co-operation in time among generations" (Rawls, 1993). [. . .] Recourse to solidarity is already justified by the disappointments of the concept of intergenerational equity: the difficulty of establishing a precise accounting for each generation, the inadequacy of the altruism invoked by "libertarians," the omission of specific relationships among family generations. [. . .] In breaking away from the narrow views of equity, generational solidarity allows for generating, within the family as in society, a mutually advantageous cooperation among generations against the inadequacies of markets. (p. 356, 362)

[18] This fact merits emphasis because traditional societies are often presented as being characterized by solidarity, which was probably the case, while contemporary societies are more individualistic, with forms of bureaucratic solidarity, anonymous, and impersonal. However, former solidarities were more justified by the regime of high mortality that prevailed, as well as by the insecurities linked to the physical and human environments. Conversely, in our developed societies, the idea often prevails that basic solidarities are less necessary because of the elevated standard of living and the disappearance or attenuation of certain risks, whereas it is precisely the opposite and the impression of individual freedom of choice is only possible because many forms of institutional solidarity exist, often unknown to us, under multiple expressions of assistance or assurance and diverse mechanisms of redistribution of collective resources.

generations. In other words, the prolongation of life has resulted in growth in the periods of life lived in common among the different age groups and has assured the contemporaneousness of generations all the more heterogeneous vis-à-vis one another as the acceleration of history has given them extraordinarily varied event trajectories.

The principal challenge already being confronted by aging societies is certainly that: to assure a stimulating cohabitation among all generations, young, adult, and elderly, each carrying their own social and cultural peculiarities, their hopes, and their contradictions. That is why future solidarity can only be global and why it must be considered in terms of the totality of the group and not in its separate fragments, or worse, in terms of individuals, who never constituted a group simply by being juxtaposed. That is also why solidarity must turn away from forms derived from assistance or assurance in which it tends to be comprised today (France, 1996) and find again the more organic forms linked to the acknowledgment of the existence of a community in which each member has a useful role to fill and an honorable status to occupy.

"Youth, adults, and elderly, healthy and ill, active and inactive, same fight!" we could say, but also "same founding principle of solidarity!." Nothing is won in advance. The contradictions are often fierce, the interests divergent, and the resistance to change is strong. But success cannot result from negation of the difficulties, nor from refusal to search for original solutions. Probably nothing will be solved via prudent policies and timid compromises, nor from measures without a long-term plan or an innovative spirit. The intelligence of the "gray revolution" (Loriaux *et al.*, 1990) will be to make all generations understand that the future depends on the acknowledgment of others, and that in a war of generations, gaining victory over other age groups would be devastating stupid, resulting in each group losing more than it gains. If it is not reason that will help us find the path to natural solidarities, it will be necessity.

As Henri de France (1996) has written,

> when the social and economic context with which we have been familiar for the past 50 years (the game of private interests corrected by the anonymous 'social safety net') will have shown that its fruitfulness is henceforth exhausted, that will then be, as in 1789, the end of one world. A collective survival reflex may then push men to admit that a social life is possible only when constituted of interdependent cells and solidarity. That of course does not prevent either inequalities or conflicts, these being normal features in an organic society. But it would lead each person to discover that other people's lives are important to him, and that it is not possible for the head to say to the feet, 'I don't need you'. The survival of human groups today depends on this radical change in perspective, and thus spiritual aspects appear to have to be 《the infrastructure》 of the material elements, if it is desired that social reproduction continue.[19] (p. 22)

Let's not make a mistake here: it's not a question of pleading for a return to traditional societies, but of advocating that all collective efforts seek henceforth to give to the social link its forgotten primacy and to reaffirm the dependence of the economy in relation to the society. Contemporary societies have entered a crucial phase of their transition, in which generations will have to declare if they wish to cooperate with one another for a better future or if they choose the path of isolation and confrontation. Examples exist almost everywhere in the world that militate in favor of each of these possibilities, and there is nothing currently that allows us to affirm with certainty that the optimistic outcome will necessarily carry the day over the pessimistic outcome. What does seem rather certain is that solidarity, which attests to the reciprocity of dependence and the necessity of intergenerational exchanges, constitutes, if not the only, then at least probably one of the better possible responses to the changes of demographic structures by age and by sex.

XII. DECLINING BIRTH RATES AND PENSION INSOLVENCY: A VICIOUS CIRCLE?

A young Swiss author, physician by training and banker by profession, David Cosandey (2003), recently published a work with a title as enigmatic as it is provocative, *The Culpable Insolvency of Retirement Pensions. How Our Old-Age Insurance Makes Fertility Decline*,[20] which counters many established ideas about retirement systems while raising strong controversies that are probably not ready to be appeased. And yet

[19] "lorsque le cadre économico-social qui nous est familier depuis 50 ans (le jeu des intérêts privés corrigé par le filet anonyme des 'assurances sociales') aura manifesté que sa fécondité est désormais épuisée, ce sera alors, comme en 1789, la fin d'un monde. Un réflexe collectif de survie peut alors pousser les hommes à admettre qu'il n'y a de vie sociale possible que constituée de cellules interdépendantes et solidaires. Cela n'empêcherait, bien entendu, ni les inégalités ni les conflits, ceux-ci ayant leur place normale au sein d'une société organique; mais cela amènerait chacun à découvrir que la vie de l'autre lui importe, et qu'il n'est pas possible à la tête de dire aux pieds 'je n'ai pas besoin de vous'. La survie des groupes humains dépend aujourd'hui de ce changement radical de perspective, et c'est ainsi que le spirituel apparaît comme devant être 'l'infrastructure》 du matériel, si l'on veut que la reproduction sociale se poursuive.' "

[20] *La faillite coupable des retraites. Comment nos assurances vieillesse font chuter la natalité.*

the author's avowed goal is none other than to reestablish the solidarity among generations that apparently was worsened precisely by the introduction of retirement systems.

The paradox is total: whereas usually the explanation is that it is the reduction in births following the baby boom and the succession of small generations compared to the previous full generations (the baby boomers) that is responsible for the financial difficulties of our retirement systems, Cosandey supports the converse argument, affirming that it is because of the institution of retirement systems that falling birthrates emerged (or were strengthened) because these social protection measures broke the natural linkages of reciprocal solidarity among the generations and caused us to lose sight of the fact that to be able to finance and pay for retirements, there must be a new generation of largely the same size as the preceding one. In the absence of this perspective, childbearing can only diminish, and its drop precipitates the breakdown of retirement systems, which self-destruct via this tragic vicious circle.

To transform this vicious circle into a *virtuous* one, there is only one remedy: stop basing the calculation of retirement pension amounts on the number of years of professional activity spent paying into the system and henceforth take into account the only important parameter, namely the number of children raised or supported by the individual, while putting into place "a structure allowing everyone, even non-parents, to contribute to the new generation, so as to have the right to a pension also." In short, make "clairvoyant" systems that are currently "blind."

And the rule is true, whatever the method of financing pensions, by redistribution (pay-as-you-go) or by capitalization (individual retirement accounts), because both are functions of economic dynamism and thus of demographic vitality. Cosandey considers, moreover, that to reform old-age insurance and make it more equitable and viable for the long term, it will hardly be necessary to upset either the institutions or their means of functioning, but only to modify some parameters and calculation rules. In sum, what is needed is simply to make current retirement systems be more similar to traditional systems of family protection, in which solidarity was expressed by virtue of a relatively simple principle: parents feed, shelter, and educate their children in exchange for which the children, after reaching adulthood, support, care for, and protect in turn their now-elderly parents.

Despite inevitable weaknesses, notably in cases of the premature death of children or family conflicts that may deprive parents of their old-age indemnity, this system functioned relatively well and according to an equitable principle.[21] By contrast, with the making of anonymous intergenerational systems of social protection that were disconnected from family linkages, intergenerational exchanges were perverted to the extent that the contributors (the parents) have to pay for the non-contributors (non-parents). They act so as to give to the oldest the condition of receiving from the youngest, whereas it is to give to the youngest that should condition receipt of a pension by the oldest.

And in such a regime, one can't reproach non-parents for showing themselves to be selfish or not being characterized by solidarity: they simply show the good sense and rationality of taking

> the most gratifying path that the authorities, the law, the society offers them. [. . .] Those without children appear to come out ahead. They can more easily travel, devote themselves to sports or reading, buy clothes, engage in their favourite hobbies. They enjoy tranquillity in their personal affairs and a superior standard of living. They symbolize affluence and success. [. . .] With everyone taking advantage of the investments realized by parents, our systems of old-age insurance encourage what can be called a tactic of "demographic parasitism." (Cosandey, 2003, p. 47, 48)[22]

[21] Several recent works show that private solidarity is far from having disappeared. Even if it has receded in the face of increased public solidarity supported by economic progress, it continues nonetheless, often in a complementary manner with regard to public solidarity. Illustrations of this can be found, notably in the research of André Masson (1995, 2004) and of Claudine Attias-Donfut and Martine Ségalen (1998), who writes:

> Solidarity between the young and the elderly in rural societies was of a private nature; this was disrupted by the effects of industrialization, which made proletarians of several generations, and by migration, which ruptured family linkages, both social and cultural. These linkages re-established themselves when the condition of workers improved, notably thanks to the appearance of retirement pensions. It is after the Second World War that economic growth, in conjunction with the social burst which appeared in the years 1945–1950, favoured the development of strong social protection instituting a national solidarity among generations. [. . .] In brief, in discharging the family from its traditional function of taking care of economic needs in old age, public solidarity has not eliminated the private mutual aid between young and old, but has changed its orientation, while at the same time transforming the respective statuses of the generations, It has also contributed to the emergence of new linkages, established on the basis of the autonomy of generations. (p. 33, 36)

[22] "le chemin le plus gratifiant que les autorités, la loi, la 'société' leur offrent. [. . .] Les inféconds font figure de gagnants. Ils peuvent plus facilement faire des voyages, se consacrer au sport ou à la lecture, s'habiller, se consacrer à leur passe-temps favoris. Ils jouissent de la tranquilité dans leur chez-soi et d'un niveau de vie supérieur. Ils symbolisent l'aisance et le success. [. . .] En faisant profiter tout le monde des investissements réalisés par les parents, nos assurances vieillesses encouragent ce qu'on pourrait appeler une tactique de 'parasitisme démographique.' "

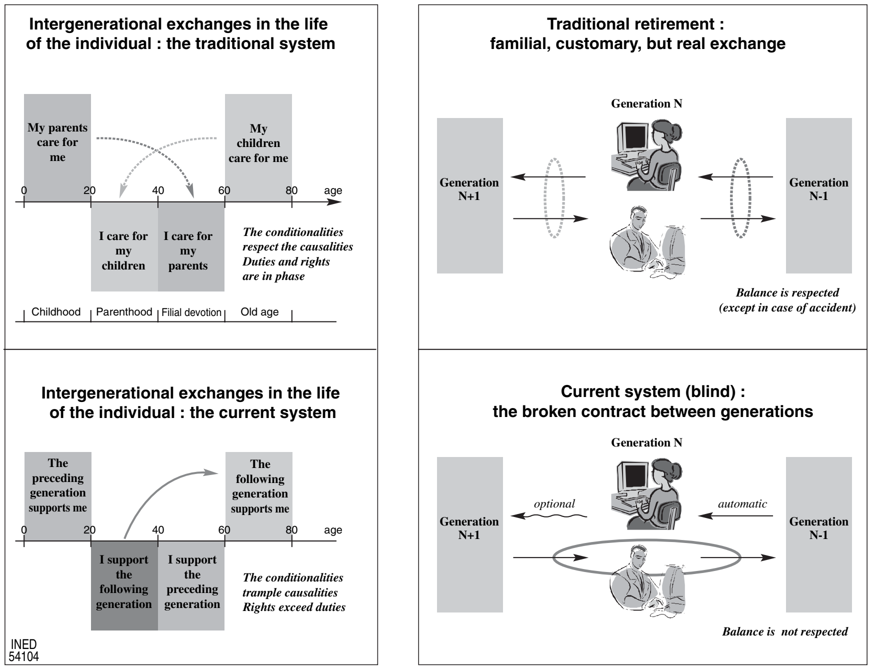


FIGURE 111–1 A vision of intergenerational exchanges, in the traditional system and in the current system, according to David Cosandey (2003). Extract from a presentation by the author of his work at a conference before the Central Council of the Economy, Brussels, June 2004.

Is it really necessary to penalize non-parents and provoke potential opposition between fecund and non-fecund couples or individuals? Not necessarily, because Cosandey imagines in his clairvoyant pension system several means permitting individuals who had no children, or only one (i.e., less than the threshold number established to benefit from a full retirement pension, being in principle two children) to contribute to the following generation and to acquire rights for their retirement (a minimal pension, financial sponsorship, fusion of retirement payments and child allocations, etc.).

As to the remedies currently evoked to confront the financial insolvency of the present systems, Cosandey rejects all of them, one after the other, because of their inefficiency and their injustice or because of the uncertainty of the expected results. This is the case for raising the retirement age, increasing taxes on persons and/or consumption, improving the productivity of work and/or increasing employment rates (most notably, of women), and resorting to immigration and investment in the economic actions and obligations of developing countries that are experiencing strong economic growth.

The solution to "furtively go back to the traditional system where grown-up children personally take care of their elderly parents as thanks for having been looked after in their own childhood" is itself rejected because, in making the role of the State disappear, where the State "offers a guarantee for parents having lost their children, thereby permitting a certain redistribution of income and avoids promiscuity among generations," this would be a step backward for society.

In sum, Cosandey's attempt merits consideration to the extent that he encourages a recasting of intergenerational solidarities on more explicit foundations, and at a time when not only are aging societies facing an imperative need for solidarity (precisely because they are aging and thus more weakened in certain segments of their populations), but also because they have never been so multigenerational as they are today, integrating numerous and strongly differentiated generations in an increasingly broad age interval, going from 0 to 100 and beyond.

The result of all this is that to manage these multigenerational (and in addition, multicultural) societies has become a real policy conundrum, if we are really willing to admit that good governance must seek to satisfy in an optimal way the needs and aspirations of all citizens, whatever their age, sex, social category, or religious or cultural affiliation. Unfortunately, it is at the moment when solidarity among generations should be recognized as one of the essential cornerstones of our societal edifice that it is under attack, disavowed everywhere and denigrated in the name of pretexts that are most often fallacious.

Exacerbated individualism and the winds of economic liberalism that have blown since after World War II in our Western societies have greatly contributed to the establishment of this disastrous climate. They created the sentiment that protection of the group or the community was less necessary from the time when essential needs (and even less essential needs or superfluities) were assured and the great risks (illness, unemployment, accident, old age) covered, leaving to each individual the possibility of managing his life in an autonomous fashion, freeing himself from collective constraints.

The discourse of insurers and bankers being increasingly along this line of thought, all of the current efforts seeking to encourage private insurance formulas or to replace, partially or totally, the systems of redistribution by systems based on capitalization are insidious ways of doing harm to the principle of intergenerational solidarity and encouraging young workers to break the great chain of solidarity among the generations. Only omitted here is a reminder that these supposedly miraculous formulas are subject to the random effects of economic fluctuations and to the risks of financial crises. Above all, young workers will run up against the challenge of the prolongation of the average duration of life of retirees (which could double, from the current 10–20 years to 30–40 years, if the legal ages of retirement are maintained).

Despite its undeniable originality, Cosandey's thesis nonetheless suffers from unavoidable flaws that compromise its chances to ever be adopted by public authorities, even if Cosandey can boast about a reform adopted in Germany in 2002[23] as a modest outline of a clairvoyant retirement system.

Indeed, the empirical analyses on which Cosandey relies to demonstrate that from the time when more or less generous retirement pensions were put into effect or strengthened in several countries (France, Germany, Switzerland, Italy, United States), childbearing declined ("each time that old-age pensions, in their current form, were strengthened, childbearing declined. And vice versa") are not entirely persuasive. They are based on a simplistic concept of causal relationships that does not take account of the transformations of the societal context that make the classic phrase, "all other things being equal," absurd. We must never lose sight of the fact that the demographic behaviors of individuals and couples are, first and foremost, social realities that only take their full and complete meaning if they are situated in the totality of changes of the systems in which they are operating.

Without a doubt, we should acknowledge that it isn't the discovery of the birth control pill or other technical reasons of this type that can explain by themselves the decline of fertility. There existed a set of convergent historical factors that contributed to the modification of reproductive behaviors, and the institution of systems of social protection constituted only one element among many of these societal changes that accompanied the industrial revolution.

Among these other elements, we can certainly allude to the changes in women's status and the increased participation of women in the market economy, the reduction of infant mortality, the preference for quality of children rather than quantity, the transformations regarding conjugal unions with the appearance of free unions and the wave of divorce. In brief, it is the entire family, with the positions occupied by its different members and its ends, that has been subject to profound transformations and which has been reprogrammed as a function of new technological, economic, social, and cultural contexts.

As a consequence, to dream of a new increase in fertility to the levels attained in the 1960s is unrealistic. Many potential parents have understood that it would be better, in the contemporary context, to be parsimonious about childbearing, so as to conserve all of the chances to endow their progeny with the best means of success possible, notably in terms of schooling in an increasingly competitive society.

In the face of persistent youth unemployment, which is not, however, incompatible with an increase in value added and in collective wealth, we might ask why there would be a return to higher fertility and investment to increase the quantity of human capital at a time when it is increasingly in competition with technical capital, despite the improvement in the productivity of labor.

But the most unfavorable argument against Cosandey's model is without doubt that, by the admission of the author himself, the system he advocates could only be operational beginning in 2050, in order to give time to new generations to reconstitute their numbers. That is, about 30 or 40 years after the advertised frontal shock among the generations will have taken place and the dreaded catastrophe (the failure of the retirement systems) has taken place (if at least one follows the forecasts of those predicting apocalypse). On the contrary, it is possible that the worst will be avoided and efficacious regulatory measures will have been taken in a timely fashion (if one adheres rather to the arguments of the optimists).

[23] The Riester reform, from the name of the minister who proposed it.

XIII. THE ULTIMATE QUESTION: RENEGOTIATE A NEW INTERGENERATIONAL CONTRACT

Starting with a question that appeared to be purely technical, highly objective, and of apparent simplicity, How do we manage the population pyramid?, we very quickly were led to ethical, normative, and ideological considerations. The reason for this is not very difficult to understand: a population is not a simple statistical construction and even less just human material, adjustable to the imperatives of the economy or of politics. The question of management of the demographic structures cannot be reduced to a sort of engineering of the human resource without inevitably falling into one or the other of the faults criticized here. Unfortunately, engineers or human resource technicians are listened to today more than philosophers, ethicists, or sociologists, and when these latter speak, they never do so with one voice.

In addition, the social sciences have never been as partitioned as they are at present, when the need for interdisciplinarity is so often expressed. Experts speak to experts, and ideally, only to those of their own discipline, or even their own "scientific chapel," and their relationships with the citizen are at the very least difficult.

> The citizen will have a good deal of difficulty appropriating what the experts tell him about the society in which he lives, either because they propose to him an image in little pieces that are impossible to put together, or because on the contrary they reconstruct it entirely in the image of their own system of thought, erected in Universal and definitive truth. (Supiot, 1996)[24]

These truths are much less neutral than indicated by the formal or informal models on which they depend. But the increase in the circulation of information and knowledge via computer and telecommunication networks has not had a parallel impact on the diversity of perspectives and points of view on objectives to pursue and the means to achieve them. Rather, it has favored concentration around dominant models, principally from the big international institutions or the big research centers that are effectively under their authority.[25]

Instead of being totally unique, thought has become institutionalized and normalized, so that there is no longer a place for an autonomous wild thought. To "preserve prosperity in an aging society," as OECD (1998) proposed in one of its recent and important research programs, is certainly not a goal to be condemned on its own merits. But it is rather disquieting to note that between the OECD, the United Nations, UNDP, the ILO, the Council of Europe, and the European Union, the convergence of solutions recommended is such that they appear to be closely under the influence of a single dominant school of thought.

There is no doubt that we can agree with the idea of encouraging an "active ageing" (Cotis, 2003), in favoring a greater flexibility of life cycles and greater freedom of choice regarding the passage from professional life to retirement. But is it necessary to automatically propose simultaneously to "slow expenditures of public retirement regimes and systems of health insurance"[26] (OCDE, 1998)? Nothing is less evident and, in any case, it would be regrettable if decisions concerning such crucial problems were to be made without a broad public debate being organized, at which the principal social partners and all those in positions of responsibility could participate.

The long-term future of our aging societies is too profoundly linked to these choices to allow them to be made hastily and left to the discretion of a small group of experts guided by political and financial or intellectual elites, who risk forgetting that the real question is not "how to preserve prosperity in an ageing society" (OCDE, 1998), but rather "how to renegotiate a new social contract among generations with solidarity that ensures to each the recognition of a valued role and status, without exclusion or marginalization, whatever his sex and his age" (Loriaux, 2002).

Acknowledgment

This chapter was translated from the French by David Shapiro.

[24] "aura bien du mal de s'approprier ce que les experts lui disent de la société où il vit, soit qu'ils lui en proposent une image en miettes impossibles à recoller, soit qu'ils le reconstruisent au contraire tout entier à l'image de leur propre système de pensée, érigé en Vérité universelle et definitive."

[25] This may be taken account of in observing to what point certain concepts are diffused like a train of powder across the international scientific community, governing powers and to the larger public, often beginning with confidential work of small groups of experts. In the field of demography and development, this was the case for the concept of integration of population and development and, more recently, for those of reproductive health and sustainable development. The criticism formulated here does not deal with their value as such, but with the way that they were diffused, without precaution nor a detailed examination of their implications, with considerable means of publicity and propaganda (worldwide diffusion in real time, translation into multiple languages, etc.) under the auspices of big organizations or powerful pressure groups that left little place for any contesting after the fact or even a simple critical reevaluation.

[26] "Why must spending on public pensions, health and long-term care be contained?" (OCDE, 1998).

References

Attias-Donfut Claudine (ed.), 1995. *Les solidarités entre générations. Vieillesse, famille, État.* Paris, Nathan, 362 p. (Essais et recherches, Série sciences sociales).

Attias-Donfut Claudine and Ségalen Martine, 1998. *Grands-parents. La famille à travers les générations*. Paris, Odile Jacob, 330 p.

Blanchet Didier, 2000. Quels liens entre politique familiale et démographie?, *in*: François Charpentier (ed.), *Encyclopédie protection sociale. Quelle refondation?*, p. 891–896. Paris, Economica and Editions Liaisons, 1494 p.

Bonnici Bernard, 1997. *Politiques et protection sociales.* Paris, Presses universitaires de France, 128 p. (*Que sais-je?*, no. 2333).

Bourdelais Patrice, 1993. *Le nouvel âge de la vieillesse: histoire du vieillissement de la population.* Paris, Odile Jacob, 441 p. (Histoire).

Caillé Alain, 1996. Pour en finir dignement avec le XX^{e} siècle: temps choisi et revenu de citoyenneté, *La Revue du MAUSS*, no. 7, "Vers un revenu minimum inconditionnel?" p. 135–150.

Charpentier François (ed.), 2000. *Encyclopédie protection sociale. Quelle refondation?* Paris, Economica/Editions Liaisons, 1494 p.

Chaunu Pierre and Legrand Jean, 1979. *Un futur sans avenir. Histoire et population.* Paris, Éditions Calman Levy, 320 p. (Archives des sciences sociales).

Chauvel Louis, 1998. *Le destin des générations, structures sociales et cohortes en France au XXe siècle.* Paris and Vendôme, Presses universitaires de France, 301 p. (Le lien social).

Chesnais Jean-Claude, 1987. La croissance démographique, frein ou moteur du développement? *in*: Sylvie Brunel (ed.), *Tiers-Monde. Controverses et réalités*, p. 119–142. Paris, Economica, 519 p.

Chesnais Jean-Claude, 1995. *Le crépuscule de l'Occident. Dénatalité, condition des femmes et immigration.* Paris, Robert Laffont, 366 p. (Notre époque).

Collomp Alain, 1992. Les formes de la famille. Approche historique, *in*: François de Singly (ed.), *La famille: l'état des savoirs*, p. 13–21. Paris, La Découverte, 449 p.

Cosandey David, 2003. *La faillite coupable des retraites. Comment nos assurances vieillesse font chuter la natalité.* Paris, L'Harmattan, 166 p.

Cotis Jean-Philippe, 2003. Vieillissement démographique: relever le défi, *L'Observateur de l'OCDE*, n° 239, p. 19.

Dumont Gérard-François, 2002. *La géographie de la France: situations humaines et économiques.* Paris, Ellipses, 300 p. (Université. Géographie).

Dumont Gérard-François, Chaunu Pierre, Legrand Jean and Sauvy Alfred, 1979. *La France ridée. Échapper à la logique du déclin.* Paris, Le livre de poche, 477 p. (Pluriel).

Dupâquier Jacques, 2000. Revanche des berceaux: mythe ou réalité?, *Panoramiques*, no. 47, "Quand on entend démographie, faut-il sortir son révolver?" p. 66–68.

Dupâquier Jacques and Biraben Jean-Noel, 1981. *Les berceaux vides de Marianne. L'avenir de la population française.* Paris, Le Seuil, 176 p.

Easterlin Richard A., 1975. An economic framework for fertility analysis, *Studies in Family Planning*, vol. 6(3), p. 54–63.

Euzéby Alain and Euzéby Chantal (eds.), 1996. Les solidarités, fondements et defis, Paris, Economica, 250 p.

Ferry Jean-Marc, 1995. *L'allocation universelle. Pour un revenu de citoyenneté.* Paris, Éditions du Cerf, 124 p. (Humanités).

Fourastié Jean, 1979. *Les Trente Glorieuses ou la révolution invisible de 1946 à 1975.* Paris, Fayard, 288 p.

France Henri de, 1996. La solidarité et ses fondements implicites: les conditions d'un retour à une vision organique de la société, *in*: Alain Euzéby and Chantal Euzéby (eds.), *Les solidarités, fondements et défis*, p. 1–22. Paris, Economica, 250 p.

Franciosi Maria Laura, Scocci Sergio and Tarini Anna (eds.), 1996. *Per un sacco di carbone (For a sack of coal). Brussels*, ACLI-Belgio, 386 p.

Gaullier Xavier, 1988. *La deuxième carrière. Ages, emplois, retraites.* Paris, Le Seuil, 408 p. (L'épreuve des faits).

Gaullier Xavier, 1999. *Les temps de la vie. Emploi et retraite.* Paris, Éditions Esprit, 258 p. (Sociétés).

Gaullier Xavier, 2000. Qu'apporterait la pluriactivité à tout âge?, *in*: François Charpentier (ed.), *Encyclopédie protection sociale. Quelle refondation?*, p. 323–333. Paris, Economica/Editions Liaisons, 1494 p.

Godet Michel, 2003. *Le choc de 2006. Démographie, croissance, emploi. Pour une société de projets.* Paris, Odile Jacob, 300 p.

Godet Michel and Monti Régine, 2000. Le lien démographie croissance existe-t-il?, *in*: François Charpentier François (ed.), *Encyclopédie protection sociale. Quelle refondation?*, p. 65–74. Paris, Economica and Editions Liaisons, 1494 p.

Guillemard Anne-Marie, 1986. *Le déclin du social. Formation et crise des politiques de la vieillesse.* Paris, Presses universitaires de France, 394 p.

Kotlikoff Laurence J. (ed.), 1992. *Generational accounting: knowing who pays and when for what we need.* New York, The Free Press, MacMillan, 262 p.

Lassonde Louise, 1996. *Les défis de la démographie: quelle qualité de vie pour le XXIe siĉcle?* Paris, La Découverte, 225 p.

Le Play Frédéric, 1871. *L'organisation de la famille suivant le vrai modèle signalé par l'histoire de toutes les races et de tous les temps.* Paris, Tequi, 318 p. (Saint Michel).

Loriaux Michel, 1986. Présentation du rapport POLIWA. Plaidoyer pour une politique démographique douce, *in: VI^{e} colloque national de démographie, La fécondité dans les pays industrialisés*, p. 295–304. Paris, Éditions du CNRS, 566 p.

Loriaux Michel, 2002. Les sociétés de la géritude et la gestion des relations intergénérationnelles, *in*: Antonio Guerci and Stefania Consigliere (eds.), *Living and "curing" old age in the world, 3rd International Conference on Anthropology and the History of Health Disease (Genova 2002), vol. 4, The Elderly in the Mirror. Perceptions and Representations of Old Age*, p. 49–74. Genève, Erga Edizioni, 426 p.

Loriaux Michel, Remy Dominique and Vilquin Éric (eds.), 1990. *Populations âgées et révolution grise: les hommes et les sociétés face à leurs vieillissements.* Louvain-la-Neuve, Éditions CIACO, Institut de démographie (UCL), 1118 p. (Chaire Quételet 1986).

Masson André, 1995. L'héritage au sein des transferts entre générations, *in*: Claudine Attias-Donfut (ed.), *Les solidarités entre générations. Vieillesse, famille, État*, p. 279–325. Paris, Nathan, 352 p. (Essais et recherches, Série sciences sociales).

Masson André, 1999. Quelle solidarité intergénérationnelle?, *Revue française d'économie*, vol. 14(1), p. 27–30.

Masson André, 2000. Equité ou solidarité intergénérationnelle, *in*: François Charpentier (ed.), *Encyclopédie protection sociale. Quelle refondation?*, p. 355–364. Paris, Economica/Editions Liaisons, 1494 p.

Masson André, 2004. Économie du débat intergénérationnel: points de vue normatif, comptable, politique, *in*: Jacques Véron, Sophie Pennec and Jacques Légaré (eds.), *Âge, générations et contrat social*, p. 15–58. Paris, INED, 314 p. (Cahier no. 153).

Monnier Alain and Prioux France, 2000. La baisse de la fécondité est-elle irréversible?, *in*: François Charpentier (ed.), *Encyclopédie protection sociale. Quelle refondation?*, p. 881–890. Paris, Economica et Editions Liaisons, 1494 p.

Minc Alain, 1987. *La machine égalitaire.* Paris, Grasset, 284 p.

Monti Régine, 2003. *La gestion prospective des âges des populations larges.* Doctoral thesis. Paris, Conservatoire national des Arts et Métiers, 329 p.

OCDE, 1998. *Préserver la prospérité dans une société vieillissante.* Paris, OCDE/OECD (Organisation de coopération et de développement économiques), 152 p. (Synthèse de l'OCDE, no. 5. This document is also available in English).

OECD, 1998. *Maintaining prosperity in an ageing society.* Paris, OECD, 152 p. (Policy brief, no. 5).

Quetelet Lambert Adolphe J., [1869] 1997. *Physique sociale ou essai sur le développement des facultés de l'homme.* New edition annotated by Eric Vilquin and Jean-Paul Sanderson. Brussels, Académie royale de Belgique, 700 p.

Rawls John, 1971. *A theory of justice.* Cambridge (Massachussetts), Belknap Press of Harvard University, 607 p.

Rawls John, 1987. *Théorie de la justice.* Paris, Le Seuil, 666 p. (La couleur des idées).

Rawls John, 1993. *Justice et démocratie.* Catherine Audard's French translation of Rawls' *Selected Papers, 1978–1989*. Paris, Le Seuil, 355 p. (La couleur des idées).

Rosental Paul-André, 2002. *Intelligence démographique. Sciences et politiques des populations en France (1930–1960).* Paris, Odile Jacob, 367 p. (Histoire).

Rosnay Joël de, 1975. *Le macroscope. Vers une vision globale.* Paris, Le Seuil, 313 p.

Sauvy Alfred and Hirsch Anita, 1987. *L'Europe submergée. Sud-Nord dans trente ans.* Paris, Dunod, 274 p.

Sullerot Evelyne, 2000. *La crise de la famille.* Paris, Fayard, 300 p. (Pluriel).

Supiot Alain, 1996. Perspectives on work: Introduction (and editor) of the special issue "Perspectives on the nature and future of work" *International Labour Review*, vol. 135(6), p. 603–614. (BIT).

Supiot Alain, 1996. Le travail en perspectives: introduction, *Revue internationale du travail*, vol. 135, no. 6, p. 663–674. (BIT.)

Tapinos Georges P., 1996. *La démographie. Population, économie et sociétés.* Paris, Le Livre de Poche, 255 p. (Inédits).

UNFPA, 1998. *The state of world population 1998. The new generations.* New York, United Nations Fund for Population Activities, 76 p.

van Parijs Philippe, 1990. Peut-on justifier une allocation universelle? Une relecture de quelques théories de la justice économique, *Futuribles*, no. 144, p. 29–42.

van Parijs Philippe, 1999. The disenfranchisement of the elderly, and other attempts to secure intergenerational justice, *Philosophy and Public Affairs*, vol. 27(4), p. 290–333.

Vekeman Victor, 1945. *Les berceaux vides.* Montréal, Editions du Levrier, 64 p.

Verrière Jacques, 1978. *Les politiques de population.* Paris, PUF, 207 p. (Le géographe).

Wisensale Steven K., 1997. Le vieillissement de la population mondiale: le débat inévitable sur l'équité entre les générations, *Bulletin sur le Vieillissement*, no. 2–3, p. 7. (Nations Unies/United Nations).

P A R T

V

COUNTRY CASE STUDIES

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Family and fertility policies, health policies, migration policies, policies on age distribution, each country and era has its own priorities, and population policies often address one or other of these aspects more than the others as circumstances vary. But regardless of the time or period, policy action tends to have planned or unplanned consequences in all these spheres, and it is important to be aware of these, get a purchase on the totality and interplay of all government measures that can influence population dynamics, intentionally or otherwise. The most appropriate way of doing this seemed to be to analyze the population policies in the broad sense of a selected group of countries.

The choice of countries was relatively straightforward. It was first decided to cover the two most populous countries in the world—China and India. Isabelle Attané studies the former in Chapter 112, Jacques Véron and S. Irudaya Rajan the latter in Chapter 113. To these two big Asian countries were added three other developing countries. First, a large country in Latin America–Mexico—whose population policy is covered in Chapter 114 by Maria-Eugenia Cosio-Zavala, then two countries in sub-Saharan Africa—Kenya and Côte d'Ivoire. Kenya, analyzed in chapter 115 by Valerie Golaz and Carolyne Wanja Njue, was among the first in this region to experience the onset of fertility decline. Côte d'Ivoire, covered in Chapter 116 by Amoakon Anoh, represents the African countries where it had still not occurred at the start of the 1990s.

As a counterpart to these five developing countries, we selected three developed countries: the United States and Russia, the two biggest, represent completely different experiences. The United States, which dominates today's world in all matters from its unassailable superpowerdom, and where free enterprise is the overarching principle in all economic and social matters with obvious consequences for population policy, is covered in Chapter 117 by Carl Haub. Russia, by contrast, represents the former communist bloc, and especially the countries of the former USSR, where the onset of demographic transition was much later than in Western Europe, and in fact only very shortly before their subjugation to communism. It is analyzed by Serguey Ivanov, Anatoly Vichnevsky and Sergei Zakharov in Chapter 118. Finally, France was also chosen as a case apart of demographic transition without population growth, and the acute awareness of low fertility problems it rapidly came to. This case is analyzed in Chapter 119 by Jean-Claude Chesnais.

Increasingly, however, notwithstanding the few supranational powers they actually possess, international organizations are stepping into population issues and having a major, if only indirect, influence on population policies. For that reason, these analyses of national experiences have been supplemented with a general review of international population policies by Jean-Claude Chasteland in Chapter 120.

CHAPTER

112

Fifty Years of Demographic Policies in China:

An Assessment

ISABELLE ATTANÉ

Institut national d'études démographiques (INED), Paris, France

After their victory over the Nationalist armies in 1949, the Chinese Communists found themselves in charge of a country exhausted and ravaged by war, its economy ruined by decades of internal strife and foreign interference. The new leaders had to put China back on its feet, rebuild its identity after a century of humiliation—in short, restore its dignity. The time had come to break with old feudal ways. Society had to be reshaped from top to bottom. The task was colossal.

The key concerns of young Communist China were the economy and society. Two groups were singled out as the prime victims of the social order: the impoverished peasantry and women. Pervasive abuses by the propertied classes had to be stopped. The traditional family became a target of the class struggle. The country set out to eradicate the "three main differences" (towns vs. countryside, manual workers vs. intellectuals, workers vs. peasants) (Attané, 2002c, p. 355). The regime preached equality, balance, a fair distribution of resources and the means of production, and harmonious economic development. Another goal was to improve the quality of the population by a campaign to combat devastating mortality and by raising the consciousness of the masses through education.

Mao Tse-tung's death in 1976 marked a radical break with egalitarian ideology. The reforms launched by Deng Xiaoping in the late 1970s opened the doors wide to individual initiative. The economy was liberalized. The winds were changing, the socialist dogma was buried, and social inequality deepened.

In both its hardline and liberal phases, this half-century of Chinese socialism relied extensively on all manner of authoritarian policies. No area was spared—the economy, the circulation of people, health, reproduction, and so on. In this chapter, we review the main demographic policies—in the broadest sense—implemented since 1950. We then assess the achievements, and their consequences, at the turn of the 21st century.

I. HOUSEHOLD REGISTRATION AND MOBILITY CONTROL

Among the most decisive population policies, the one that had the most visible and lasting impact on Chinese demographics—along with birth-control policies—was the control of personal mobility. As early as 1955, a household registration system was set up under the authority of the Interior Minister. But its implementation remained very partial, most notably because of staff shortages. Control of population movements was not enforced until the Regulations on Household Registration of 1958. One of the main goals was to end illegal migrations to urban areas (Blayo, 1997, p. 28). To justify the introduction of this rigid and restrictive system, the authorities invoked the need to spare China from the informal urbanization typical of other developing countries. Curbing urban growth soon became a major priority. This, it was hoped, would prevent the worsening of urban employment conditions and channel population movements to better serve the regime's interests. At the advent of the People's Republic, urbanization was a marginal trend. One in ten Chinese (10.6%) lived in a town or city, a far smaller proportion than the worldwide average, which had already reached 30%. (The definition of the urban population has changed frequently since the 1950s. For more details, see Blayo, 1997, p. 101–111.) The rate remained modest until after Deng's economic reforms, which led to a gradual easing of restrictions on internal migration. In 1978, urbanites still accounted for only 18% of the total Chinese population. Until the 1980s, therefore, population migration and worker mobility—the keystones of the planned economy—were tightly controlled. The Public Security offices kept a special watch on interprovince migration; except for compulsory transfers, it was nearly impossible to migrate from rural areas to urban areas without official permission (Cai Fang, 2002, p. 169). Even today, despite its steadily declining effectiveness, control is exercised via household registration: the *hukou* system. The *hukou* is a residence booklet that includes, among other information, the holder's date and place of birth, ethnic group, and marital status. A veritable internal passport, it obliges each individual, from cradle to grave, to reside in his or her birthplace, making migration virtually impossible. For many years, legal migration was a one-way ticket, each type of locality being assigned to a specific position in a very rigid hierarchy. A person could leave a large city for a smaller one, a town, a district or township capital, or a village (Figure 112–1). Horizontal migration was tolerated for people wishing to move to a locality of identical rank; so was vertical migration, but only from the larger unit to the smaller unit, not vice versa. A person born in a village could move to another village in the same province, subject to permission by the authorities of both localities. This registration system served for many years as a key instrument in the arsenal deployed by the regime to subjugate society (Trolliet and Béja, 1986, p. 155).

Even marriage did not afford an escape from *hukou* control. If a peasant woman married an urban dweller, the husband could resettle in his wife's native village, as the locality was of lower rank. But there was no hope of transferring the wife's *hukou* to the city. An urban dweller married to a peasant woman could not even look forward to being joined by his children one day, as the urban or rural status enshrined in the *hukou* was transmitted by the mother. Today, a child is free to choose between mother's or father's *hukou.* As a result, the system spawned a Chinese Communist invention that was—to say the least—unusual: couples separated for administrative reasons and saw each other only once a year, usually during the spring festival vacation. The only ways of being freed from this compulsory residence order were to enlist in the People's Liberation Army, to become a Party member, or to obtain a university degree, which entitled the holder to a job in a State-owned company.

FIGURE 112–1 Structure of Chinese territorial administration. (Source: Attané, 2002b, p. xvii.)

The system helped to turn China into a frozen society. Far from closing the gap between urban and rural areas, the system fostered the perception that the right to live in urban areas was a privilege reserved for a minority. This feeling was exacerbated by the rationing system introduced after the Revolution. Ration coupons, for which only urban *hukou* holders were eligible, gave access to grain in the State-run market. Peasants were obviously excluded, and this posed an insuperable obstacle to their settling in cities. (Until 1985, it was illegal to sell grain on the free market.) As the coupons could be exchanged only in the issuing town or city, they were also a means for controlling personal travel and thus complemented the *hukou* system (Trolliet and Béja, 1986, p. 157). Rationing of staple commodities was abolished in the early 1990s.

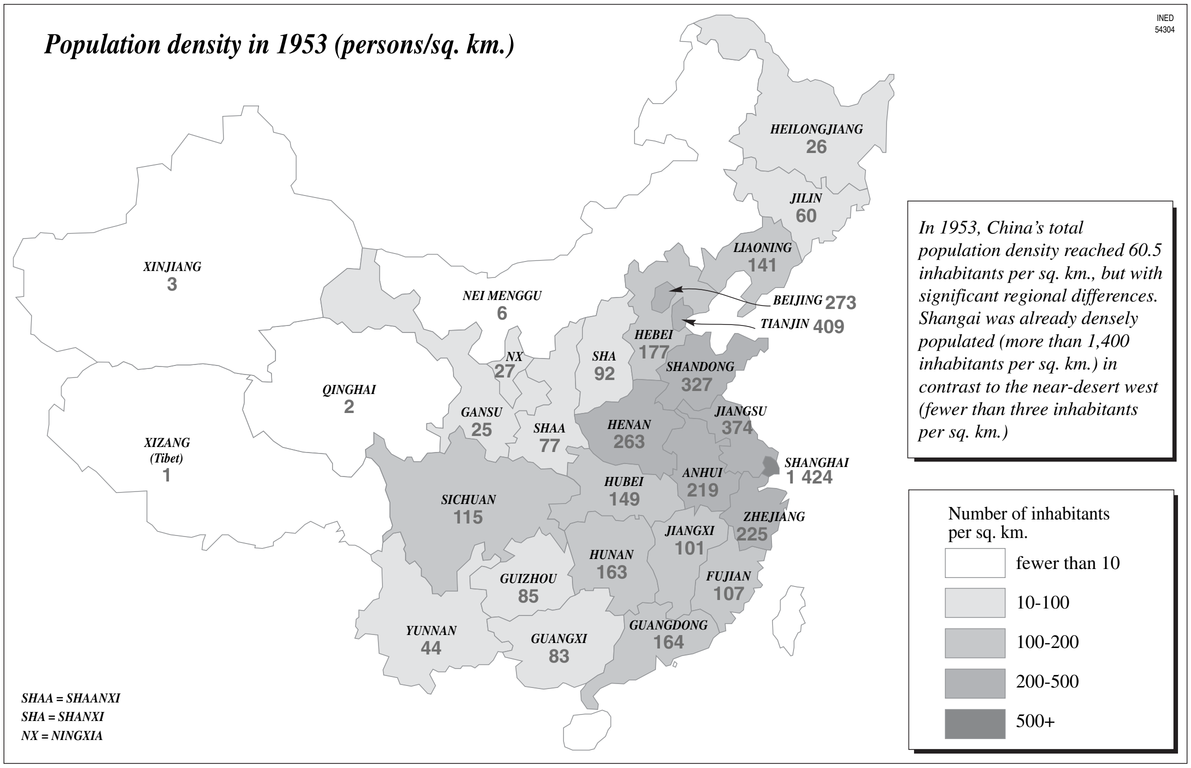
Out of exhaustion rather than because of a political decision, the regime gradually relinquished control over all personal movement. Rural decollectivization—the first step in economic reform—freed an abundant labor supply from the obligation to work on farms. The excess agricultural workforce was forced to turn to nonfarming activities in order to survive. The businesses in cantons and towns (*xiangzhen qiye*) specially created to absorb this glut became increasingly incapable of doing so. Very often, the excess rural labor had other no choice than migration in order to meet the needs of new private-sector urban companies, which required large amounts of cheap labor. Meanwhile, the restructuring of State-owned companies put millions of people out of work. Having failed to reassign these workers itself, the State turned a blind eye to many individual migrations, the only solution for finding new work. The dichotomy generated by the personal registration system thus became less biased against peasants.

The paradox is that the job seekers were forced into illegality solely because of the registration system's inconsistency. The system defines a migrant as anyone who legally transfers his or her *hukou* to a place of residence other than his or her place of birth. Now this legal migration, while ever less restrictive, remains subject to annual quotas well below requirements: 17 to 19 million people are allowed to migrate per year, "une population flottante nécessaire mais exclue" (Attané, 2002b, p. 185).

Migration with transfer of residence to the host locality thus accounts for a minimal share of actual migration. Most flows are generated by illegal migration, that of the "floating population" (*liudong renkou*). This phenomenon, specific to China, denotes temporary or permanent movements of persons in a province or between provinces, very often from rural areas to urban areas, without *hukou* transfer. Even if they relocate at several hundred kilometers' distance, these persons are officially regarded as still residing in their place of origin. In practice, they therefore avoid any form of control, including on the number of births. This illegal migration is, by definition, hard to quantify. Recent estimates give a broad range of 80 to 130 million people—a significant mass of about 10% of the total population. Therein lies one of China's most blatant contradictions in its march toward globalization. The *hukou* system, while totally inappropriate to today's economic requirements, has undergone no substantial change since the late 1950s.

Rural-urban migration remains the strongest driver of urbanization, although the natural increase is still rather brisk in the urban population at 0.7% in 1999, versus 0.9% in the countryside. The loss due to rural exodus has considerably reduced the overall growth of the rural population, which has become negative in the latest intercensal period: there were 783.8 million rural dwellers in 2000 versus 834.4 million in 1990, or an annual average decline of 0.6%. By contrast, the total growth of the urban population (cities and towns) has become truly explosive, averaging 4.4% per annum in the past 10 years. It should be borne in mind, however, that the changes in the definition of urban localities and the incorporation into these localities of some rural areas have increased the urban population in relatively artificial ways in the past two decades (see Blayo, 1997). The tripling of the number of cities since 1982 has also played a decisive role. The urban explosion has therefore not been as strong as the raw figures imply. With an official urban population of 459 million, or more than one in three inhabitants (37%), China lags behind the rest of the world: the urban population stands at 47% on a global average and 76% in what the United Nations classifies as the "most developed countries" (Figure 112–2).

But China will soon catch up. According to official projections, one in two Chinese (47%) will be urban dwellers by 2020, and the proportion will rise to at least 60% by 2040–2045 (Li Jianmin, 2002, p. 283). The number of towns and cities tripled in the past two decades, from 236 in 1982 to 667 in 1999. Three very large cities have more than 7 million inhabitants: Beijing, Tianjin, and Shanghai. For a while, therefore, China will have succeeded in curbing the size of sprawling megalopolises: with 11.5 million and 14.3 million inhabitants respectively in 2000, Beijing and Shanghai remain modest by comparison with such giants as São Paulo, Mexico, and Bombay (18 million) and Tokyo (26 million).



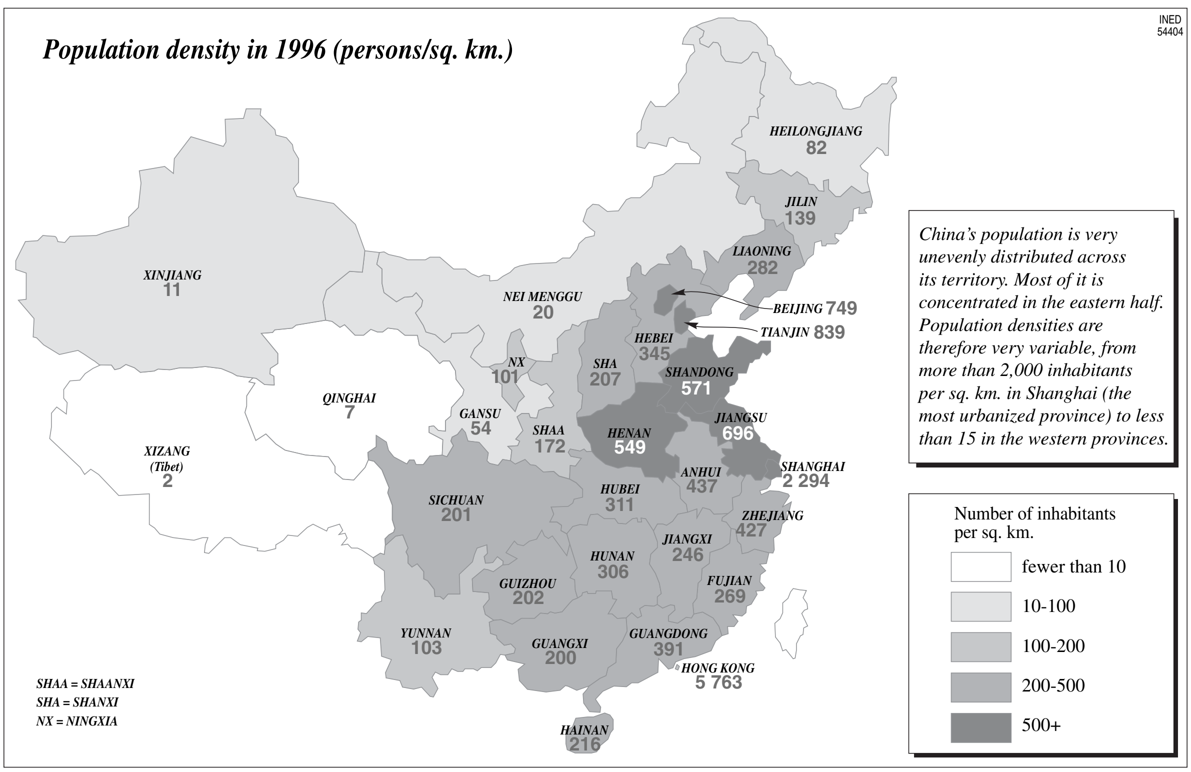


FIGURE 112–2 Population density by province in 1953 and 1996. (Source: Attané, 2002b, p. xviii, xix.)

II. BALANCING REGIONAL DEVELOPMENT AND STRENGTHENING BORDER REGIONS

For all socialist countries, central economic planning implies the planned allocation of all factors of production. Rational reproduction of the population must be ensured, particularly to allow optimal use of labor. By making industrial development the top priority of its first five-year plan (1953–1957), young Communist China forced itself to strictly manage its human resources. To meet the needs of the socialist economy more effectively, migration had to be controlled. As part of a planned process, migration would serve to balance labor requirements among regions. The authorities acted swiftly to curb urbanization and restrict population movements inside China—particularly the spontaneous migration of peasants to towns and cities.

In the early days of the regime, China's rulers saw a large population as a powerful asset. Although it reflected the varying hospitality of the soil, the terrain, and the climate, the population was unevenly distributed across the national territory. The eastern half of the country consists of plains and hills; the western half, with a rigorous climate, consists of plateaus and mountains. The west, which accounts for more than one-half of China's surface area (including Inner Mongolia, Xinjiang, Tibet, Qinghai, Ningxia, Gansu, Shaanxi, Yunnan, Guizhou, Guangxi, and Sichuan) contained only 28% of the population in 1953. These proportions have barely changed. In 2000, the year of the latest census, the west, with 319 million inhabitants, accounted for only 25.7% of the total population. Just after the Revolution, therefore, almost three-quarters of Chinese lived on less than one-third of the territory. The corollary was a highly disparate range of population densities: from over 300 people per square kilometer in Jiangsu and Shandong to fewer than 5 in Xinjiang (3/sq. km.), Qinghai (2/sq. km.), and Tibet (1/sq. km.). Concentrated in the eastern part of the country, the population lived far from the national reserves of natural resources needed by industry. There, infrastructure was lacking and workforce skills were low. Most modern plants were located in the coastal regions; heavy industry—which the Japanese began setting up in 1931 as part of their expansionist policy—was concentrated in southern Manchuria. In the rest of the country, the forces of production were weak and industrial facilities obsolete (Meissner, 1990, p. 9–33). To cancel these imbalances, China acted promptly to redistribute its population. Another goal was to foster the development of remote rural areas by sending managers and skilled labor there. Achieving balanced regional development required better management of human resources and local resources such as raw materials. A senior Chinese official declared in 1964:

> China has enough space . . . We have rich mineral resources and huge tracts of land that are just asking to be cultivated. Let us take the example of the north-west: Xinjiang province. Only 6 million people live there, but Xinjiang could easily provide enough living space and food for 50 million. (qtd. in Tien, 1973, p. 130)

In the very first five-year plan (1953–1957), the government implemented an interregional migration policy. The stated goal was to deconcentrate industry by promoting its expansion in the hinterland and to place new land under cultivation. But there was another purpose as well: to stem the growth of towns and cities, where rural-urban migration, then in full swing, was creating ever greater problems, that is, housing and food shortages coupled with substantial underemployment. The measures adopted accordingly sought to keep the rural population in place (thanks to the *hukou* system) and reduce the urban population. In all, 1.3 million blue-collar workers and skilled white-collar workers were transferred from urban areas to rural areas, of whom 950,000 between 1953 and 1962 and 200,000 between 1963 and 1981 (*China Daily*, April 25, 1982, qtd. in Blayo, 1997, p. 55). Massive projects were launched to build new infrastructure and clear arable land. Between 1950 and 1958, China built 94,000 kilometers of roads and 5,400 kilometers of railroads. In the same period, the surface area under cultivation grew by 11%, with the largest expansions in Xinjiang, Inner Mongolia, and the northeastern regions (Meissner, 1990, p. 31).

The policy of planned migrations thus served two purposes: the first was economic, the second demographic, namely, to correct unbalanced population distribution. The policy relied on the multiple systems for controlling individuals instituted in the early decades of the regime.

The central government supported the development of remote regions, with a special emphasis on border areas. The policy also had an ethnic objective, albeit not always clearly stated: these peripheral areas, sparsely inhabited but rich in natural resources, were also those in which ethnic minorities were concentrated. Because of their crucial strategic importance, the main targets were the north and the west: Inner Mongolia, bordering on the People's Republic of Mongolia and sharing a short border with the USSR; Xinjiang, two-thirds of whose borders touch the Soviet Union and Pakistan; and Tibet, neighbor of China's timeless rival, India. Strengthening the presence of the Han (the majority ethnic group, now representing over

90% of the total population) in these areas was also a way to dilute the indigenous populations and so dampen any separatist aspirations. Thus, to reinforce central authority, guarantee stability on its frontiers, and promote economic development, China set out to strengthen these sensitive regions. A Chinese general declared in 1965: "In recent years, rifle in one hand and shovel in another, officers and men [. . .] have preserved security on the borders of the Motherland and have worked to build socialism by engaging in agriculture," (according to an article published by *New China News Agency*, January 2, 1965, qtd. in Tien, 1973, p. 129).

China very quickly turned Xinjiang into one of its main pioneer lands, most notably because, during the phase of Sino-Soviet friendship, it was the link that was supposed to ensure continuity between economic spaces (Larivière and Sigwalt, 1996, p. 178). Xinjiang (literally "new frontier") was not incorporated into China's sphere until 1759, date of its conquest by the Manchu dynasty of the Qing. In the two centuries that followed, its ties to the central government remained very loose—so much so that, on several occasions, it found itself in a state of near-independence. Since 1949, the year marking the conquest of power by the Communist Party and the inauguration of effective nationwide control, Xinjiang has been ever more firmly anchored to China. Xinjiang is an object of both desire and fear. It is desired for the wealth of its subsoil and its strategic position, but loathed for its indigenous population: the rebellious, untamed Uigurs. As Turkophone Muslims, Uigurs are fixated on the pinnacles of Muslim civilization that are Istanbul, Samarkand, and Bokhara, rather than on Beijing. They possess the rich cultural traditions of Central Asian Muslims, but have nothing in common with the Han. Uigurs also have a strong sense of identity, largely tied to their geographic concentration on a territory of their own, where they form a majority, but of which they are gradually being dispossessed. In 2000, Uigurs, Kazakhs, and Kirghizes accounted for 53% of the total Xinjiang population. Hence ethnic grievances and separatist claims have risen, driven by rejection of Chinese supremacy (Becquelin, 1997, p. 10–21).

With respect to its ethnic minorities, the Communist regime installed in 1949 broke, at least in its rhetoric, from the narrowly homogenizing views of the Nationalists by emphasizing China's multinational character. Nevertheless, the goal of strengthening national unity persisted. In the 1950s, for economic and strategic purposes, the Communists implemented a Sinification policy to be achieved through population transfers. The radio and press appealed to the socialist consciousness, patriotism, and spirit of adventure by encouraging young people to migrate. There were several forms of planned migration: transfers of entire families for a defined period; transfers of married men, young or mature, sent off alone and to be joined later by their families; and transfers of young unmarried men and women for permanent settlement in the region in order to realize long-term projects (Blayo, 1997, p. 47). Migration was encouraged particularly to Xinjiang (Tien, 1973, p. 123), where the construction of irrigation systems and establishment of State-owned farms created a major need for migrant labor. Thousands of soldiers discharged from military duty (by becoming "peasant soldiers") were sent to the peripheral provinces to occupy key positions and disseminate the technical knowledge needed to improve agricultural and industrial production. As an incentive to compensate for their isolation and distance from their native regions, the settlers were offered various benefits, especially financial. These typically consisted of bonuses (Tien, 1973, p. 124–125). By 1959, 220 State-owned farms had been set up on 13 million *mu* (one *mu* is equal to 0.067 hectares) in Xinjiang, accounting for 30% of the province's farmland at the time. Around 1960, between 150,000 and 200,000 peasant soldiers were reportedly stationed in Xinjiang (Tien, 1973, p. 127).

Since 1949, the most sensitive regions have been massively settled. The number of Han in Inner Mongolia doubled between 1953 and 1964 from 5.1 million to 10.7 million. In Xinjiang, it rose sevenfold from 330,000 to 2.3 million in the same period—an annual average growth rate of 18%—and doubled between 1964 and 1982. Planned migrations were curtailed from late 1958 onward. But this did not spell the end of population transfers for economic, hegemonic, and strategic purposes. Today, 18.5 million Han are settled in Inner Mongolia, where they make up 80% of the province's population. Another 7.5 million Han live in Xinjiang, accounting for 40% of the inhabitants of the province—an increase of 34% since 1982 (Attané and Courbage, 2000, p. 257–280). The phenomenon is harder to measure in Tibet, where military personnel, who represent a sizable proportion of the Han presence, are excluded from census statistics. Whatever the validity of the figures, the Han surge is manifest, averaging 5% a year in 1964–1982 and 3% in 1982–2000. This is well above the increase achievable through natural growth, even at a brisk pace. In 2000, 160,000 Han were living in Tibet, twice as many as in 1990. (The Han population in Tibet fell between 1982 and 1990, before the revival of China's political and economic interest in its western provinces.) Officially, therefore, 6% of the Tibetan population is Han (Table 112–1). We should emphasize that, according to inde-

TABLE 112–1 Han Population (in Millions and %) in the Most Sensitive Border Provinces, According to Censuses, 1953–2000

Province	1953	1964	1982	1990	2000
Inner Mongolia	5.120 (83.9%)	10.743 (87.0%)	16.278 (84.5%)	17.290 (80.6%)	18.466 (79.2%)
Xinjiang	0.332 (6.9%)	2.321 (31.9%)	5.284 (40.4%)	5.695 (37.6%)	7.489 (40.6%)
Tibet (Xizang)	—	0.037 (2.9%)	0.091 (4.9%)	0.081 (3.7%)	0.159 (6.1%)

Sources: Chinese censuses (*Zhongo renkou pucha*) for corresponding years (NBS, 1985, 1993, 2002). For 1953 and 1964: 1949–1985 Yearbook (NBS, 1988).

pendent sources, these statistics leave out several hundred thousand military personnel, whose inclusion would raise the proportion of Han in Tibet to about 15%. Better than speeches, these colossal transfers of Han population to buffer zones reflect the strategic fears as well as the hegemonic designs of the central government, in the name of national unification.

Xinjiang—like all the other provinces of China's "Far West," for that matter—has not finished with the Han. After 20 years of economic liberalization, with the eastern coast being recognized as the sole beneficiary of the reforms, the government launched an ambitious program to "develop the West" in 2000. This strategy is not only economic but also political. Its aim is to buttress the power of the Beijing authorities through economic growth driven by the central government. Poverty in the remote regions is to be reduced by encouraging local investment in capital and technological advances. The program priorities are infrastructure, environment, and training. Large-scale projects are under way in transportation and communications, such as airports, rail links, highways, and telephone networks. Their unifying effect is unquestionable. But Xinjiang's value to the Chinese economy is largely due to its oil-production facilities (Domenach, 2002, p. 319–321). Behind this strategy also lies a calculation with regard to Buddhism and Islam, which the Chinese Communist Party (CCP) does not deny: "When the economy develops, the attention of the people will gradually shift from religion to the joys and pleasures of this world" (article in *Far Eastern Economic Review*, August 21, 2001, qtd. in Domenach, 2002, p. 319–321).

III. HEALTH POLICIES

Since 1949, as we have seen, China has gone through two distinct phases: (1) the hardline socialist era, marked by a quest for social equality (1950s to 1970s), and (2) the liberal era, which saw a transition toward the strangely syncretistic system of a socialist-style market economy (1980s and 1990s). The two crucial reforms were the decollectivization of agriculture and the restructuring of State-owned enterprises. Hitherto, collective institutions had taken charge of every Chinese from cradle to grave. By disrupting the overall social organization, the dismantling of those institutions had repercussions well beyond the economic sphere. Previously, each citizen, through his or her work unit, was strictly dependent on the State in all areas of everyday life: employment, housing, health, child education, retirement, social insurance, and so on. By taking down its collective structures, the State has, in effect, partially or totally handed over to the private sector a set of functions that it could no longer perform.

Public service, as developed by Mao Tse-tung in the 1950s and 1960s, was informed by the principles of egalitarianism and massification. The focus was on public health, with three main goals: rapid improvement of access to health care, prevention and mass campaigns, and creation of a comprehensive insurance system. Priority was assigned to training medical workers in charge of dispensing primary care, and to pharmaceutical production. This strategy enabled China to achieve remarkable progress—particularly, as we have seen, in combating infant mortality and improving life expectancy at birth. In less than three decades, rural areas were provided with infrastructure offering access to primary health care for all local residents. By the eve of the reforms, mass vaccination campaigns, most notably against the most lethal child diseases, had eradicated large-scale epidemics and the worst health scourges. Patriotic health campaigns (*aiguo weisheng yundong*), were launched, in particular to eliminate agricultural pests. One of these campaigns was aimed at eradicating the "four pests" (*si hai*): flies, mosquitoes, rats, and grain-eating sparrows. But when the elimination of sparrows turned out to have serious ecological consequences, they were replaced by bugs as targets.

Another example is the campaign against schistosomiasis. Known as "snail fever" in Chinese, this parasitic disease is caught by walking barefoot in contaminated water. The authorities informed the population about the disease's transmission, helped to screen for the disease, and facilitated treatment. A number of community initiatives, such as sewer construction, were taken to improve sanitation standards. Drawing inspiration from the Soviet system, China established centers for disease and epidemic prevention supervised by the Health Ministry in various regions (Liu and Hsiao, 2002, p. 391).

Other initiatives launched in the 1950s and early 1960s were largely modeled on the USSR, which provided a large share of technical assistance at the time. Several schools were opened to offer "university-level" medical training (typically in 6-year programs); some of the older schools were transferred from the cities on the eastern coast to regions with urgent needs, particularly in the west. Schools providing intermediate-level medical education (2- to 3-year programs) opened to train medical assistants, nurses, midwives, pharmacists, technicians, and other categories. In 15 years, these schools reportedly turned out more than 100,000 doctors, about 170,000 medical assistants, and 185,000 nurses—a more than 250% increase in the health care workforce. Between 1949 and 1965, the number of hospital beds rose from 84,000 (1 per 6,500 inhabitants) to 650,000 (1 per 1,200 inhabitants).

As with birth control, the health policies implemented from the 1950s onward owe their success largely to China's administrative organization. After the agrarian reform, collective property and central control by different echelons of government became the predominant type of social and political organization. The three-tier health care system was modeled on the administrative system. In rural areas, clinics were set up in villages, medical centers in towns, and hospitals in district capitals. Village clinics were staffed by low-skilled doctors, who had only 1 year of training after middle school. Town medical centers, which had a capacity of 10 to 20 beds, were supervised by doctors with 3 years of medical training after 6 years of secondary schooling. The practitioner was generally assisted by midwives, nurses specializing in maternal and child care, and village doctors. The district hospitals were far larger facilities (200–300 beds) staffed by fully qualified doctors with 4 to 5 years of medical training. This vertical organization allowed efficient treatment of patients. At the time, public hospitals charged for procedures and medicines, but the State-set prices were affordable for most people.

During the Cultural Revolution, which began in 1966, the role of Party leaders and the bureaucracy was reassessed. The widening disparities between urban and rural areas, between intellectuals and manual workers, and between Party leaders and the people were among the major dysfunctions denounced by the Cultural Revolution's chief instigator, Mao Tse-tung. In a directive of 1965, he reproached the Health Ministry for giving precedence to urbanites and neglecting the needs of the peasantry. In the years that followed, tens of thousands of urban health care workers were sent to the countryside on a permanent basis or as part of mobile medical teams. In 1975, rural areas numbered 1.6 million "barefoot doctors" who—despite their shoddy medical training (a 3-month program before being sent into the field) and the political turmoil of the Cultural Revolution—achieved a substantial improvement in health conditions and public hygiene.

TABLE 112–2 Life Expectancy at Birth and Infant Mortality per 1,000 since the 1950s

	1950–1955	1960–1965	1970–1975	1981	1989–1990	1995
e_0	40.8	49.5	63.2	67.7	69.6	69.6
m_0	138.4	85.6	51.5	37.7	33.4	34.0[a]

[a] In 1997.

Sources: 1st row: Peng Fei and Cambois, 2002. 2nd row: 1950–1955 to 1981: Huang and Liu, 1995; 1989–1990 and 1995: Attané, 2002c, p. 372.

A few years after Mao Tse-tung's death, the Chinese experienced a decisive turning point in their lives: the advent of economic reforms. The outstanding productivity gains that they induced have led to sizable improvements in real income, nutrition, and housing—and hence a decline in overall poverty. But this progress has benefited only a minority: the mainly urban populations of the eastern seaboard. At the national level, inequality has widened. Moreover, economic growth has not favored the pursuit of the health transition. The egalitarian ideology, whose top goal was to serve the people (*wei renmin fuwu*), soon gave way to the notion that it was better to let some people get rich. Once under the responsibility of people's communes, the health care system, now decentralized and on the way to privatization, has ceased to operate as a public service: patients are required to pay for treatment, while doctors and hospital managers must meet profitability criteria. Rural medical cooperatives, which provided medical insurance for a flat fee to 90% of the rural population in the late 1960s, now cover less than 10%. The State's fiscal allocation to rural health care shrank from 21% of the total health care budget in 1978 to 10% in 1991 (Cailliez, 1998, p. 38–45). Medical care has become unaffordable for the poorest citizens. The gap is widening: in 1997, a rural household spent an average 62 yuan per person on health care—three times less than urban dwellers (179 yuan). In 2001, the outlays came to 97 and 343 yuan respectively, an increase of 56% for the first group and 92% for the second (NBS, 2002). Decentralization and tax reform have transferred health care–system financing to local governments. Faced with a deep cut in subsidies, the system has been privatized in anarchic fashion. As most of its resources come from patients, it has neglected the unprofitable aspects of its mission. As a result, inequality has increased. Urban residents in the developed regions are getting the lion's share: an estimated 15% of the population absorbs 60% of public health care expenditures. China has no shortage of pharmaceuticals or doctors, but the improvements of the past 20 years are concentrated in cities. Most peasants have access only to village and canton facilities. Regional disparities in health statuses are extreme. According to World Bank data, in 1993, health care expenditures per capita in urban Guangdong were 33 times as high as those in rural Tibet. The differences measured by the main health care indicators are naturally just as large. For example, in the late 1980s, the maternal mortality rate was 18 per 100,000 in Shanghai, but 108 per 100,000 in Ningxia. Rising costs and the total loss of medical insurance have deprived a sizable proportion of the rural population of access to health care (Cailliez, 2002, p. 409–420).

Inequalities exist not only in access to health care, but in the quality of care. The canton medical centers (*xiang*)—a legacy of the people's communes—are usually able to treat serious cases and perform surgical procedures. But their medical personnel typically have no university training, only "hands-on" experience. A 1990 Health Ministry survey of 90,300 village doctors revealed that 73% did not go beyond secondary school. Many of them are former barefoot doctors whose title, for those who passed their formal exam, was officially changed to village doctors in 1981. Those who failed the exam became medical assistants. The care they provide is not always appropriate. For patients, the risk exposure can be high. Moreover, since 1985, clinics no longer have the monopoly of drug supply. It is now in the hands of profit-conscious private-sector producers, and there is no longer any way to effectively control the quality of pharmaceutical production (Chen, 1989).

The health care statistics compiled for the entire country are fairly reassuring. The average number of doctors per 1,000 inhabitants is rising steadily, having moved from 1.16 in 1982 to 1.67 in 2000. But the rural fringe of the population has been largely bypassed by these advances: the average number of doctors measured against the total population is 2.7 times smaller than in urban areas. The countryside gained 106,000 doctors between 1982 and 1990, but only 23,000 between 1990 and 2000. The total increase was 19% between 1982 and 2000 in rural areas, versus a doubling in urban areas; 187,000 hospital beds have been eliminated in rural areas since 1982, while their number was multiplied by 2.3 in urban areas: the ratio per capita in the countryside is at one-third the level in urban localities (Table 112–3).

TABLE 112–3 Medical Infrastructure in Urban Areas (Towns and Cities) and Rural Areas (Districts), 1952–2000

Type of settlement area	1952	1970	1982	1990	2000
Number of doctors per 1,000 inhabitants					
All China	0.74	0.85	1.16	1.54	1.67
Towns and cities			3.04	3.30	2.76
Districts			0.85	0.94	1.03
Number of hospital beds per 1,000 inhabitants					
All China	0.28	1.33	2.01	2.30	2.38
Towns and cities			4.03	4.68	4.17
Districts			1.53	1.48	1.32
Number of doctors (thousands)					
All China	424	702	1,307	1,763	2,076
Towns and cities	81	241	628	978	1,268
Districts	343	461	679	785	808
Number of hospital beds (thousands)					
All China	160	1,105	2,053	2,624	2,948
Towns and cities	121	510	832	1,387	1,914
Districts	39	595	1,221	1,237	1,034

Source: NBS, 2002.

The overall health status of the population has been stagnant since the start of economic reform. A decline in the effectiveness of preventive and remedial care thus seems to have offset the improvements due to the rise in general living standards. Having underestimated the consequences of its laissez-faire policy, China is struggling to cope with the negative effects of economic reform on the health care sector. The health gap between social categories is widening. The ratio of the infant mortality rate in rural areas to that of urban areas climbed from 1.67 in 1981 to 2.21 in 1989 and 3.17 in 1998; the ratio for life expectancy at birth fell from 0.94 in 1981 to 0.92 in 1998 (Attané, 2002b, p. 108–109, 372). The health of rural households, especially in the low-income bracket, is therefore feeling the impact of trends in the health care sector: a decline in prevention and quality of care, collapse of the social security

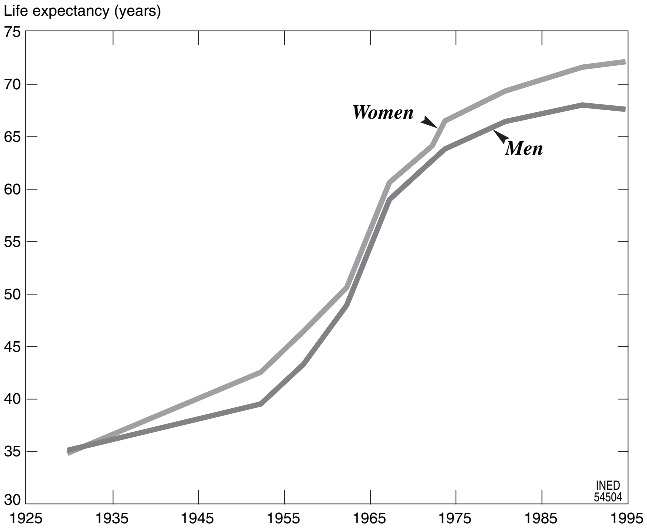


FIGURE 112–3 Chinese life expectancy since the 1950s. (Source: Peng and Cambois, 2002.)

system, and escalating medical costs. A growing share of the population now lacks any form of social protection. By 1983, only 40% to 45% of villages were still covered by the collective medical system. Between 1993 and 1998, the proportion of uninsured urban and rural residents rose from 46% to 58% and from 87% to 90% respectively. The poorly managed collective medical systems have lost credibility. Many people preferred not to subscribe when membership became optional in the 1980s, as the benefits offered usually proved inadequate. Nearly everywhere, the insurance covers only minor expenses, and families must bear the largest share of catastrophic medical expenses.

The government, increasingly aware of the problems created by the transfer of Welfare-State functions to the private sector, has recently conducted a major review of its health policies. This has resulted in embryonic reforms that, for the moment, consist merely of local initiatives tested in pilot projects.

IV. CONTROLLING FERTILITY

On taking power, the Communists promptly set up an authoritarian system designed to serve the regime's interests by regulating individuals and controlling their mobility. At the time, however, the Communists had no intention whatever of regulating population growth. Far from viewing it as problematic, socialist ideologues saw a large population as a force for economic prosperity: unemployment—incompatible with socialism—was not to be feared; collectivism would make food, health, and education available to all. Moreover, even if the Communists had considered such an intervention, it would have lacked a scientific basis. Absent reliable sources, very little was known about the demographics of the world's most populous country. In 1949, China was believed to have fewer than 500 million inhabitants—well below its maximum carrying capacity, arbitrarily estimated at 800 million. The early years of the regime were therefore marked by openly pronatalist rhetoric. Regulating fertility was "reactionary" and "capitalist," and the most precious capital of revolutionary China was its vast population.

The first republican census (1953) revealed that the population was 590 million, or 100 million more than expected. This sparked fears that excessively rapid population growth would compromise economic development. Only then did young socialist China engage in a Malthusian debate. Temporarily convinced by birth-control advocates, Mao Tse-tung declared in 1956: "Except for areas inhabited by national minorities, it is necessary to publicize and popularize fertility regulation and promote birth control in all densely populated regions" (Zou, 1986, p. 4). A few months later, the first birth-control campaign was launched. The contradiction between population growth and economic development was abruptly denounced. China started producing contraceptives and liberalized sterilization and abortion.

Owing to the lack of efficient methods, the shortage of skilled personnel, and the traditional Chinese reluctance to discuss sexual matters, family-planning propaganda remained confined to urban areas. This initial attempt had no visible effect on fertility. The campaign against right-wingers unleashed in 1957 put intellectuals—and birth-control militants along with them—on trial. Demography, like the other social sciences and the humanities, became a taboo science again. Population growth ceased to be regarded as a problem per se; it was now a phenomenon that should self-regulate. Industrial development was restored as a priority. With the Great Leap Forward, launched in 1958, the authorities feared that the labor supply would fall short of what was needed to build socialism. The link between population and production triggered a violent attack on family-planning advocates. All reference to population control was viewed as a reactionary opinion. Efforts to regulate fertility suddenly stopped, and contraceptive production was halted.

A few years later, however, birth control was back on the agenda. In the Black Years (1959–1961) that followed the Great Leap Forward, a terrible famine caused an estimated 30 million deaths, exacerbating the imbalance between food supply and population. The birthrate plunged. In 1960—the darkest year—

deaths exceeded births, causing a net population decline of 3 million people. Then came the recovery. The millions of children whose conception had been postponed because of famine, political instability, and the economic crisis swelled the ranks of already large cohorts. China registered an exceptional baby boom in the 1960s, with 25 to 30 million children born every year. The government discreetly drew the lessons from the Great Leap Forward. It made agriculture a priority again, and the problems of population growth were no longer ignored.

In 1962, a second campaign was launched. To reduce the number of births, the authorities opted for "soft" methods: delayed marriage and wider distribution of contraceptives. The number of children was limited to two or three, with a mandatory spacing of 3 to 5 years. The propaganda machine swung into action. Young people were warned of the negative effects of overly precocious sexual activity on their physical and intellectual health. Early love and marriage were stigmatized for undermining revolutionary enthusiasm and so jeopardizing the construction of socialism. Abortion was still liberalized in 1962: it could now be performed at the sole request of the pregnant mother and it was free if she was married. The intra-uterine device (IUD or loop) was introduced and vasectomy strongly encouraged. This second attempt at birth control, better organized and more pragmatic, was relatively successful in some big cities. But before it had a chance to spread to rural areas, it was swept away by the fervor and turmoil of the Cultural Revolution in summer 1966.

In the first two decades of socialism, official positions therefore oscillated between ideological dogmatism and economic realism, each outweighing the other in turn. Not until the end of the Cultural Revolution did a consistent discourse finally emerge. Malthusian pragmatism prevailed—lastingly.

At the start of the 1970s, the demographic transition was still modest. Mortality was declining. Life expectancy had gained 20 years since the 1949 Revolution, from 41 years in 1950–1955 to 60 years in 1965–1970 (Huang and Liu, 1995, p. 4) (Figure 112–3). But the birthrate kept rising, to all-time highs. With fewer deaths and more births, population growth peaked at more than 2% per year, hitting 2.8% in 1968. Twenty million people were being added each year. After having been forgotten for some years, birth control became a national priority again. In 1971, the Council of State's Directive 51 marked the official launch of the third birth-control campaign, which would be pursued relentlessly in the following decades. Three directives were issued in 1973, summarized by the slogan "late, spaced, few" ("*wan*, *xi*, *shao*"), that is, marry late, space births, and have fewer children (Attané, 2002). Annual birth quotas were imposed. However, the infinite diversity of settlement patterns, cultures, contexts, constraints, modes of production, and other factors did not escape the promoters of this third campaign. The population was officially classified into three groups: urban, rural, and ethnic minorities. From the outset, urbanites were subjected to the strictest rules: no marriage before 25 for women, before 28 for men, and no more than two children. By comparison, peasants were privileged: they were allowed three children and their minimum legal age at marriage was set at 23 for women, 25 for men. Both urban and rural dwellers, however, were required to space births by at least 3 to 4 years. No prescribed conduct was defined for ethnic minorities. In any event, as small groups mostly confined to low-density peripheral areas, they did not have a major role to play in fulfilling the birth-control objective. (For more information on the implementation of birth control among ethnic minorities, see Attané and Courbage, 2000; Courbage, 2002.)

In less than 10 years, fertility was halved—from an average 5.7 children per woman in 1970 to 2.8 in 1979 (Table 112–5 and Figure 112–4). This was the steepest decline ever recorded in the world in so short a time span. In urban areas, women were already having fewer children than needed for cohort replacement: the number had fallen to 1.4, compared with the 2.1 required to ensure that two children replace their two parents, allowing for mortality. The results were so good that observers later concluded—correctly, in

TABLE 112–4 Crude Birth and Death Rates (per 1,000), 1953–2000

	1953	1960	1970	1980	1990	2000
CBR	37.0	20.9	33.4	18.2	21.1	14.0
CDR	14.0	25.4	7.6	6.3	6.7	6.4

Source: Attané (2000a).

TABLE 112–5 Total Fertility Rate, 1970–2000

Type of settlement area	1970	1975	1980	1985	1990	2000
All China	5.7	3.6	2.3	2.1	2.3	1.8
Urban areas	3.2	1.8	1.2	1.3	1.6	1.2
Rural areas	6.3	4.0	2.6	2.4	2.8	2.0

Sources: 1970–1985: Chen and Coale, 1993; 1990: Peng Xizhe, 2002, p. 61; 2000: Gu, 2003, p. 1–2.

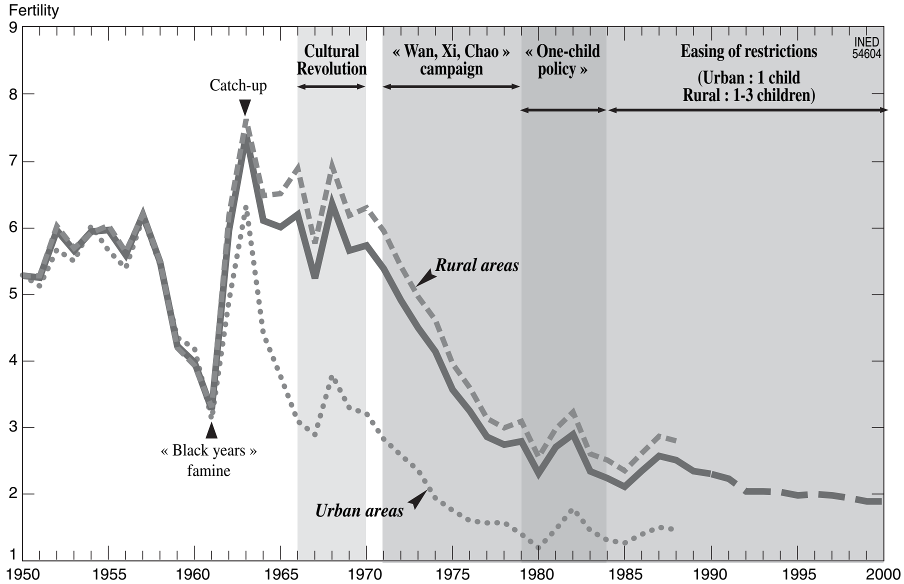


FIGURE 112–4 Total fertility rate since 1950. (Sources: 1950–1988: Chen and Coale, 1993; 1989–2000: Attané, 2000a.)

fact—that the Chinese population, both rural and urban, was simply awaiting a modest incentive from the authorities (as it turned out, better access to modern contraception) to have fewer children. Infant mortality had fallen sharply, from 175% in 1953 to 76% in 1969 (Peng and Cambois, 2002, p. 99). Parents thus no longer needed to maintain as large a "store" of children to offset early deaths and provide for their old age.

China owed this extraordinary success also—indeed, mainly—to the tremendous efficiency of its collective institutions. Nobody could escape the strict control and extreme vigilance of people's communes and neighborhood committees. Their arsenal included intimidation visits to private homes, public denunciations, collective pressure, and deprivation of ration coupons. Women's menstruation calendars were posted at factory gates; any woman suspected of pregnancy was hounded. The work unit conditioned every aspect of private life—not only work, of course, but also health, children's education, housing, food supply, and so on. The means of coercion were formidable. Whoever tried to evade the family-planning injunctions took huge risks. This also explains why birth-control measures were obeyed so unanimously at the time.

At his death in 1976, Mao Tse-tung left China with an ailing economy, in the wake of a Cultural Revolution fraught with heavy economic and social consequences. In the demographic sphere, by contrast, the birth-control program had already achieved remarkable results and dispelled the threat of a population boom that the country could not have supported. But this was still not enough. Signs pointed to a resurgence of the birthrate that threatened to jeopardize these fragile gains. Birth control had to be tightened even further. The reform and opening policy (*gaige kaifang zhengce*), inaugurated in 1978 by Mao Tse-tung's successor Deng Xiaoping, was therefore assigned a second objective: to curb population growth in order to finally ensure economic take-off. The then-Premier, Hua Guofeng, declared in 1979: "We must control population growth effectively and plan births. This is a prerequisite for the development of the national economy, and a problem that we need to solve so that we can achieve the Four Modernizations" (Chen, 1979). To meet this goal, the new policy—officially announced in January 1979—imposed the draconian rule of the single child. Moreover, the rule applied to almost all of China: 95% of urban dwellers and 90% of rural dwellers were supposed to comply. To promote the measure, the authorities established a reward system. Couples had to pledge to have only one child by signing the one-child certificate (*du sheng zinü zheng*). In exchange, they would receive various compensations, which varied considerably from one locality to another: monthly bonuses of a few yuan, free medical care and education for the child, easier access to housing, allocation of an additional small plot for peasant families, retirement bonus for employees of State-owned businesses, and so on. These benefits were not negligible, at least in theory (in some of the poorer regions, such commitments were hard to honor in practice). Some certificate signers merely received a thermos bottle or a blanket. Couples who resisted the government injunctions—particularly those who gave birth to a second child despite having signed the certificate—faced a variety of penalties, such as the obligation to reimburse the bonuses received, income deductions, fines, partial confiscation of the family plot, and dismissal for recidivist employees of State-owned businesses. (For more details on the implementation of the one-child policy, see Croll *et al*., 1985; Banister, 1987; Bianco, 1989; Aird, 1990.)

A quota system was introduced in 1973. Defined at the national level, the quotas were quite simply permissions to have children. They were distributed among the regions, down to the smallest administrative units: the village or, in towns and cities, the neighborhood. The cadres at these lowest levels were in charge of allocating the quotas among couples who wanted to have a child. As a rule, priority was assigned to young married couples, eager to make their unions fertile. Once all the quotas had been handed out, attempts were made to convince the other couples to give up their plans to have children, at least temporarily. If the birth nevertheless occurred, it was counted as an "out-of-quota" birth (*jihua wai shengyu*). In the late 1980s, 40% to 45% of rural births, or 7 to 8 million a year, were unauthorized. The cohorts of women born in the early 1960s (1959–1963) had reached childbearing age when the one-child policy came into effect. After only10 years of fertile life, they had already reached or exceeded the maximum allowed number of children in three-quarters of all provinces (Table 112–6) (Attané, 2002a, p. 103–113).

The one-child requirement was in total conflict with the interests of families, particularly in rural areas. It sparked stiff grassroots resistance (Attané, 2000b). As a result, the government was very soon forced to relax the conditions it had imposed on rural couples wanting a second child. The decision was officially announced in April 1984 in the Central Committee's Document no. 7 (Blayo, 1997, p. 184). While the stated goal was to allow a greater proportion of couples to have a second child, the main aim was to "open a small breach to plug a bigger one" (*kai xiao kou du da kou*), that is, to legitimate a few exceptions to the one-child rule in order to avoid the largest number of them. Since then, China's birth-control policy has been a two-speed process.

Limiting births became a constitutional duty in 1982: "The State encourages family planning to ensure harmony between demographic growth and economic and social development plans" (art. 25); "Husband and wife alike have the duty to practice family planning" (excerpt from art. 49). (For more information on the causes and manifestations of this resistance, see Attané, 2000b.) This obligation is also stipulated in the 1980 Marriage Act: "Family planning is practiced" (excerpt from the 1982 Constitution of the People's Republic of China [*1982 zhonghua renmin gongheguo xianfa*]) (excerpt from art. 2). However, until very recently (September 2002), no legislation had been passed to set out the principles of birth-control policy at the national level. (The provisions of the act are described in the Appendix.) Because of China's immensity and diversity, its leaders have generally preferred the adoption

TABLE 112–6 Officially Stipulated Completed Fertility (*Zhengce Yaoqiu Zhongshen Zong Shengyu Lü* [ZSSYL]) and Actual Fertility for Women Aged 25–29 in 1988

Province	ZSSYL (1)	D_{P25-29} (2)	Ratio (2)/(1)	Province	ZSSYL (1)	D_{P25-29} (2)	Ratio (2)/(1)
Beijing	1.33	1.11	0.831	Henan	1.56	2.00	1.279
Tianjin	1.35	1.12	0.830	Hubei	1.55	2.16	1.394
Hebei	1.67	1.87	1.117	Hunan	1.64	2.16	1.314
Shanxi	1.69	1.92	1.133	Guangdong	1.85	2.05	1.108
Inner Mongolia	1.80	1.80	1.000	Guangxi	1.57	2.21	1.404
Liaoning	1.50	1.29	0.857	Hainan	1.97	2.20	1.117
Jilin	1.50	1.39	0.927	Sichuan	1.57	1.48	0.939
Heilongjiang	1.44	1.54	1.069	Guizhou	1.74	2.35	1.348
Shanghai	1.28	0.95	0.742	Yunnan	2.13	2.26	1.059
Jiangsu	1.52	1.47	0.964	Shanxi	1.64	2.20	1.338
Zhejiang	1.54	1.61	1.042	Gansu	1.58	2.36	1.494
Anhui	1.61	2.11	1.311	Qinghai	2.08	2.12	1.019
Fujian	1.61	2.16	1.339	Ningxia	2.06	2.57	1.248
Jiangxi	1.52	2.36	1.549	Xinjiang	2.40	2.58	1.073
Shandong	1.55	1.62	1.042	**China**	**1.62**	**1.85**	**1.142**

Source: Attané 2002a, p. 106.

of provincial regulations that take local conditions into account. Accordingly, each province enacted its own birth-control regulations in the 1980s. They are unanimous as regards the minimum legal age at marriage: in compliance with the 1980 Marriage Act, it is set at 20 years for women and 22 years for men in all provinces. All the regulations also stipulate that anyone "who is of reproductive age and is not preparing to conceive a child legally must use a reliable contraceptive method." Some even prescribe the method suited to each situation: "a couple with one child must use contraceptive methods with long-term efficiency [*chang xiao*]" such as the IUD (loop), and "a couple with two or more children must undergo sterilization." In the event of unplanned pregnancy, most regulations require taking remedial measures (*bujiu cuoshi*), namely, immediate abortion. (For the texts of these regulations, see *Zhongguo renkou ziliao shouce* [Chinese population data], 1985.) Other regulations, as in Sichuan, also stipulate a contribution to be paid monthly starting in the early months of an unplanned pregnancy. Most provide for economic penalties—in the form of fines (50–200 yuan in the mid-1980s) or withholdings on wages (10%–30%)—for women falling pregnant without permission. These sums are to be fully refunded, however, if the couple undergoes an abortion. Fines for out-of-plan births now run into the thousands or even tens of thousands of yuan—sometimes more than the household's annual income (1 yuan is currently worth about 0.11 euros).

Since 1984, the one-child rule is thus no longer systematically enforced in rural areas. There, families—in particular among certain ethnic minorities—are generally allowed to have a second and even a third child. The eligibility criteria for a second child, however, can vary from one province to another and sometimes even from one district or village to another. Residents of the southern province of Guangdong, for example, are not necessarily subject to the same measures as their compatriots in the north or west, in Shanxi or Sichuan. Each administrative division has enacted its own child-quota regulations and defined rewards and penalties—all, in theory, respectful of the local context. China is administratively divided into 22 provinces, 5 autonomous regions, 4 municipalities, and 2 Special Autonomous Regions (not included here): Hong Kong and Macao.

The official indicator produced by the central government in the mid-1980s spells out birth-control targets at the provincial level (Table 112–4). This indicator, called "completed fertility officially stipulated in the framework of the birth-control policy" (*zhengce yaoqiu zhongshen zong shengyu lü* [ZSSYL]) (Yin, 1995), theoretically defines the ceiling for the cohort fertility quantum. However, it has not been used as such, but has been translated into a period indicator setting the maximum mean number of children per woman at 1.62. Its values range from a low of 1.28 in the heavily urbanized municipality of Shanghai to a high of 2.40 in the autonomous Uigur region of Xinjiang, where urbanization is weak and most inhabitants belong to national minorities.

Today, more than 20 years after the adoption of the one-child policy, family planning still relies on the same weapons: persuasion, coercion, and fines. But reforms have led families to gradually emancipate themselves from collective institutions. Birth control is slipping out of the hands of the regime's cadres, and coercive measures are failing. To remain effective, regulations need to be adjusted. By promoting local initiatives, decentralization has made it possible to institute new types of penalties more directly targeted at family interests. A district in Guangdong, for example, developed a family monitoring system in 1989. Every couple that meets the requirements of the family-planning authorities receives a birth-control certificate (*jihua shengyu zheng*) that must be produced when applying for certain administrative authorizations. Rule-breakers may be denied a driver's license for a motorcycle or the mandatory permit for starting a retail business. If a woman conceives without permission, cadres have a right to cut off her water or electricity supply. In the same district, the fine for an out-of-plan birth came to 17,000 yuan (1,800 euros) in the mid-1990s. Couples who fail to comply with the required birth spacing pay a fine of 5,000 to 8,000 yuan. The mere presumption of prenuptial conception (any child born to a couple married for less than nine months) also entails a 5,000-yuan fine. A woman who has not made one of the four annual visits to check her IUD is liable to a 1,000-yuan fine; if she has not been sterilized within four days of the birth of her second child, she must pay 10 yuan per day of delay. (On this topic, see Guangdong Licheng qiangya fünu [Women forcefully taken away in Licheng, Guangdong], *Ouzhou ribao,* March 26, 2002.)

The Population and Birth Control Act passed in September 2002 (see Appendix) reasserts the goal of strict limits on the number of children. However, recent developments suggest changes in the means used to reach the objective. The program's emphasis has been shifting toward voluntarism for two reasons: (1) because it is increasingly difficult for the authorities to interfere in couples' private lives and (2) because the threat of unsustainable population growth has now been eliminated. The new focus is on health: reproductive health, education, and information are the main priorities.

V. TOWARD REPRODUCTIVE HEALTH POLICIES

The International Conference on Population and Development (Cairo, 1994) and the World Conference on Women (Beijing, 1995) were milestones in China's approach to its demographic issues. Its natural growth is now curbed, at a rate of less than 1% a year since 1998 and only 0.7% in 2001. Accordingly, China has decided to redefine the objectives of its family-planning program, hitherto exclusively centered on the changes in demographic indicators. Without abandoning its strictly demographic concerns, the program is gradually shifting toward the broader concept of reproductive health, focused on individual needs and the quality of services.

With the participation of the Chinese Family Planning Association and its local branches, pilot experiments are being tested in several regions. There, birth quotas have been scrapped, births are no longer subject to prior approval, and couples can decide by themselves when they want to have children. Contraception is no longer mandatory: couples are now free to choose the contraceptive method they prefer. Financial punishments for couples violating birth-control regulations have been abolished. The emphasis now is on information about reproductive health and child education. Demographic indicators such as the birthrate are no longer the sole criteria for assessing the program's performance. At present, these experiments have been extended to more than 800 districts (*xian*) and prefectures nationwide (see Gu *et al*., 1999, 2001).

For many years, the single objective was to curb population growth—at the expense of reproductive health. That is one of the reasons for the strong current prevalence of various infections of the reproductive system and sexually transmitted diseases. In many rural regions such as Yunnan, more than half the women suffer from such conditions. With financial support from the International Planned Parenthood Federation (IPPF), the local branches of the Chinese Family Planning Association offer help for diagnosing and treating these diseases. Other programs are designed to promote awareness of reproductive health and AIDS-related issues among adolescents. In 2002, more than 150 districts launched projects on teenage reproductive health, but these initiatives are still in their infancy.

As regards AIDS, China has started to launch information campaigns to warn about the risk of transmission. For example, in the Hunan city of Huaihua—a major center of emigration of young workers, particularly to Guangdong—the Chinese Family Planning Association has distributed information manuals for migrants and their families. A prevention campaign was launched in Henan, one of the provinces hardest hit by AIDS.

TABLE 112–7 Awareness of Sexually Transmitted Diseases and Access to Preventive Care, 2001

Women's level of awareness	Proportions of women (%)			
	All China	East	Center	West
Have heard about sexually transmitted diseases	63.6	69.8	65.2	53.4
Have heard about AIDS	72.7	78.4	75.1	62.1
Have been tested for infections of the reproductive system	42.8	50.1	40.3	35.6

Source: Gu Baochang, 2003, p. 6.

The wide interregional disparities in access to information and services reflect economic gaps and the unequal access to medical care. Here as well, the central and western regions lag the east, with a lower public awareness of sexually transmitted diseases and reproductive diseases, and less access to medical care (Table 112–7).

In every respect, the central regions and, even more so, the western regions are disadvantaged in terms of reproductive health. Access to contraception is much more restricted, and childbirth in hospital—the norm in the developed eastern provinces—is very infrequent: one in four births in Guizhou, one in five in Tibet (Table 112–8). While infant mortality rates in Beijing and Shanghai converge toward those of the most developed countries (approximately 5%), in Xinjiang more than 1 in 20 children die before their first anniversary. Maternal deaths have reached residual values in Beijing and Shanghai (fewer than 10 per 100,000 women), but are10 to 40 times as frequent in the western provinces: over 140 per 100,000 in Guizhou, Qinghai, and Xinjiang, and 466 in Tibet. By emphasizing reproductive health, the government hopes to eventually cancel these huge disparities.

VI. ASSESSMENT

In the past 50 years, China has experienced drastic upheavals—not only political and economic, but also social and demographic. In many ways, today's China bears no comparison to Mao's, which, in turn, saw itself as the antithesis of imperial China. These upheavals, driven by all manner of authoritarian policies, have been anything but spontaneous. In regard to

TABLE 112–8 Reproductive Health Indicators in Selected Eastern and Western Provinces, 2000

Province	Contraceptive use (%)	Childbirth in hospital (%)	Infant mortality rate (per 1,000)	Maternal mortality rate (per 100,000)
Eastern provinces				
Beijing	87.8	99.6	5.4	9.7
Tianjin	91.1	99.0	9.4	18.6
Shanghai	91.0	100.0	5.1	9.6
Jiangsu	91.1	98.9	11.2	28.5
Zhejiang	91.1	98.7	15.6	19.6
Western provinces				
Guizhou	90.1	25.8	38.8	141.7
Tibet	71.1	20.1	35.3	466.3
Gansu	87.5	55.3	28.9	108.8
Qinghai	86.9	38.0	41.0	142.0
Xinjiang	82.7	56.6	55.5	161.4

Sources: column 1: National Birth Planning Commission; columns 2–4: CWC, (see Gu, 2003).

migration, health, and the demographic transition, the present situation is largely the result of government initiatives. In the past half-century, China has thus resorted extensively to population policies, which have left deep marks on the current demographic landscape.

Assessing these policies is difficult because of their all too frequent lack of clearly defined initial objectives. The goal of controlling individual mobility was unquestionably met—at least for as long as the measures were strictly enforced. The urbanization process was largely curtailed between 1950 and 1978, sparing China the negative effects of the rural exodus that was already afflicting many developing countries at the time. But it is hard to say as much for the recent period. To satisfy the labor demand generated by economic reforms, but also to provide new jobs for the excess workers freed up by agrarian decollectivization and the restructuring of State-owned companies, China is now turning a blind eye to massive internal migrations. As many as 130 million Chinese (1 in 10) are believed to be working and living—at least temporarily—in a locality other than their registered place of residence. Yet, in the formal regulations, almost nothing has changed. Officially, every Chinese is required to reside in his or her place of birth, despite the fact that this rigidity is in total contradiction with the interests of the new market economy.

The reason for this caution is obvious: the government is reluctant to loosen its policy because it fears having to cope with an even more massive influx of rural dwellers into towns and cities—with all the attendant consequences. Rather than legislate on the issue, it is content to tolerate these internal migrations, which, in any event, serve its economic objectives; in so doing, the authorities reserve the right to tighten their control if and when needed.

The same can be said about birth control. Instead of openly relaxing its policy, China has opted for a measure of laissez-faire—even though a liberalization would, in the longer run, slow the aging process and attenuate the sex imbalance at birth due to the strong preference for sons. For now, the regime prefers to tolerate some resistance and is merely setting up pilot projects without changing its official rhetoric. The Population and Birth Control Act (2002) remains unambiguous, its wording identical to that of earlier regulations: "Couples are encouraged to have only one child" (see Appendix). The policy has ultimately met the official objectives. In 1980, Hua Guofeng announced that the population should not exceed 1.2 billion by the end of the century. The official figure from the 2000 census is a not too distant 1.265 billion. The target has since been revised upward: the latest official declarations state that the population should be kept at under 1.4 billion between now and 2010, which will manifestly be the case. (Population goals: less than 1.4 billion by 2010, *China Daily*, November 6, 2000). The goal of strict controlled population growth may now be regarded as having been reached.

But can we speak of clear-cut success when we see the extent to which individual freedoms have been violated in order to obtain it? (see Aird, 1990).

Or when we see the extent to which the reduction in fertility has been achieved at the expense of girls, whose percentage of the total population is becoming ever more abnormally small in the youngest age groups? The sex ratio of births was normal on the eve of the implementation of the one-child policy, at 105.8 boys per 100 girls in 1979. By the 2000 census, it had risen to 116.9. Or when we observe that nothing has been done, concurrently, to try to improve the health and status of women? (see Attané, 2002a).

We shall not take the risk, therefore, of describing Chinese population policies as failures or successes. What we can say is that, until now, China has never spared any effort to attain its goals. It sought to curb population growth: it succeeded. It set out to develop its economy: it is doing so. But at what price? The world's most populous country is now also one of the most inegalitarian. But it is also a country in which, despite a lingering totalitarianism, authoritarian measures are being accepted with ever-greater difficulty when they seek to regulate the family and personal intimacy. China is opening up to the world, and the world is getting involved in China. Therein lies the hope that, in future, Chinese population policies will be less coercive and more respectful of people.

APPENDIX

Description of Population and Birth Control Act of the People's Republic of China (enacted September 1, 2002)

The first allusion to the need to promulgate such a law dates from 1978, a few months before the implementation of the one-child policy. But the diversity of local conditions and the difficulty of legislating at the national level on so sensitive an issue delayed its drafting and adoption for more than two decades. The Act was eventually completed in December 2001, consolidating the provincial regulations that had formed the basis for Chinese birth-control policy since the 1980s.

The Act chiefly provides a legal framework, and no longer a merely regulatory framework, for family-planning activities. But it does not introduce any major change in the goals and implementation of birth-control policy.

Birth control remains a "fundamental national policy" (excerpt from art. 2).

Decentralization is still a guiding principle: "People's local governments supervise family-planning work in their administrative jurisdictions" (excerpt from art. 5).

"People's local governments at the district level and above, based on the population development plan, shall prepare a population and birth-control plan and shall be in charge of applying it at the local level. The departments of family-planning administration in the People's local governments at the district level and above shall be responsible for the field implementation of the daily family-planning activities set out in the population and birth-control plan" (excerpt from art. 10).

The Act's injunctions concerning the proper behavior for limiting births are virtually unchanged from those of the provincial regulations.

> Couples must practice late marriage and procreation [*wanhun*, *wanyu*]. They are encouraged to have only one child. Those who meet the requirements spelled out in legal and regulatory provisions may have a second child. [. . .] Ethnic minorities are also required to practice birth control, in keeping with the specific measures in effect in the provinces, autonomous regions, and municipalities [to which they belong]. (excerpt from art. 18)

"Birth control rests on the use of contraception. [. . .] The State must create the conditions that will enable citizens to choose and use safe, effective, and suitable means of contraception in order to control births" (excerpt from art. 19).

"Couples that already have a child are encouraged to use a contraceptive method with long-term effectiveness [*changxiao*, i.e., IUD or sterilization]" (excerpt from art. 34).

"Couples of reproductive age shall use family-planning methods to prevent pregnancies and control births. Couples shall prevent and reduce the number of unwanted pregnancies" (excerpt from art. 20).

"Couples of reproductive age shall have free access to contraceptive methods and family-planning technical services. The State shall reward couples practicing birth control, in accordance with existing regulations" (excerpt from art. 21).

Couples volunteering to have only one child shall be issued an "honor certificate for one-child parents" (excerpt from art. 27).

The subsequent articles contain a set of provisions already included in other legislation, such as the Protection of Minors Act (1991), the Women's Rights and Interests Protection Act (1992), and the Mothers' and Children's Health Act (1994).

"It is strictly forbidden to use ultrasound scans or any other technique to detect the sex of the fetus, except for medical purposes. It is strictly forbidden to terminate a pregnancy [for the sole purpose of selecting the child's sex], except for medical reasons" (art. 35).

The following crimes shall be punishable in accordance with law:

1. Illegal practice of surgical family-planning procedures on a third party;

2. Performance of ultrasound scans or use of any other technique on a third party to detect the sex of the fetus, except for medical purposes, and the voluntary termination of a pregnancy for the purpose of selecting the child's sex;
3. "Performance of false family-planning procedures, issuance of false medical certificates or false family-planning certificates" (excerpt from art. 36).

A State civil servant who commits one of the acts listed below in the course of his or her family-planning activities [. . .] shall be prosecuted, in accordance with law:
[. . .]
1. requesting or accepting a bribe [guilty of corruption]
2. falsifying birth statistics, under-reporting, or refusing to register a birth. (excerpt from art. 39)

Acknowledgement

This chapter was translated from the French by Jonathan Mandelbaum.

References

Aird John, 1990. *Slaughter of the innocents*. Washington, DC, AEI Press, 196 p.

Attané Isabelle, 2000a. La fécondité chinoise à l'aube du XXI[e] siècle: constats et incertitudes, *Population*, vol. 55(2), p. 233–264.

Attané Isabelle, 2000b. *La planification familiale en Chine, pour ou contre la femme?* Paris, CEPED, 55 p. (Les dossiers du CEPED, no. 60).

Attané Isabelle, 2002a. China's family planning policy: an overview of its past and future, *Studies in Family Planning*, vol. 33(1), p. 103–113.

Attané Isabelle (ed.), 2002b. *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société.* Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Attané Isabelle, 2002c. Un demi-siècle de socialisme chinois: heurs et malheurs de la famille paysanne, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société*, p. 355–386. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Attané Isabelle and Courbage Youssef, 2000. Transitional stages and identity boundaries: the case of ethnic minorities in China, *Population and Environment*, vol. 21(3), p. 257–280.

Banister Judith, 1987. *China's changing population.* Palo Alto (California), Stanford University Press, 488 p.

Becquelin Nicolas, 1997. Pékin et l'Asie centrale après la fin de l'URSS, *Perspectives chinoises*, no. 44, p. 10–21.

Bianco Lucien, 1989. La population chinoise face à la règle de l'enfant unique, *Actes de la recherche en sciences sociales*, no. 78, p. 31–40.

Blayo Yves, 1997. *Des politiques démographiques en Chine.* Paris, INED and PUF, 410 p. (Travaux et documents, Cahier no. 137).

Cai Fang, 2002. Les migrations internes en Chine: mains "invisibles" et pieds "visibles," *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société,* p. 169–184. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Cailliez Charlotte, 1998. L'effondrement du système de santé rural, *Perspectives chinoises*, no. 47, p. 38–45.

Cailliez Charlotte, 2002. Le système de santé rural: de l'utopie à l'effondrement, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société*, p. 409–420. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Chen C. and Bunge Frederica M., 1989. *Medicine in rural China: a personal account.* Berkeley, University of California Press, 218 p.

Chen Muhua, 1979. Shixian si ge xiandaihua, bixu you jihua de kongzhi renkou zengzhang (Achieving the Four Modernizations requires planned control of population growth), *Renmin ribao*, no. 11 (August).

Chen Shengli and Coale Ansley, 1993. *Zhongguo ge sheng shengyu lü shouce, 1940–1990 (Fertility data for Chinese provinces, 1940–1990).* Beijing, Zhongguo renkou chubanshe, 215 p.

Courbage Youssef, 2002. Les minorités nationales en Chine: l'exception démographique?, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société*, p. 215–254. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Croll Elizabeth, Davin Delia, and Kane Penny, 1985. *China's one-child family policy.* London, The Macmillan Press, 237 p.

CWC, 2002. *Final evaluation report on China's women development program (1995–2000)*. Beijing, The State Council, Committee on Women and Children.

Domenach Jean-Luc, 2002. *Où va la Chine?* Paris, Arthème Fayard, 392 p.

Gu Baochang, 2003. Population, reproductive health and poverty in China, in Richard Leete, Rabbi Royan, and Mickie Schoch (eds.), *Population and poverty. Achieving equity, equality, and sustainability*, p. 63–79. New York, UNFPA, 150 p. (Population and development strategies, no. 8).

Gu Baochang, Simmons Ruth, and Szatkowski Diana, 2001. Offering a choice of contraceptive methods in Deqing county, China: changing practice in the family planning program since 1995, *in*: Nicole Haberland and Diana M. Measham (eds.), *Responding to Cairo: case studies of changing practice in reproductive health and family planning*, p. 58–73. New York, Population Council, 462 p.

Gu Baochang, Zhang Erli, and Xie Zhengming, 1999. Toward a quality of care approach: reorientation of the family planning programme in China, *in*: Jay Satia, Patricia Mathews, and Aun Ting Lim (eds.), *Institutionalising reproductive health programmes. Innovations.* vol. 7–8, p. 39–52. Kuala Lumpur (Malaysia), ICOMP, 239 p.

Huang Rongqing and Liu Yan, 1995. *Zhongguo renkou siwang shujubiao (Data on mortality in China).* Beijing, Zhongguo renkou chubanshe, 307 p.

Larivière Jean-Pierre and Sigwalt Pierre, 1996. *La Chine.* Paris, Masson and Armand Colin, 288 p.

Li Jianmin, 2002. La population de la Chine à l'horizon 2050, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société*, p. 255–326. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Liu Yuanli and Hsiao William, 2002. Les soins de santé en Chine: histoire de deux époques, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI[e] siècle. Questions de population, questions de société*, p. 389–407. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Meissner Werner, 1990. La voie orthodoxe (1949–1955), *in*: Marie-Claire Bergère, Lucien Bianco, and Jürgen Domes (eds.), *La Chine au XX[e] siècle de 1949 à aujourd'hui*, vol. 2, p. 9–33. Paris, Arthème Fayard, 448 p.

NBS (National Bureau of Statistics [China]), 1985. *Zhongguo 1982 renkou pucha ziliao (Data on the 1982 Chinese census).* Beijing, Zhongguo tongji chubanshe, 674 p.

NBS (National Bureau of Statistics [China]), 1988. *(Yearbook of statistical data on population of People's Republic of China, 1949–1985).* Beijing, Zhongguo caizheng jingji chubanshe, 940 p.

NBS (National Bureau of Statistics [China]), 1991. *Zhongguo renkou ziliao shouce (Data on the Chinese population).* Beijing, Zhongguo renkou qingbao ziliao zhongxin, 258 p.

NBS (National Bureau of Statistics [China]), 1993. *Zhongguo 1990 renkou pucha ziliao (Data on the 1990 Chinese census), 4 vol.* Beijing, Zhongguo tongji chubanshe, 1101, 919, 1147, and 558 p.

NBS (National Bureau of Statistics [China]), 2000. *Data collection of 1997 National Population and Reproductive Health Survey.* Beijing, China Population Publishing, 496 p.

NBS (National Bureau of Statistics [China]), 2002. *China statistical yearbook.* Beijing, National Bureau of Statistics, 909 p.

NBS (National Bureau of Statistics [China]), 2002. *Zhongguo 2000 renkou pucha ziliao (Data on 2000 Chinese census), 3 vol.* Beijing, Zhongguo tongji chubanshe (Beijing Institute of Economics), 1,903 p.

NBS (National Bureau of Statistics [China]), 2003. *2002 zhongguo tongji nianjian (China statistical yearbook).* Beijing, Zhongguo tongji chubanshe, 935 p.

Peng Fei and Cambois Emmanuelle, 2002. Mortalité et transition épidémiologique en Chine depuis 1950, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI^e siècle. Questions de population, questions de société*, p. 87–109. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Peng Xizhe, 2002. La fécondité chinoise, constat et perspectives, *in*: Isabelle Attané (ed.), *La Chine au seuil du XXI^e siècle. Questions de population, questions de société*, p. 59–77. Paris, INED and PUF, 602 p. (Les Cahiers de l'INED, no. 148).

Tien H. Yuan, 1973. *China's population struggle.* Columbus, Ohio State University Press, 405 p.

Trolliet Pierre and Béja Jean-Philippe, 1986. *L'empire du milliard.* Paris, Armand Colin, 317 p.

Yin Wenyao, 1995. Lun zhongguo shengqu renkou kongzhi yu zhengce zhuanhuan de fei tongbu xing (Lack of synchronization between political change and population control in Chinese provinces), *Renkou xuekan*, vol. 2(3–8).

Zou Ping, 1986. Guanyu jianguo chuqi woguo renkou zhengce zhuanbian de hui gu yu sikao (Examination of changes in Chinese population policy after the Revolution), *Renkou yu jingji*, vol. 6, p. 8–11.

C H A P T E R

113

India's Population Policy in the Face of Demographic Momentum

S. IRUDAYA RAJAN AND JACQUES VÉRON

Center for Development Studies (CDS), Kerala, India

Institut national d'études démographiques (INED), Paris, France

In 1951, India's population numbered 361 million. In the same year, not long after gaining independence in 1947, the country launched its First Five-Year Plan, one section of which dealt with population policy implementation. Already the Indian government was stressing the importance of the "population problem" and the difficulty of boosting the country's development in a context of "rapid population growth" (GOI/PC, 1951). The fact that this population growth accelerated during the 30 years preceding the plan's adoption only heightened awareness of the demographic problem.

Half a century later, over one billion people are living in India. Does this mean that India's population policy was a failure? Admittedly, the measures adopted initially did not have the expected results; population policy changed course (sometimes radically); and the country's extreme economic, social and demographic heterogeneity, made it difficult to apply a uniform policy. However, we should not conclude that family planning measures were completely ineffective, as the tripling of India's population might suggest. Those responsible for family planning had (and still have) to contend with demographic inertia in a country that has acquired considerable growth potential in its population age structure.

When redefining national population policy in 2000, India restated its demographic objectives: a total fertility rate of 2.1 children per woman in 2010 and stabilization of India's population by 2045. In reality, however, the Indian government has been pursuing the goal of zero population growth for over 50 years. Although fertility is indeed falling throughout India, and is only just above, or even below, replacement level in the south, the most populous states are still those where family size remains the highest.

Population policy is not only concerned with measures or programs intending to reduce fertility but it also may deal with the population's health or spatial distribution, or with living conditions among the elderly. Moreover, any development policy has a bearing on the type of population policy that can be implemented. In India, although the government's main thrust has been toward fertility control, it has also introduced other types of policy, such as those concerning health, in 1983 and 2002. Incidentally, although several of India's plans have mentioned the excessive population growth in its large cities, India has not yet adopted a policy on migration.

I. INDIAN POPULATION POLICY: STAKES AND DIRECTIONS

1. Indian Population Dynamics before Independence and the Beginning of Demographic Malthusianism

India's family planning history began in 1916, with the publication of a pamphlet by Shri Pyare Krishen Wattal, titled *The Population Problem in India*, in which its author campaigned for family planning (Sinha, 1979). The first family planning clinic opened in 1925. V. C. Sinha explained that the Neo-Malthusian League was formed in Madras several years later. Margaret Sanger also encouraged family planning during her visit to India in 1935–1936.

As early as 1940, before independence, Jawaharlal Nehru was already conscious of the need for a fertility reduction policy. President of the Indian National Congress at the time, he declared:

> In the interests of social economy, family happiness, national planning, family planning and limitation of children are essential and the state should adopt a policy to encourage these. It is desirable to lay stress on self control as well to spread knowledge of cheap and safe methods of birth control. (qtd. in Chandrasekhar, 1976, p. 38)

It is hardly surprising, therefore, that Nehru ensured that a national population policy was adopted as soon as he became Prime Minister of India. Meeting to draft the policy's outline, the Indian Planning Commission over which he presided restated the importance of a family planning program: without a reduction in the population growth rate, it asserted, increasing efforts would be needed just to maintain consumption at its existing level.

Fertility was around 45 per 1,000 (Table 113–1) at the time Nehru and the Planning Commission were stressing the importance of getting it under control. With the crude death rate still just above 30 per 1,000, India's population growth rate was only 1.4% during the years 1931–1941, and even slightly lower in the following decade, apparently as a result of even higher mortality. Since the early 20th century, however, overall mortality tended to decrease: between 1901–1911 and 1941–1951, life expectancy at birth rose by 10 years for men, and nearly 9 years for women. Nonetheless, at 32.5 years for men and 31.7 years for women, the average expectation of life in India was still short (Table 113–2). During the period preceding independence, fertility hardly changed: the total fertility rate, which oscillated around six children per woman in the period 1881–1891, remained at this level until the 1960s (Véron, 1997; Naïr and Véron, 2002).

2. The First Five-Year Plan: Recognizing a Complex Population Problem and Adopting a Family Planning Program

The First Five-Year Plan (1951–1955), presented to India's parliament in December 1952, specifies the nature of India's demographic problem (GOI/PC, 1951). With mortality falling during the previous decades, mainly (but not solely) because of advances in hygiene and medicine, and without any reduction in fertility, population growth sped up. In the Planning Commission's view, whereas science and medicine made it possible to expect growing production capacity, "it is clear that, under present conditions, an increase in manpower 'resources' does not strengthen the economy but, in fact, weakens it" (GOI/PC, 1951, p. 16). This Commission recognized the complexity of the population problem as well as its diverse economic and social components. Reducing the birthrate appeared indispensable, if only to meet basic needs (food and clothing in particular) at a level comparable with pre-war conditions. The planning document contains a reminder that growing pressure on inevitably limited natural resources could only compromise economic and social progress. It emphasizes the importance of reducing the birthrate, although without establishing a quantitative goal.

India's First Five-Year Plan clearly states the need for population policies. The section dealing with methods for curbing demographic growth, however, is only titled "Family Planning." This was perceived as a priority for slowing population growth, but also as only the first step, as it would later be important to be able to improve children's health in particular. If it initially focused on educating public opinion and on the need to limit numbers, this was because "a programme of limitation must be a social movement" (GOI/PC, 1951, p. 206), mobilizing both families and public

TABLE 113–1 Crude Birth and Death Rates, India, 1901–1951

Decade	Crude birthrate (per 1000)	Crude death rate (per 1000)	Natural growth rate (%)
1901–1911	49.2	42.6	0.66
1911–1921	48.1	48.6	–0.05
1921–1931	46.4	36.3	1.01
1931–1941	45.2	31.2	1.40
1941–1951	39.9	27.4	1.25

Source: Davis, 1951.

TABLE 113–2 Life Expectancy at Birth, by Sex, India, 1901–1951

Decade	Men	Women	Difference
1901–1910	22.6	23.3	+0.7
1911–1920	19.4	20.9	+1.5
1921–1930	26.9	26.6	–0.3
1931–1940	32.1	31.4	–0.7
1941–1950	32.5	31.7	–0.5

Source: Davis, 1951.

opinion. The Planning Commission believed the State could play only a limited role in this sphere.

Thus, developing cheap, reliable, and effective contraceptive methods suitable for all population categories was a key objective. It was also necessary to explore other ways of reducing family size besides contraceptives—by increasing girls' age at marriage in particular. From 1951 onward, making sterilization more accessible was also added to the agenda.

3. Educating the Public: A New Priority from the 1960s Onward

The Second Five-Year Plan (1956–1961) was an extension of the first (GOI/PC, 1956). More dynamic action was taken, notably through making contraceptives more widely available, but the strategy remained the same: it gave priority to family planning clinics responsible for providing services and sensitizing the population to birth control ("clinic oriented approach," Srinivasan, 1995). The first family planning programs were based on the premise that a demand for contraception exists, and that providing contraceptives to satisfy the demand will actually reinforce it. It quickly became apparent, however, that the development of essentially hospital-based services was not enough; the adoption of the Third Five-Year Plan (1961–1966) marked a shift in family planning strategy (GOI/PC, 1960).

Educating the public became the priority. As spelled out in the Third Five-Year Plan, this was essential to create the social conditions for a broad-based family planning program. The plan also envisaged integrating family planning into the health services, while continuing its policy of making contraceptives available and encouraging sterilization. During the course of the Third Plan, it was decided to extend the education approach (GOI/PC, 1960) to encourage public support for the small family norm. The years 1966–1969, a period of transition, were not covered by a plan per se. The Fourth Five-Year Plan (1969–1974) was written in the spirit of the previous plan, attempting especially to intensify mass education and programs of incentives (GOI/PC, 1969). All contraceptive methods were promoted, although the intra-uterine device was expected to play a major role in controlling births because it had the advantage of being effective, reversible, and acceptable. In reality, few couples chose this method to limit the size of their family.

In addition, a law legalizing abortion (Medical Termination of Pregnancy Act) was passed in 1971 and revised in 1975. This law authorized abortion for medical reasons (the health of mother or child), or for social and economic ones (if the mother's economic and social environment threatened the health of the child to be born). It also allowed abortion in the case of contraceptive failure, but only among married mothers (Retherford and Roy, 2003).

After this, during the mid-1970s, Indian population policy passed through various episodes that tended to discredit it.

4. From Family Planning to Family Well-Being

As K. Srinivasan noted (1982), it became apparent how limited was the success of family planning programs from the start of the Fifth Five-Year Plan (1974–1979). In reality, Indian fertility had begun its decline (Figure 113–1) but, largely because of the relative success of female sterilization, mortality was falling even more rapidly (Figure 113–2), and population growth was accelerating (Table 113–3).

In June 1975, the then–Prime Minister, Indira Gandhi, declared a state of emergency. The following year, the government launched a National Population Policy through a statement to the press made by Dr. Karan Singh, then Minister of Health and Family Planning. In his communiqué of April 16, 1976, Dr. Singh declared that the population problem should rank among national priorities (Singh, 1976), and that it was no longer possible to wait for the effect that education, as well as economic and social developments, would have on fertility—all the less so because of the vicious circle linking population and economic development. "Our real enemy is poverty," Karan Singh also declared, "it is as a frontal assault of the citadels of poverty that the Five-Year Plan has included the Minimum Needs Programme" (Singh, 1976, p. 309–310). He considered necessary to resort to

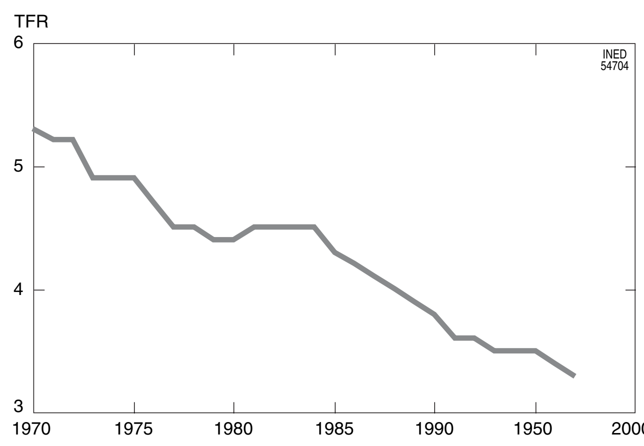


FIGURE 113–1 Total fertility rate (TFR) trends since 1970. (Source: Sample Registration System annual publications.)

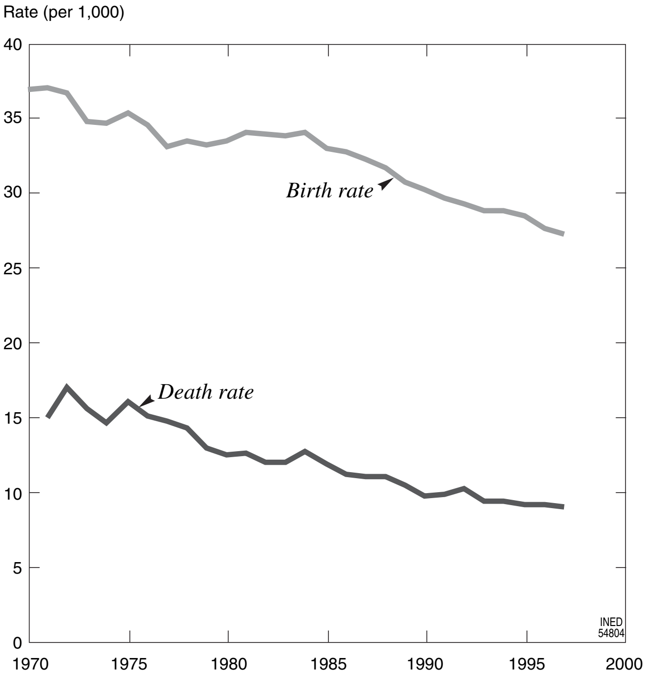


FIGURE 113–2 Trends in birth and death rates since 1970. (Source: Sample Registration System annual publications.)

TABLE 113–3 Population Growth in India, 1901–2001

Census year	Number of persons	Annual growth (absolute numbers)	Annual exponential growth rate (%)
1901	238,396,327	—	—
1911	252,093,390	13,697,063	0.56
1921	251,321,213	–772,177	–0.03
1931	278,977,238	27,656,025	1.04
1941	318,660,580	39,683,342	1.33
1951	361,088,090	42,427,510	1.25
1961	439,234,771	78,146,681	1.96
1971	548,159,652	108,924,881	2.20
1981	683,329,097	135,169,445	2.22
1991	843,387,888	163,058,791	2.14
2001	1,027,015,247	180,627,359	1.93

Source: 2001 census (Registrar General and Census Commissioner, 2001).

all sorts of methods to gain control over demographic growth.

Among the methods put forward to reduce population growth were a higher age at marriage, female literacy ("informal" education should be promoted; the effect of female literacy on fertility was assumed), and child nutrition (taking into account the association between mortality, ill health, and fertility). In reality, these were all direct and indirect methods for fertility control. The young were also encouraged to plan their family size. The population policy contained "multimedia motivational strategies," directed especially at the rural population. It also increased the financial rewards granted to couples limiting their family size to two or three children. The idea of compulsory sterilization was discussed, and the Indian government authorized the states in the Federation to pass a law making sterilization obligatory. However, when the State of Maharashtra proposed compulsory sterilization for certain categories of couples, through the 1976 Maharashtra Family Act, it did not obtain the Federal Government's authorization (Gwatkin, 1979). The fact that the population's alarming growth makes it impossible to stamp out poverty is one of the preambles to this law, which should guarantee that couples have no more than three children or four, if the first three are of the same sex (Chanrasekhar, 1976).

In this same year, 1976, the government's stated objective was to reduce the crude birthrate from 34 per 1,000 to 25 per 1,000 by 1984, a goal that was not in fact reached until 2002. Taking mortality decline into account, the population growth rate should have been 1.4% in 1984; the actual annual growth rate for the most recent intercensal period (1991–2001) was still close to 2%.

The fall of Indira Gandhi's government appears to mark a new era in population growth control, with wide condemnation of the abuses committed during the state of emergency, especially with respect to sterilization. The Family Planning Program thus became the Family Welfare Program. But the Janata government, the Congress Party's successor, was itself greatly concerned by the population problem. As Nortman (1978) noted, "like its predecessor the Desai government is alarmed by the possibility of an Indian population of one billion by the year 2000, a level implied by the current 2 percent per annum growth rate" (p. 277). As Gwatkin (1979) observed, with unobtrusive but crucial measures, it was striving "to keep the government's family planning machinery intact" in order to be able to use it once the mistrust toward the previous population policy had evaporated. Asok Mitra (1977) noted that the national population policy declaration of June 29, 1977 broadly replicated, in its substance, that of April 16, 1976.

In power once more, Indira Gandhi presented the family planning program's priorities in terms very different from those used during the Emergency era:

> The real answer [to curb population growth rate], of course, is development from both angles: first the development to provide what the people need, to open up areas, and secondly, because development itself curbs families, I mean

> people who reach a particular standard are more conscious of their duty to the child. That is, they want something more for their children than they had; and, therefore, they themselves think about smaller families. But ultimately, of course, size of the population has to be tackled at various levels not just through development, but through persuasion of people (Gandhi, 1981; p. 168)

Nehru's daughter thus stressed the fact that the birth of a child should be a question of choice for the two parents, and that persuasion is not coercion. During her Inaugural Address at the first national conference of the Indian Association of Parliamentarians on Population and Development, Indira Gandhi spelled out her vision of family planning: "Family planning must become a people's movement—of the people, by the people, for the people—only then can our hopes be realised" (Gandhi, 1981; p. 559). Beyond these public speeches on reducing family size, and despite all the effort invested, the Indian population continued to expand during the 1990s at an annual rate of around 2%. In 2000, a new population policy emerged.

5. A New Population Policy Approach

The policy approach taken in the year 2000 shares some features with independent India's first planning program, as it still aims to satisfy the demand for contraceptives and, as in the 1960s programs, it also stresses the need to "promote vigorously the small family norm." Some of the stated objectives in 2000 are strictly demographic in nature, while others are more closely related to the health or well-being of the population. The government announced targets of replacement level fertility by 2010, as well as infant mortality at less than 30 per 1,000 and maternal mortality under 100 per 100,000. It also aimed to immunize children against all illnesses that could be avoided through vaccination, and to raise age at marriage for girls—not before the age of 18 years, and preferably after 21. The 2000 population policy also specified that schooling should be free and compulsory for all children until 14 years of age.

Fertility in India has fallen substantially, particularly since the early 1980s. Very strong demographic inertia in India's population structure means, however, that the demographic growth rate has rarely fallen, and was still at 1.93% during the last intercensal period (1991–2001). Moreover, fertility decline has been very uneven across the different states, although it has taken place in such a way that makes it impossible to distinguish the impact of more or less dynamic family planning programs at the individual state level, from the effects of economic and social development.

II. FERTILITY, FAMILY PLANNING AND DEVELOPMENT IN INDIA: NATIONAL AND STATE LEVELS

1. National Population Policy: Frequently Revised Goals

In the early 1950s, the Indian government expressed its desire to see the birthrate fall "to the extent necessary to stabilize the population at a level consistent with the requirements of the national economy" (Srinivasan, 1995). The population growth rate was only 1.25% when the first family planning program was adopted (Table 113–3). Between 1951 and 1981, however, population growth accelerated (Figure 113–3), with a one-point increase in the growth rate (which reached 2.22% during the period 1971–1981). The pace slowed later, giving a natural growth rate of 1.69% in 2002, according to the Sample Registration System (see Table 113–7).

With population growth speeding up while family planning programs were active, and with growth rates reluctant to fall, must the conclusion be that India's population policy has been a failure?

India's fertility has dropped significantly during the last 25 years, but, because of its young population, the decline has had only a partial effect on lowering birthrates (P. Visaria, 1995). During the same period, birth and death rates decreased in parallel (Figure 113–2). With hindsight, it thus appears that, in its desire to stabilize the population as quickly as possible, the Indian government established overambitious or overoptimistic goals, as it had failed to take into account the constraints imposed by population dynamics (Table 113–4 and Figure 113–4). Each new Five-Year Plan provided the opportunity to evaluate the preceding one; the limited success of family planning programs was regularly emphasized. Thus, referring to the Seventh Plan, the Eighth Plan (GOI/PC, 1992) observed, "In spite of massive efforts in the form of budgetary support and infrastructure development, the performance of the family welfare program has not been commensurate with the inputs." The plan then specified that "containing population growth" was one of its six main objectives. The next plan repeated that "population stabilization is an essential prerequisite for sustainability of the development process" (GOI/PC, 1992).

Demographic inertia is therefore largely responsible for the difficulty in achieving the birthrate targets established regularly by the Planning Commission. The population age structure effect plays an even greater role in accelerating growth than would mortality if it declined faster than fertility. According to a

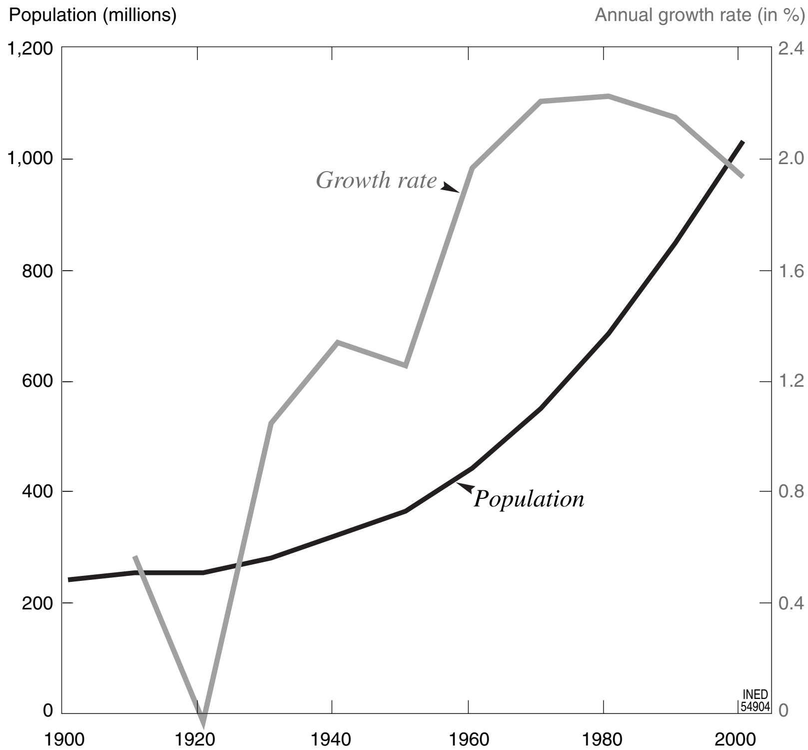


FIGURE 113–3 Population Momentum: absolute numbers and growth rate. (Source: Registrar General and Census Commissioner, 2001.)

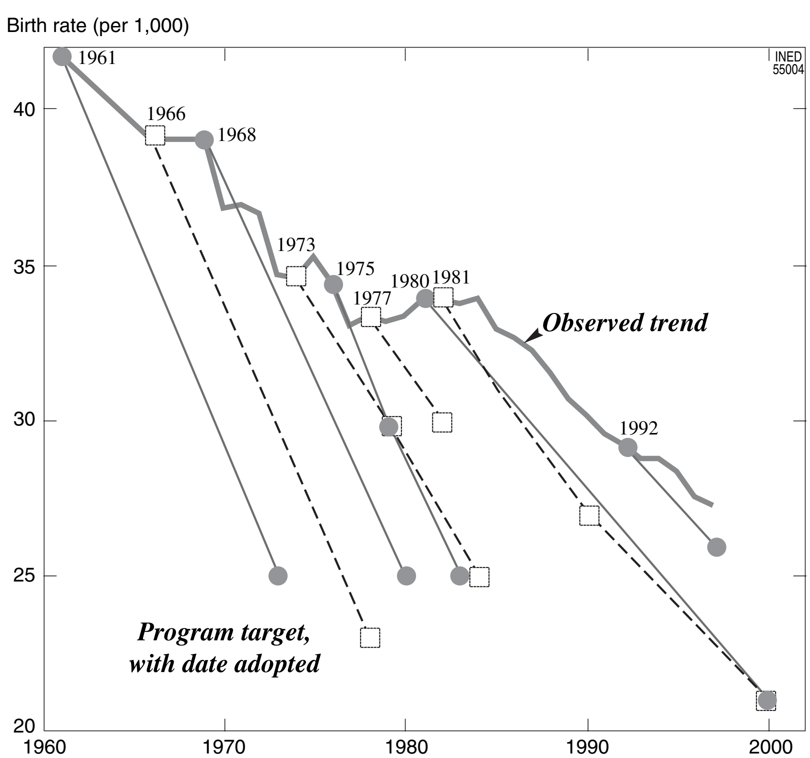


FIGURE 113–4 Evolution of crude birthrate and targets for family planning programs. (Source: update of a representation by Nortman, 1978.)

TABLE 113–4 Crude Birthrate Targets and the Degree to Which They Were Satisfied

Year	Desired crude birthrate	Year target should be reached	Actual birthrate for the same year
1950s decade	No particular target		
1962	25	1973	34.6
1966	25	As fast as possible	
1968	23	1978–1979	
1969	32	1974–1975	
Fourth Plan (1969–1974)	25	1978–1981	33.8
Fifth Plan (1974–1979)	30	1979	33.1
	25	1984	33.8
National population policy (April 1976)	30	1978–1979	
	25	1983–1984	
Second national population policy (April 1977)	30	1978–1979	
	25	1983–1984	
Sixth Plan (1980–1985)	21	2000	
National Health Policy (1983)	31	1985	32.9
	27	1990	29.9
	21	2000	25.8
Seventh Plan (1985–1990)	29.1	1990	29.9
Eighth Plan (1992–1997)	26.0	1997	27.2
National population policy for the year 2000	Total fertility rate at replacement level in 2010		

Sources: Singh, 1976; Srinivasan, 1995; GOI/PC, 1997, 2002; GOI, 2000; Registrar General, 2002, 2004; Various *Sample Registration System Bulletins.*

recent estimate for the early 2000s, once the level of infant mortality and other socioeconomic factors are taken into account, 60% of the population growth rate stems from the importance of the population of reproductive age, 20% from a higher fertility than desired due to an unmet demand for contraception, and the other 20% from desired fertility levels (GOI/PC, 2002).

Aware of the force of demographic inertia, the Indian government distinguished three types of goals in its year 2000 Population Policy, according to the time horizons with which they were associated:

1. An immediate goal, involving the satisfaction of contraceptive needs, health infrastructure and medical personnel, and the provision of integrated reproductive and child health services;
2. A medium-term target of a total fertility rate of 2.1 children per woman by 2010;
3. A long-term goal of Indian population stabilization by 2045.

If fertility reaches replacement level by 2010, India's population will continue to grow during the following 35 years because of demographic inertia; the number of Indians will first overtake the number of Chinese, and then reach 1.5 billion in the middle of the 21st century.

2. Development and the Demographic Transition

Family planning programs are not the only type of change likely to have an effect on fertility in India. During the 50 or so years since independence, age at marriage has increased, rates of literacy have improved, and health policies have been implemented. It is impossible to separate the population policy effect from the development effect with any sort of precision, since the two are clearly not independent, as the regional analysis demonstrates.

Governments in India have long been preoccupied with raising the age at marriage. In 1951, the average age at marriage for women in Andhra Pradesh was 12.6 years; in Uttar Pradesh it was 13.8 years, and in India as a whole it was $15\frac{1}{2}$ years (Agarwala, 1967). Marriage frequently used to be celebrated before this age, as is shown by the fact that the 1955 Hindu Marriage Act set the minimum age for celebrating a religious ceremony at 18 years for boys and 15 years for girls. The 1976 National Population Policy emphasized the need to change India's marriage model, not only because of its effect on population growth, but also to assist the move toward "more responsible parenthood" and improve the health of women and children (Singh, 1976). In 1976, the Child Marriage Restraint Act raised the minimum age at marriage to 21 years for boys and 18 years for girls. Although both men's and women's average age at marriage rose over the course of the last decades, according to the National Family Health Survey (NFHS) in 1998–1999, the median age at marriage was hardly above 16 years (Table 113–5). As noted previously, the National Population Policy (2000) set an age at marriage for women of not less than 18 years, and preferably above 20 years, as one of its sociodemographic objectives for the year 2010.

Other changes are also likely to influence fertility trends; this is especially true for progress in education for girls, because it promotes higher status for women. The 2000 Population Policy reported on the need to develop education for the young. Although it is true that female literacy rates increased fivefold between 1951 and 2001, only just over half of Indian women were literate in 2001 (Table 113–6). Substantial regional differences remained, with a female literacy rate of 87% in Kerala compared with 34% in Bihar.

In all its Five-Year Plans, the Indian government emphasized the importance of reducing maternal, infant, and child mortality. At the start of the 1970s, the level of infant mortality was still at 130 per 1,000. It fell steadily afterward (Figure 113–5), halving the rate 30 years later (63 per 1,000 in 2002).

Adopted in 1983, the First National Health Policy aimed to develop a primary health care network well dispersed throughout India, while at the same time promoting health education; it called on health volunteers to play in intermediation role, and aimed at treating patients at a decentralized level as far as possible. A new health policy was implemented in 2002. It restated the infant mortality target of 30 per 1,000. It also aimed to halve mortality from tuberculosis and malaria by the year 2010, and check the spread of the AIDS epidemic by 2007. From the demographic point of view, this 2002 health policy was an extension of the 2000 population policy.

With strong demographic, economic, and social heterogeneity in India, a regional analysis provides a complementary view of the respective effects of development and population policy on fertility. Given India's great demographic diversity and heterogeneity, Leela and Pravin Visaria suggest a multifaceted approach to strengthen family planning and slow population growth (Visaria and Visaria, 1995).

TABLE 113–5 Average Age at Marriage for Men and Women

Year	Men	Women
1901	20.0	13.1
1951	21.3	15.6
1961	22.7	15.5
1981	23.3	18.3
1991	23.9	19.5
1997	—	19.5
1998–1999 (median age)		16.4

Sources: Raghavachari *et al.*, 1974; NFHS-2 of 1998–1999 (IIPS and ORC Macro, 2000); Tenth Five-Year Plan 2002–2003 (GOI/PC, 2002).

TABLE 113–6 Trends in Literacy Rates (%) between 1951 and 2001

Year	Total	Men	Women
1951	18.33	27.16	8.86
1961	28.30	40.40	15.35
1971	34.45	45.96	21.97
1981	43.57	56.38	29.76
1991	52.21	64.13	39.29
2001	65.38	75.85	54.16

Source: 2001 Census (Registrar General and Census Commissioner, 2001).

3. Regional Variations in Fertility: What Role Did State-Level Population Policies Play?

In 2002, demographic growth rates still varied greatly from one state to another (Table 113–7). Growth was slowest in the south of the country, particularly in Kerala and Tamil Nadu, where annual rates were barely

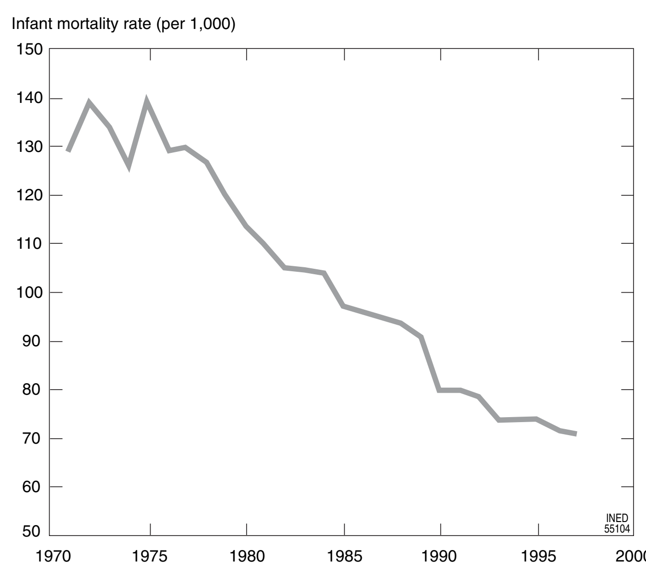


FIGURE 113–5 Infant mortality trends since 1970. (Annual Sample Registration System data.)

TABLE 113–7 Birth, Death, and Natural Growth Rates for India's Largest States, 2002

States	Birthrate (per 1000)	Death rate (per 1000)	Natural growth rate (%)
Andhra Pradesh	20.7	8.1	1.26
Assam	26.6	9.2	1.74
West Bengal	20.5	6.7	1.38
Bihar	30.9	7.9	2.30
Gujerat	24.7	7.7	1.70
Haryana	26.6	7.1	1.95
Karnataka	22.1	7.2	1.49
Kerala	16.9	6.4	1.05
Madhya Pradesh	30.4	9.8	2.06
Maharashtra	20.3	7.3	1.30
Orissa	23.2	9.8	1.34
Punjab	20.8	7.1	1.37
Rajasthan	30.6	7.7	2.29
Tamil Nadu	18.5	7.7	1.08
Uttar Pradesh	31.6	9.7	2.19
India	25.0	8.1	1.69

Source: Sample Registration System (Registrar General, 2004).

over 1% (Rajan, 2005). Rapid population increase continued in the country's most populous states ("Bimaru states"), on the other hand, with growth rates exceeding 2% per year. One of them, Bihar, is still increasing by 2.3% annually. These states are going through the demographic transition, a process that is already complete in the states of southern India: birthrates in Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh were still over 30 per 1,000 in 2002, and fertility well over 3 children per woman.

Kerala's demographic evolution is worth mentioning, as fertility and birthrates plummeted particularly quickly in this state: the crude birthrate fell from 44 per 1,000 in the 1950s to 17 per 1,000 in 2002, and the total fertility rate (TFR), at 5.6 children per woman in the 1950s, is now under 2 (Mari Bhat and Rajan, 1997). In his analysis of Kerala's rapid and extreme fertility decline, K. C. Zachariah (1997) explored the role played by official population policies. He suggests that various factors may have contributed to the fall in fertility: an increase in age at marriage and in the prevalence of contraception, sterilization, maternal and child health programs, rising female education levels, and the State's political involvement in social action. Zachariah estimates that 30% of fertility decline between 1968 and 1978 was due to the increase in marriage age. He also notes that contraceptive prevalence rates rose from 37% to more than 72% between 1981 and 1992, and that the universal vaccination program, launched in 1985 and reaching half of the children, reduced infant mortality. In his opinion, Kerala's high population density sharpened its awareness of the population problem. Finally, this state implemented reforms leading to greater equality; according to Zachariah, this also played a part in triggering the demographic transition, and in its timing. In this Indian demographer's opinion, however, such rapid fertility decline is explained above all by the strong interaction between the provision of family planning services and changes in the socioeconomic context. Improvements in child health through the Maternal and Child Health program also made an important contribution to this reduction in family size.

Fertility decline was also strong in Tamil Nadu state: today, the average number of children per woman is around replacement level. As in Kerala, a higher age at marriage contributed to the drop in fertility up to the 1980s. According to S. Krishnamoorthy *et al*. (2005), the importance of poverty, women's high status, the existence of a family well-being program, and a movement in favor of social reform and media action are among the other possible causes of this reduction in family size.

It is generally acknowledged that the family well-being program and the ideas spread by the media helped popularize the small family size norm. Couples also cite the reduction in infant mortality as a reason for their lower fertility. Although the state of Tamil Nadu has not experienced any real economic development, it appears that people feel that the cost of educating their children has increased. Parents, aspiring to a better life for their children, wish to send them to school more systematically: the expense of educating children has thus increased the cost of raising them. In the case of female births, the dowry also represents a burden for family. Finally, as in Kerala's case, the interaction between family planning programs and social change explains why fertility fell so much and so quickly. Conversely, in the Bimaru states, where social change has been slower and less substantial, family planning programs cannot experience the same degree of success. These states accumulate vast populations who are impoverished, uneducated and the like.

In Kerala and Tamil Nadu, as everywhere else in India, sterilization remains the main means of controlling fertility, known by all and preferred to other birth control methods.

III. FIFTY YEARS OF BIRTH CONTROL IN INDIA

1. Fertility Control Based on Sterilization

From the early 1950s, the Indian government has encouraged couples to use sterilization as a way of limiting births. At that time, the main emphasis was on male sterilization (vasectomy). Economic incentives were set aside for men undergoing sterilization, or for those who encouraged others to do so. Then, in the 1960s and 1970s, the attempt to encourage sterilization intensified (Véron *et al*., forthcoming). From 1966 onward, those responsible for family planning estimated the annual number of sterilizations needed to achieve demographic targets. The degree to which these objectives were attained varied: 119% in 1967–1968 and less than 42% in 1973–1974. Between the mid-1960s and mid-1970s, there was an upward, but irregular, trend in the number of sterilizations (Figure 113–6).

The 1970s was an important decade for sterilization. At the start of it, two mass sterilization camps, each lasting one month, were set up at Ernakulam in Kerala, and led to almost 20,000 sterilizations being carried out. The idea that an effective family planning program was possible was proved (Krishnakumar, 1972). Shortly afterward, the decision was taken to launch a massive vasectomy campaign at the level of the whole Gujarat State (Thakor and Patel, 1972). The initial goal was 150,000 vasectomies but, during the 8-week campaign, over 220,000 men were finally sterilized. On the other hand, although much energy was invested in encouraging them, the decision to be sterilized was to be a matter of choice, at least in principle. In 1975, however, Indira Gandhi proclaimed a state of emergency and, influenced by her son, Sanjay, who was extremely concerned by India's population problem, the Indian Prime Minister lent her support to authoritarian fertility reduction measures, which in the end were applied very little (Gwatkin, 1979). Nevertheless, while 4.3 million sterilizations was the target for the 1976–1977 financial year, the actual number reached 8.26 million—three-quarters of which were vasectomies—because of certain particularly zealous states. The question of making sterilization compulsory was even raised; during his presentation of the 1976 Population Policy, Karan Singh asserted that the population was ready to accept more coercive family planning measures. In the same year, through the Maharastra Family Act, the government of Maharashtra intended to impose sterilization on parents with at least three children, under certain conditions, but it did not receive the Indian government's authorization to implement this document (Chandrasekahar, 1976, p. 131–143).

The electoral defeat of Indira Gandhi's party in 1977 appears to mark the return to couples' free choice over birth control. The need to promote methods for spacing births was restated regularly. At under a million in 1977–1978, nevertheless, the number of sterilizations started to rise again, reaching an average of 4 million annually. Today, however, women are almost the only ones sterilized (Figure 113–7). From 1970 to the present day, the rate of contraceptive use has evolved remarkably, more than quadrupling. Although the gap between the curves for all contraceptive methods combined and for couples protected by sterilization has gradually widened, since the mid-1980s, the latter nonetheless remains the most common means of fertility control (Figure 113–8).

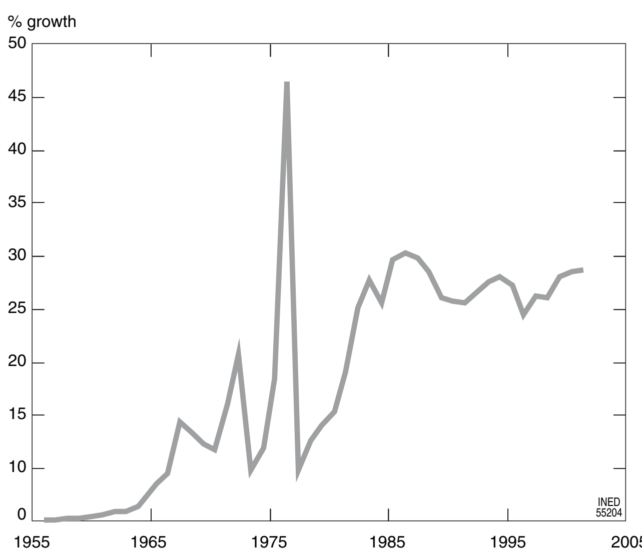


FIGURE 113–6 Annual evolution of number of sterilizations. (Source: Véron *et al*., forthcoming.)

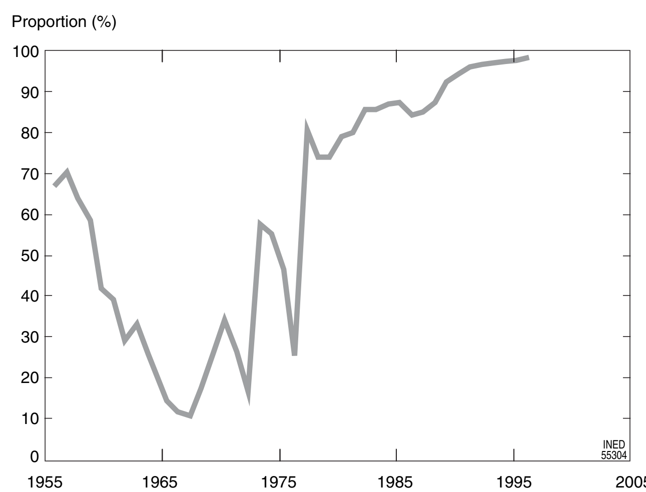


FIGURE 113–7 Proportion of female sterilizations as a proportion of the total number of sterilizations. (Source: Véron *et al*., forthcoming.)

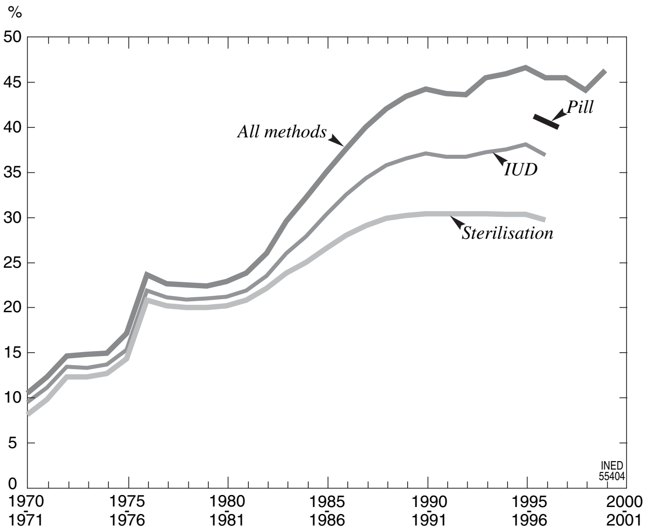


FIGURE 113–8 Contraceptive use by type of method. (Source: Véron *et al.*, forthcoming.)

2. Sterilization Creating Its Own Demand?

According to the NFHS and NFHS-2 surveys, over 80% of sterilized women had used no birth-spacing method prior to sterilization; in addition, at the NFHS-2, 65% of women who were not using any form of contraceptive at the survey stated their preference for female sterilization as their future means of birth control. Can we conclude from this that a real demand for sterilization emanates from couples themselves? Is it a real choice? To what extent do women decide on sterilization?

India's most developed states are also those where sterilization is the most widespread. Thus, the highest proportions of couples protected by sterilization are found in the southern states (Kerala, Tamil Nadu, Karnataka, Andhra Pradesh), in Maharashtra, Punjab, and, to a lesser extent, in Haryana. The percentage of sterilized women lessens with their level of education, but close to 40% of the most educated women in Andhra Pradesh have been sterilized. Women's average age at sterilization is also lowest in the South Indian states, where the average number of children per woman is also the lowest. Couples appear to create the desired number of children and then opt for the terminal contraception that sterilization constitutes.

The fact that a demand for sterilization arises from couples themselves should not make one ignore the "offer" effect. Improvements in sterilization techniques (minilaparotomy and laparoscopy) have played a role in the high prevalence of this birth control method, as has the possibility for a woman to be sterilized in hospital immediately after her child's birth. Fertility control appears to remain women's business, even if it does not express freedom of choice. The Tenth Five-Year Plan intends to promote male sterilization, considered to be more reliable than female sterilization.

IV. THE OTHER COMPONENTS OF POPULATION POLICY

1. Population Health

The 1983 National Health Policy stressed the importance of encouraging the small family size norm (GOI, 2000). When adopting this policy, Parliament emphasized the need to separate national population policy, focused primarily on fertility reduction, from health policy. The 2000 National Population Policy set health objectives (with respect to maternal and infant mortality) because higher child survival rates and better maternal health favor fertility decline.

The 2002 Health Policy (GOI, 2002), taking a "holistic" approach, established its priorities at the same time. It sought to fight both "historical" diseases, such as tuberculosis and malaria, and emergent ones, like AIDS.

In this 2002 Health Policy, the Indian government also showed concern for equal access to health services. Regional and social inequalities, as well as rural-urban differences, were so marked that placing the accent on primary health care was the only way to make this equity goal even slightly realistic. Improving the quality of health services was another challenge; doubling the percentage of the GDP allocated to health was an additional target for 2010.

The evolution of life expectancy at birth gives an indication of progress in the health sphere since the adoption of the first population policy in the early 1950s. At that time, the average life span was around 30 years; a half-century later, it had doubled (Table 113–8). Differences in life duration between men and women are minor. Even though urban-rural disparities are still substantial, they are less marked than they were in the early 1970s.

2. Population Policy and Development Policy

It is difficult to isolate the respective effects of population policies, development policies, and development itself on demographic change; the analysis of Kerala's demographic transition is a prime example. Population policies, in their broader acceptation, are in

TABLE 113–8 Life Expectancy at Birth since the Early 1970s

Period	Men			Women		
	Total	Rural	Urban	Total	Rural	Urban
1971–1975	50.5	48.9	58.8	49.0	47.1	59.2
1976–1980	52.5	51.0	59.6	52.1	50.3	60.8
1981–1985	55.4	54.0	61.6	55.7	53.6	64.1
1986–1990	57.7	56.1	62.0	58.1	56.2	64.9
1991–1995	59.7	58.5	64.5	60.9	59.3	67.3
1995–1999	60.8	59.7	65.5	62.5	61.0	65.5

Source: L. Visaria, 2004.

any case close to development policies. Legislation on age at marriage, or advances in the country's education level, can at least partly explain the drop in both mortality and the birthrate.

On the other hand, the national population policy adopted in the year 2000, whose main goal is still to stabilize India's population, also includes such socioeconomic objectives as compulsory free schooling until the age of 14 or vaccinations for all children. In addition, the holistic approach that inspired the elaboration of this population policy further recognizes the importance of improving women's status. It also mentions the circumstances of several hundred million shantytown residents; without ensuring these individuals access to drinking water, and without the benefit of primary health care, it is illusory to aim at a sizeable reduction in infant mortality.

3. Population Density and Urbanization

India's remarkable century-long population growth translated into a considerable increase in density, particularly in some states (Figure 113–9). At the 1901 census, the country's average density stood at 77 persons per square kilometer; 100 years later it reached 324 persons per square kilometer (Registrar General and Census Commissioner, 2001). The population's spatial distribution was very uneven, with densities varying from 1 to 9. Excluding the special case of Delhi's almost exclusively urban population, the highest densities are found in Kerala (819 persons per km^2), Bihar (880/km^2) and West Bengal (904/km^2). The presence of a huge city can explain extremely high density—as is the case for West Bengal, with Calcutta's urban agglomeration—but it can also occur independently of urbanization, as in Kerala. Managing extreme densities raises serious problems for development and the environment, and the same is true for urbanization.

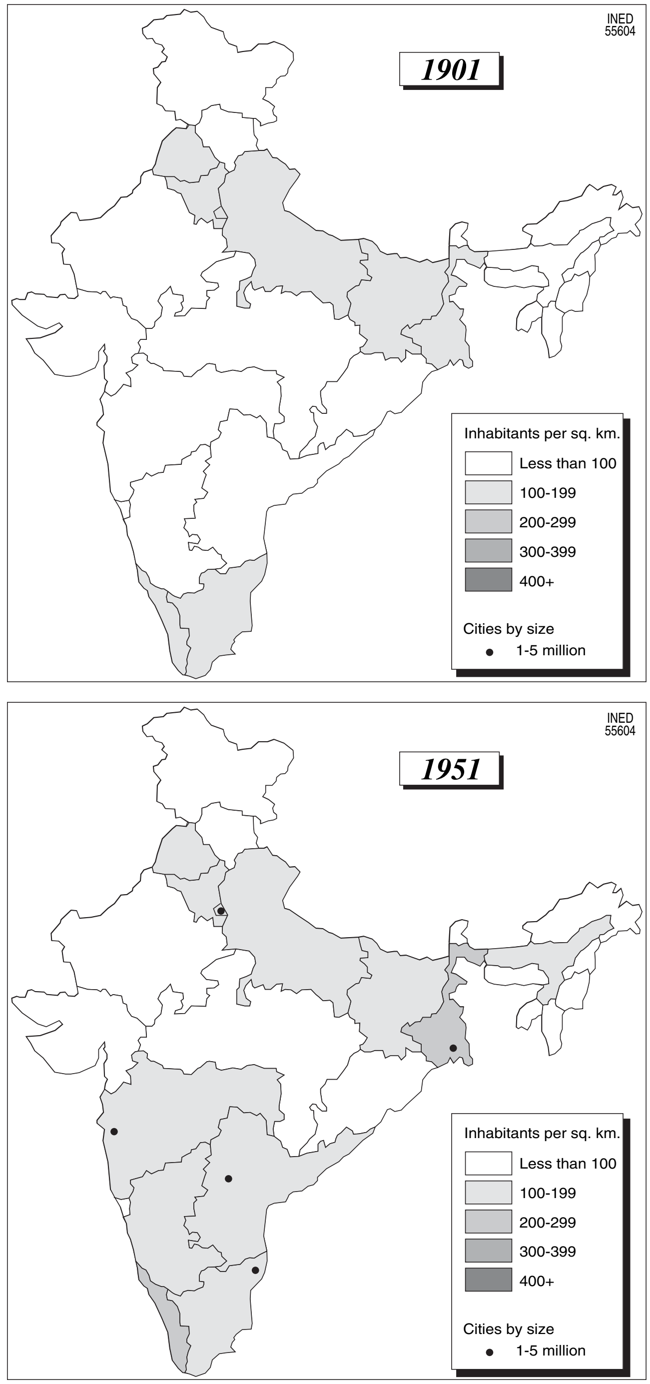


FIGURE 113–9 Population density by State and principal Indian cities by size, since the beginning of the 20th century.

India is characterized by a relatively low level of urbanization and a sizeable urban population (Table 113–9). At the 2001 census, there were more urban residents in India than there had been Indians in 1931. No policy in India explicitly aimed to limit the rural exodus, for example, or curb growth in the very large urban agglomerations. The Eighth Five-Year Plan

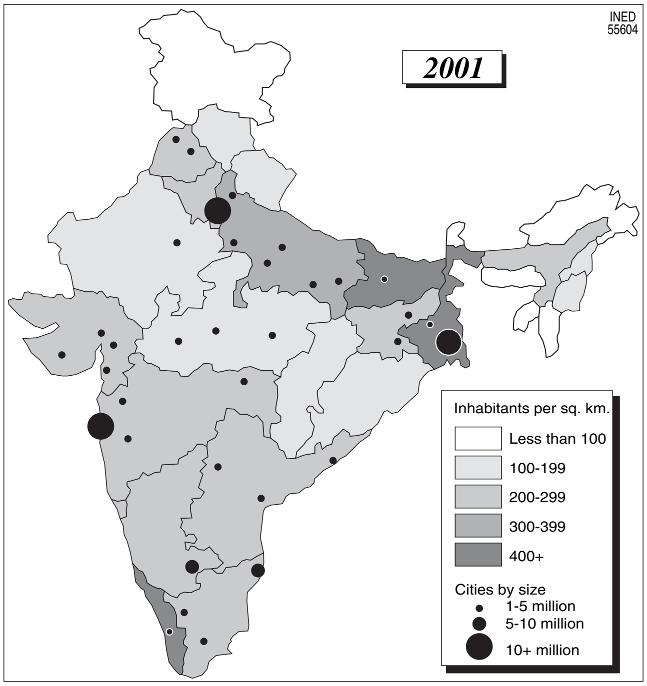


FIGURE 113–9 (continued) Population density by State and principal Indian cities by size, since the beginning of the century.

TABLE 113–9 Percent Urban and the Urban Population in India from 1901 to 2001

Year	Percent Urban	Urban population (millions)
1901	10.8	25.9
1911	10.3	25.9
1921	11.2	28.1
1931	12.0	33.5
1941	13.9	44.1
1951	17.3	62.4
1961	18.0	78.9
1971	19.9	109.1
1981	23.3	159.5
1991	25.7	217.6
2001	27.8	285.3

Source: Registrar General and Census Commissioner, 2001.

(GOI/PC, 1992), while recognizing urbanization as the inevitable consequence of development, nevertheless revealed a deterioration of the physical environment and quality of life in the large cities, and promoted the creation of small and medium-sized towns. To be effective, urban development policies had to be linked with policies for industrial development policy, as labor-force migrations were involved. The Ninth Plan (GOI/PC, 1997) mentions the global aim of "development of urban areas as economically efficient, socially equitable and environmentally sustainable entities," and more precise ones, such as improving the habitat, fighting poverty, developing an effective and cheap public transport system, and the like.

CONCLUSION

This analysis of 50 years of population policy in India, a country characterized by very strong demographic, economic, social, and cultural heterogeneity, shows that family planning programs were only truly effective when accompanied by development: some southern states experienced a particularly rapid fertility decline. In the very populous northern states, the demographic transition is still in progress: fertility has declined substantially, but demographic inertia is so strong that annual growth rates still exceed 2%.

Regularly obliged to revise its population objectives, the Indian government has come to understand the meaning of demographic inertia; replacement level fertility in 2010, and population stabilization 35 years later, were the stated goals when India's new population policy was introduced in 2000.

Nowadays, population growth is no longer the Indian government's only concern, as new problems have arisen. Curbing the spread of AIDS has become a priority, and the government also needs to fight the strong discrimination against little girls; the 2001 census revealed the presence of such discrimination in some states, the Punjab and Haryana in particular, leading to frequent abortions of female fetuses (Nanda and Véron, 2005).

As for the states that have completed the demographic transition, and practically resolved their problem of population growth, they need to start worrying about demographic aging.

Acknowledgment

This chapter was translated from the French by Heather Juby.

References

Agarwala S. N., 1967. *Population.* New Delhi, National Book Trust, India, 141 p.

Bose Asish, 1981. Population policy in India: two comments, *Population and Development Review*, vol. 7(1), p. 170–171.

Chandrasekhar Sripati, 1976. *Population and Law in India.* Bombay, Calcutta, Madras and New Delhi, Blackie & Son Ltd, 157 p.

Chaudry Mahinder, 1988. *Population policy and fertility in India*. Kingston (Canada), Royal Military College of Canada, 122 p.

Davis Kingsley, 1951. *The population of India and Pakistan*. Princeton (New Jersey), Princeton University Press, 263 p.

Gandhi Indira, 1981. Inaugural Address by The Prime Minister, Smt. Indira Gandhi, *Population and Development Review*, 7(1), p. 168–169.

GOI, 2000. *National Population Policy*. New Delhi, Government of India, 37 p.
GOI, 2002. *National Health Policy*. New Delhi, Government of India, 39 p.
GOI/PC, 1951. *First Five-Year Plan: a draft outline*. Delhi, Government of India, Planning Commission, Manager of Publication, 295 p.
GOI/PC, 1956. *Second Five-Year Plan.* Delhi, Government of India, Planning Commission, Manager of Publication, 653 p.
GOI/PC, 1960. *Third Five-Year Plan: a draft outline.* Delhi, Government of India, Planning Commission, 263 p.
GOI/PC, 1969. *Fourth Five-Year Plan: a draft outline.* Delhi, Government of India, Planning Commission, 430 p.
GOI/PC, 1978. *Fifth Five-Year Plan.* Delhi, Government of India, Planning Commission, 276 p.
GOI/PC, 1980. *Sixth Five-Year Plan.* Delhi, Government of India, Planning Commission, 463 p.
GOI/PC, 1992. *Eighth Five-Year Plan. Volume II: Sectoral programmes of development.* Delhi, Government of India, Planning Commission, 480 p.
GOI/PC, 1997. *Ninth Five-Year Plan 1997–2002. Volume II: Thematic issues and sectoral programmes.* New Delhi, Government of India, Planning Commission, 256 p.
GOI/PC, 2002. *Tenth Five-Year Plan, 2002–2007. Sectoral policies and programmes.* New Delhi, Government of India, Planning Commission, 1116 p.
Gwatkin Davidson R., 1979. Political will and family planning: the implications of India's emergency experience, *Population and development review*, vol. 5(1), p. 29–59.
IIPS, 1995. *National Family Health Survey, India 1992–1993.* Bombay, International Institute for Population Sciences, 402 p.
IIPS and ORC Macro, 2000. *National Family Health Survey (NFHS-2), India 1998–1999: Andhra Pradesh.* Mumbai (Inde) and Claverton (Maryland), International Institute for Population Sciences and ORC Macro, 310 p.
Krishnakumar S., 1972. Kerala's pioneering experiment in massive vasectomy camps, *Studies in Family Planning*, vol. 3(8), p. 177–185.
Krishnamoorty S., Kulkarni P. M. and Audinarayana N., 2005. Causes of fertility transition in Tamil Nadu, *in:* Christophe Z. Guilmoto and S. Irudaya Rajan (eds.), *Currents of demographic change in South India,* p. 227–247. New York, Sage Publications, 452 p. + maps.
Mari Bhat P. N. and Rajan S. Irudaya, 1997. Demographic transition since independence, *in:* Kunniparampil C. Zachariah and S. Irudaya Rajan (eds.), *Kerala's demographic transition, determinants and consequences,* p. 33–78. New Delhi, Sage Publications, 367 p.
Mitra Asok, 1977. National population policy in relation to national planning in India, *Population and Development Review*, vol. 3(3), p. 297–306.
Naïr Sadasivan P. and Véron Jacques, 2002. L'Inde : un milliard d'habitants en 2001, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde. Géants démographiques et défis internationaux,* p. 227–243. Paris, INED-PUF, 768 p. (Les Cahiers de l'INED, no. 149, 2nd ed. rev.).
Nanda Aswini and Véron Jacques, 2005. Child sex ratio imbalances, fertility behaviour and development in India: recent evidence from Haryana and Punjab, *in*: Isabelle Attané and Jacques Véron (eds.), *Gender discrimination among young children in Asia*, p. 91–131. French Institute, Pondicherry, India (Collection Sciences sociales no. 9, IFP and CEPED).
Nortman Dorothy L., 1978. India's new birth rate target: an analysis, *Population and Development Review*, vol. 4(2), p. 277–312.
Raghavachari S., Natarajan K. S., Biswas A. K. and Bawa Satinder S., 1974. *The population of India.* New Delhi, Paris, Office of Registrar General and Census Commissioner, CICRED, 152 p.
Rajan S. Irudaya, 2005. Introduction: emerging demographic change in South India, *in:* Christophe Z. Guilmoto and S. Irudaya Rajan (eds.), *Fertility change in South India,* p. 23–52. New York, Sage Publications, 452 p. + maps.
Registrar General, 2002. *Sample Registration System Bulletin,* vol. 36(2). New Delhi (India), Registrar General, India Vital Statistics Division, 4 p.
Registrar General, 2004. *Sample Registration System Bulletin*, vol. 38(1). New Delhi (India), Registrar General, India Vital Statistics Division, 6 p.
Registrar General and Census Commissioner, 2001. *Census of India, 2001.* New Delhi, Aravali Printers & Publishers, 183 p. (Series 1, Paper 1).
Retherford Robert D. and Roy Tarun K., 2003. *Factors affecting sex-selective abortion in India and 17 major states.* Mumbai (India) and Honolulu (Hawaii), International Institute for Population Sciences and East West Center Program on Population, 80 p. (National Family Health Survey Subject Reports, no. 21).
Singh Karan, 1976. National population policy: a statement of the Government of India, *Population and Development Review*, vol. 2(2), p. 309–312.
Sinha V. C., 1979. *Dynamics of India's population growth.* New Delhi, National Publishing House, 506 p.
Srinivasan Krishnamurty, 1982. Population policy and programme, *in:* ESCAP (ed.), *Population of India,* p. 158–188. Bangkok and New York, ESCAP, United Nations, 414 p. (Country Monograph Series, no. 10).
Srinivasan Krishnamurty, 1995. *Regulating reproduction in India's population: efforts, results, and recommendations.* New Delhi and Thousand Oaks (California), Sage Publications, 329 p.
Thakor V. H. and Patel Vinod M., 1972. The Gujarat State Massive Vasectomy Campaign, *Studies in Family Planning*, vol. 3(8), p. 186–192.
United Nations, 2003. *World population prospect. The 2002 revision,* 2 vol. New York, United Nations, Department for Economic and Social Affairs, Population Division, 781 p., 939 p., and CD-Rom (ST/ESA/Ser.A/225).
Véron Jacques, 1997. La transition démographique en Inde, *Espace, populations, sociétés*, no. 2–3, p. 135–144.
Véron Jacques, Rajan S. Irudaya, and Padmavathi Pulavarty, forthcoming. La régulation des naissances en Inde: d'une offre de stérilisation par l'Étatà une demande des couples.
Visaria Leela, 2004. Mortality trends and the health transition, *in:* Tim Dyson, Robert Cassen, and Leela Visaria (eds.), *Twenty-first century India. Population, economy, human development, and the environment*, p. 32–56. New York, Oxford University Press, 414 p.
Visaria Leela and Visaria Pravin, 1995. India's population in transition, *Population Bulletin (Population Reference Bureau, Washington)*, vol. 50(3), p. 1–51.
Visaria Pravin, 1995. Demographic transition and policy responses in India, *Demography India*, vol. 24(1), p. 1–12.
Zachariah Kunniparampil C., 1997. Demographic transition: a response to official policies and programmes, *in:* Kunniparampil C. Zachariah and S. Irudaya Rajan (eds.), *Demographic transition, determinants and consequences,* p. 79–111. New Delhi, Sage Publications, 367 p.
Zachariah Kunniparampil C., Mathew Elangikal Thomas, and Rajan S. Irudaya, 2003. *Dynamics of migration in Kerala: dimensions, differentials and consequences.* Hyderabad, Orient Longman Pvt Limited, 476 p.

CHAPTER

114

Population Policies in Mexico

MARIA E. COSIO-ZAVALA
Université de Paris X-Nanterre, Nanterre Cedex, France

Two periods can be distinguished in the development of population policies in Mexico: 1974 to 1994, the period of the first steps and the implementation of a Malthusian policy, and from 1994 to the present day, the revision period of the methods and objectives. It must be stressed that these two different population policy models were largely inspired by the discussions and the changing objectives within the international community, Mexico having been particularly active in the preparation of three world conferences on population: Bucharest in 1974 (with Antonio Carrillo Flores, a Mexican diplomat, as Secretary General), Mexico City in 1984, and Cairo in 1994. The turning point of Cairo was particularly distinct and coincided in Mexico with progression from the Malthusian model, which, in the Mexican population policy, resulted in the adoption of quantitative objectives, to the new inspired model of Cairo where protection of reproductive health and individual care for men, women, and families was central.

The decline in the total fertility rate (TFR) was considerable, falling from an initial level of 6.5 children per woman in 1975 to 2.8 in 2000. This rapidly entailed a decline in the natural growth rate, which in the same period fell from 3.1% to 1.6% (Table 114–1). On the other hand, the improvement in the population's standard of living, the complementary objective in population programs, suffered from two lost decades in Mexican economic development, with negative gross domestic product (GDP) growth rates (–0.44%, between 1980 and 1990 and since 1999), and an increase in poverty (which was estimated as affecting 54% of the population in 2000).

I. GENESIS AND IMPLEMENTATION OF A MALTHUSIAN POPULATION POLICY IN MEXICO[1]

After the results of the 1960 census revealed that Mexico's population was growing at a rate close to 3.5% per year, the country adopted the United Nations viewpoint and became one of its best pupils. The first Mexican Population Law was adopted by Parliament in December 1973, just prior to the United Nations International Conference on Population in Bucharest (1974). Previously, a debate had been started by some Mexican economists regarding the relationship between population growth and economic growth. Debates developed on this subject, notably from 1960, when those in favor of interventionism, including the most authoritative, clashed with the supporters of laissez-faire (Cosio-Zavala, 1994).

Following the example of Mexico, the issue of population policy was then discussed by all Latin American countries at the government level, during preparations for the Bucharest conference. As Léon Tabah, Director of the United Nations Population Division, stated, the Conference was doubtless the clearest demonstration at a global level of the concept of intervention in population issues. The most comprehensive text on this point was the *World Population Plan of Action* adopted by 136 governments (Tabah, 1976). In all of these discussions, Mexico appeared convinced of the importance of population policies,

[1] This section reiterates the basic points of an earlier article on population policies in Mexico up to 1988 (Cosio, 1989).

TABLE 114–1 Population Indicators, Mexico, 1975–2000

Indicator	1975	1985	1995	2000
Population				
Total population (thousands)	59,083	75,462	91,138	98,872
Annual rate of natural growth (%)	3.1	2.2	1.8	1.6
Proportion of 0–14 years (%)	46	42	36	33
Proportion of 60 years and over (%)	6	6	6	7
Fertility and contraception				
Total fertility rate	6.5	4.2	3.1	2.8
Fertility rate at 15–19 years (per 1000)	116	95	77	70
Proportion (%) of contraceptive users (all types)	30	53	67	71[a]
Proportion (%) of modern contraceptive users	23	45	58	n/a
Mortality and life expectancy				
Female life expectancy at birth	65	71	75	75
Male life expectancy at birth	60	64	69	70
Infant mortality rate (per 1000)	69	47	34	31
Urbanization				
Proportion of urban population (%)	63	70	73	74
Urban growth rate (%)	4.3	3.2	2.1	1.9
Percentage of population aged 6–14 years attending primary and secondary schools				
Primary/secondary school attendance (%) of women	72	87	88	90
Primary/secondary school attendance (%) of men	83	90	89	90

n/a = not available; [a]Tuirán *et al*., 2002; Source: United Nations, 2002.

and very advanced in the definition of their implementation, in a context of integrating population issues into development projects. The Mexican population policy thus served as an example for other countries in Latin America, Asia, and Africa until 1994.

1. The Population-Resource Debate

The western neo-Malthusian stream was echoed in Benjamin Viel's resounding book on the population explosion, published in 1966 in Chile. This Chilean physician sounded the first alarm in Latin America, sensitized by his daily confrontation with poverty. Abroad, mainly in the United States, neo-Malthusian thinking developed in a climate which Pierre Pradervand (1974) described as fear caused by the population growth of the Third World in general, and of Latin America in particular. In his preface to Adelyne More's publication (1917), William Robinson, by proposing to "preach the gospel of birth control among the peoples of Russia, China, Japan, India, and Mexico," demonstrates the obvious political agenda behind the choice of the countries mentioned. For example in Mexico, population growth was still very low at the beginning of the 20th century, and an immigration policy was considered necessary. However, in 1974, Pradervand wrote,

> Fear, which is the main motivation for neo-Malthusian thinking—and for many Western reactions vis-à-vis the Third World [. . .] will only reinforce the illusion that the rich countries can continue to prosper materially while the Third World becomes poorer (relatively), and delay the moment when the truth will break out, all the more violently as we have sought to repress it.

In Mexico City, in 1970, the first Latin American regional population conference provided the opportunity to develop the different points of view that are the bases of Latin American population policies. Economists such as Raúl Prebisch, who at the time was Director General of the Latin American Institute for Economic and Social Planning (ILPES), lobbied for structural social reforms, for vigorous economic growth, and for a more egalitarian distribution of revenue, for which it is necessary to support family planning,

> because the Latin American economic system in general, as it operates has demonstrated its incapacity to cope with the phenomenon of the productive absorption of the extraordinary increase in the work force which results . . . from the population growth that started 15 or 20 years before. (Prebisch, 1970)[2]

During this conference, senior officials, such as Carmen Miró, director of the Latin American and Caribbean Demographic Center (CELADE), Victor Urquidi, a Mexican economist and chairman of the Conference's organizing committee, and representatives of the Mexican government of President Díaz Ordaz, stressed the need for a coordination of the various sectoral development policies that would be able to take into account the interdependence of the problems in question. The proposal of integrating population policies into economic development policies and of not locking them into the narrow framework of family planning was at the core of Latin American thinking in general, and Mexican thinking in particular.

Nevertheless, at the beginning of the 1970s, a section of scientific opinion remained largely hostile to direct intervention in family planning supported by the United States and considered as a North American neo-imperialist instrument, based on the neo-Malthusian assumption of population growth as a major obstacle to economic development. However, the official North American theses, and especially the foundation of economic theory that supports them, were refuted in Latin America by economists (Fucaraccio, 1970; Herrera *et al*., 1976) who notably criticized the analyses of Ansley Coale and Edgar Hoover (1958) on India and Mexico, according to which population growth was preventing economic growth. These authors notably demonstrated that the neo-classical hypothesis according to which economic growth depended on investments and growth in household savings could hardly be applied in Latin America, where poverty was largely structural, as a direct result of a very unequal distribution of income.

Indeed, between 1941 and the end of the 1960s, high economic growth and high population growth went hand in hand in Mexico, and also in Brazil, Venezuela, Colombia, Peru, Chile, and other countries. These were the years of the "Mexican miracle" (Ham Chande, 2003). All the studies on the empirical comparative analysis of the relationship between population growth and economic growth reached the same conclusion: that there was no significant negative correlation between the two phenomena, which destroyed one of the basic arguments of the neo-Malthusian theories. In particular, according to the United Nations' Economic Commission for Latin America and the Caribbean, unequal development modes were far more responsible for poverty, unemployment, and the urban problem, than high population growth and the rural exodus toward the large metropolitan areas.

However, it was generally acknowledged by economists and demographers that the accelerated growth of the Latin American population was an exacerbating factor in the problems of underdevelopment and that a reduction in mortality alone should not be counted on to cause a spontaneous reduction in fertility (Tabah, 1987). Social differences and inequalities in the attribution of development benefits were to blame for the extremely high fertility of the end of the 1960s (Cosio-Zavala, 1994). However, Latin American economists and demographers were determined to dissociate themselves from the institutions that supported policies based only on family planning, and which were justified due to their low cost and high effectiveness (Fucaraccio, 1970).

Thus, Carmen Miró demanded a population policy "because the success [. . .] of government programs to ensure free and egalitarian access to the benefits of development for everyone should not be delivered randomly from a so-called systematic adaptation of demographic behavior"[3] (Miró, 1970). But, for Carmen Miró, family planning measures were not population policies, because in Latin America of the 1970s, they were neither coordinated, coherent, nor integrated into national development policies. The definition of a population policy for the Latin American nations was so complicated and had "such a wide meaning that it included nearly all the objectives of development"[4] (Miró, 1970).

Furthermore, foreign intervention in family planning clashed with highly developed nationalist feelings, both on the right and the left. In Brazil, there was a scandal in 1968 regarding sterilizations carried out

[2] "El sistema económico latinoamericano en general, tal como está funcionando, ha demostrado su incapacidad para hacer frente a este fenómeno de absorción productiva del incremento extraordinario de la fuerza de trabajo, que ha resultado y que viene resultando . . . adel incremento de la población iniciado 15 o 20 años antes"(Prebisch, 1970).

[3] "Asegurar el acceso libre e igualitario de los beneficios del desarrollo a todos los habitantes [. . .] no es posible dejar librado al azar de una supuesta adaptación automática del comportamiento demográfico, el logro de las metas específicas de la política estatal" (Miró, 1970).

[4] "Una acepción tan amplia que practicamente quedaron englobados dentro de ella todos los objetivos del desarrollo" (Miró, 1970).

on Amazonian Indian tribes, which resulted in the expulsion of North American missionaries. The Bolivian film *La sangre del Condor* (The Blood of the Condor), by Jorge Sanjines (born in La Paz in 1936) and co-signed by physicians, was also a protest against the sterilization programs carried out in Bolivia on indigenous women without their knowledge.

2. Adoption of a Population Policy in Mexico

Despite these different points of view, the principles used by the Mexican government as a basis for implementing a population policy were reinforced by the emergence of a Latin American position regarding population policies. In Caracas in 1967, the ministerial conference organized by the Organization of American States (OAS) and the Pan-American Health Organization on population and development policies in Latin America tacitly accepted that "there is no population policy but rather a development strategy, and there can be no development without a population policy" (Concepción, 1970). This was surely a reflection of a compromise between the pro-birth stance, which was that of the majority (including Mexico) at the time and the more Malthusian stance of other countries (e.g., Colombia), closer to that of the large multilateral agencies and some of the people in charge of economic and social policies.

However, the declaration in 1972 of Mexico's president, Luis Echeverría, at the third UN Conference on Trade and Development (UNCTAD), in Santiago, Chile, was quickly followed in 1973 by the promulgation of the Mexican law on population, the aim of which was to reduce the rate of population growth, thus making Mexico the first Latin American country to officially adopt an explicit population policy.

President Luis Echeverría's decision, although he displayed populationist opinions at the beginning of his mandate, was finally made during 1971, under the influence of arguments by economists who demonstrated that the Mexican population would experience high growth with direct damaging effects on a labor market incapable of absorbing the annual flow of young people starting their working lives. According to Raúl Benítez and Gustavo Cabrera's population forecasts (1966), Mexico's population would double in the 20 years between 1960 and 1980, and this is exactly what happened.

Luis Echeverría thus declared in Santiago: "We should not forget that the human race will double in a single generation. Although reducing the population growth rate is desirable, the first task for the Third World is to give its economic development a vigorous push"[5] (United Nations, 1972). In this context, the means of achieving this objective was through the reduction of population growth.

Following the adoption of the World Population Plan of Action in Bucharest in 1974, in spite of the divergences between different Latin American countries, a common attitude regarding population policies began to develop. The second Latin American population meeting, held in Mexico City in 1975, recognized the rights of individuals and families to decide on the number and spacing of their children and the state's duty to provide them with the information and the means to practice family planning. Put in these terms of equality and individual freedom, the concept of guaranteed access to family planning became not only acceptable, but also desirable for governments. In 1975, the Mexico Declaration (CEPAL, 1975) also proclaimed each nation's sovereignty regarding population policies.

The principle of integrating population policies into development planning was reaffirmed in Mexico City in 1984, during the second International Population Conference (CEPAL, 1984). The Latin American position therefore benefited from a general consensus among the developing countries during various regional and international meetings, a position that was defined notably during the Regional Latin American preparatory meeting for the Mexico City conference held in Havana in November 1983.

It can be summarized as follows:

To respect the national sovereignty to freely make decisions regarding population policies.
To integrate the population policy into economic and social development planning.
To respect the rights of individuals and families to freely decide the number and spacing of their children;
To provide the necessary information and means to give individuals and families freedom in their decisions regarding reproduction.

In these general terms, it can be said that Mexico's political doctrine on population served as a basis for the common Latin American position developed during the various international population conferences which, apart from these principles, also included an explicit reference to solving the food problem, to reducing mortality, to the right to health care, and to

[5] "Tampoco hay que olvidar que la raza humana va a duplicarse en una sola generación. Si entonces es deseable una reducción de la tasa de crecimiento de la población, la primera tarea que incumbe al Tercer Mundo es la de dar a su desarrollo un impulso viguroso" (United Nations, 1972).

planned regional development. The Havana Declaration (CEPAL, 1984) was adopted in November 1983 during the Latin American regional preparatory meeting for the International Population Conference held in Mexico City in 1984. Since 1984, the state's principle of intervention in population problems was thus widely accepted in Latin America, particularly with regard to guaranteed access to modern contraceptive methods for couples and the necessity of modifying the spatial distribution of the population, which a large majority of the country considered as being significantly imbalanced.

3. Implementing the Mexican Population Policy (1974–1994)

During the 10 years between the United Nations Population Conference in Bucharest in 1974 and the one in Mexico City in 1984, awareness of the primary importance of population problems in economic and social development strategies, as well as planning, had increased. Mexico thus equipped itself with the legal and institutional means to implement a population policy and integrate it into various national development plans. It was based mainly on a national family planning program and an internal migration policy. Programs concerning health, mortality, and morbidity also had demographic consequences, as did both educational programs relating to population and legal provisions relating to international migration, but they were not an explicit part of the Mexican population policy. Their consequences on the population were indirect, resulting from other priorities pertaining to health, education, or foreign policy. Furthermore, the growing concentration of wealth that accompanied Mexican economic development resulted in an exacerbation of the economic and social gaps between rich and poor regions. The reduction of these differences has never specifically been taken into account in population policies, either directly or indirectly.

The continuous monitoring of government perceptions and policies relating to the demographic situation and to its evolution in the development context resulted in a series of UN surveys which made it possible to trace the governments' positions on population policies every two years. Yet, Mexico declared that life expectancy levels in 1976 and 1986 were "acceptable" and did not provide a quantitative objective for mortality (Table 114–2).

It is true that the regular and significant increase in life expectancy at birth since 1930 (Figure 114–1), obtained mainly due to the decline in infant mortality (Figure 114–2), provided proof of the adequacy of health policies, which relied on the large public health bodies (particularly the Mexican Institute of Social Security and the Ministry of Health).

The objectives of the health programs pertained only to specific actions targeting certain categories of the population (infants and children aged under 5 and their mothers), or certain particular causes of mortality (e.g., infectious diseases). Health program objectives were generally limited to the health sector, with

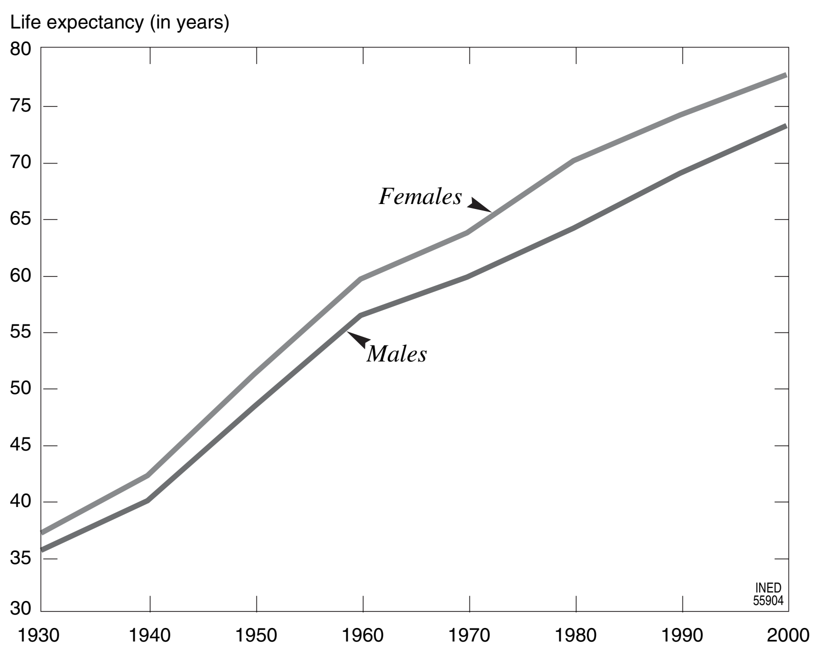


FIGURE 114–1 Trend since 1930 in life expectancy at birth. (Source: United Nations, 2002.)

TABLE 114–2 Opinions and Policies in Mexico, 1976–2001

Policies	1976	1986	1996	2001
Population growth policies				
Opinion	Too high	Too high	Too high	Too high
Growth objective	Reduce	Reduce	Reduce	Reduce
Fertility and family planning				
Opinion	Too high	Too high	Too high	Too high
Objective	Reduce	Reduce	Reduce	Reduce
Access to contraception	Direct aid	Direct aid	Direct aid	Direct aid
Adolescent fertility			Very concerned	Very concerned
Programs	—	—	IEC	IEC
Health and mortality policies				
Life expectancy	Acceptable	Acceptable	Acceptable	Acceptable
Mortality before age of 5	—	—	Unacceptable	Unacceptable
Opinion on AIDS	—	—	Very concerned	
Induced abortion, permitted in the case of			Saving the mother's life; Rape	Saving the mother's life; Rape
Spatial distribution policies				
Opinion	Change as much as possible	Change as much as possible	Change as much as possible	Change as much as possible
Regional planning policy	—	Yes	Yes	—
Urban-rural planning	—	No intervention	No intervention	—
Internal migration policies				
To metropolitan areas	No intervention	Reduce	Reduce	—
From rural zones	No intervention	Reduce	Increase	—
International migration policies				
Immigration				
Opinion	Satisfactory	Satisfactory	Satisfactory	Satisfactory
Policy	Maintain	Maintain	Reduce	Maintain
Emigration				
Opinion	Too high	Satisfactory	Too high	Too high
Policy	Reduce	Maintain	Reduce	Reduce
Encourage emigrants to return	No	—	Yes	Yes

IEC = Information, education, and communication; Source: United Nations, National Population Policies, 2002.

no reference to the general population and development policy. Regarding international migration, the Mexican authorities considered that the flows observed were satisfactory and should be generally maintained at their levels (Table 114–2).

On the other hand, regarding fertility, recognized as an essential component of population growth, Mexico began constitutional and legislative reform. The changes made in 1973 to Article 4 of the Constitution concerned the right of individuals to choose the number and spacing of their children and expressly encouraged responsible paternity. Development planning referred explicitly to rapid population growth as an impediment to economic and social development.

The population policy was coordinated by the *Consejo Nacional de Población* (National Population Council [NPC]), which was created in 1976. The NPC's missions are to develop, monitor, and coordinate the national population program. Up to the early 1980s in Mexico, both the NPC and the Council for the Coordination of National Family Planning Program within the Ministry of Health were responsible for the family

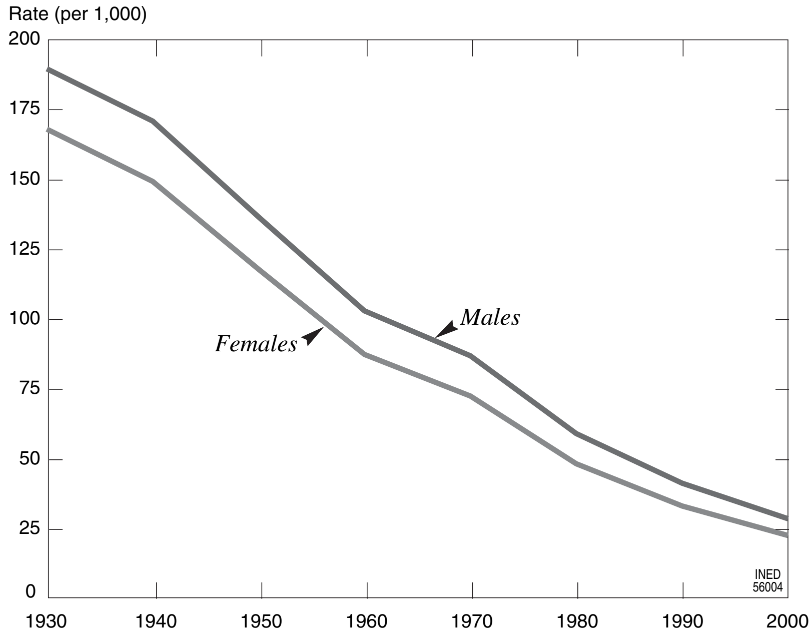


FIGURE 114–2 Trend in infant mortality rate since 1930. (Source: United Nations, 2002.)

planning programs of all institutions. Of the two councils, the only one that still exists today is the NPC, which since 1983 relies on the State Population Councils of the 32 states which make up the Mexican federation and which regulate the Ministry of Health's reproductive health programs.

At the beginning of the 1980s, the international community encouraged the government to intervene in population policies. Furthermore, physicians and economists were largely in favor of this. The only institutional opposition came from the upper hierarchy of the Catholic Church, following the objection that was officially made by the Vatican on July 25, 1968—The Encyclical of Pope Paul VI on Birth Control. However, surveys have proved that religious practices have very little influence on the reproductive behavior of the Mexican population (Brito, 1971; Leñero, 1979).

4. Laws, Regulations, Institutions

Once the presidential decision to authorize a population policy and a birth control policy (in 1972) had been taken, legislative changes made its application possible. Article 4 of the United Mexican States' political constitution was amended on December 31, 1973. This amendment established the equality of men and women in the eyes of the law, the protection of the family, and the free and informed choice of each individual regarding the number and spacing of his/her children.

The government's basic principles were as follows:

The population policy does not replace the economic and social development policy; it is fully integrated into it and its objective is a long-term one.

The State has a duty to inform, educate, and guarantee the freedom of individuals and families regarding procreation. Responsible paternity is a right, guaranteed by access, free of charge if possible, to family planning.

The State has a duty to protect the family; in particular it should protect women and children in their health and education, and their physical and social well-being.

The General Law on Population of December 11, 1973 was published in the *Official Journal* on January 7, 1974. The decree specifying how the law should be enforced (*Reglamento*) dates from November 1976 and the programs were implemented beginning in 1977, within the new development planning system of President José López Portillo (1976–1982). The first global development plan covered the 1980–1982 period, and the second covered the 1983–1988 period. The national family planning program was presented on October 28, 1977, with the objective of reducing the rate of population growth to 2.5% in 1982, and then to 1% in 2000. It gave the responsibility of extending access to contraceptive methods to the entire population mainly to the national public health institutes, such as the Mexican Institute of Social Security (IMSS) and the Ministry of Health. Furthermore, the first global development plan also defined economic and social measures in the fields of employment, health, education, housing, food, industry, trade, agriculture, and stock farming. But no explicit link was established with the demographic objectives, as these were considered as useful per se.

The results of this policy became rapidly visible in the massive improvement of the entire population's awareness and of access to family planning services; the use of contraceptive methods, considered lawful, became generalized within public health bodies. The proportion of women of child-bearing age using contraception thus increased from 30% in 1976 to 71% in 2000 (Tuirán *et al*., 2002).

Figure 114–3 shows the trend in TFR and completed fertility between 1900 and 1995, which confirms that fertility had already begun to decline from around 1968 onward, that is, 10 years before the promulgation of the national family planning program. The latter had therefore been well accepted, with contraceptive use preceding legislation, notably among the small urban and educated minority group known as the "family planning pioneers" in Mexico (Cosio-Zavala, 1994). Indeed, this is why the Population Law received the sanction of the members of Parliament, as they themselves belonged to these pioneer groups.

5. Spatial Distribution and Internal Migration Policies

During the first period of development planning in Mexico, government bodies were mainly concerned with planning economic activities, without explicitly taking into account the demographic consequences of actions undertaken at regional and local levels. During the preparation of the Mexico City conference in 1984, the associations between development policies and population policies were the subject of numerous studies, and consideration of this theme led the Mexican government to take a position. Mexico considered that the distribution of its population in the national territory was inadequate at a regional level and wished to modify it. Among the main social problems in the country were unemployment and poverty, associated with the rapid and poorly controlled development of urbanization. The urban growth rate was estimated at 3.2% in 1985 (United Nations, 2002).Policies for the spatial redistribution of the population were therefore developed. They were based on four general recommendations:

1. Policies aimed at maintaining and retaining the population in the departure zones, encouraging the maintenance of a significant proportion of the population in rural zones with the aid of productive investments, an increase in the quality of human resources

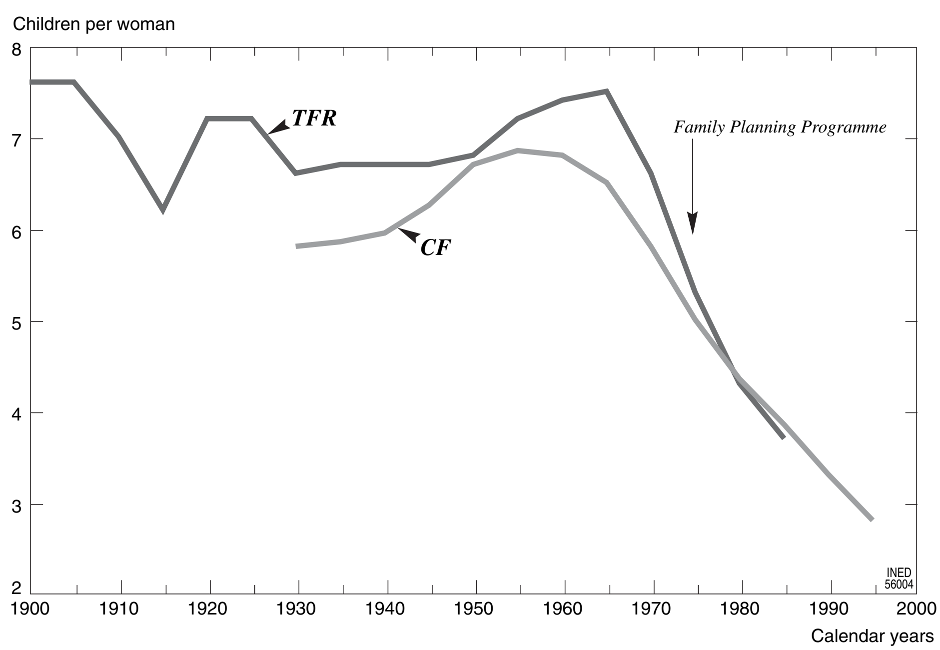


FIGURE 114–3 Trends in total fertility rate (TFR) and completed fertility (CF) during the 20th century (CF is represented with a difference of 26 years which corresponds to average age at maternity). (Source: Cosio-Zavala, 1994, p. 207.)

in agriculture, and the creation of nonagricultural employment in rural regions;

2. Policies for the reorientation of migratory flows to divert migration from the large metropolitan areas to other towns, in particular medium-sized towns; examples of these policies were provided by strategies for development areas, in particular surrounding industrial areas, ports, and mining areas; a coastal development program was prepared, in a country where the population is largely concentrated at an altitude above 1500 meters, a historical inheritance linked to climatic and geographical factors.

3. Rural development projects, projects for the agricultural colonization of new zones or the development of new agricultural techniques or new produce in traditional rural zones; these projects frequently invited rural populations from the most underprivileged or overpopulated areas.

4. Decentralization policies, with the goal of slowing the growth of the large urban zones, in particular by limiting the creation of new employment in the major cities, through restrictive measures for setting up companies, and for the public sector and administrative services, by moving civil servants and administrative services out of the large cities. Mexico wanted to place some federal administrative services outside of Mexico City, but this was only approached timidly by moving the National Institute for Public Health to Cuernavaca, only 70 kilometers from Mexico City, and INEGI (*Instituto Nacional de Estadística, Geografía e Informática* [National Institute of Statistics, Geography, and Computer Science]) based since 1987 in Aguascalientes.

The four principles stated above have defined an explicit policy for internal migration since 1979 called "the three R policy" (*retención, reorientación, reubicación,* [retention, reorientation, and decentralization]) (CONAPO, 1978), which is the program for population distribution in the national territory, and has been the subject of specific and coordinated measures with the sectoral programs of the National Development Plan. (In Mexico, planning is implemented in large sectoral programs, which correspond partly to social policy and partly to economic activity sectors.) The various intervention zones were defined according to regional demographic characteristics, through the State Population Council of each of the 31 states and the federal district. Relations between the population policy and economic and social planning were identified in order to integrate population and development. However, the NPC did not have the necessary tools nor the specific resources to implement its policy, which implied the population policy's subordination within the sectoral plans, which were sometimes in opposition in terms of results. For example, certain aspects of the regional industrial policy, boosted by the governors of dynamic states, were likely to be in opposition to the regional policy of decentralization defined by the NPC at a federal level. This is why the results of the policy for the spatial distribution of the Mexican population are generally regarded as disappointing. The programs implemented had few positive results. Control of migratory movement was difficult to ensure while the components essential to regional and international economy were beyond the control of national planning, notably due to the principle of laissez-faire applied to productive investments. This is why population distribution continues to be extremely polarized and concentrated in the very large cities (Figure 114–4).

Indeed, policies regarding population distribution in Latin America (as in almost everywhere in the world, including in developed countries) have never been very successful. Even with the existence of political good will, such as in Mexico, the intervention methods were frequently missing. The expected results of these redistribution policies were often thwarted by the side effects of macro-economic and sectoral development policies (UNFPA, 1984).

6. Assessment of the First Phase of the Population Policy in Mexico (1974–1994)

The population policy in Mexico was developed under extremely varied forms and with a strong impetus and an explicit will for a population policy in all fields (population structure and growth, fertility and birth control, internal migratory flows, and spatial distribution). However, it is the policy instruments more than the policy formulation that determine the results. Regarding fertility, the progress in medical techniques and the intensity of interventions in the area of contraception, in the context of health systems, had a significant effect on family planning programs, particularly for the most underprivileged sections of the society, with low incomes and low levels of education (Brugeilles, 2002).

The huge intervention of the public health bodies' family planning programs quickly caught up with the NPC's policy of education regarding population, promoting women, and social communication. The activities carried out throughout the health system were based on highly effective, and even radical, medical techniques such as sterilization, not dissimilar to the route taken by the mortality decline, in a period before the emergence of the concept of a population policy, when the Pan-American Health Organization had

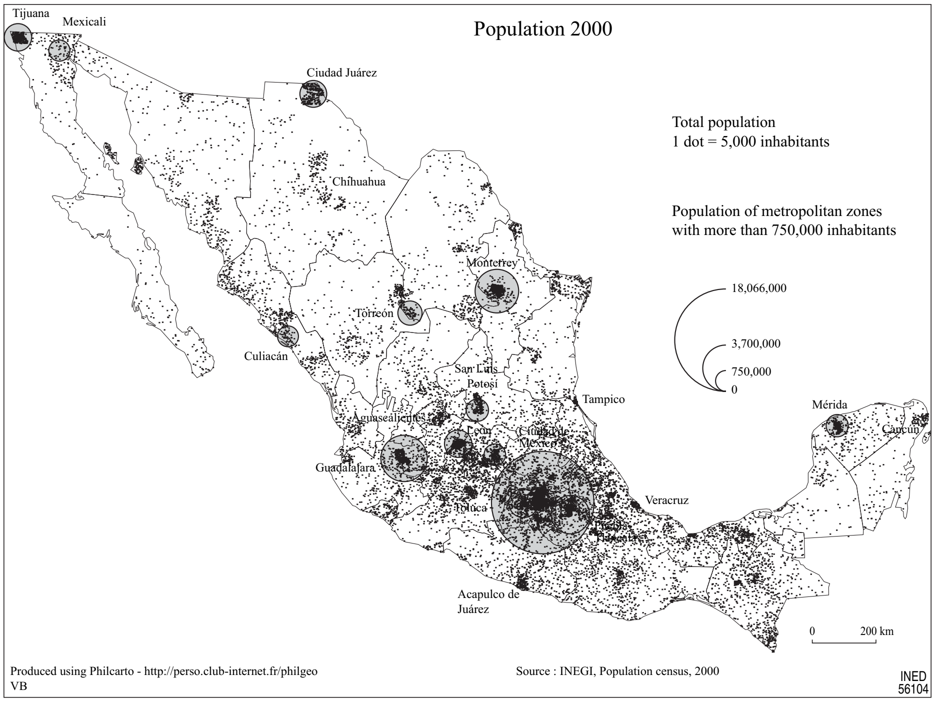


FIGURE 114–4 Geographic distribution of the Mexican population (dots), and metropolitan zones with more than 750,000 inhabitants in 2000.

already been implementing coordinated public health policies for several decades (since the 1930s).

The experiments on the spatial distribution of the population were considered rather disappointing, due to lack of effective tools for controlling internal migratory movement.

Finally, it seems that, until the 1990s, Mexico's population policy only had a real effect on contraceptive practices, despite the more ambitious starting principles of raising the standard of living and reducing social inequalities. Steps were taken after 1994 to widen this policy of reducing fertility by generalizing access to both information and contraceptive methods, and by integrating the policy into social and educational policies and policies for increasing women's status.

II. THE POPULATION POLICY IN MEXICO AFTER 1994

The objectives of the 1977 National Family Planning Program—a natural population growth rate of 2.5% in 1982 and 1% in 2000—were not achieved, as evidence by a natural growth rate of 1.6 in 2000, 60% higher than the 1% hoped. The objectives were actually badly defined, as the annual rate of natural growth depends on age structure which, in a young country, in itself contains an immense growth momentum (see Chapter 4, Volume I). The 1995–2000 National Population Program modified these objectives by aiming to reach replacement level by 2005 (i.e., a TFR of 2.1 children per woman by that date). This objective was confirmed by the 2001–2006 National Population Program. In order to achieve this, the proportion of women using contraception must reach 73.5% of all women of childbearing age living with a spouse. This forecast seems realistic, as the progress required is lower than that observed in the 1997–2001 period. The extrapolation of recent fertility trends is even leading the NPC to forecast a TFR of 1.68 children per woman in 2030 (Tuirán *et al*., 2002).

After the 1994 Cairo Conference

Since the International Conference on Population and Development held in Cairo in 1994, the World Population Plan of Action was oriented in new directions and Mexico, like many other countries, modified its population programs. As Louise Lassonde (1996) explained, the reproductive health and family planning programs became independent from demographic targets (population stabilization) and acquired a legitimacy based on the fact that individual well-being and health are ends in themselves. Could this change in perspective divert attention from regional or global demographic issues as Louise Lassonde fears? In any case, it has, to its credit, placed the emphasis on new themes, such as preserving the environment and improving women's status.

Reproductive health is one of these new themes. It includes the fight against sexually transmitted diseases, AIDS, maternal and child mortality, illegal induced abortion on the one hand, and on the other hand, promoting reproductive and sexual rights, that is, the consensual realization within a couple of each other's reproductive and sexual aspirations, for both men and women. Family planning is part of this, but is not the only aim of the programs, of which quality and adaptation to the couple's needs take precedence. In order to conform to these objectives, the programs implemented so far have had to be significantly reorganized. In a continued effort to improve the status of women, the fight against inequality in the social relations of gender, violence, and the denial of women's rights are also part of the new programs.

Education regarding population has also become a priority, within formal education on the one hand, and within mass communication on the other hand. It is addressed first and foremost to young people, by developing positive values and civil behavior in the context of the family and the community. It also aims at overcoming retrograde attitudes, notably with regard to adolescents' sexual education and information on contraception. *Gente jóven*, an experimental program, was developed with young people from Mexico City's working-class neighborhoods, based on their peers as promoters of reproductive health (Marston, 2001).

Another important theme discussed in Cairo was the issue of the population's well-being. In a context of neo-liberal, economic policies and the regulation of employment through market laws, we observe the state's retreat in the fight against social inequality and a significant deterioration in the standard of living of the increasingly numerous poorest sections of society. Mexico, as well as the other Latin American countries, is looking for political alternatives to redefine the economic and social role of governments in the current context of globalization (Welti, 1997), as centralized economic and social development planning has reached its limits and become largely obsolete.

The concept of sustainable development emerged with a view to preserving the environment, nonrenewable resources, quality of life, and conserving the natural environment for future generations. After the 1992 Rio de Janeiro Summit on the environment, the ecological damage in the Amazon could no longer be ignored, no more than could the general deforestation, the great fires caused by illegal clearing (such as in northern Mexico during the summer of 1998), the insalubrities of the slums of the large cities (called *tugurios* in Mexico), or the serious atmospheric pollution in Mexico City, as well as that in many other large industrial cities.

The themes of migration and urbanization are also part of the priority actions for a readjustment of the population's spatial distribution, urban decentralization, and also the reception of migrants in arrival zones, notably those who have been victims of forced migration or are looking for refuge from political violence (such as the Guatemalans in southern Mexico during two decades). International migration has also become a priority political issue, even though its effect on population numbers remains low. Thus, the net Mexican emigration to the United States estimated at 300,000 per year, only corresponds to a net emigration rate of 0.3%, which is however the highest rate in Latin America. Since 1996, the Mexican government has attempted to control migratory flows and to encourage the emigrants to return, in clear discontinuity with the previous policies of nonintervention in the face of migration levels that were previously considered as satisfactory (Table 114–2). However, the Mexican government's willingness to negotiate a veritable international migration policy with its powerful North American neighbor came up against the events of September 11, 2001, and with President Bush's consequent refusal to discuss it further.

In the context of the discussion of the state's role and the questioning of governmental monopoly of social policies, there is strong support and mobilization in favor of Cairo's objectives in Mexico, insofar as they are concerned with the well-being of the population and the protection of human rights and women's rights. Private institutions and nongovernmental organizations have in the past demonstrated their effectiveness and their capacity to intervene alone or in collaboration with state social programs (e.g., in the area of the family or child health). This is what Benítez (1999) calls "a new demographic culture." It will have positive effects due to the participation of all citizens encouraged by the current context of political democratization in nearly every country (Welti, 1997).

However, a major theme of discussion in the Cairo conference was the Catholic Church's reaffirmed traditional opposition to contraception and induced abortion and many Latin American countries officially supported this position, notably Argentina. Indeed, the governments formally accepted this point of view, including Mexico, a secular country which has accepted the separation between the Church and the State for a long time but had not dared to publicly oppose the Catholic hierarchy on the issue of legalizing induced abortion when it was publicly debated at the beginning of the 1980s. On the other hand, the Catholic Church is far less vehement regarding official birth control programs. The Catholic Church is the only institution that is opposed to modern contraceptive methods, but generally it turns a blind eye to the issue, particularly in the poorest sectors. Finally, zoning by sectors of economic activity or by branches of social policy has been replaced by zoning according to the socioeconomic and cultural characteristics of the population to be protected in their reproductive and sexual rights. Groups with different socioeconomic extremes coexist in Mexico: some of them are well integrated into economic and social development (in the large cities, sectors with high cultural and economic capital), and some of them are very marginalized (such as rural, indigenous, uneducated population groups with high infant mortality). Furthermore, a significant increase in the differences between the poorest and the richest groups is observed. In this context, differences in fertility are considerable during the same period, with for example in 1995–1997, TFRs ranging from 1.66 children per woman among executives to 4.7 among uneducated women (Tuirán *et al.*, 2002). Despite these differences, the growing economic cost of children to parents in an increasingly developed consumer society where there exists a generalized medical culture, exposure to mass media, generalized social protection, and improvement in women's social conditions in both education and health has led to the spread of the reduced family model, including in the more marginalized population groups. As for the medical culture in Mexico, reported that 82% of deliveries were carried out by physicians in 1994–1997 (Tuirán *et al.*, 2002). As for Mexicans' exposure to mass media, in 2000, 86% and 85% of the population had a television and a radio, respectively, compared with 31% and 76% in 1970. With regard to generalized social protection, the Mexican Institute of Social Security (IMSS) covered 47 million people in 2000, compared with less than 10 million in 1970, which is respectively 47% and 19% of the national population.

Since 1995 (and up to 2006), the Mexican population policy includes specific actions for the protection of vulnerable population groups, notably in the fields of education and health. It is associated with the poverty reduction program PROGRESA (*Programa nacional de Educación, Salud y Alimentación*), which covered 300,000 families in 1997 (Tuirán *et al.*, 2002) and by 2004 covered 4 million. It aims to promote opportunities for children to succeed and to improve women's participation in society, providing them with more egalitarian living conditions and eliciting more autonomy for women from men with regard to education, employment, and health. This occurs in particular through the provision of study grants for children, particularly girls, so that they can attend school and prepare for working life. PROGRESA benefits for the vaccination, feeding, and education of children are paid directly to the mothers.

CONCLUSION

Since 1995, the Mexican population policy has therefore become above all a poverty reduction policy, partly decentralized, with family benefits based on the poorest families that it aims to protect in areas of health, feeding, and education. It is concerned with bridging the gap that separates the poorest social groups from the richest, with the objective of attaining replacement-level fertility (2.1 children per woman) or even below (1.68 children per woman forecast by the NPC in 2030), by developing access to contraception. The NPC therefore continues to coordinate the family planning policy at a national level, despite the decentralization of health policies at the state level.

The implemented actions are aimed at improving the standard of living of the most vulnerable social groups. Priority to the Malthusian objectives of the 1980s has therefore been abandoned, and the idea that their achievement would automatically result in the social progress desired as a consequence of the moderation in the rate of population growth has also been abandoned. This simplistic idea has been contradicted by the facts. The reduction in the population's growth rate has not reduced the considerable social differences; on the contrary. The government has reoriented its population policy in order to achieve this objective, through new poverty reduction programs (such as PROGRESA), which, far from aiming only to control fertility through family planning, also aim at directly reducing the gaps between the poorest sections and the rest of the population in terms of education, children's health, and the promotion of women's economic activity and their participation in social life.

Acknowledgment

This chapter was translated from the French by Accenta Ltd.

References

Benitez Zenteno Raúl, 1999. Editorial, *in:* IISUNAM (ed.), *DEMOS, Carta demógrafica sobre México,* Mexico, IISUNAM, 24 p.

Benitez Zenteno Raúl and Cabrera Acevedo Gustavo, 1966. *Proyecciones de la población de México 1960–1980.* Mexico, Banco de México S.A., 247 p.

Brito Velásquez Enrique, 1971. *¿Quién escucha al Papa? Sondeo efectuado sobre las actitudes ante la encíclica Humanae vitae.* Mexico, IMES, 199 p.

Brugeilles Carole, 2002. *Les politiques de population au Mexique (1974–2000).* Paris, CEPED, 57 p. (Les dossiers du CEPED, no. 71).

CEPAL, 1975. *Actas de la Segunda Reunión Latinoamericana de Población.* Mexico, Comisión Económica para América Latina.

CEPAL, 1984. América latina y el Caribe en la Conferencia internacional de Población, *Notas sobre la economía y el desarrollo de América latina*, no. 397–398, July 1984, 4 p. (Comisión Económica para América Latina).

Coale Ansley J. and Hoover Edgar M., 1958. *Population growth and economic development in low income countries.* Princeton, Princeton University Press, 389 p.

CONAPO, 1978. *Política demográfica nacional y regional 1978–1982.* Mexico, Consejo Nacional de Población.

Concepcion Mercedes, 1970. Políticas de población en el contexto latinoamericano, *in:* International Union for the Scientific Study of Population (ed.), *Conferencia Regional Latinoamericana de Población, Mexico City, El Colegio de Mexico,* vol. 2, p. 308–312. Liège, IUSSP, 519 p.

Cosio-Zavala Maria-Eugenia, 1989. Politiques de population au Mexique, *Politiques de population, études et documents*, vol. 4(1), p. 1–133. (Louvain-la-Neuve, CIACO éditeur).

Cosio-Zavala Maria-Eugenia, 1994. *Changements de fécondité au Mexique et politiques de population.* Paris, L'Harmattan and IHEAL, 256 p. (Recherche et documents Amérique Latine).

Fucaraccio Angel, 1970. El control de la natalidad y el argumento del ahorro y la inversión, *in:* International Union for the Scientific Study of Population (ed.), *Conferencia Regional Latinoamericana de Población, Mexico City, El Colegio de Mexico,* vol. 2, p. 57–63.—Liège, IUSSP, 519 p.

Ham Chande Roberto, 2003. *El envejecimiento en México: el siguiente reto de la transición demográfica.* Mexico, El Colegio de la Frontera norte, Ed. Miguel Angel Porrua, 332 p.

Herrera Amilcar O., Scolnik Hugo G. *et al.*, 1976. *Catastrophe ou nouvelle société? Un modèle mondial latino-américain.* Ottawa, CRDI/IRDC, 108 p.

Lassonde Louise, 1996. *Les défis de la démographie: quelle qualité de vie pour le XXIe siècle?* Paris, La Découverte, 225 p.

Leñero Otero Luis, 1979. *Población, Iglesia y Cultura: sistemas en conflicto.* Mexico, IMES-FERES, 526 p.

Marston Cicely, 2001. *Youth reproductive health in Mexico: can peer leaders make the difference?* Doctoral thesis. London, London School of Medicine and Tropical Hygiene, 285 p.

Miro Carmen, 1970. Política de población: ¿qué? ¿por qué? ¿para qué? ¿cómo?, *in:* International Union for the Scientific Study of Population (ed.), *Conferencia Regional Latinoamericana de Población, Mexico City, El Colegio de Mexico,* vol. 2, p. 276–279. Liège, IUSSP, 519 p.

More Adelyne, 1917. *Uncontrolled breeding, or fecundity versus civilization. A contribution to the study of over-population as the cause of war and the chief obstacle to the emancipation of women*. Introduction by Arnold Bennett; preface and notes by William Robinson. New York, Critic and Guide Co., 108 p.

Nations Unies, 1972. Résumé de l'allocution du Président Echeverría à la 92e session plénière, *in: Troisième conférence des Nations unies sur le commerce et le développement (CNUCED),* vol. IA, p. 185. United Nations. (Sales number: 73.II.D; French version).

Pradervand Pierre, 1974. La peur démographique, *Économie et humanisme*, January–February, p. 61–70.

Prebisch Raúl, 1970. Intervención, *in:* International Union for the Scientific Study of Population (ed.), *Conferencia Regional Latinoamericana de Población, Mexico,* vol. 2, p. 272–275. Liège, IUSSP, 519 p.

Robinson William, 1917. Preface, *in:* Adelyne More, *Uncontrolled breeding, or fecundity versus civilization. A contribution to the study of over-population as the cause of war and the chief obstacle to the emancipation of women*. New York, Critic and Guide Co., 108 p.

Tabah Léon, 1976. La recherche démographique en mutation après Bucarest? *Bulletin démographique des Nations unies*, no. 8, p. 1–16.

Tabah Léon, 1987. Bucarest et Mexico: d'une Conférence de Population à l'autre, *in:* EHESS and ORSTOM (eds.), *Changements démographiques en Afrique et en Amérique latine, Année 1984–1985:* p. 9–15. Paris, EHESS and ORSTOM, 198 p. (Séminaire EHESS; mimeo).

Tuiran Rodolfo, Partida Virgilio, Mojarro Octavio, and Zuñiga Elena, 2002. *Fertility in Mexico: trends and forecast.* Mexico, CONAPO, 27 p. (Paper presented at the Expert Group Meeting on Completing the Fertility Transition, New York, Population Division, 11–14 March 2002).

UNFPA, 1984. *Rapport de la mission d'évaluation. Projet "Intégration de la politique de population aux plans et programmes de développement, MEX/79/P04."* New York, United Nations Population Fund, 69 p.

United Nations, 2002. *National population policies, 2001.* New York, United Nations, Department of Economic and Social Affairs, Population Division, 427 p.

Viel Benjamín, 1966. *La explosión demográfica ¿Cuántos son demasiados?* Santiago, Ediciones de la Universidad de Chile, 241 p.

Welti Carlos (coord.), 1997. *Población y desarrollo: una perspectiva latinoamericana después de El Cairo-94.* Mexico, United Nations Population Fund, MacArtur and IISUNAM, 302 p.

C H A P T E R

115

Population Policy in Kenya

VALÉRIE GOLAZ AND CAROLYNE WANJA NJUE

Institut national d'études démographiques (INED), Paris, France

Population Council, Nairobi I, Kenya

Few African countries adopted an explicit population policy before the 1990s; Kenya is one of the rare exceptions. At different moments of its existence this country engaged in voluntarist demographic policies, in relation to the population's spatial distribution and to health and fertility. Many studies have focused on the two latter issues, particularly since the 1980s. No synopsis on population policy, however, relates health and fertility policies to those concerned with population distribution. In Kenya, nonetheless, the justification for both types of intervention draws on the relationship between population and resources, particularly land.

One particularity of this country stems from its physical characteristics, with a vast semi-arid area contrasting sharply with very fertile and densely populated regions in the highlands and along the coastal strip. This physical polarity has important implications on the settlement patterns, lifestyle, and health of the inhabitants of different regions. But, quite independently of these contrasts, conflicts over resources have long existed in most of these ecosystems: conflicts over fishing in Lake Victoria, over access to water among pastoral groups in northern Kenya, and over access to land among agro-pastoralists in more populated regions. If the existence of these conflicts testifies to deep-seated defense and attack strategies within traditional Kenyan societies, their frequency and violence reflect the fact that Kenya, unlike most other African countries, has a weak ratio of resources to population.

The population's spatial distribution derives not only from the country's physical characteristics, but also from its administrative history. Most present-day conflicts can be seen as a direct consequence of administrative measures taken by the colonial government, rather than of population growth as such. In this chapter, it is therefore essential to assign an important place to the colonial era.

The history of the state of Kenya has three distinct stages (Figure 115–1): its birth, with the early days of British colonization (1895–1945); followed by its transition to, and the first decades of, independence (1945–1984); and, finally, the period of consolidation of Daniel arap Moi's rule (1984–2002) up to the arrival in power of Mwai Kibaki in December 2002. It is certainly too early as yet to say whether this recent political turnaround will be reflected in the demographic arena. While fully aware of the artificial nature of all boundaries, we have nonetheless chosen to use political markers to trace the evolution of population policy, as the timing of political change appears to coincide with modifications in health and fertility.

I. COLONIAL UPHEAVALS

Most of Kenya's current problems originated during the colonial period. Without wishing to lay the entire responsibility for Kenya's contemporary social and economic problems on British colonization, it is nonetheless important at this point to recall the policies adopted by the colonial authorities and the place reserved for the African population. After presenting the problems encountered by the colonizers on arrival, we discuss their handling of population

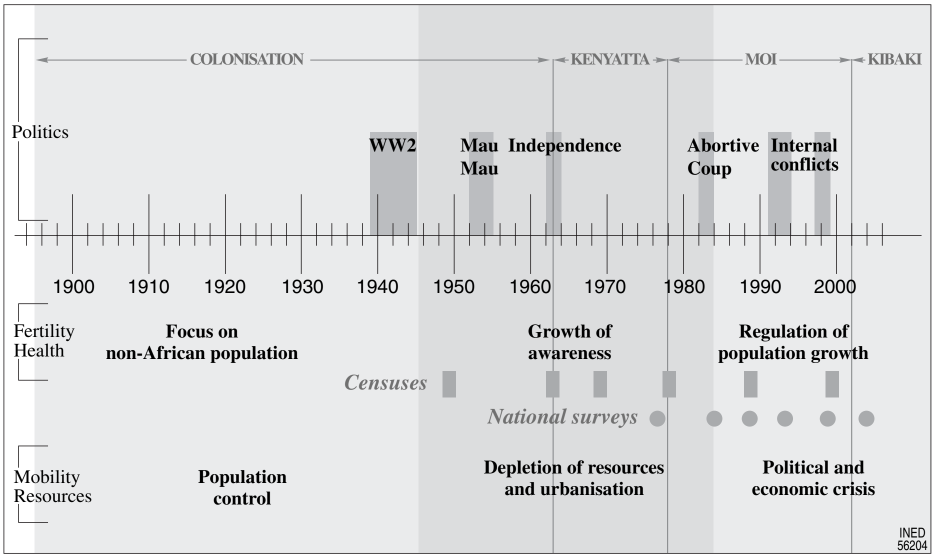


FIGURE 115–1 Chronology of political events and attitudes toward population issues.

settlement and their ignorance of African population demographics.

1. The Colonial Government's Initial Difficulties

On arrival, in 1895, the colonial authorities had to administer as a single entity a set of relatively distinct societies, with diverse ways of operating, different forms of political organization, and with some more closely linked than others. Thus, highly stratified societies (such as the kingdom of Wanga) were grouped with completely acephalous societies into a single state. Furthermore, the end of the 19th century was a period of demographic devastation: epidemics, wars, and famine followed one another, decimating East Africa's cattle and human populations.[1] One might expect, therefore, that the government's first public policies would be directed toward improving health, on one hand, and agricultural production, on the other. But before there could be any question of population policy, the British colonial authorities had to put all their energy into controlling and administering the territory, while ensuring their own survival, a common problem in the early years of colonial Africa. Except on the coast, missionaries only began their work in Kenya after the first settlers arrived and work on the railway, which was to cross the country from east to west, had started in earnest (Maupeu, 1998, p. 144). Under Maasai control, the country's heartland remained for many years an impassable barrier for 19th-century caravans, who therefore preferred to use more southern routes across what is now Tanzania. Only in the 1890s was a new commercial route established between Mombasa and Uganda, following closely the construction of the railway linking the coast to Uganda between 1897 and 1902. Unlike Uganda, where the kingdom of Buganda provided the colonial administration with a ready-made infrastucture, no Kenyan group stood out in terms of its ability to dominate, or even influence, the entire country.

The British colonial administration generally employed a system of indirect rule (which transformed into local government during the 1940s), seeking its base among precolonial political authorities. In Kenya, such authorities simply did not exist in most of the country (Bourmaud, 1988, p. 7–31). Colonial chiefs were created within the groups to be governed, but their efficiency was questioned because they lacked legitimacy.

2. Controlling the African Labor Force and Creating a Special Territorial Framework

After pacifying "rebel populations," such as the Nandi, the colonial government's first priority was to

[1] During this period occurred a smallpox epidemic in 1892 and cattle plague in 1890–1891 and 1898. Harvests were destroyed by locusts in 1894 and 1899 and the drought of 1900–1901. In addition, intense warfare between different Maasai groups and their neighbors was observed (Lonsdale, 1992, p. 17; Muchoki, 1998, p. 79–82).

revive agricultural production; however, the colonial administrators were so deeply dissatisfied with the attitude of the African peasants and the local leaders they had selected that they excluded the African population from export production. With policies encouraging foreign settlers, the country's second administrator, Sir Charles Elliot, succeeded in attracting land-hungry Finns and Boers, as well as those from the Punjab (Lonsdale, 1992, p. 34). In this way, an alternative population base emerged in the colony at the very moment that African population numbers were particularly weak; these settlers occupied interstitial areas traditionally separating the African populations, thus serving as buffer zones by filling the territorial gaps between the African ethnic groups (Lonsdale, 1992, p. 34–36). The new colony's economic basis rested on the export of agricultural products from its large plantations, with wheat dominating initially. But, like the country's general administration, these estates operated through the existing African male labor force, for both agricultural work and road building.

Thus emerged very rapidly a dual system, based on a physical and racial separation between micro- and macro-exploitations, with Africans restricted to subsistence agriculture while the foreign settlers took charge of cash crop cultivation and exports. Pastoral groups were the main victims of this system in which they had no legitimacy. In practice, the administration was built on a segregationist model: racial segregation (social and spatial) between Africans, Indians, and Europeans, as well as separation of the various African ethnic groups within the reserves. Already, European settlement in the margins of land between the African groups led in that direction but, beyond this, the permanent goal of all successive administrators was to govern territories with homogeneous populations. Internal boundaries were therefore drawn up. This ethno-territorial definition of populations went hand in hand with the identification of individuals with a single ethnic group and a single location. Consequently, place of birth grew to have an unparalleled importance for African populations (Peatrik, 1998, p. 66).

The Maasai's displacement southward from the northern section of the Maasai Reserve was the first element of population policy to have a direct impact on the African population in Kenya. The first waves of this transfer and relocation were organized between 1905 and 1911, although the relocation process of the Maasai sections, Uasin Gishu and Moitanik, extended over a period of more than 40 years. In reality, individuals from very diverse origins found themselves within these relocated populations, forced to migrate by the government or having chosen to do so for a variety of reasons. The officials responsible for organizing this population movement described the mix of origins in, for example, the 1929 Annual Report produced by the District Commissioner of Narok (DC/NRK/1/1/2, 1929): "They came from Timboroa and are in fact the sweepings of many tribes who were squatting on farms near Eldoret. A large number of them are cave-dwellers from Mt. Elgon. Some cannot even speak Maasai."

The transfers of population continued despite complaints from the Trans Mara administration. Moreover, the migrants' mixed origins, coupled with the fact that most had already worked on European farms, meant that the so-called Maasai cooperated better with the local administration than the others. They were willing to return to work on European farms, and most spoke ki-Swahili, which facilitated communication (DC/NRK/1/1/2, 1926).

At the start, land pressure in the African reserves was not enough in itself to incite men to look for paid work, and the colonists were hard put to find farm laborers. A taxation system was put in place, therefore, to refill the coffers (emptied by the construction of the railway line to Uganda) and encourage the departure of men toward the large white plantations in search of wages.

After World War I, in 1920, a system of permits, or identity cards—*kipande*—was instituted to control the employment of native workers on European farms. Only workers were allowed to live on these estates. In Nairobi, where segregation was strong, Africans were not allowed to stay more than 24 hours (Médard, in press). *Kipande* was also necessary to go from one African reserve to another, and forced relocations were frequent.

The colonial government increasingly used the *kipande* system with the express aim of restricting internal migration in the country which, according to the archives, appears to have been a major preoccupation. Necessary for employment outside African reserves and, in theory, indispensable for crossing any administrative boundary, *kipande* was compulsory for all men over the age of 16. In practice, however, the system encountered some problems: not all men were registered, and some were registered several times so as to conceal their true identity (Kuczynski, 1949, p. 139). What could theoretically have been a very precise source for population estimation suffered from additional problems, arising principally from the fact that the registered population was significantly higher than the adult male population possessing a *kipande*, because deaths and emigration were not systematically reported; in addition, the population was not registered at all in certain parts of the country (Kuczynski, 1949, p. 140–141).

In practice, although the *kipande* system did not always provide correct population estimates, it nonetheless permitted a real control of migration, in particular toward European farms. This control was also established at the boundaries of African reserves, and transfers from one reserve back to another were frequent.

To solve the problem of rapidly increasing population densities, mainly in the Kikuyu reserve, a system of squatting that had been practiced for many years was legalized: in exchange for 180 days of labor in a year, a man could settle with his family on a large estate and cultivate a small piece of land for his own subsistence (*kaffir farming*). This preserved the dual agricultural system (large plantations vs. small farming), the aim of the Kenya Land Commission (Ghai and McAuslan, 1970).

3. The Government's Ignorance of, and Indifference toward, the African Population

Paradoxically, the African population itself received very little attention during this period, as illustrated by the fact that the first censuses apply only to the country's non-African minority (four censuses were taken in Kenya between 1911 and 1931). The possibility of a complete population census had been envisaged as early as 1925 (Kuczynski, 1949, p. 130–132), but the idea was discarded several times for fear of achieving even less satisfactory results than those obtained by the local administration. In 1931, however, one section of the indigenous population was included: residents in the municipalities of Nairobi and Mombasa, plus some living on the European plantations, although they comprised only 2% or 3% of the total African population. The pessimistic view of the probable outcome of such an operation in the African reserves was linked to fears that the census would be badly received by the population as a whole and that its harmful effects would outweigh any benefits to the government. There were certainly many who deplored the absence of reliable data on births and deaths in the African reserves; chiefly, however, a census would have made it possible to improve taxation among the African population and gain insight into the impact of illegal migrations, such as the overflow from densely populated reserves into less populated ones (Kuczynski, 1949, p. 130; Kuczynski quotes McLellan Wilson on the Kikuyu settled in a Maasai reserve near Narok). The tax system, based on a hut count in the reserves, was hampered by the extensive demolition of homes that preceded the count. Yet even this prospect, completely in line with government actions, was not enough to launch an operation whose successful outcome appeared difficult to achieve.

Locally, however, population growth was raising questions for colonial administrators to deal with, quite outside the national policy sphere. Among the Meru, for instance, no territorial expansion was possible (bordered on all sides by enemy groups or by Mount Kenya's infertile lands). The very strict traditional social system acted as a powerful restraint on fertility, through late initiation and marriage, a prohibition on reproduction outside marriage, and a prolonged period of abstinence (2 years) after each birth. The frequency of abortions, the ensuing maternal mortality, and the low level of Meru fertility worried successive administrators; their efforts to extend the reproduction period by reducing the age at initiation, on the one hand, and to limit maternal deaths following abortions, on the other, brought about a marked social evolution (Beck, 1970, p. 143–144; Peatrik, 2002).

In contrast, among the sedentary populations in most other Kenyan regions (e.g., in Kisii), it was rapid population growth that alarmed administrators although they were concerned more with controlling the population's territorial expansion than with limiting its growth.

Although interest in population growth was restricted to the local context, the need for medical personnel and educated chiefs became increasingly obvious, and led to the creation of schools and public health services. In the early years, health and education remained the prerogative of missionaries, while the colonial government dealt exclusively with the immigrant population.

The colonial ideology reinforced social differences between ethnic groups (Eshiwani, 1990, p. 1–2). Public schools were gradually established, however, if only to provide an educated workforce for local administration. The situation within the health sector can be described in similar terms. At the start of the 20th century, western medicine was practiced only by Europeans for Europeans. In rural areas, medicine was left in the hands of the missionaries (Iliffe, 2002, p. 39). Ann Beck (1970, p. 145–146) showed that the colonial government was ready to join the struggle against the high level of African mortality, but only up to a certain point, for fear of the ensuing increase in population. It was World War I that alerted the colonial authorities to the poor standards of health in the African reserves. The government, therefore, decided to subsidize certain mission hospitals and then established a small number of its own health facilities (Iliffe, 2002, p. 39–40).

The example of the colonial authority's attitude toward the Maasai is striking. Assuming that venereal diseases, such as syphilis, were rife, colonial administrators feared the extinction of the Maasai; yet nothing was done to improve the situation, attributed to cultural norms strongly condemned by officials (Coast, 2000, p. 192). A single government intervention is worthy of note: in 1939, treatment was offered at a price, in the sole hope that the Maasai would sell cattle in order to obtain it, thereby stocking the Nairobi market (White, 1990, also quoted by Ernestina Coast, 2000). This can hardly be described as a public health policy, however.

At that time, the extension of health care to the African population was hampered by a lack of personnel. No medical training was available locally until 1929 (Iliffe, 2002, p. 44). In practice, a sick African seeking treatment had three options: to fall back on traditional medicine, to visit a mission hospital (which might not be possible because of the distances involved), or to find a small private clinic, generally run by former mission hospital employees (even a former sweeper, at times) (Peatrik, 1998, p. 72–73; Iliffe, 2002, p. 41–42). At that time, most Europeans preferred to employ uneducated Africans, fearing confidence and arrogance among those who were the first to be educated.

This half-century of colonial rule was characterized by significant economic development in the country, through land alienation, taxation, and paid agricultural labor. Largely in the hands of the colonizers, this economy allowed Kenya to enter the international market. During this period, however, the colonial government had failed to take account of the demographic factor, whose power became apparent from the 1950s.

II. TRANSITION AND INDEPENDENCE

On the eve of independence the situation began to evolve. On one hand, the state's control of population movements weakened and, on the other, the African population started to gain the upper hand, politically and socially. The first family planning clinics emerged, African education developed, and the infrastructure and policies for a public health system were set up.

1. The Decline of Colonial State Control

Little by little, the population's spatial distribution evaded state control. Arising out of the growing pressure on land in Kikuyu country, the Mau Mau movement (1952–1954) overturned the colonial land order and paved the way to independence. Until the 1930s, the settlement of squatters and their families on large plantations matched the European colonizers' high demand for labor. As mechanization progressed, however, it was no longer possible to justify the presence of so many dependents, particularly as the squatters' strong demographic regime tended to reinforce their numbers on these estates (Buijtenhuijs, 1998). The colonial government had no hesitation in forcefully repatriating unwanted squatters into the Kikuyu reserve, a place completely unknown to some (Médard, 1999, p. 353). Migration to cities, especially the capital, increased and the population of shantytowns, which had already existed illegally for several decades (Rodriguez-Torrès, 1995), grew sharply from the 1940s onwards. A massive urbanization emerged. As the government no longer had the authority to tackle these issues directly, indirect attempts were made to limit them. Fear of a new rebellion redirected the government's policy interests toward the African reserves. The need to make reserve lands support growing numbers of people led to a policy of agricultural development involving the individualization of land tenure, sanctioned by the registration of small farms (called the Swynnerton Plan, 1954) (Médard, 1999, p. 310–311). Cash crops were gradually introduced, 30 years later than on European farms, but always on a small scale so as to limit competition between African reserves and European states. Restrictions on the number of trees or the size of holdings depended on the place and the era (see, e.g., Barnes, 1976, p. 124). In the political arena, the period preceding independence was one of debate on the national representativeness of a president. Above and beyond the choice of an individual, questions arose about how the country should be administered. Access to resources was at stake, particularly to former European holdings that were in the process of being sold off to Africans. Most of these lands were situated in the Rift Valley province, and Kalenjin groups were making claims to them in the name of reinvented historical rights (Médard, 1999, p. 353). Ancient Kikuyu rights to these lands were equally well established, however; Kikuyu settlement in the area had been consolidated by the kaffir farming system in place during a large part of the colonial era.

The first Kenyan political parties were formed along ethnic lines, founded essentially on the cleavage between minority and majority groups. KANU, the alliance party between Kikuyu and Luo, the two largest ethnic groups in Kenya, stood in opposition to KADU, a coalition of parties representing the smaller ethnic groups in the population, principally Kalenjin, Luhya, Maasai, and Mijikenda populations. KADU advocated a federal constitution granting the administration of land to the regions (the *Majimbo* constitution, which means "regions" in Kiswahili) (Médard, 1999, p. 88–89), and the massive decentralization it wished for received the backing of the Europeans, Indians, and the British government, who feared that the birth of a centralized independent state might be contrary to their interests (Bourmaud, 1988, p. 81–82).

From 1959 onward, the population of European origin became gradually less involved in the country (Leys, 1975, p. 55). The policy of redistribution of their lands was long debated: should large plantations be preserved at the expense of small ones? Should the estates be split up or, on the contrary, should the dual agrarian economy put in place at the start of the colonial period be preserved? (This agrarian debate has been much argued by agricultural development specialists, economists, and political scientists; see, e.g., Orvis, 1993.) The government attempted to ensure the continuity of the two-tier agricultural system, hoping that Africans would take over the large European farms. But European departures from the country were hasty at times, necessitating a speedy reorganization of their farms. Faced with a lack of potential African buyers, some states, especially those close to a reserve, were divided up, constituting the first land settlement schemes from 1961 onward. This operation, which later proved to be a financial failure, allowed a small fraction of squatters to legalize their situation and some of the better-off reserve inhabitants to flee the pressure on land by settling on European properties. The underlying idea was to meet the demands of landless peasants while at the same time providing land to curb the process of urbanization.

In May 1963, the KANU party won the elections and, from 1964 onward, the centralization of Jomo Kenyatta's regime put an end to the devolution of any significant powers to the regions. Little by little, KADU party members associated themselves with the party in power.

Transfer of former European farms continued up to the 1970s. Under Kenyatta, "land companies" were created. These associations acquired land in their own name and then redistributed it among their members. The majority were failures, however. In practice, social networks determined how land was distributed, and people were not equal in this regard. Thus, some powerful men benefited more than the others, and this system facilitated Kikuyu migration within a strictly legal framework (Médard, 1999, p. 355).

Some regions retained special status at independence, however, with the aim of protecting their population's access to resources: one example is Trans Mara district, which remained a "closed district" until 1967 (Matampash, 1993), open only to individuals possessing a *kipande* until 1972, and where strangers to the region were forbidden to buy land.

Outside these protected districts, the transition period and first decades of independence are characterized by successive waves of settlement on land alienated for European settlers at the start of British colonization, almost all of them situated in Rift Valley province.

2. Early Public Policies for the African Family

From the point of view of African families, westernization was gaining momentum. In 1949–1950, 26% of Kenyan children of primary school age were educated (compared with 6% in French-speaking West Africa) (Iliffe, 1995, p. 222). The urban employment market was developing, fed by a migration that was henceforth tolerated.

Population flows toward the city exacerbated urban economic and social problems. Extremely high population densities in the Kikuyu reserve and the pressing need for land in the rest of the country presaged an even stronger growth in internal migration. Rather than focusing exclusively on resources, as it had earlier on, the government also turned its attention to the country's African populations. Knowing these populations better seemed to be the key to improved planning.

At the end of World War II, however, nobody could provide an estimate of the African reserves population or even sketch demographic characteristics. Only in 1945 did the government start preparations for the first general census (Kuczinski, 1949), finally carried out in Kenya and its two neighboring countries, Tanzania and Uganda, in 1948. Fears about difficulties with data collection were justified, and these first national censuses are of low quality. The population count, especially of young children, is underestimated (Thomas, 1998). These problems are relative, however, and, compared with earlier estimates based on hut counts, the 1948 census results are appreciable. Moreover, this first attempt served as a basis for subsequent censuses, which could only improve on it. The whole operation was therefore hailed as a great success (Martin, 1949).

Probably because the population had been significantly underestimated until then, the 1948 census figures were alarming. In 1952, the first family planning associations were created in Nairobi and Mombasa. These associations provided the public with the whole gamut of services, from information to the distribution of contraceptives (Uche, 1974, p. 9). They amalgamated in 1961 to form the Family Planning Association of Kenya, the first such association in tropical Africa to be affiliated with the International Planned Parenthood Federation in 1962.

Despite creating a national program much early on, Kenya reached a natural growth rate of over 4% by the end of the 1970s, with a total fertility rate of over 8 children per woman.

In 1966, Kenya became one of the first African countries to profess an official view on this subject (Uche, 1974; Chesnais, 1979, p. 207; Greeley, 1988, p. 212–213; Locoh and Makdessi, 1996, p. 12). As early as 1967, a national program was created, in association with the already existing program for maternal and infant health improvement. Its main aim was to establish a network of clinics and health centers to deal with maternity and birth control (Uche, 1974). The program achieved its objective, but had no immediate effect on fertility; quite the opposite, in fact, as fertility levels rose between 1969 and 1979. It was not until the 1980s, 15 years later, that government family planning policies began to have an impact. The government's emphasis on family planning services and its lack of attention to the population's indifference to these services explain the program's lack of success. Nothing had been done to promote a reduction in the demand for children (Frank and McNicoll, 1987, p. 222). At that time, the program was entirely responsible to the Ministry of Health, whose focus was maternal and infant health (Kelley and Nobbe, 1990, p. 41).

In parallel to government policies, the more targeted efforts of several nongovernmental agencies (NGOs) (Family Planning Association of Kenya, Family Planning Private Sector, and the National Council of Churches of Kenya, in particular) should be mentioned. Until the 1980s, access to this type of service depended entirely on these organizations.

This period of political transition was also a period of transition in population policy. More and more, the Kenyan population was becoming the object of analysis and measurement, though these studies remained disconnected from the population's real needs. Presented at the macro-demographic level, the family planning program appeared to serve the interests of the State more than those of families. Population growth reduction was still viewed from the macroeconomic perspective. Economically speaking, the transition to independence was one of continuity, with the unequal redistribution of alienated lands perpetuating the dual colonial system.

III. A DIFFICULT END TO THE CENTURY

The year 1984 marked a turning point in Kenya's history. Several events made it an exceptional year. A failed coup confirmed the authority of Daniel arap Moi, in power since 1979, and his ideas increasingly shaped national policy. It was also the year of the great East African drought, which had direct repercussions on the Kenyan economy. In addition, the first cases of AIDS and a new strand of malaria, resistant to quinine, appeared toward the end of the 1980s and grew more and more devastating each year. Finally, with the country's record population growth sounding the alarm bells at the international level, a more assertive approach was taken to policies for fertility control: a national survey was conducted that year, and the government took a stronger line on this issue and invested more in it. For instance, the population action program was partly decentralized, with the creation of the District Population Program.

During the 1980s and 1990s, the Kenyan population's economic and demographic situation evolved rapidly: mortality increased, sending life expectancy at birth back to 45 years; fertility plunged by around 50% in the country as a whole; primary education became quasi-universal; interethnic tensions, already perceptible at independence, came to the forefront of the national political scene; forced migrations and refugee waves became recurrent.

1. Public Health Unable to Keep up with Demand

At the end of the 1970s, Kenya possessed an extraordinary public health system for an African country. The system had not managed to keep pace with population growth, however. Even though reproductive health care generally continued to improve through projects developed specifically in this domain by both the government and NGOs, the rapid emergence of new diseases—AIDS evidently, but also malaria, resistant to chloroquine by the end of the 1980s—seriously undermined the health system. The vaccination rate, which had reached 80% of children in 1993, for example, had dropped to 65% by 1999.

From Policy to Practice: The Case of the AIDS Epidemic in Kenya

AIDS did not receive the attention it deserved in Kenya for a long period, which explains why the illness spread so widely during the late 1980s and early 1990s. At this time, AIDS was viewed as an "unclean" disease, affecting unclean people who used unclean needles or indulged in unclean sexual practices. The conservative nature of Kenyan society, coupled with the fact that men controlled the sexual sphere, tended to stifle the question of AIDS. Screening was uncommon; often unavailable, it was in any case too expensive for most of the population and, as a result, HIV carriers remained unaware of it. Publicity about birth control methods was nonexistent, and the young, potentially the most receptive group to safer sex practices, were totally ignored by policymakers. The rate of condom use even decreased between 1984 and 1989 (Kelley and Nobbe, 1990, p. 48). During the same period, Kenya's economic recession contributed to the development of the prostitution market. In the 1990s, AIDS was above all a concern for the health authorities. Yet, according to John Iliffe, a study conducted in the mid-1990s shows that 30% of AIDS victims died without ever having benefited from the health system (Iliffe, 2002, p. 235). Gradually, however, a large part of the population became actively involved in the struggle against the pandemic. Politicians, religious leaders, and the media started to condemn various cultural practices that could contribute to the spread of AIDS (such as the sexual purification of widows, circumcision using the same knife, etc.) and called for them to be abandoned to reduce the risk of transmission. At this time, despite a high level of awareness of the disease, most policymakers within the government and other sectors chose not to define policies to help control the spread of the disease and moderate its effect on society and the economy. Only at the end of the 1990s did the government realize that the AIDS epidemic could be detrimental to development projects and, from then on, every possible effort was made to fight the disease.

AIDS continues to be a powerful threat in people's lives. It is no longer perceived today uniquely as a question of public health, but also as a problem for development; a whole chapter in the current national development plan is devoted to HIV and the impact of AIDS on social well-being and economic growth. In a country where AIDS was scarcely mentioned in public discourse only a few years ago, it is now commonplace to hear politicians refer to the disease's socioeconomic consequences, citing examples from the regions they represent. The Kenyan government, aided by NGOs, missions, and private institutions, continues to extend its activities in the realm of prevention and to improve treatment of sexually transmitted diseases and health care. The AIDS epidemic's negative effect on Kenya's economic and social development has provided an added incentive for the government and its partners to become actively involved in promoting awareness of the problem, and increased the need for strong, well-coordinated policies. Such measures were essential to prevent additional disastrous socioeconomic consequences of the epidemic, and to limit those that had already started. Toward 1997, strategic and collaborative policy development, coupled with concerted efforts made to reinforce NGO influence over national policy in the HIV/AIDS battle, started to show tangible results in Kenya. The government, with aid from a broad spectrum of disciplines—technical, legal, and ethical—prepared a national policy for the fight against AIDS, which led to significant improvements in Kenya, better mechanisms for reinforcing HIV prevention and care, as well as specific changes in official policies.

Moreover, in response to its increasingly profound effect on society, the Kenyan government integrated priorities and strategies for the fight against the epidemic and its catastrophic consequences. Official strategies were of two types: (1) to prevent to spread of HIV and (2) to reduce the disease's impact on society. Working through the existing network, improvements in HIV prevention and treatment gradually started to appear. Recently (October, 2002) the *Kenya National HIV/AIDS Strategic Plan*, developed by the *National AIDS and STD Control and Prevention Program* (NASCOP), was elaborated, covering different aspects of AIDS. Among other things, it provides national guidelines for home care services and patient counseling. Private companies and NGOs adopted policies, or made plans to do so, aiming to curb the spread of the disease among their workers, and certain religious groups made a public commitment to provide support for their adherents with regard to prevention. Some gaps and ambiguities of the Kenyan law were also a focus, such as women's inheritance rights, HIV screening, confidentiality, rape and sexual abuse, and so on. Legislation was debated in parliament to arrive at new policies for HIV/AIDS prevention, disease tracking, treatment, and counseling.

The newly elected government has intensified its efforts to reduce the spread of the AIDS epidemic. The latest actions have been the institution of the Cabinet Committee on HIV/AIDS, with the President of the Republic as president, and the creation of the AIDS Initiatives by the President's wife, an association offering assistance to AIDS orphans and disadvantaged young people. Advocacy networks, NGO coalitions, and religious groups have been created or have also intensified their efforts: promoting awareness, integrating HIV prevention, establishing Voluntary Counseling and Testing Centers (VCTs) throughout the country, taking charge of counseling and sexually transmitted disease management in both the public and private sector. Private sector employers are also developing on-site programs that include discussions with their department directors on the policies and projects to adopt in the fight against AIDS, and peer-group counseling. In 2001, the statistics indicated that the national HIV prevalence rate had fallen to 10.2%, and by 2003, the prevalence had fallen to 7%.

2. Growing Emphasis on Population Policy

Pressure from financial backers prompted the Kenyan government to modify its population policy. In so doing, the government, other policy decision makers and planners proved that they were well aware of the importance of population policies, designed to alleviate the weight of demographic growth, for the country's economic and social development. The creation of this new population policy went hand in hand with a considerable drop in fertility. On the national scene, nonetheless, population growth remains a sensitive and controversial topic.

a. The Decline in Fertility

Although the government had adopted a national program of family planning as early as 1967, its *Sessional Paper No. 4*, a guide spelling out the implementation of national population policy didn't appear until 1984, 17 years later. Meetings were then organized between policy decision makers, local consultants, and university experts to ensure as comprehensive an approach as possible during the preparation of policy projects by sector. Debates taking place in population conferences (e.g., the First African Population Conference in Accra, Ghana, in 1971 and the World Population Plan of Action in Bucharest in 1974) had a strong influence on this policy. The strategy document and the national population policy, henceforth multisectorial, were presented to the public during a national conference and revised in the light of criticisms before being definitively adopted as law. It should be noted that, at that time, Kenya was still relatively advanced compared with other African countries; it was not until the 1984 Arusha conference that African political representatives as a whole accepted the validity of family planning for the first time.

In publishing this sessional paper, the government finally recognized that the country's population constituted its most precious resource, acknowledging that the population factor—centered on the assumption that fertility decline was largely dependent on social and economic development and that family planning programs could not succeed unless a certain development threshold was reached—could not be isolated from its socioeconomic context and that purely demographic interventions should be avoided; it also suggested a much improved population program.

Implementing these policies brought substantial results, especially with regard to campaigns encouraging Kenyans to adopt a small family size as the norm. Quite remarkably, the population's annual growth rate dropped from 3.8% in 1979, to 3.3% in 1989, and to 2.8% in 1999. The decline in fertility was as late as it was radical, perhaps precisely because fertility levels started from so high. Estimated to have reached 8 children per woman at the end of the 70s, the total fertility rate fell below 4.7 children per woman according to the Kenya Demographic and Health Survey (KDHS) of 1998, a drop as unforeseen by international organizations as had been its rise during the 1960s and 1970s.

This decline in fertility can be associated with the increase in contraceptive use throughout the country. The Contraceptive Prevalence Survey (1984) and Demographic and Health Surveys (1989, 1993, and 1998) carried out in Kenya showed an increasing use of birth control. The proportion of women aware of at least of one method of birth control increased from 81% to 97% between 1984 and 1998. The rate of contraceptive use among women in a couple, all methods combined, rose from 7% in 1978 to nearly 40% in 1998. Also among women in a couple, ideal family size has oscillated around four children since 1989 (Table 115–1).

Numerous studies have examined the factors responsible for the Kenyan fertility decline (Van de Walle and Foster, 1990; Kizito *et al*., 1991; Robinson, 1992; Caldwell *et al*., 1992; Brass and Jolly, 1993; Egerö and Mburugu, 1994). The debate continues, however: what prompted this speedy decline—the politico-economic-health crisis or public policies? At the macro level, Patrice Vimard (1996) makes the association between the Kenyan fertility decline and the transition from a period of economic growth to one of stagnation. Progress during the period of growth (in terms of education and access to information and birth control) makes it possible for value systems to evolve during a period of economic stress. In Kenya, the decline in fertility can be linked in particular to the increasing scarcity of land—the primary resource in a predominantly rural population (Caldwell *et al*., 1992).

TABLE 115–1 Fertility Decline in Kenya

	KFS 1979	KCPS 1984	KDHS-I 1989	KDHS-II 1993	KDHS-III 1998	KDHS-IV 2003
Estimated population	15.3 m	—	21.4 m	25 m	28 m	32 m
TFR	7.9	—	6.7	5.4	4.7	5.0
Knowledge of contraceptive method	—	81%	—	—	97%	97%
Prevalence of contraceptive use	7%	17%	27%	33%	39%	39%

TFR = total fertility rate.

Despite the government's efforts to set up a suitable political environment, rapid population growth in the 1970s and 1980s hindered economic growth and led to an unprecedented level of population pressure on agricultural land that reduced the advantage of large family size. Higher rates of female literacy engendered new attitudes, especially with regard to family size. Increasingly, parents sought a better education for their children, despite rising school fees. These trends stimulated a demand for family planning. The government became highly involved in this issue and committed itself to providing accessible and diverse methods of birth control, a crucial factor in the success of any population program. As a consequence, use of these services increased and fertility rates dropped substantially. More than three-quarters of the population have access to health services, and more than three quarters of men and women aged 15 and over can read.

Factors that have had a positive influence on the policies of the last decade include the political involvement of the government, decision–makers, and higher-level partners; the increasing availability of well-trained personnel; and the inclusion of population issues in programs relating to development, the environment, the economy, and society. The whole structure being developed also encouraged these policies: legal reforms setting the possibilities and limits of success of national policy, national seminars on population and development, recommendations based on population and development research conducted by the international community, and the creation of pertinent guides and manuals for efficient policies.

Despite the successes achieved so far, however, there are areas in which progress is still possible. For instance, family planning needs in Kenya are not completely satisfied and birthrates will certainly go on decreasing as those needs are met. According to the *Economic Survey* (MPND, 2003), lack of resources is a key constraint on the national action program, with the government having nowhere near the funds necessary to implement it. The root of the problem lies in the withdrawal, by international monetary agencies, of part of their funding in the 1990s because Kenya failed to fulfill the necessary conditions—zero tolerance toward corruption, in particular.

Some questions remain particularly tricky, such as young people's access to information and services concerning sexual and reproductive health, still a controversial issue. Religious and secular groups have failed to reach a consensus on school syllabus content and on condom use, with some believing in the importance of talking to sexually active adolescents about condoms and sexual fidelity, and others insisting that only the value of abstinence should be taught. The Catholic Church, opposed to all initiatives that promote condom use, has a strong influence on Kenyan media and public opinion in general (see Ajayi and Kekovole, 1998). Thus, Kenyan religious authorities even exaggerate positions taken by the Vatican; certain eminent members of the religious hierarchy, from simple priests to the archbishop of Nairobi, regularly express their views in public against condom use, going so far as to deny their efficiency, and even invert the relationship between the diffusion of condoms and HIV. Campaigns organized by the government and NGOs to educate the general public are followed by campaigns of disinformation organized by religious leaders. According to Steve Bradshaw, reporter for the BBC,

> In Kenya—where an estimated 20% of people have the HIV virus—the church condemns condoms for promoting promiscuity and repeats the claim about permeability. The archbishop of Nairobi, Raphael Ndingi Nzeki, said: "AIDS [. . .] has grown so fast because of the availability of condoms." [. . .] In Lwak, near Lake Victoria, the director of an AIDS testing center says he cannot distribute condoms because of church opposition. Gordon Wambi told the program: "Some priests have even been saying that condoms are laced with HIV/AIDS."

The criminal polemics initiated by some Catholic groups have spread to other religious groups since the mid-1990s. Muslims and Adventists have rallied to the fight against condom use, so much so that a vast number of Kenyans share these views, as numerous references in the national press testify, and this despite the high incidence of HIV in the country. As a result, the whole debate on sexual education in schools is far from being resolved.

Fortunately for the national action program and for the Kenyan population in general, the Church does not control the funds used to fight AIDS or teen pregnancies. Most Kenyan population programs are financed by foreign funding agencies, which made it possible to maintain existing activities and create new ones in the area of reproductive health and in the fight against HIV/AIDS throughout the 1990s, despite pressure from the religious lobby.

Donors have the right to decide how their money will be spent, however, and as a result, many programs are centered on objectives chosen by the funding agencies, regardless of the population's real needs and cultural norms. For instance, the American government has offered vast sums of money for family planning programs, on condition that they promote abstinence rather than condom use in the fight against HIV/AIDS and teen pregnancy, joining ranks with the most extremist religious movements. Rechanneling the funds in this way meant that the supply of contraceptives (condoms included) to many NGOs was halted. Marie Stopes International, for example, closed its reproductive health clinic in Nyanza province, the province with the highest incidence of HIV in Kenya, and many other health services were reduced in scale. Such demands and restrictions from funding agencies can have an influence on national population policy, and many countries could be forced to set up inefficient programs that may even put the lives of women, children, and whole families in jeopardy. According to Amy Coen (2003), president of Population Action International,

> The Global Gag Rule is yet another example of how the Bush Administration is allowing political ideology to trump science. The policy shows no respect for scientific evidence and proven public health practices, and no compassion for the millions of women around the world engaged in a daily struggle for existence. Under this policy, foreign NGOs are forced to choose between desperately needed family planning funds and the ability to provide medically ethical information or participate in public debates over their countries' abortion policies.

b. The Political Face of Demographic Growth

The issue of population policy needs to be viewed within the context of national construction and budget distribution. In Africa, numbers are often perceived both as a means to power and as its justification. In Nigeria, budgets are allocated according to the size of the population administered in a given region, and this led to an escalating problem of population overestimation until the 1981 census (Omoluabi and Lévy, 1992).

The dichotomy between minority and majority ethnic groups is Kenya's main source of antagonism. During the period of consolidation in power of Daniel arap Moi, independent Kenya's second president (1984–2002), this rivalry heightened. Although the regional *majimbo* ideology, spearheaded by Moi in the early 1960s while a member of KADU, advocates the rights of minority groups with respect to majority ones, the predominance of the latter nonetheless remained an obstacle to his authority. So Moi played a double game: on one hand, he defined himself as a minority group member while, on the other, he sought to compete with majority groups on their own ground, namely numbers. He made use of a variety of methods to ensure the numerical strength of the Kamatusa groups (Kalenjin, Maasai, Turkana, and Samburu), his main electoral basis.

First, he juggled ethnic categories to strengthen his position. From 1979 onward, the category "kalenjin" was created and used officially in population censuses to merge seven ethnic subgroups with related languages: the Kipsigis, Nandi, Keiyo, Marakwet, Tugen, Sabaot, and Pokot groups, of which five were already the product of a regrouping of distinct communities. This made it possible to raise the rank, numerically speaking, of one of the Kamatusa components to fifth place among Kenyan ethnic groups. Finally, the very emergence and use of the term *Kamatusa* made it possible to swell numbers even more.

Second, the 1989 census was manipulated in such a way as to raise the Kalenjin group from fifth to fourth rank in Kenya's population composition. Not only was data quality not up to the mark, but a new step had also been taken: no longer a simple matter of juggling categories, it had become a straightforward case of data manipulation, poorly camouflaged in the published document, whose aim was to consolidate the Kalenjin position in the country as a whole and in the Rift Valley province in particular. Recalculating census figures shows that the Kalenjin's rise in demographic rank at the national scale is patently due to the addition of 20,000 Kalenjin in the Rift Valley province (Golaz, 1997, p. 115, 120).

Third, although the official position fed to funding agencies was one of fertility reduction throughout the country, some Kamatusa groups were, on the contrary, encouraged to maintain a high birthrate. In 1993 the Minister of Public Works, Jonathan Ng'eno, advised the Maasai and other minority ethnic groups to give up family planning "since the last general elections showed that we cannot command if we don't have the numbers" (DN, 01/08/93). The following year, the

Maasai political leader, Ntimama, asked the Maasai to reject birth control (DN, 05/12/94). When influential politicians offer a public invitation to certain ethnic groups to put no limits on their fertility, such advice can undoubtedly sway them; it can also trigger off a chain reaction among other ethnic groups who were not involved, regardless of current public policies. The same speech delivered to the Maasai was later repeated to the Kuria, a Bantu-speaking ethnic group, by their deputy, Shadrack Manga (DN, 23/06/97).

Even though practice does not necessarily correspond with official discourse, this type of attitude betrays a two-tier policy that can only be understood within a framework much wider than that of the simple diffusion of contraception.

3. Attempts at Spatial Redistribution of the Population

Even after fertility has started to decline, however, population structures evolve slowly and demographic imbalance aggravates economic deterioration, as a report from the International Labor Office had foreseen long ago (ILO, 1972). From the late 1980s, overcrowding and the unequal distribution of land within the population were used to justify disputes rooted in much more complex issues. Thus, although not part of the government's official policy, people close to the President, eminent members of the government, ministers, deputies, and so on, were involved in the 1990s conflicts that caused, in addition to thousands of deaths, waves of internal refugees unprecedented in Kenya.

Moi's doctrine, "*nyayo*," which means "in the footsteps" in ki-Swahili, implied political continuity between Kenyatta's government and his own; the reality, however, corresponds to a highly distorted sense of the expression: any continuity with Kenyatta's policy exists only insofar as it serves Moi's community and the Kalenjin group in general. Regionalism resurfaced with his arrival in power, as did Kamatusa leaders' land claims to the whole Rift Valley province. In the wake of the privatization of former European lands, the Rift Valley has become the most mixed province in Kenya; to adopt a *majimbo* system would mean expelling millions of people from the Rift Valley (HRW, 1993, p. 12–13). The "majimboism" of the 1990s can be viewed as a warped version of the regionalism that marked the beginning of the 1960s.

At the end of 1991, under pressure from the population and international bodies, Moi legalized the multiparty system and thus consolidated his position at the head of the Kenyan state (Grignon, 1993, p. 20). In December 1992, thanks to the protectionist power endowed on him by the State (i.e., purchasing of votes; in the Mathare Valley shantytowns, many people describe how the authorities tried to bribe them in the months preceding the elections) (Rodriguez-Torres, 1995, p. 66), but mainly because of divisions in the opposition, he was re-elected head of state. From 1991 to 1994, however, as economic growth hit its lowest level since independence after yet another drought, politicians close to power fueled conflicts between communities of different ethnic origins in and around the Rift Valley. Generally, these were Kalenjin and Maasai attacks on the Luhya, Luo, Kikuyu, and other groups settled in the region, in the hope of making them flee, or retaliation by these groups (National Assembly, 1992; HRW, 1993; Médard, 1996, 1999). This attempted ethnic cleansing in the Rift Valley created more than 300,000 refugees (NCCK, 1994, p. 9), not including those who were able to return directly to their family, and left at least 1,500 dead (Médard, 1999, p. 193–194). In 1997, as the following elections approached, hostilities flared up again in two stages. First, on the coast, the town of Likoni was hit by a wave of violence toward immigrants from the west and center of the country, forcing hundreds of thousands, mostly Luo and Luhya, to flee (Médard, 1999, p. 242). This scandal, in one of the most popular tourist resorts in Kenya, hit the country's main source of foreign currency hard. In the interior, on the other hand, pre-electoral tensions and scheming by local politicians rekindled old, unforgotten disputes, giving rise to a wave of violence comparable to that of the first half of the decade. Division in the opposition, partly the result of these conflicts, were such that Moi won the ballot once more.

In addition to Daniel arap Moi's two successive re-elections in 1992 and 1997, the displacement of Kikuyu, Luhya, Luo, and Gusii populations from lands that had been settled by some for many decades, and even many generations, meant that the government could use access to this land to buy the support of some Kamatusa communities from the Rift Valley province (Médard, 1998).

These refugee migrations within Kenya went unnoticed at the international level, however, partly because of the lack of publicity surrounding them. When looking globally at refugee flows, Kenya undeniably appears as a host country rather than as a source of refugees. Lacking any formal infrastructure, most refugees from the 1990s conflicts, registered by the National Council of Churches of Kenya, disappeared somewhere between the cities, their region of residence, or the region from which their family originated. This is not the case, however, for refugees from the Horn of Africa who, since the early 1990s, have found political asylum in Kenya where vast refugee camps were set up under the aegis of the UNHCR (United Nations High Commission for Refugees). Since 1992, more than 100,000 international refugees have been hosted in camps in the Dadaab region (Garissa district) in northwestern Kenya (Cambrezy, 2001). Another refugee camp, Kakuma, in Turkana district, also in northwestern Kenya, mostly shelters Sudanese refugees. According to the UNHCR's last calculations, 128,700 Somalis are currently living in Dadaab, and 67,300 Sudanese in Kakuma (UNHCR, 2003). Many other camps also operated in the 1990s, but are now closed. It is impossible to estimate the number of refugees living outside these organized camps, but it certainly exceeds the 10,000 or so people declared by the UNHCR for the city of Nairobi, for instance.

What are the consequences of these many years of economic degradation and political conflict on urban growth? Urbanization in Kenya never reached the alarmist levels predicted by urban planners in the 1970s. It is not possible to calculate the country's rate of urbanization from the most recent census of 1999. In addition to the classical problems of population underestimation in the poorest urban areas, the creation of new, bizarrely defined municipalities[2] in the 1990s prevents from any relevant interpretation of aggregate urban figures. The fact that the rural-urban dichotomy is undermined by the extremely high population density of some primarily agricultural parts of the country (800 inhabitants per sq. km. in Vihiga district, 700 in Gucha district) (Figure 115–2), makes it all the more difficult to reach any conclusion on the issue.

Apart from the asylum given to international refugees, mostly orchestrated by the UNHCR, Moi's regime ended in bloody internal conflicts. Globally, the whole settlement of Rift Valley province was questioned and, locally, hundreds of thousands of people were chased out. But no specific policy ever addressed the question of the population's spatial redistribution in Kenya, and the time may have come to place access to land among the priorities of Kenyan politics.

The procedure for formulating explicit population policy is now familiar territory to the Kenyan players. Still too often, however, creating efficient programs to meet policy objectives remains problematic. Integrating population policy and action plans into government priorities is not helped by the urgency of the country's politico-economic situation. Moreover, each ministry is responsible for its own sphere of policy implementation and at times it proves extremely difficult to maintain any coherence between them; this is despite both the multi-sectorial approach, whose aim is to harmonies the different facets of national policy, often closely linked to one another, and the efforts of National Council Program for Development (NCPD), mandated to coordinate them, on the other.

The NCPD's Unit of Population Planning, for example, is attached to the Ministry of Planning and National Development. It is the technical body responsible for national population policy, in terms of both the global coordination of policy formulation and its application. The Ministry of Health, on the other hand, is responsible for coordinating technical issues linked to reproductive health (i.e., it coordinates projects, evaluates results, and monitors spending at all levels of the population program). Another case, more serious still, the AIDS epidemic falls under the control of the Office of the President, though many believe the Health Ministry should be responsible; this creates confusion and conflict, over funding in particular, worsened still by the country's growing interest and financial investment in the struggle against the epidemic.

In the face of such massive challenges, the government needs to implement clear and efficient projects if the current policy guidelines are to prove their worth. Kenya's economic development perspectives are largely dependent on the government's success in its attempt to speed up efforts to expand family planning services and to fight AIDS and maternal and infant mortality. In a decentralized system like Kenya's, achieving this would require integration and cooperation between actors in the different ministries, donors, NGOs, the electorate, researchers, and politicians, all of whom influence government priorities.

Further decentralization of the policy formulation and implementation process could be envisaged at both national and district levels, giving more power to population management teams in each district, whose role it is to develop programs of action on the ground. This would make policies and the projects themselves better adapted to local needs. Even so, it should be mentioned that responsibilities were, at times, transferred before local authorities were sufficiently competent to take charge, which spelled failure for the programs. There is also, now and permanently, a need for relevant, accurate data on which to base new population policies, for qualified personnel, and for

[2] For example, Kehancha municipality includes the whole district of Kuria (Nyanza province), which is more than 90% rural. Our thanks to Claire Médard (IRD) for drawing our attention to this.

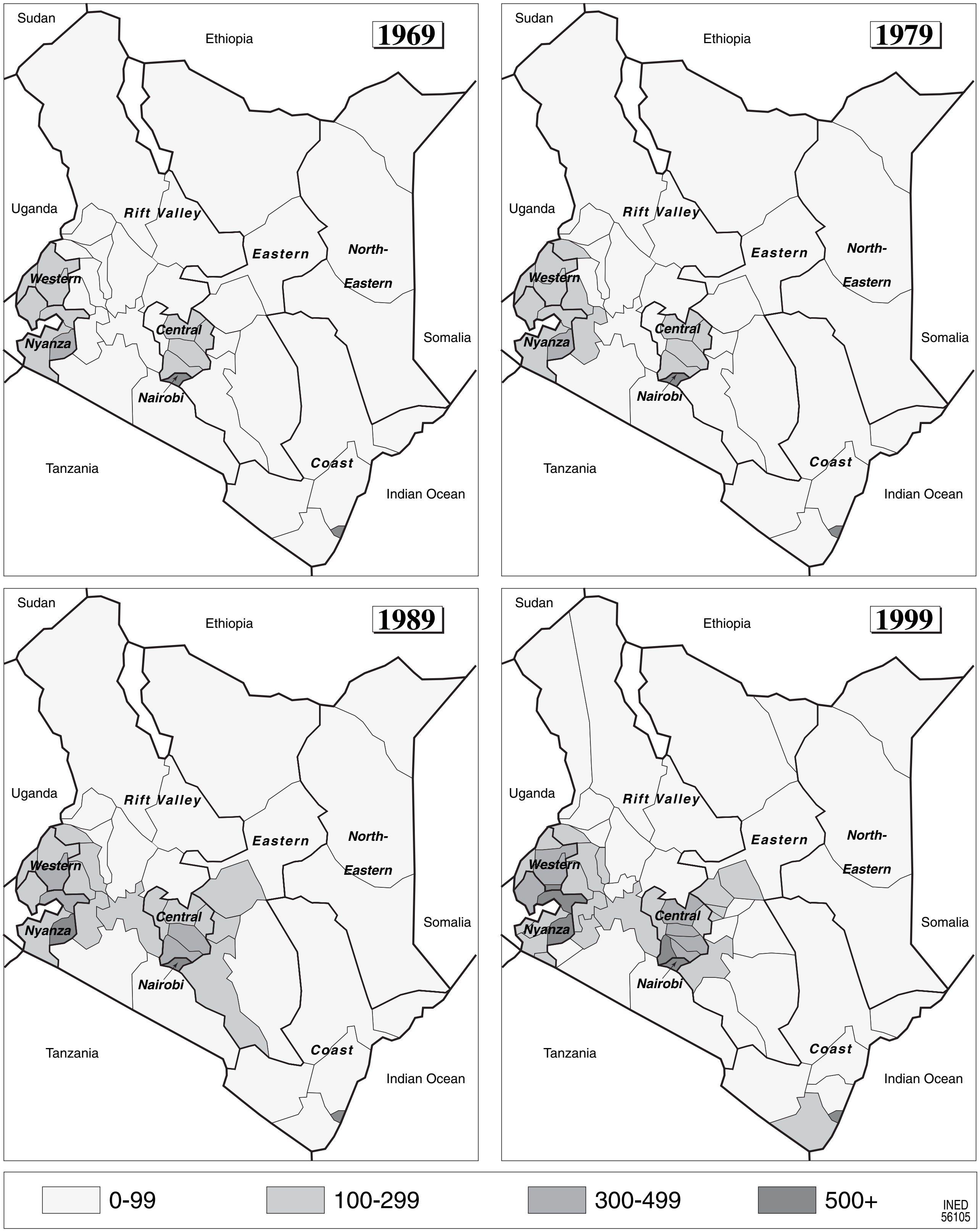


FIGURE 115–2 The evolution of population density in Kenya. (Sources: Republic of Kenya, *Population Census 1969* [1970]; *Population Census 1979* [1980]; *Population Census 1989* [1994]; *Population Census 1999* [2001].)

additional studies that analyze both the interaction between population and development and the real needs of the population.

CONCLUSION

The National Rainbow Coalition (NARC), the new party elected to power in December 2002, is certainly attempting to meet these challenges, offering free primary schooling to all Kenyans, for example, and taking up the torch in the fight against corruption, a measure that enabled Kenya to benefit once more from International Monetary Fund (IMF) funding in 2003. By reformulating its public health programs of action, by improving the health system particularly with respect to reproductive health and by investing in information campaigns, the new government hopes to improve awareness of family planning, AIDS and other reproductive health problems. Dynamic campaigns for HIV prevention and for treatment of sexually transmitted diseases and AIDS are emerging. By offering free education, making preventive methods such as Information, Education and Communication (IEC) and condoms available at low cost (subsidized), and by giving attention to the problems of women and other vulnerable groups, the NARC government is tackling the most fundamental social problems.

The Kenyan population's demographic growth remains high. Even though fertility levels have dropped considerably, the population's age structure means that numbers will continue to rise uninterrupted for several decades to come. In some parts of the country, the very rapid rise in rural and urban population densities (Figure 115–2) will bring a proportional increase in their needs. Development of the labor market, and that of other facilities (e.g., health and education), however, has suffered from the country's years of politico-economic drift.

Like the health situation, the problem of access to resources appears extremely difficult to resolve. Kenya may be one of the few African countries to have started registering by the end of the 1960s, but some areas have still not been surveyed yet. The first plots to be registered have since been subdivided, sometimes several times, usually without official backing. As a result, there are various land ownership statuses. The same plot of land may even have been legally acquired by several people, if one includes the owners before and after the 1990 conflicts. In a country with a primarily rural population it is essential to solve the land question as fairly as possible—a formidable challenge for the new government.

Whether from the perspective of planners, the government, or researchers from the international community, the lack of high-quality demographic data remains a problem even though, compared with the rest of sub-Saharan Africa, Kenya's data are relatively reliable, plentiful and well preserved. Demographic and health surveys, of which four have been carried out since 1989, are only moderately well adapted to the country's diversity of social contexts and only cover the areas of health and reproduction. Censuses have suffered from the country's political drift since the 1980s. It is time to reinstate this indispensable tool, the basis for all population projections, as well as policy and program planning.

Acknowledgment

This chapter was translated from the French by Heather Juby.

References

Ajayi Ayorinde and Kekovole John, 1998. Kenya's population policy: from apathy to effectiveness, *in*: Anrudh Jain (ed.), *Do population policies matter? Fertility and politics in Egypt, India, Kenya and Mexico*, p. 113–156. New York, Population Council, 203 p.

Barnes Carolyn, 1976. *An experiment with African coffee growing in Kenya: the Gusii, 1933–1950.* PhD thesis. Ann Arbor, Michigan State University, 266 p.

Beck Ann, 1970. *A History of the British medical administration of East Africa, 1900–1950.* Cambridge (MA), Harvard University Press, 271 p.

Bourmaud Daniel, 1988. *Histoire politique du Kenya. État et pouvoir local.* Paris, CREDU and Karthala, 326 p.

Brass William and Jolly Carole L. (eds.), 1993. *Population dynamics of Kenya.* Washington, National Academy Press, 183 p.

Caldwell John C., Orubuloye Israël O., and Caldwell Pat, 1992. Fertility decline in Africa: a new type of transition?, *Population and Development Review*, vol. 18(2), p. 211–242.

Cambrezy Luc, 2001. *Réfugiés et exilés. Crise des sociétés, crise des territoires.* Paris, Éditions des archives contemporaines, 216 p.

Chesnais Jean-Claude, 1979. Un cas de sextuplement en cinquante ans: la population du Kenya, *Population*, vol. 34(1), p. 206–211.

Coast Ernestina, 2000. *Maasai demography.* PhD thesis. London, University College, 293 p.

Coen Amy, 2003. Access denied: US restrictions on international family planning. Bush Administration's Global Gag Rule jeopardizing health care, weakening HIV/AIDS prevention and endangering lives, *Population Action International*, Special issue, 24 September, p. 1–12. (www.globalgagrule.org)

Egerö Bertil and Mburugu Edward, 1994. Kenya: reproductive change under strain, *in*: Bertil Egerö and Mikael Hammarskjöld (eds.), *Understanding reproductive change, Kenya, Tamil Nadu, Punjab, Costa Rica*, p. 65–100. Lund (Sweden), Lund University Press and Chartwell-Bratt, 167 p.

Eshiwani George S., 1990. *Implementing educational policies in Kenya.* Washington DC, World Bank, 33 p.

Frank Odile and McNicoll Geoffrey, 1987. An interpretation of fertility and population policy in Kenya, *Population and Development Review*, vol. 13(2), p. 209–242.

Ghai Yash P. and McAuslan J. Patrick W. B., 1970. *Public law and political change in Kenya. A study of the legal framework of government from colonial times to present.* Nairobi, London and New York, Oxford University Press, 536 p.

Golaz Valérie, 1997. Les enjeux ethniques des recensements: le recensement de 1989 au Kenya, *Politique Africaine*, no. 67 ("La France et les migrants"), p. 113–121.

Greeley Edward H., 1988. Planning for population change in Kenya: an anthropological perspective, *in*: David W. Brokensha and Peter D. Little (eds.), *Anthropology of development and change in East Africa*, p. 201–216. Boulder (Colorado) and London, Westview Press, 258 p. (IDA Monographs in Development Anthropology).

Grignon François, 1993. *Le multipartisme au Kenya? Reproduction autoritaire, légitimation et culture politique en mutation (1990–1992).* Nairobi, IFRA, 81 p. (Working Paper no. 12).

HRW (Human Rights Watch), 1993. Divide and rule: state-sponsored ethnic violence in Kenya. New York, Africa Watch, 97 p.

Iliffe John, 1995. Africans. The history of a continent. Cambridge (UK), Cambridge University Press, 323 p.

Iliffe John, 2002. *East African Doctors. A history of the modern profession*, 2nd ed. Kampala, Fountain Publishers, 336 p.

ILO, 1972. *Employment, incomes and equality. A strategy for increasing productive employment in Kenya.* Geneva, International Labour Office, 600 p.

Kelley Allen C. and Nobbe Charles E., 1990. *Kenya at the demographic turning point? Hypotheses and a proposed research agenda.* Washington DC, World Bank, 97 p.

Kizito Paul M. L., Obunju Walter, Kibet Moses, and Njogu Wamucii, 1991. Fertility transition in Kenya. Calverton (Maryland), Macro International, 23 p. (DHS Further Analysis Studies, no. 11).

Kuczynski Robert R., 1949. *Demographic survey of the British colonial Empire.* London, Oxford University Press, 950 p.

Leys Colin, 1975. *Underdevelopment in Kenya: the political economy of neo-colonialism 1964–71.* Nairobi, EAEP, 284 p.

Locoh Thérèse and Makdessi Yara, 1996. *Politique de population et baisse de la fécondité en Afrique sub-saharienne.* Paris, CEPED, 40 p. (Les dossiers du CEPED no. 44).

Lonsdale John, 1992. –The Conquest State of Kenya 1895–1905, *in*: Bruce Berman and John Lonsdale (eds.), *Unhappy Valley, conflict in Kenya and Africa. Book one: State and class*, p. 13–44. London, Nairobi, and Athens (Ohio), J. Currey, Heinemann Kenya, and Ohio University Press.

Martin Cyril J., 1949. The East African population census, 1948. Planning and enumeration, *Population Studies*, vol. 3(3), p. 303–320.

Matampash Kenny, 1993. The Maasai of Kenya, *in*: Shelton H. Davis (ed.), *Indigenous views of land and the environment*, p. 31–44. Washington, World Bank, 91 p. (World Bank Discussion Paper no. 188).

Maupeu Hervé, 1998. Eglises et société, *in*: François Grignon and Gérard Prunier, *Le Kenya contemporain*, p. 143–168. Paris and Nairobi, Karthala and IFRA, 394 p.

Médard Claire, 1996. Les conflits "ethniques" au Kenya: une question de votes ou de terres, *Afrique contemporaine*, no. 189, p. 62–74.

Médard Claire, 1998. Dispositifs électoraux et violences ethniques: réflexions sur quelques stratégies territoriales du régime kenyan, *Politique Africaine*, no. 70: Le Kenya: le contrat social à l'abandon, p. 32–39.

Médard Claire, 1999. *Territoires de l'ethnicité: encadrement, revendications et conflits territoriaux au Kenya*, 2 vol. PhD thesis. Paris, Université de Paris-I Panthéon Sorbonne, 527 p. + 47 p.

Médard Claire, in press. *Nairobi: croissance urbaine et accès à la terre.* Paris, Karthala, 34 p.

Muchoki Francis, 1998. La période coloniale. 1890–1940, *in*: François Grignon and Gérard Prunier (eds.), *Le Kenya contemporain*, p. 77–93. Paris and Nairobi, Karthala and IFRA, 394 p.

National Assembly, 1992. *Report of the parliamentary select committee to investigate ethics clashes in Western and other parts of Kenya*. Nairobi, Government Printer, 238 p.

NCCK, 1994. Review report on the registered land clashes affected persons in Western Kenya. Nairobi, National Council of Churches of Kenya, 102 p.

NCPD, 1989. Kenya Demographic and Health Survey 1989. Columbia (Maryland), NCPD and Institute for Resource Development/Macro Systems, Inc., 121 p.

NCPD, 1994. Kenya Demographic and Health Survey 1993. Columbia (Maryland), NCPD and Institute for Resource Development/Macro Systems, Inc., 219 p.

NCPD, 1999. Kenya Demographic and Health Survey 1998. Columbia (Maryland), NCPD and Institute for Resource Development/Macro Systems, Inc., 211 p.

NCPD, 2003. Kenya Demographic and Health Survey 2003. Preliminary report. Columbia (Maryland), NCPD and Institute for Resource Development/Macro Systems, Inc., 38 p.

Omoluabi Elisabeth and Lévy Michel L., 1992. Le recensement du Nigeria, *Population et Sociétés*, no. 272, p. 1–3.

Orvis Stephen, 1993. The Kenyan agrarian debate: a reappraisal, *African Studies Review*, vol. 36(3), p. 23–48.

Peatrik Anne-Marie, 1998. Les sociétés anciennes et leur héritage, *in*: François Grignon and Gérard Prunier, (eds.), *Le Kenya contemporain*, p. 53–76. Paris and Nairobi, Karthala and IFRA, 394 p.

Peatrik Anne-Marie, 2002. *La vie à pas contés.* Paris, Société d'ethnologie, 571 p.

Republic of Kenya, 1970. *Kenya Population Census 1969*. Nairobi, Ministry of Finance and Ecnomic Planning, Statistics Division, 123 p.

Republic of Kenya, 1981. *Kenya Population Census 1979*. Nairobi, Central Bureau of Statistics, 326 p.

Republic of Kenya, 1994. *Kenya Population Census 1989*. Nairobi, Central Bureau of Statistics, Office of the Vice-President and Ministry for Planning and National Development.

Republic of Kenya, 2001. *Population and Housing Census 1999*. Nairobi, Central Bureau of Statistics, Ministry of Finance and Economic Planning.

Robinson Warren C., 1992. Kenya enters the fertility transition, *Population Studies*, vol. 46(3), p. 445–457.

Rodriguez-Torres Deyssi, 1995. *Nairobi: Le bidonville face à la ville.* Louvain-la-Neuve and Paris, Academia and L'Harmattan, 83 p. (Cahiers du CIDEP n° 24).

Thomas Lynn M., 1998. *Colonial "knowledge" and counting "superstitions": population discourses and the 1948 East African census*. Communication presented at the African Studies Association, Chicago, November 1, 1998. 9 p.

Uche U. U., 1974. *Law and population growth in Kenya.* Medford (MA), Tufts University, 40 p. (Law and population Monograph Series, no. 22).

UNHCR, 2003. *Refugees and returnees in the East and Horn of Africa*, Geneva, GIMU (Geographic Information and Mapping Unit, PGDS (Population and Geographic Data Section), 1 p. (map). (www.unhcr.ch) van de Walle Etienne and Foster Andrew D., 1990. *Fertility decline in Africa: assessment and prospects.* Washington, World Bank, 63 p. (World Bank Technical Paper).

Vimard Patrice, 1996. Évolution de la fécondité et crises africaines, *in*: Jean Coussy and Jacques Vallin (eds.), *Crise et population en Afrique: crises économiques, politiques d'ajustement et dynamiques démographiques*, p. 293–318. Paris, CEPED, 580 p. (Les Études du CEPED, no. 13).

White Luise, 1990. *The comforts of home: prostitution in colonial Nairobi.* Chicago, University of Chicago Press, 286 p.

C H A P T E R

116

Population Policies in Côte d'Ivoire

AMOAKON ANOH

École nationale supérieure de statistique et d'économie appliquée (ENSEA), Abidjan, Côte d'Ivoire

INTRODUCTION

Côte d'Ivoire, in West Africa, lies along the Guinea Gulf, over a surface of 322,463 square kilometers. The natural environment has favored the development of an export-oriented agricultural economy based on coffee and cocoa production on the one hand, and forestry on the other hand. The country experienced exceptional economic development before it was hit by the economic crisis that affected many African countries in the 1980s, and has been subjected to structural adjustment programs since 1981.

Due to a decrease in economic investments since the early 1980s, the prevalence of poverty has been increasing, and social inequality becoming more serious: poverty rose from 10% in 1985 to 37% in 1995 and 38% in 2002. Poverty affects all regions in the country, though at various levels. According to the 2002 survey on living standards, poverty prevalence varies from 15% in the city of Abidjan to 64% in the western part of the country.

As in many other Francophone African countries, demographic policies were predominantly pronatalist in Côte d'Ivoire up until the early 1990s. However, the first official population program of action, which was endorsed in March 2002, was aimed at reducing demographic growth, thus confirming a complete ideological reversal—quite a belated one since, as Joseph Chamie (1988) recalled, awareness of the problems due to fast demographic growth first became apparent through an official statement made in Accra in 1971, during the first African Conference on Population.

The main purpose of this chapter is to describe the evolution of the population policy in Côte d'Ivoire. A summary of the demographic evolution in the country from the 1950s will be presented in the first section. In the second section the demographic ideas and the measures taken in conformity with the pronatalist principles that prevailed before 1991 will be analyzed. The third section covers how the authorities' attitudes and policies radically changed beginning in the early 1990s. In the last section the new reproductive health policy will be described in detail.

I. THE MAIN DEMOGRAPHIC TRENDS SINCE 1950

According to the most recent census, in 1998, the population of Côte d'Ivoire was 15.4 million. United Nations sources suggest that the population increased from 2.8 million in 1950 to 15.8 million in 2000 (Table 116–1).

The growth rate, though currently declining, is still high: 3.3% a year over the period between the last two censuses (1988–1998). The rapid demographic growth results from high natural growth and high levels of immigration (in 1998, 26% of the population were foreign born). Both factors, however, have been substantially reduced in recent times.

1. A Recent Halt in the Decline of Mortality

The average life expectancy has considerably improved from the late 1950s to the late 1980s. According to UN sources, it rose from a bare 35 years in 1950 to over 50 years in the late 1980s, which means a 15 years improvement within thirty years (Figure 116–1). However, the progress was arrested by the emergence of the AIDS epidemic and since then, there has been a significant decline in life expectancy. According to the United Nations, the life expectancy at birth was only 43.2 years in 1995–2000 (United Nations, 2003).

TABLE 116–1 The Evolution of the Population of Côte d'Ivoire since 1950

Years	Population (in thousands)
United Nations estimates (United Nations, 2003)	
1950	2775
1955	3229
1960	3803
1965	4527
1970	5521
1975	6754
1980	8427
1985	10,460
1990	12,505
1995	14,365
2000	15,827
Population censuses[a]	
1975	6,709,600
1988	10,815,694
1998	15,366,672

[a] Three population census operations have been conducted in Côte-d'Ivoire: the General Population Census in 1975 (15 April to 15 May), the General Population and Housing Census in 1988 (known as RGPH, March 1–31) and the 1998 RGPH (November 21–December 21). Concerning the more distant past, some population estimates were provided through the administrative censuses during the colonial period and the regional surveys conducted from 1962 to 1965 by the Ministry of Finance, Economic Affairs, and Planning to serve as a basis for the regional planning of economic development. A summary of such regional surveys was made by Louis Roussel (1967).

During the phase of progress, improved life expectancy was due mainly to a decline in child mortality. The evolution of the various indicators concerning the latter clearly mirrored that of the life expectancy up to the mid-1980s (Table 116–2) (see also Vimard *et al*., 2002). Since then, the part played by child mortality in the fresh decrease of life expectancy has not been so clear. In particular, judging from the UN estimates (Figure 116–2), in recent years, infant mortality seems to have remained unchanged since the early 1990s, which would not have been enough to account for the decline in life expectancy. Though the results of the 1998–1999 Demographic and Health Survey (DHS) are far worse than those of the 1994 DHS, and seem to suggest a sharp rise in infant mortality in very recent times, the decline in life expectancy, as observed from the late 1980s, was probably due more to a fresh upsurge of adult mortality related to the AIDS epidemic.

It is not unlikely, however, that besides the AIDS epidemic, the reversal in mortality trends observed in recent years may be related to a deterioration of social and economic conditions, which among other consequences may have affected child mortality. This, however, came later than the epidemic. And it has been proved that, at least in its first phases, the economic crisis in the 1980s did not have much impact over child mortality (Barbieri and Vallin, 1996). However, as difficulties deepened and health resources grew ever scarcer, it seemed that the crisis began to affect children in Côte d'Ivoire. That, combined with the effects of mother-to-child HIV transmission, probably caused infant mortality to rise sharply again in the late 1990s. Nevertheless, the fresh upsurge of AIDS-related adult mortality is what initially halted the progress in life expectancy and later caused it to decline.

There is no precise estimate of the extent of the pandemic because not all sites are equipped to monitor cases of AIDS. The number of reported cases (the first case of AIDS having been diagnosed in 1985) is conse-

TABLE 116–2 The Evolution of the Three Indicators of Child Mortality (per thousand)

Mortality indicators	1960[a]	MRS* 1978–1979[b]	CIFS 1980–1981[b]	DHS* 1994[c]	DHS* 1998–1999[d]
Infant mortality rate ($_1q_0$)	157	103	97	88.5	112
Child mortality rate ($_4q_1$)	na	85	59	67	77
Total child mortality rate ($_5q_0$)	201	179	176	150	181

*indices for the 5 years before the survey; na = not available. Sources: [a] Ministry of Planning, 1980; [b] Multiround Surveys, CIFS: Côte d'Ivoire Fertility Survey (Ahonzo *et al*., 1984); [c] N'cho *et al.*, 1995; [d] INS and Macro Intern., 2001.

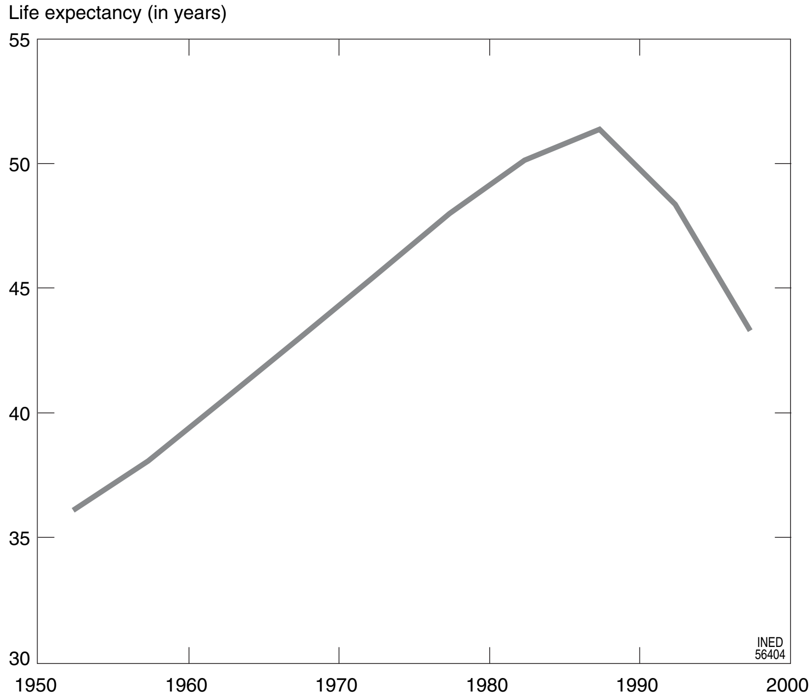


FIGURE 116–1 The evolution of life expectancy at birth since 1950. (Source: United Nations, 2003.)

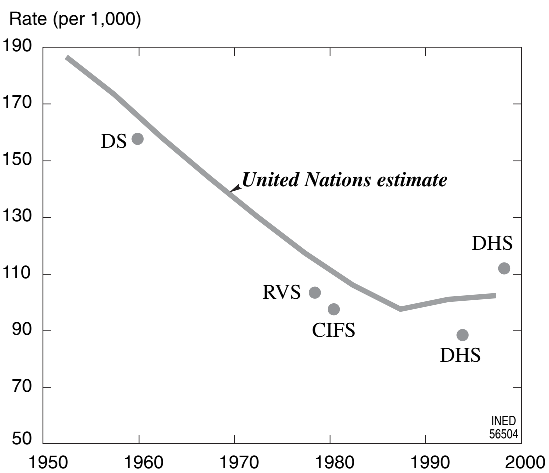


FIGURE 116–2 The evolution of child mortality rates since 1950. (Source: see Table 116–2.)

quently below the actual level (Vimard *et al*., 2002). According to the data available, Côte d'Ivoire is one of the countries most severely affected in Western Africa, and AIDS has been the primary cause of death in the country since the early 1990s. According to the newsletter published by the Ministry for AIDS Control, there are currently 1 million HIV-positive people in Côte d'Ivoire, with the prevalence rate ranging from 10% to 12% (Ministry for AIDS Control, 2003).

2. A Slight Rise before a Rapid Decline in Fertility

There are various estimates of fertility in Côte d'Ivoire in the 1950s. According to the surveys from that time, the total fertility rate (TFR) was 6.5 in 1958 (Roussel, 1967, quoted in Antoine, 1985) and 6.4 in 1962–1964 (Roussel, 1967) (data drawn from the regional studies). According to a current estimate by the United Nations, it was 7 in 1950 (United Nations, 2003). UN estimates, as well as comparison between past surveys and those conducted during the 1980s 7.2 according to the 1980–1981 Côte d'Ivoire Fertility Survey (Ahonzo *et al*., 1984) and 6.8 according to the 1988 General Population and Housing Census (N'cho *et al*., 1995), clearly suggest that there was a slight rise in fertility in the 1960s and 1970s (Figure 116–3). In the past, the level of fertility resulted from a combination of early first marriages among women, rapid remarriage among widows and divorcees—which meant that virtually all women were married—and traditions of breast-feeding (which lengthened the period of postpartum amenorrhea) and sexual abstinence after delivery (which entailed longer intervals between births) (Caldwell, 1993).

The pattern began to change in the 1950s and 1960s, with reduced intervals between births, as a result of the decline of those traditions. The fertility rates by age after 25 increased up to the early 1980s (Figure 116–4).

On the other hand, the effects of the reduced intervals between births were combined with a decline in sterility and a longer life span among couples, as a result of improved health conditions. All of these factors contributed to an increase in fertility and a boost in the natural growth of the population, though the latter was mainly due to reduced mortality, particularly among children.

Beginning in the late 1980s, however, fertility began to decline rapidly, as UN estimates and the results of recent surveys show. From 7.4 children per woman in 1980–1985, according to the United Nations, the fertility level dropped down to 5.3 in 1995–2000, and even to 5.2 in 1998–1999 according to the latest DHS. Though the decline includes all age groups, it is particularly sharp among the youngest (Figure 116–4), which is a sign that the use of contraception has expanded.

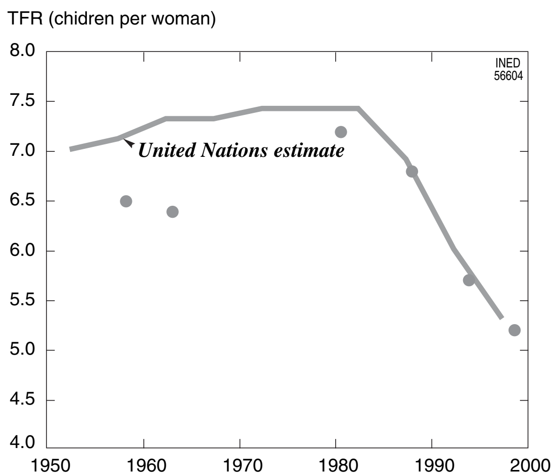


FIGURE 116–3 The evolution of the total fertility rate since 1950. (Sources: 1958: Roussel 1967, quoted in Antoine, 1985; later: N'cho *et al*., 1995; INS and ORC Macro, 2001.)

3. Migration Flows Abruptly Reversing

It is not easy to trace the evolution of migratory flows accurately. To make things as simple as possible in this brief outline of demographic trends in Côte d'Ivoire, we refer to the UN estimates for the annual rate of net immigration (Figure 116–5). From 1950 to 1985, net rates of immigration were 4 to 11 per 1,000. According to Georges Tapinos (2002), the annual rate of net immigration was even as high as 1.3% in 1978–1979. This period of high immigration levels was followed by a period of complete reversal of the situation, with a net immigration rate that was almost nil in 1995–2000.

4. The Incipient Evolution in the Population Pyramid

Due to the influence of those various factors, the age distribution of the population has changed since the

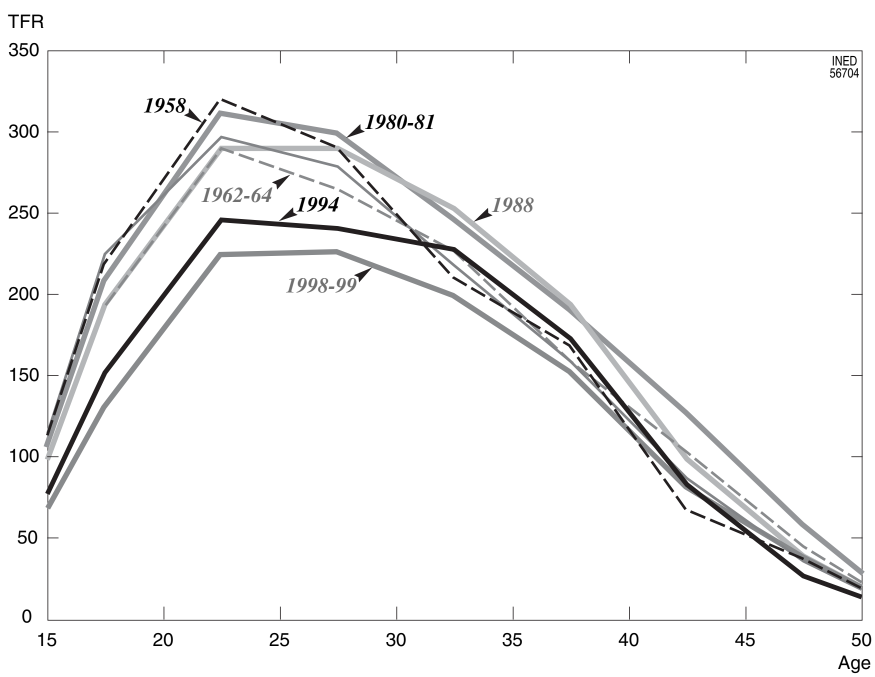


FIGURE 116–4 The evolution of fertility by age from 1958 to 1999. (Sources: 1958: Roussel 1967, quoted in Antoine, 1985; later: N'cho *et al*., 1995; INS and ORC Macro, 2001.)

1950s. The population pyramid, the base of which was very broad in 1950, had evolved by 2000 into quite different a pyramid, with a base clearly narrower and adult age brackets clearly more substantial (Figure 116–6). However, the evolution went through several contrasted phases. This is easily made apparent by tracing the evolution in time of the proportions of people aged under 15 and of adults aged 15 to 64. (The proportion of elderly people aged over 65 remained very small throughout the period.)

From the 1950s to the mid-1960s, the proportion of young people decreased while that of adults increased (Figure 116–7) because the effects of declining child mortality and increasing fertility—which result in making the population younger—were not as strong as the expansion of the adult population, which occurred as a result of immigration. From the mid-1960s to the late 1980s, however, the two factors were perfectly balanced. Though immigration was at its highest, the high birthrate among the native population was heightened by that of immigrant families. In the last phase, from the early 1990s, with the halt in immigration and despite the fall in fertility and the fresh upsurge in child mortality, the proportion of young people has risen again. In fact, by cursorily comparing the 1950 and the 2000 pyramids, we might have come to a wrong conclusion: the aging of the population, due to the recent decline in fertility, is certainly not the main factor of change, and its effects still largely remain to be felt.

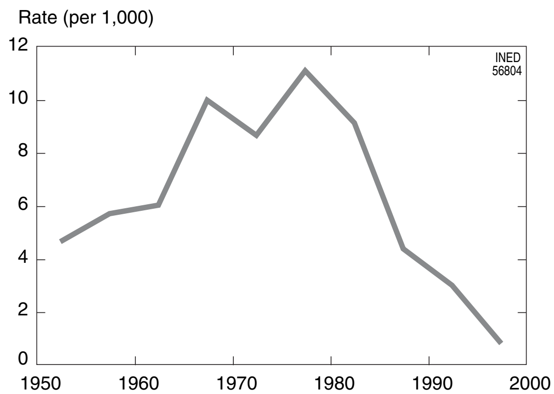


FIGURE 116–5 The evolution of net immigration rates from 1950 according to United Nations estimates. (Source: United Nations, 2003.)

5. The Arrested Urbanization Process

One consequence of the growth of the population was that the density of the population also increased. From 13 people per square kilometer in 1965, the density rose to 21 in 1975 and 48 in 1998. The evolution, however, was very unequal from one region to another (Figure 116–8). For example, the forest areas had their population grow faster, and they are currently more densely populated than the savannah areas. This is the result of intense immigration from the North of the country and from Burkina Faso. The migration began during the colonial period, when it was encouraged by the colonial authorities in a bid to meet the demand for labor in southern Côte d'Ivoire, where plantation economy was being developed, as well as various economic infrastructures. The migration continued after independence, as the various construction projects launched in the South required plenty of labor. Moreover, a succession of dry seasons in the North urged large numbers of people to leave the North and move to the South.

At first the migrants mostly headed for the Southeast, where the plantation economy and the forest industry had begun. Later, they changed their destination, spreading along the Cocoa Belt and more generally following the drift of the plantation economy and the forest industry, which had moved from the East to the West in the early 1970s. Thus it came about that in certain areas in the Southwest, which were practically empty up to the mid-1970s,

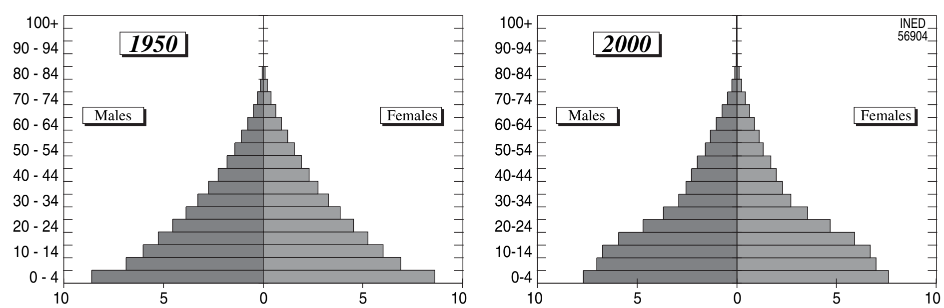


FIGURE 116–6 The distribution of the Côte d'Ivoire population by age and by sex in 1950 and in 2000. (Source: United Nations, 2003.)

TABLE 116–3 The Evolution of the Urban Population since 1921

Years	Towns with a population over 5,000		Urban population		
	Numbers	% of the total population	Numbers	% of the total population	Growth rate (%)
1921	13,000	0.8	32,000	2.1	—
1926	33,000	1.9	—	—	—
1932	—	—	50,000	2.9	4.1
1936	55,000	2.8	—	—	
1945	115,000	5.4	—	—	
1948	—	—	186,000	8.9	8.5
1955	296,000	11.4	330,000	12.9	8.5
1958	—	—	493,800	15.4	14.3
1965	893,000	22.3	980,000	24.5	10.3
1975	2,349,815	35.0	2,146,293	32.0	8.2
1988	4,972,450	46.0	4,220,535	39.0	5.5
1998	6,660,945	43.3	6,529,138	42.5	4.1

Sources: Dureau, 1985; Doudou, 2001.

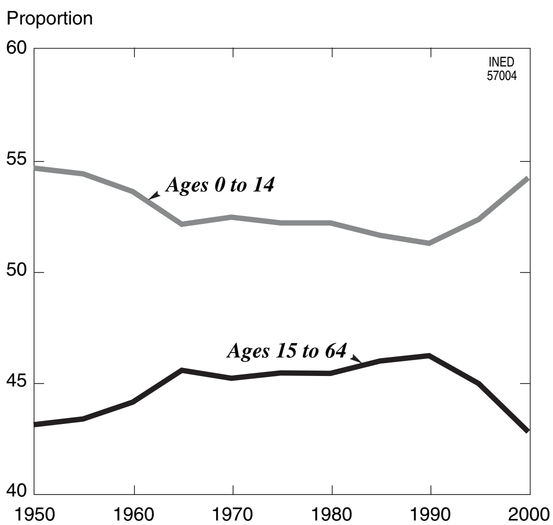


FIGURE 116–7 The evolution of the proportion of people aged under 15 and of those aged 15 to 64 in the population of Côte d'Ivoire, from 1950 to 2000. (Source: United Nations, 2003.)

the population densities are currently higher than the national average.

Urbanization[1] remained negligible until the mid-20th century. In 1948 only 9% of the population of Côte d'Ivoire was living in urban settlements (Table 116–3). However, urbanization was on the way, and in the early 1920s, when the lands of the colony were first developed, the urban population was a bare 2%. Urban growth subsequently gathered speed. It even soared in the 1950s, 1960s, and 1970s, as a combined result of migration, natural growth, and the inclusion of some rural areas among urban settlements. However, the pace of the urban population growth has been slowing down since the 1980s, mostly as a result of migratory flows changing their destination, which in turn was the result of a decline in economic investment.

There are wide disparities from one *département* to another. According to the data from the 1998 population census, some had been through urban transition (Abidjan, Bouaké, Katiola, Grand-Bassam, Dimbokro, Yamoussokro), whereas urbanization was still insignificant in more than half the *départements*.

The urban structure of Côte d'Ivoire is characterized by the overwhelming predominance of the huge city of Abidjan. In 1998 its population was about 2.8 million, whereas that of Bouaké, the second largest town, was under half a million. The number of towns with a population of over 100,000 rose from two in 1975 to five in 1988 and eight in 1998 (Figure 116–8 and Appendix 116–1). Half those larger towns are located in the southwest and the midwest regions of the country, and expanded only as recently as the 1980s and 1990s. Located as they are in the new Cocoa Belt, their expansion is mostly due to high levels of immigration there.

[1] The way "urban population" is defined with respect to Côte d'Ivoire has varied. During the colonial period the definition was mostly an administrative one. From 1960 the size of the population too was taken into consideration. In 1975 a more sophisticated definition was chosen as a result of a consultation process: settlements with a population of over 10,000 were considered urban settlements, but so too were those with a population ranging from 4,000 to 10,000 but where over 50% of the heads of households were engaged in nonagricultural activities. In the 1998 census, the urban world was made up of administrative centers at the level of *départements*, as well as *sous-préfectures* whose population is either over 10,000 or between 4,000 and 10,000 but with over 50% of the heads of households in nonagricultural activities.

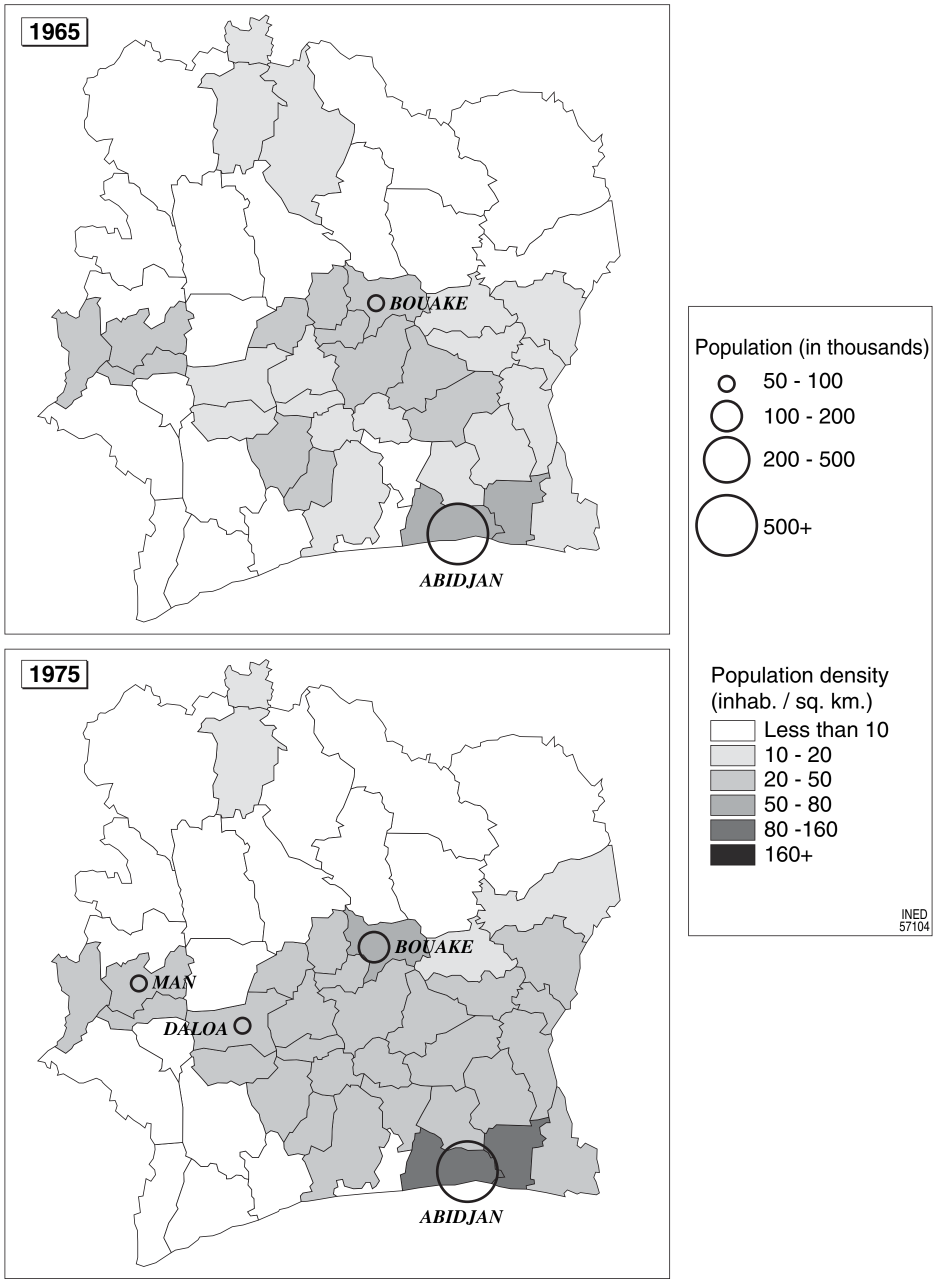


FIGURE 116–8 Population densities by *département* and location of major towns by size in 1965 and 1975. (Source: INS.)

II. A DECLAREDLY PRONATALIST POLICY UP TO 1991

1. The Views in Favor of Population Growth and Their Foundations

The UN permanent surveys with governments suggest that in 1974 Côte d'Ivoire, along with such countries as Gabon, Libya, Morocco, the Central African Republic, Cameroon, and Cape Verde (Makinwa-Adebusoye, 1993), considered that the growth rate of its population was too low. In 1976, however, the government of Côte d'Ivoire stated that the level of fertility and the rate of natural growth were satisfactory and that it did not intend to take any measures to influence them in any way. Most of the African countries who had participated in the UN survey had the same position, particularly the Francophone ones, including neighboring countries such as Burkina Faso, Mali, Guinea, Togo, and Benin. As far as Côte d'Ivoire

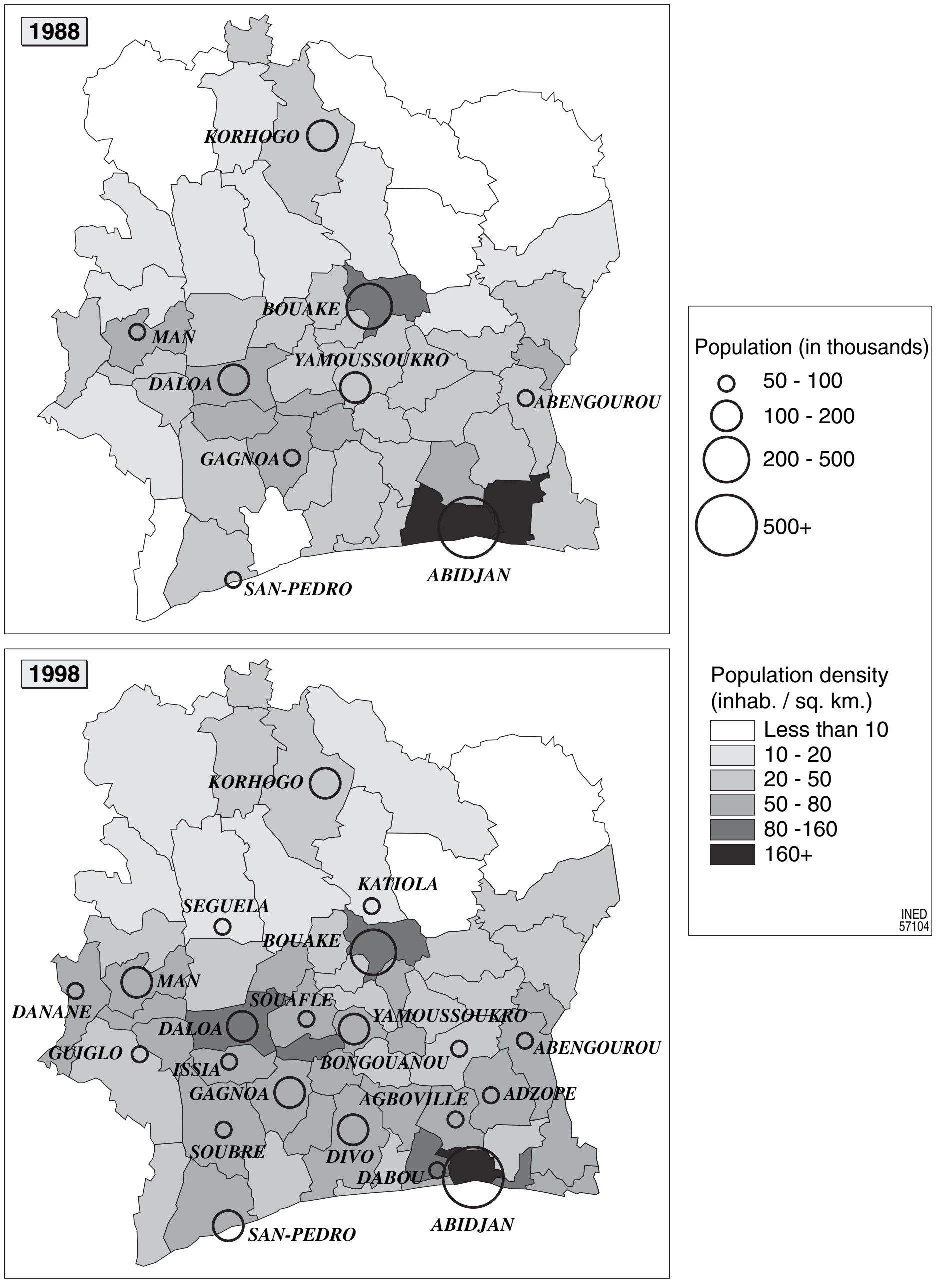


FIGURE 116–8 (*Continued*) Population densities by *département* and location of major towns by size in 1988 and 1998 (Source: INS.)

was concerned, this was a turnaround in policy. The change in attitude was probably due to improved knowledge of demographic facts, which in turn entailed awareness of the fact that, in view of declining mortality, the fertility level was still very high, and thus it was unrealistic to say that the rate was too low.

Through the results of subsequent UN surveys, which were published in the 1980s (1983, 1986, 1989), it appeared that although the government of Côte d'Ivoire was still content with the fertility level and the natural growth rate, it was considering new actions to keep them from growing any higher. In that respect, Côte d'Ivoire was sharing the same position as Mali, Togo, and São Tomé and Principe. Among its Francophone neighbor countries, however, Mali alone officially maintained an identical position as that of Côte d'Ivoire, as Guinea (in 1986) and Burkina Faso (in 1989) had begun to find that the fertility and natural growth were exceedingly high, and stated that they intended to intervene so as to reduce them (Sala Diakanda, 2000).

However, the government's official statements, as reported in UN surveys, were somewhat different from the statements made in such technical documents as the five-year development plans, in 1976–1980 and 1981–1985. In the 1976–1980 planning document, significantly demographic events were taken into account in development planning for the first time, and one objective was to "achieve increased awareness and harmony in population growth, both at family and at national level." According to the planners, such an objective was intended to entail a "quantitative increase in the number of children alive for the benefit of all mothers," on the one hand, and "qualitative improvement in the health level of mothers as well as children," on the other hand (Ministry of Planning, 1980, p. 474–477).

That the government persistently intended the population to keep rapidly growing can be illustrated through three examples of positions expressed in the 1980s. These examples are revealing of how the ideology prevailing among the political class had remained unchanged since the 1960s, in a single-party regime characterized by single political thought and discourse.

In December 1982, in his closing speech at the seminar "Population and Development in Côte d'Ivoire," held in Abidjan, the Minister of Health declared:[2]

> As we have said, the population is the primary element in development. And as you experts have established, the population in our country, which is currently estimated about 8 million, including the native and the immigrant population, is not large enough if compared to the surface of the country. Comparing our population with those of countries with similar surfaces, in 1980 the population of Italy was estimated 57 million for a surface of 310,000 square kilometers. During the same period, the population of Japan was 117 million for a surface of 372,000 square kilometers. If we take only that factor into account, we realize that there is no threat of demographic explosion in our country. (*Fraternité Matin*, December 21, 1982)

Along the same line, the director for Statistics at the Ministry of Planning said, in July 1987, on the occasion of the celebration of population of the planet reaching 5 billion:

> There is no population problem in Côte d'Ivoire. Though certain areas have been identified [. . .] where the population density is very high, with all the social problems it entails, other areas are virtually empty. [. . .] There is no demographic problem; the problem is rather how to use our population. (*Fraternité Matin*, July 13, 1987)

In January 1989, the Minister for Women's Advancement also stated how the authorities thought it desirable that the population grew larger, in her opening speech at the seminar in Abidjan on approaches to birth spacing programs:

> For a long time to come, there will be no overpopulation problem in Côte d'Ivoire. [. . .] It is unnecessary to try to limit childbirths in order to lift concerns that may result from a growth rate among the African population that may be deemed over high, so as to create an obsessive fear of "the Black peril," for example.

In the view of the national authorities of the time, there was no such thing as a population problem, nor was there any need to slow down the population growth in the country. Far from that, a high rate of population growth was deemed desirable. Such views in favor of population growth in Côte d'Ivoire were similar to those previously expressed by civil society representatives in "*Côte d'Ivoire 2000*, the first National Long-Term Perspective Study, which was conducted from October 1973 to June 1974.

The authorities' position was underlain by several factors. Firstly, the government believed that population growth would have positive effects on economic and social development. The alleged advantages of a large population were of various kinds—economy of scale, an urge to work hard and be creative, and so forth. Population density was perceived as a means through which population growth determined the wealth of nations. In that view, low population density at national level (17 people per sq. km. in 1970, 25 in 1980, and 36 in 1990), combined with wide regional disparities, contributed to the authorities continuing to think of population growth as desirable. The abovementioned comparison between Côte d'Ivoire and a few countries such as Japan or Italy, which had a far different point of view concerning population density, and still more so from that of economic power, is revealing of the terms of the argument of the time, in which population density was believed to be synonymous with economic development.

Such an attitude among the authorities was further confirmed by ideological and socioeconomic motivations among individuals. In the common people's view, the birth of a child is always perceived as a positive event. From a cultural point of view, in the people's beliefs, a child bears the soul of some member of the community of ancestors. This is the reason why birth limitation remained all but taboo. Also, from a socioeconomic point of view, the parents wanted to

[2] The seminar was held in Abidjan December 13–19, 1982, and jointly organized by the Ministries of Planning and Industry, and of Public Health and Population, on the one hand, and the United Nations Economic Commission for Africa (ECA) on the other hand. The Minister's speech was read by the Minister of State for Industry.

have a large family because children represent ready-made labor, a guarantee of security for their old age, and social prestige (attached to the large number of their descendants).

One more factor that accounted for continued pronatalist attitudes in Côte d'Ivoire was the fact that it belonged to the group of former French colonies. Not only had France left its colonies with a pronatalist legislation very much like its own, but most importantly, it tended to hold a qualified position in international conferences on population and development. For example in Bucharest in 1974 the French delegate, Michel Durafour, then Minister of Labor and Population, steered a middle course between those countries that advocated family planning programs (such as the United States, India, Indonesia, and Bangladesh) and those that believed that development was the best of contraceptives (Algeria and Argentina, among others) (see *Fraternité Matin*, August 22, 1974).

The pronatalist position of Côte d'Ivoire thus appears to be underlain by many different reasons, not only economic and material, but also political, cultural, and religious ones. Such a variety of factors account for the fact that, despite improved knowledge of the relations between rapid population growth and economic and social development, on the one hand, and the economic crisis and the consequences of the structural adjustment programs that have been enforced since the early 1980s, on the other, it took a long time for the political authorities to change their official position.

The pattern for development in Côte d'Ivoire, which was particularly interventionist, largely relied on planning, which required good knowledge of the socioeconomic and demographic situation. The combined efforts of the Ministry of Planning and multilateral and bilateral cooperation made it possible to improve demographic knowledge. For example, regional surveys were conducted all over the country from 1962 to 1964 in preparation for the first five-year plan. Also, Côte d'Ivoire participated in the African Census Program (population censuses in 1975, 1988, and 1998), the World Fertility Survey (with a fertility survey in 1980), and the Demographic and Health Surveys (1984, 1998–1999).

2. The Social Measures Intended to Encourage the Natural Growth of the Population

The first social measures taken immediately after the independence of Côte d'Ivoire, particularly in the areas of women's education and health, for "the promotion of man in Côte d'Ivoire and improved adequacy of the economic, social, and cultural system" (Ministry of Planning, 1980, p. 455–487) were reinforced from the second five-year plan (1976–1980) with a view to promoting "more awareness and harmony in demographic growth, both at family and at national level" (Ministry of Planning, 1980).

As far as education was concerned, an ambitious policy had been decided on immediately after independence, with a school enrollment rate of 100% by 1980 as its main objective. (For more on the development of education in Côte d'Ivoire, see, e.g., Anygbayere, 1992). Measures were taken to urge families to send their children to school. A slogan was launched—"A man who cannot read is like a blind man in the forest" (Any-Gbayere, 1992, p. 2)—and the government contributed to expenses for children's schooling. In addition, infrastructure was expanded, particularly during the 1970s, salaries were made more attractive (by giving teachers special rates, different from those of other civil servants), and teachers were supplied with lodgings, which ensured that educational and training services were available. The budget for the sector was constantly on the rise, from 12.5% in 1960 to 40% in 1980 (Ministry of National Education, quoted by Zanou, 1994).

Those efforts resulted in significant reduction of illiteracy among the adult population, particularly among women. The literacy rate among adult women rose from 5% in 1970 to 24% in 1990. During the same period, the rate among men increased from 35% to 44%.

The influence of education on fertility depends on supply and demand. Education changed people's notions about fertility-related matters and provided better knowledge of biomedical norms of procreation. Education also mechanically postponed the beginning of reproductive life and broadened access to health care, which was favorable to mothers' and children's health. Besides, it broadened access to information on the existence and proper use of modern contraceptive methods.

Health-related measures were also taken with a view toward encouraging people to have many descendants while improving child survival. From the 1976–1980 plan, the health system was firmly geared to promote preventive medicine, social hygiene, and health education, based mostly on the approach to basic health care advocated by the international community in the 1970s. Although the priority was preventive medicine, curative medicine was not neglected. The health infrastructure was thus expanded, particularly in Abidjan and certain regional administrative centers such as Bouaké, Abengourou, Korhogo, Yamoussoukro, and Daloa. Specific programs in mothers' and children's protection were implemented. For example, the Expanded Immunization Program was developed, with a view to immunizing children from the most serious transmittable diseases and pregnant women from tetanus.

The pronatalist legislation inherited from France, particularly the act passed in France just after the war (July 31, 1920) in an effort to stop the fall in fertility, banning abortion as well as advertisements of contraceptive products and their importation and sales in Côte d'Ivoire, was relaxed in the early 1980s, with a view to ensuring that women and children were in better health. The 1920 act was repealed through Article 11 in Act 81-640 (dated July 31, 1981), instituting the new penal code (*Journal Officiel de la République de Côte d'Ivoire*, January 4, 1982). Access to reversible contraceptive methods was made easier, and new laws were enforced that slightly relaxed the former laws on abortion. The complete ban on abortion was lifted, only to be replaced with an extremely restrictive system, in which abortion was allowed only to save the life of a pregnant woman. Two doctors, besides the general practitioner, were required to confirm that the abortion was necessary to save the mother's threatened life. In all other conditions, abortion was considered a crime, and so were advertisements on and sales of abortive substances (Articles 366 to 369 in the penal code). Sterilization remained prohibited. Although the ban on the more common reversible methods, such as the pill, injections, intrauterine devices, and condoms, was lifted, they received no direct support from the government until the late 1980s. Neighboring countries such as Ghana and Mali, and pioneering ones like Kenya, had moved far ahead of Côte d'Ivoire in that respect, as they have been directly supporting the distribution of contraceptive methods since the 1970s.

Though the repressive pronatalist legislation was relaxed, because of the prevailing pronatalist ideology, no family planning program was implemented in Côte d'Ivoire during that period. In the view of the authorities, family planning in Côte d'Ivoire, as the Minister for the Promotion of Women put it, could only be

> synonymous with birth spacing, with the preservation of mothers' and children's health as its main objective, as defined by WHO, i.e. such general condition of physical, social and psychological welfare as is likely to preserve family balance and provide youth with natural means for their protection. (Aka-Anghui, 1989)

A few dissenters were sternly brought back into line, for example, during the Population and Development in Côte d'Ivoire seminar in Abidjan in 1982 (Zanou, 1994), and after the speech on family planning by the Chairman of the Côte d'Ivoire Family Welfare Association (known as AIBEF) at the pilot center at Port-Bouët (in the Commune of Abidjan) for the young people of the commune on August 18, 1988. Following the report in the August 24, 1988, issue of the daily newspaper *Fraternité Matin*, the Minister of Health, as directed by the President of the Republic, called to order AIBEF Chairman, Professor Samba Diarra, and had a conference scheduled for the members of the Economic and Social Council cancelled (Samba Diarra, personal communication, March 1999).

Among those interventions implemented to actualize the pronatalist attitude, women's education, the new laws on the family, and the health measures were ambivalent, as they did not necessarily induce such higher natural growth rate as the national authorities were bent on achieving.

The new laws on the family provide a good example of the ambivalent measures taken during the pronatalist period in Côte d'Ivoire. Beginning in the early 1960s, the authorities undertook to develop laws on the family, with a view to unifying the existing legal rules at the national level and also to adjusting customary and traditional rules to economic and social evolution. The new laws, which were inspired by the French Civil Code and which instituted the national civil law, were passed in 1964. The laws, which were passed in the National Assembly on October 7, 1964, and promulgated on December 8, 1964, were published in the *Journal Officiel de la République de Côte d'Ivoire* on October 27 and December 17, 1964 (Raulin, 1968).

The provisions concerning marriage, divorce, and inheritance granted women more autonomy and laid the basis for the notion of the family to change from "extended" to "conjugal." The law on marriage (Law 64–375), in which legal age for first marriage was 18 for women and 20 for men, made the personal consent of each of the prospective spouses mandatory, prohibited polygamy, and abolished matrimonial compensation (which consisted of offering material benefits to the prospective bride's extended family). Those provisions were intended to allow time for girls to achieve a degree of maturity before they got married. It was now impossible to marry girls off without their consent, which meant a complete break from the traditions of early, forced marriage, particularly of the polygamous type. In 1983 the reform was complemented, as men and women were granted the same rights and duties in the family (Law 83–800, August 2, 1983). Also, in the 1964 law on divorce (Law 64–376), infertility and impotence were barred as grounds for divorce. (According to the Ten-Year Perspective Study 1960–1970, there was a significant proportion of infertile women in the early 1960s.) This was a significant change, in which procreation was no longer viewed as the primary objective of marriage; however, in the traditional system, infertility did not necessarily entail divorce, as it usually encouraged men to become polygamous. Last, the conditions for inheritance were also amended (Law 4–379). The children and the wife of the deceased were now among his heirs, which was unthinkable in traditional systems, among patrilineal groups (the savannah and forest people, West of Bandama) as well as matrilineal ones (the forest people East of Bandama). However, concerning the means to enforce the new laws, the legislators did not provide for any sanction for trespassers, preferring educational methods to repressive ones (Raulin, 1968).

3. An Ultraliberal Immigration Policy

Immigration was the other component of the propopulationist policy in Côte d'Ivoire. That policy enforced during the first two decades after independence has often been termed ultraliberal. It was reflected in the speech by President Felix Houphouët-Boigny at the Sixth Congress of the Democratic Party of Côte d'Ivoire in 1985 (quoted by Brou and Charbit, 1994, p. 35):

> We have welcomed over two million of our brothers here. Many of them have settled for good, and will never go back home again. What we do in Côte d'Ivoire is welcome them. And we do so in a spirit of an African-style humanity inspired with a sense of brotherhood.

That liberal policy made Côte d'Ivoire the major immigration country in West Africa (United Nations, 1997, quoted by Tapinos, 2002). The flows were particularly high during the first two decades of independence. The net amount of international migration was estimated 450,000 people during the five years between 1976 and 1980. According to the data of the recurrent visit survey in 1978–1979, the net annual immigration rate was 1.3% in 1978–1979. However, due to the crisis and consequently scarcer economic investment, the immigration flows sharply decreased during the 1980s, and the net annual immigration rate was brought down to a bare 0.5% in the 1988–1992 period.

The data available show that although the flows were decreasing, the migratory stock still continued to go up. From 700,000 people in 1965—17% of the total population—the foreign population rose to 1.5 million (22%) in 1975, to 3.0 million (28%) in 1988 and 4 million (26%) according to the latest census in 1998.

Such immigration is not exactly the direct consequence of any ad hoc measures taken by the authorities. The two main interventions in that respect (the March 1960 Agreement between Côte d'Ivoire and Haute-Volta, and the Agreement of Cooperation between France and Côte d'Ivoire in 1961) (see Brou and Charbit, 1994) did not really have much effect in themselves. Nor did multilateral agreements such as those made by the *Conseil de l'Entente* (the Council of Accord, a West African organization for regional cooperation with five Francophone member countries: Benin, Burkina Faso, Côte d'Ivoire, Niger, and Togo) or the Economic Community of West African States (ECOWAS), which advocated free movement of people and goods. Immigration in Côte d'Ivoire is rooted in colonial strategies of land development as well as in the pattern for development decided on by the national authorities after independence. The migration was mostly labor based. At first, migrants came of their own free will. However, to thwart the tendency of the populations from the North, particularly from Haute-Volta, to go to the Gold Coast (now Ghana), where wages were more attractive, the colonial administration soon conscripted men from Haute-Volta for forced labor. The process was later reinforced during the war effort.

After forced labor was abolished in 1944, the employers formed an association, the Inter-Professional Union for Labor Recruitment (known as SIAMO), which started signing contracts with Mossi chieftains. The annual number of laborers recruited by SIAMO is estimated to be 20,000 people between 1953 and 1959. From 1919 to 1959 a total of 776,418 migrants from Haute-Volta were thus transplanted to Côte d'Ivoire.

After SIAMO stopped working in 1960, immigration continued, as encouraged in the March 1960 Agreement, in which foreign immigrants were recognized as an important component of the economy that the colonial power had left for the national authorities to manage. However, much of the migration went on without the institutional framework provided by the agreement, as spontaneous movements expanded, based on powerful social networks in the larger towns as well as in rural areas. After the immigrants were expelled from Ghana in 1969 and a crisis affected that country from the 1970s, Côte d'Ivoire became the first country for immigration in West Africa, superseding Ghana, which had been the main one in the subregion up until the early 1970s.

The ultraliberal, loudly advertised policy not only made it easy for immigrants to come to Côte d'Ivoire, but also to have access to employment, both in the private and the public sector, to enter the professions or to create agricultural concerns, particularly in the South where the forests were. Access to land property was also made easier for foreigners. However, the law published on March 20, 1963, according to which unregistered land should be considered as government property, could not be enforced, due to strong opposition from the traditional landholders, and was replaced with a mere statement by the President, who said that "lands belong to those who work on them, provided they are observant of the current laws and rules." In practice lands were distributed in accordance with customary rules, in which land is collectively owned, inalienable property. It cannot be sold off, but a portion of it can be allotted to a member of the clan or any other person who then holds the right to work on it. As a token of his gratitude, a person who acquires land in this way must offer a few bottles of liquor, a very small sum of money, and so on. Thus anyone could have access to land property without being particularly rich. This is how many foreigners who settled in Côte d'Ivoire created agricultural concerns, particularly in the new Cocoa Belt, to which most international and internal migrants rushed in the 1970s.

The political class, however, could never agree over that liberal option, about which reservations were made concerning certain aspects of the immigrants' integration. (The presence of foreigners has never been a problem in itself in Côte d'Ivoire society, which is reputedly very hospitable. It is the degree of integration that may cause reluctance.) This is the reason why in 1966 the Parliament of Côte d'Ivoire firmly opposed President Houphouët Boigny's proposition to allow nationals of the former French West Africa (AOF) to apply for double nationality (Dozon, 2000). And the reservations concerning the immigrants' integration were considerably heightened as the economic crisis affected the country and unemployment rose among educated youth in the 1980s.

The crisis and the structural adjustment programs (SAPs) caused a radical change to be effected in the population policy, reflected in both migration and fertility policies.

III. A SHIFT IN POPULATION POLICIES FROM 1991

1. From the Policy Statement on Human Resource Development to That on Population

The change in the government's attitude toward demographic issues was shown in the *Policy Statement on Human Resources Development* issued in 1991. The policy was one of the main components of the SAP during the period 1991–1993. There were three components in that phase of the SAP in Côte d'Ivoire: the Financial Sector Adjustment Program, the Competitiveness Adjustment Program, and the Human Resources Development Program. According to the government's policy statement, the human resources development policy consisted in implementing a population policy geared toward sustainable development (Republic of Côte d'Ivoire, Monitoring Committee on Human Resources Promotion, 1991).

Although the population density remained comparatively low and many factors that favored fertility were still enforced, the crisis, combined with pressure from the international institutions, caused the national authorities to take up new, neo-Malthusian views. However, besides the SAP, a conjunction of internal and external factors also urged the authorities to change tack in population policy.

At the national level, the advocacy by the national Association for Family Welfare (AIBEF), the seminars and colloquiums on population, the improvement of demographic knowledge, and population projections all influenced the public authorities' perception of population issues. For example, a national seminar was held for the members of Parliament on the theme of Health, Population, and Development in Abidjan in November 1987. The same theme was discussed in several other seminars, including the district conference for community leaders held in Bouaké on January 18, 1990. At the regional level, the African population conferences, both scientific and political (Accra 1971; Arusha 1984; Dakar 1988; Dakar/N'gor 1992), as well as the setting up of interest groups, such as the Council of African members of Parliaments on Population and Development (created in 1986), contributed to promote and strengthen a new understanding of population issues. At the global level, the international institutions had a decisive influence on the government's positions, particularly through the world population conferences organized by the United Nations, but also, in a more concrete way, through the action of the World Bank or the International Monetary Fund. Côte d'Ivoire did not attend the Bucharest conference, but it was present in Mexico in 1984, when it was represented by its ambassador to Mexico. Whereas its participation was still timid at the time, it was more determined in 1994 in Cairo, after the government had changed its position: that delegation, led by the Minister of Culture, included a dozen government and associations executives and five representatives from the media.

Both the persuasion strategies developed by the United Nations agencies and the AIBEF and the World Bank's human resources development policy clearly accelerated the change in the national authorities' perception; but it was the deterioration of the socioeconomic situation that basically provided favorable ground for the change in policy.

The human resources development policy resulted in the adoption, in March 1997, of a statement on the National Population Policy (Republic of Côte d'Ivoire, Ministère délégué auprès du Premier ministre chargé du Plan et du Développement industriel, 1997). Henceforth the situation of Côte d'Ivoire was similar to those of many other Francophone countries, the majority of which had defined a population policy no sooner than the 1990s.

The objectives and broad outline of the new National Population Policy (NPP) of Côte d'Ivoire were largely inspired by the principles and recommendations in the World Population Action Plan which, since the 1984 Conference in Mexico City, had been advocating an approach in which population programs were to be integrated in economic and social development strategies. The general objectives of the NPP were thus organized around three themes: the population movement; the economic, social, cultural, and environmental background; and the capacities in planning and managing population policies and programs.

At the institutional level, four bodies were created that were intended to manage population issues:

1. A decision body made up of the Council of Government ministers.
2. A consultative body, the National Population Council (known as CONAPO) the members of which were those Ministries involved in population issues and nongovernmental organizations (NGOs). At the regional level CONAPO was represented by Regional Population Councils (the COREPOs), which were made up of the regional representatives of the technical government departments who were members of the CONAPO.
3. An executive body made up of 14 technical government departments and those NGOs active in impacting on the demographic movement.
4. A technical body, the National Population Bureau (the BUNAP), which served as Secretariat to the CONAPO, coordinated the activities of the various sectors, provided technical support to the Government departments in resource mobilization, and prepared and submitted projects to development partners for funding.

As part of the implementation of the National Population Policy, a National Population Program of Action (known as the PNA) covering the period 2002–2006 was adopted in March 2002 (BUNAP, 2002).

2. The 2002–2006 National Population Program of Action (PNA)

The National Population Program of Action (PNA) for 2002–2006 was adopted too recently for any assessment of its implementation to be made yet. However, it is possible to assess its objectives, priority actions, and investments, as well as the institutional framework for the implementation of the program.

The objectives of the PNA were defined based on the population issues mentioned in the National Population Policy Statement and those identified during the report and diagnosis made while the PNA was being developed. A number of major demographic challenges were thus identified, for the 2002–2006 PNA to meet (BUNAP, 2002, p. 52):

Manage demographic growth
Manage migratory movements and refugees
Alleviate poverty
Alleviate the impact of AIDS
Ensure that the family's fundamental values are respected, reinforce women's autonomy, ensure that youth are the spearhead of the nation, and provide for the aged
Control urbanization, which is exceedingly rapid
Mobilize resources for the implementation of the program
Ensure efficient management of population programs

The five general objectives selected in the National Population Policy Statement have induced five operational subprograms to be defined, each of which is placed under the responsibility of a government department. The National Population Bureau (BUNAP) is in charge of the cross-sector coordination of activities (Figure 116–9). Each subprogram includes a number of Priority Actions and Investments (PAI), which in turn include immediate objectives, the term of which is 2006. The PAIs are the basic elements for the implementation of the National Action Program. Implementing the PAIs requires collaboration between several government departments, private associations, and NGOs committed to population issues.

We will take two subprograms related to demographic movements to illustrate how they work: controlling natural population growth and developing capacities in migratory movements management. The former is placed under the responsibility of the Ministry of Health and Population, and the latter of the Ministry of Home Affairs and

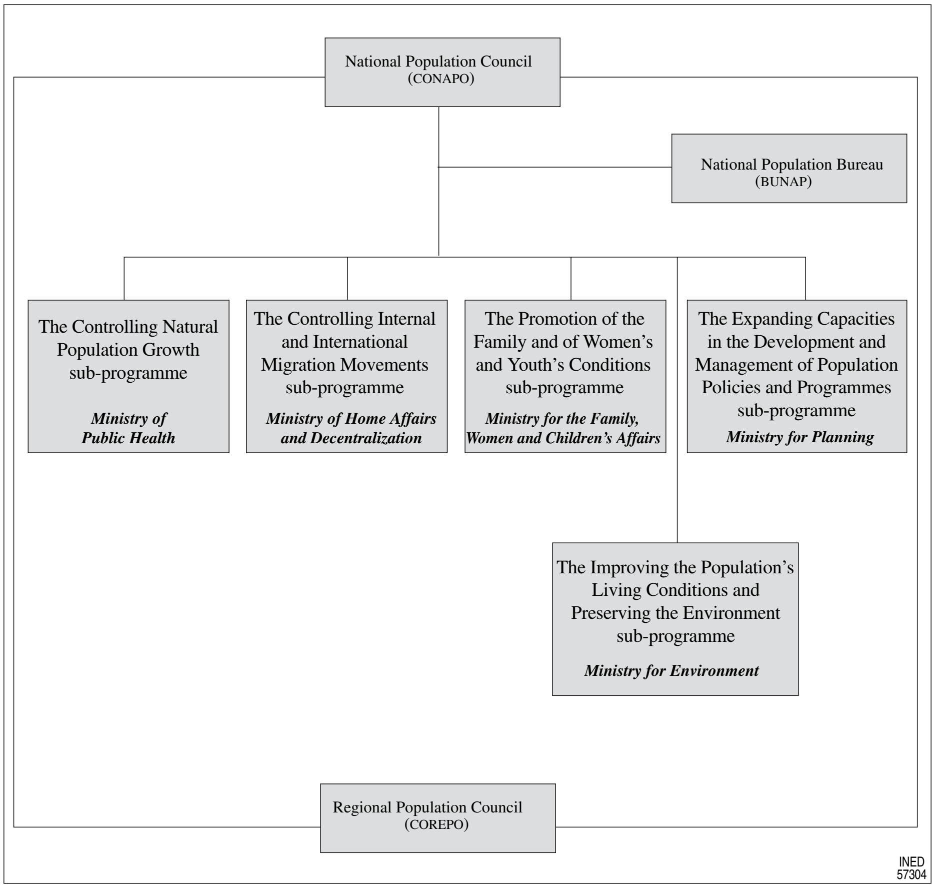


FIGURE 116–9 The organization of the Population Program.

Decentralization (now renamed Ministry of National Administration).

Controlling the Natural Population Growth

The point on which the newly adopted population policy diverged from the former was primarily the question of fertility control. The subprogram on Control of the Natural Growth of the Population includes five PAIs, one of which is concerned with "improved supply, accessibility and use of reproductive health services (including the sexual and reproductive health of adolescents and youth)."

Seven immediate objectives are included, implying a number of activities the implementation of which requires the implication of the executive directions (EDs) of various programs, including the National Reproductive Health Program (NRHP). (There will be further discussion on the part played by the ED/NRHP as we analyze the reproductive health policy.) Technical government departments (the Ministry of Youth, of Defense, etc.) are also involved, as well as NGOs and associations.

By examining the operation plans of the various subprograms, it turns out that the immediate objectives are usually defined for the whole of the national territory. Does this mean that those objectives will be uniformly implemented in all the regions in the country? It is not mentioned, but considering that demographic behaviors are not homogeneous, the heterogeneousness should be taken into account.

The relations between the National Population Bureau (BUNAP) and the implementing structures must also be studied. At first BUNAP was no more than a minor department within the Prime Minister's office, but it has just been upgraded into a central department in the Ministry of Planning and Development. Such institutional position makes it easier for BUNAP to coordinate the activities of the other central departments involved in the national population policy. At the present stage of the process of developing and implementing the national population policy, it has been able to carry out its tasks without any major difficulties, probably thanks to the participatory approach being used, but there is no knowing how long such non-conflict collaboration can be continued between BUNAP, the technical government departments, and the NGOs working on population issues. The question is justified because, due to lack of any centralized planning, the government departments tend to develop their own individual sectoral development plans and compete in the search for funding.

A few examples can be mentioned: the 1996–2005 National Health Development Plan (NHDP); the 1996–2010 National Environment Action Plan (NEAP), the 1997–2010 National Plan for the Development of the Education and Training Sector (known as PNDEF), the 2002 National Action Plan for Women, and the 2002 Provisional Poverty Alleviation Strategy Document.

The armed conflict that broke out in Côte d'Ivoire on September 19, 2002 has slowed down the implementation of the National Action Plan, in that the development partners, who very actively participated in the implementation of the population policy, have redirected their support in favor of humanitarian aid, particularly assistance to displaced populations.

However, a number of PAIs are beginning to be implemented because they were defined based on actions that were already ongoing within sectoral programs conducted by government departments. For example, the PAIs for the Control of the Natural Growth of the Population subprogram are being implemented within the National Health Development Plan (NHDP), precisely within the National Program for Reproductive Health and Family Planning.

3. Immigration Control and Foreigners' Integration

The next major turnaround in Côte d'Ivoire's population policy concerns immigration. The ongoing changes include the three main components of migratory policies: the conditions for foreigners to enter and stay in the country, foreigners' integration, and international cooperation.

The law of May 29, 1990, under which the entry and stay of foreigners in Côte d'Ivoire are presently regulated, is more restrictive. Though, according to the law, any foreigner is supposed to have free access to the national territory, provided they produce official identity and health documents. The identity documents can be used to enter Côte d'Ivoire only if a visa has been delivered, and only the nationals of a few countries, namely those who are members of the Economic Community of West African States (ECOWAS), do not have to have a visa (Aggrey, 1999). Thus, incoming foreigners are now controlled as a rule. However, the law is not always wholly enforced because the borders can be easily crossed, and illegal situations can be perpetuated because the notion of residence is quite vague in Côte d'Ivoire, inasmuch as there is no precise system of residential address as in developed countries.

The migratory flows were first inverted during the latter half of the 1980s, due to the growing crisis in employment, as a result of scarcer investment and its consequences on the migratory exchange between Côte d'Ivoire and Burkina Faso (Vimard *et al*., 2002). President Felix Houphouet-Boigny had established a liberal system, in which immigrants had the same rights as nationals, particularly as far as access to land ownership and public offices and participation in various elections were concerned. Would the change have been so great, if that policy had not been called into question? Probably not—still, most of the arguments in favor of nationalism in Côte d'Ivoire, which is sometimes considered the impulse behind the principle of *Ivorianness*,[3] are built on the demographic weight of foreigners. However, there are not enough data available for us to distinguish with any accuracy, in the reversal of migratory flows, between what is related to policy change and what is due to the crisis in employment that resulted from the economic crisis.

As a matter of fact, though migration flows have been inverted, significant flows are continuing to come into Côte d'Ivoire, as discussed in a recent article in the daily newspaper *Fraternité Matin* (Zobo, 2003). Such flows are supported by networks established by the immigrants who have previously settled in the country.

Another important component of the migration policy is the set of measures taken to ensure better integration of immigrants in the country of residence. The issue is crucial because although immigrants keep close ties with their native countries, the majority of them have come to settle for good (Tapinos, 2002). According to the data from the 1998 national population census, the population pyramid is becoming more regular: the proportion of foreigners born in Côte d'Ivoire, of immigrant parents, was 47%, and the number of foreigners is bound to keep increasing because of its natural

[3] "Ivorianness" is quite vague a concept. Those who developed the concept say that it is related to culture, whereas those who oppose it think that it means exclusion and xenophobia (see, e.g., *Politique Africaine*, no. 78, June 2000).

dynamic. Integration, however, is a long-term process, and involves a number of mechanisms including education, employment, and nationality.

As far as education is concerned, the school enrollment rate among children of immigrants, though quite high, is lower than among the Côte d'Ivoire natives; this applies to both girls and boys. According to the 1993 Migration and Urbanization Survey (known as EIMU) (N'Cho *et al*., 2001), the school enrollment rate among the daughters of immigrant parents was 35.5%, compared with 51.7% among those of Côte d'Ivoire natives; among boys, the rates are 49.7% and 61.2%, respectively (N'Cho *et al*., 2001). And yet, all young people are supposed to have equal access to schooling and training in Côte d'Ivoire, whatever their nationalities. The variations observed between the various nationalities are thus due not so much to public policies as to the parents' assessment of the costs and benefits related to schooling. On the other hand, some immigrants' children return to their parents' native countries for further training because the cost of education is lower there than in Côte d'Ivoire (Zanou, 2001).

Virtually all foreigners of working age have an occupation. They are mostly in two sectors: the primary sector (agriculture, cattle breeding, fishing and forestry) and trade, though on a smaller scale. Such orientation in the foreigners' activity is determined by that of three communities: the natives of Burkina Faso, Mali, and Guinea, who represent approximately 82% of the foreign population. In the agricultural sector, the foreigners work as foremen, unskilled workers, or mothers' helpers; in trade and the craft industry, most immigrants are self-employed. It seems that foreigners have a monopoly over certain sectors, such as retailing. The foreigners' level of economic integration thus appears to be quite exceptional, or in fact unique in West Africa. Economic integration is achieved despite scarce use of the citizenship process, which is admittedly a means to facilitate or speed up the integration of immigrants in their country of residence.

In 1996 a Department for National Integration and Immigration was created within the Ministry of Home Affairs, which showed the authorities' intention to regulate migratory flows. The intention was reinforced in the Refoundation (the economic and social development program of the *Front populaire ivoirien* ([FPI], the People's Front), which came into office in October 2000), when Law 2002–03 was passed (on January 2, 2002) relating to the identification of persons and foreigners' visits in Côte d'Ivoire, a National Identification Office was created, and an operation was launched in June 2002 for new national identity cards and residents' permits to be delivered.[4] The operation was slowed down due to the armed conflict that broke out during the night of September 18, 2002. There are not enough data to assess the effect of such policy change on immigration. However, the effect is probably small, considering that the measures taken were only aimed at distinguishing among individuals, as in the developed countries. The operation is expected to make it possible for large numbers of people to obtain identity cards, though they had never had any because the registry offices system was inadequate.

The annual number of persons who applied for and were granted citizenship rose from 15,146 in 1988 to 88,714 in 1998, which implies an average annual increasing rate of 5.3% during the period between the two censuses. If such a rate were to continue, the number of people obtaining citizenship would increase twofold within 13 years. Based on the data from the latest population census, an estimated 1.3 million people are likely to apply for citizenship, in conformity with the Côte d'Ivoire Nationality Code (Law 72–852, dated December 21, 1972).

However, appropriate communication strategies should be implemented so that large numbers of that population, which is mostly illiterate, actually apply for citizenship (Anoh *et al*., 2002).

Though economic integration of foreigners is a reality in Cote d'Ivoire, political and cultural integration sets limits that remain difficult to overcome. The more obvious illustrations include withdrawal of voting rights for foreigners, after the death of President Felix Houphouet Boigny, and strict enforcement of regulations concerning access to public offices and posts. In principle, "political rights are exclusively reserved for nationals. This entails that foreigners cannot exercise any political rights in Côte d'Ivoire, nor participate in political elections, either as electors or candidates" (Aggrey, 1999). It should also be mentioned that in traditional societies in Côte d'Ivoire, foreigners have no right to leadership. It seems this is also true in the native countries of most of the immigrants in Côte d'Ivoire. This shows that in fact the turnaround in the policy is no more than reverting to standards that were prevalent in West African traditional societies and are still enforced today.

The third component of the migration policy includes the measures in international cooperation

[4] Foreign residents' documents include (1) a free movement permit for nationals of the ECOWAS, when their visit does not exceed 3 months and the person does not have a passport; (2) a resident's permit for visits over 3 months; (3) a provisional resident's permit or a refugee's card for asylum seekers (Article 9, Law 2002–03 dated January 2, 2002, on the identification of persons and foreigners' staying conditions).

taken at bilateral or regional level with a view to improving control of migratory flows and promoting development in the emigration countries. Efforts have been made in that respect. Among other things, an agreement was signed in 1979 concerning free movement of goods and people and, more recently, in May 2000, the Conference of Presidents decided to launch the ECOWAS passport.

The controversy over Ivorianness broke out in the mid-1990s, as a consequence of the rivalry between Prime Minister Allassane Dramane Ouattara and the constitutional heir apparent, Henri Konan Bédié. The notion of Ivorianness has been termed xenophobic and is loaded with high emotional charge, which accounts for the many various constructions around it. As a matter of fact, the notion implies national preference, but also breaking away from President Houphouet Boigny's ultraliberal immigration policy; however, those who invented it did not mean it to become an element of division. According to ethnologist Niangoran Bouah,

> [it is] all the social and historical, geographical and linguistic data through which an individual can be defined as a citizen of Côte d'Ivoire. Individuals who claim that they are citizens of Côte d'Ivoire are supposed to have Côte d'Ivoire for their country, to be born of parents who belong to one of the native ethnic groups of Côte d'Ivoire. (Jarret and Mahieu, 2002)

In a speech at the PDCI-RDA's 10th Ordinary Congress, on October 28–31, 1996 in Abidjan, President Henri Konan Bédié declared:

> Whatever our ethnic group, religion, region or race, Ivorianness is meant for all of us, even for foreigners, as long as they take up the national culture for their own, as many natives of Côte d'Ivoire do once they have settled for good in their country of residence abroad, for example in France. In other words, Ivorianness, as the distinctive culture of the natives of Côte d'Ivoire, of all of them, without any exclusion whatsoever, is the opposite altogether of tribalism, ethnocentrism or racism. (quoted by Adé Adiaffi, 1998)

To those who promoted it, Ivorianness thus means assertion of a cultural specificity. The notion later took on a political connotation, based on the fact that certain foreigners were blamed for making the most of their privileges as citizens of Cote d'Ivoire, while keeping their initial nationality. However things stand, the fact is that the concept has been used in a way that went far beyond the process of cultural and political integration of foreigners in Côte d'Ivoire, which it was supposed to respect.

4. Internal Migration

Population policies are also concerned with internal migration. Some of the objectives have begun to be implemented as part of the population redistribution policy that has been going on since the 1980s. As a result of the objective of reducing regional disparities, vast land development and regional development programs were implemented as part of development planning. Three examples can be mentioned: the Bandama Valley Development project, the Southwest Region Development project, and sectoral development projects such as those on palm, rubber, or coconut trees in the forest-covered South, or those on cotton and sugarcane in the savannah regions in the North. The aim of those programs is to slow down urban drift and redirect the migration flows from the forest-covered South to the North and Center of the country. Underlying the policy are decentralization and such measures as may encourage young people to work in agriculture (Brou and Charbit, 1994). The rotating national celebrations, which implied substantial investments, have contributed to the promotion of the inland towns by opening them up and modernizing them. This, however, did not hamper rapid expansion of the city of Abidjan, which benefited from substantial investment, as the main target of the hypermodernistic urban project of the authorities (Le Bris, 1998). Nevertheless, decentralization has been selected among the major components of Refoundation, which is likely to reinforce such population redistribution policy.

As an illustration, we now take a closer look at the Reproductive Health and Family Planning Program, the component that is most advanced in the implementation of Côte d'Ivoire's National Population Policy.

IV. THE CONSTITUTION OF THE REPRODUCTIVE HEALTH SUPPLY AND THE EVOLUTION OF FAMILY PLANNING

1. From the Human Resources Development Policy Statement to the Expansion of Family Planning Services

According to the Human Resources Development Policy Statement issued in June 1991, the public sector was to get more involved in family planning activities. A national seminar on family planning was held at Grand Bassam in April 1991, when the bases for a coherent framework for family planning activities in Côte d'Ivoire were first laid, and three major policy objectives were selected:

Promote maternal and child health through prevention of risky pregnancy and birth spacing

Prevent sterility by means of detection and treatment of sexually transmittable diseases (STDs) and infertility
Contribute to economic and social development through birth regulation (Bamsié and Kouamé, 1992)

Out of that basis a first pilot family planning program was initiated in 1992 in the département of Aboisso, in the Southwest, with support form UNFPA. The evolution, which had been initiated earlier in the decade, was accelerated when the program of action of the International Conference on Population and Development (Cairo, September 5–16, 1994) was adopted, and ratified by Côte d'Ivoire. A sign of the evolution was that a national symposium on reproductive health was held in June 1996, during which all stakeholders could agree over the components of the reproductive health concept and define the different segments of the target population. Four components or areas were thus established: mothers' and women's health, children's health, young people's health, and men's health. The method was in conformity with the widespread belief that the approach to reproductive health should take an individual's whole life cycle into account.

Each area has specific components. However, there are components that the different areas have in common: family planning, STDs and HIV/AIDS, and communication for behavior change program and project management. Moreover, related components were taken into account (education, nutrition, environment), which shows that social objectives were combined with purely health-related ones.

Besides the public sector, various private parties and multilateral and bilateral cooperation organizations were also mobilized. The predominant agent was AIBEF, which conducted its own activities at the same time as it provided technical assistance and supplies to some of the centers of the Ministry of Public Health (Bamsié and Kouamé, 1992). In 2002 it was working in 84 public sector service centers and 31 more in the private and semi-public sectors. Besides its Abidjan headquarters, it has seven branches inland (Bondoukou, Bouaké, Daloa, Man, San-Pedro, Korhogo, Boundiali), which are managed by regional volunteers. It also has nine clinics.

2. The 1999–2003 National Reproductive Health and Family Planning Program

Since 1999 Côte d'Ivoire has had a reproductive health and family planning policy document, in conformity with which the various service-providing organizations are supposed to work, whether they are public, semi-public, or private. The document is used as a reference, because in it are "defined the objectives, strategies and complementary activities to be developed in order to reach the objectives in the larger framework of the National Reproductive Health Policy" (République de Côte d'Ivoire, 1999, p. 8).

A National Program for the period 1999 to 2003 was first developed. The implementation of that policy was slow, due to instability in its institutional framework (Anoh *et al*., 2002).

The Reproductive Health Policy includes four programs: the National Program for the Control of AIDS, STDs, and Tuberculosis; the National Child Health Program; the Expanded Immunization Program; and the National Reproductive Health and Family Planning Program.

The latter is focused on reproductive health care; other important dimensions, such as reproductive rights or women's empowerment, are not taken into account. Nevertheless, these have not been overlooked. A National Committee on Violence against Women was created in 2000. The committee, in which several Ministries are members, is intended to assist the Minister for the Family, Women and Children's Affairs in defining the programs and actions to be conducted with a view to reducing social inequality. Although there are no specific programs, improvement has been observed in the areas of reproductive rights and women's empowerment (Table 116–4).

Thanks to the changes brought to matrimonial rules, which were adopted in 1964, and the broadened access to reversible contraceptive methods in 1982, the autonomy of women has been improved. However, the laws on abortion and sterilization have remained extremely restrictive, and conflicting views are now being expressed about relaxing them.

Besides, for 10 years now, all forms of exclusion against women have been legally banned, since Côte d'Ivoire has ratified the United Nations Convention on the elimination of discrimination. Certain reproductive rights are even mentioned in the new Constitution: the government must ensure that all citizens have equal access to health, education, culture, information, vocational training, and employment (Article 7); mutilations and all forms of degradation of human beings are prohibited and punished by the law (Article 3); and gender equality is affirmed (Article 17): "Every person has a right freely to choose his/her occupation or employment. Access to public or private positions is equal for all. Any discrimination in access to or exercise of a position, based on sex, political, religious or philosophical opinions, is prohibited."

The increasing popularity of women's NGOs in the last decade is likely to boost enforcement of reproductive rights and observance of women's new powers.

TABLE 116–4 Some of the Achievements in Terms of Reproductive Rights

Areas	Improvements
Matrimonial system	According to Law 64–375 concerning marriage, the legal minimum age at first marriage is 18 for women and 20 for men; the personal consent of each of the prospective spouses is mandatory, and polygamy and bride price are prohibited, 1964.
Access to contraception	Article 11 in Law 81–640, dated July 31, 1981, authorizes free access to the more common contraceptive methods.
Access to sterilization	Article 343 in the 1982 Penal Code stipulates that whoever is guilty of sterilization shall be sentenced to death.[a]
Access to abortion	Abortion is regulated by Articles 366–369 in the 1982 Penal Code. Therapeutic abortion is permitted, provided that it is meant to save the life of a pregnant woman. In all other circumstances, abortion is a crime, and so are advertising and sales of abortive substances.
Sexual violence	Rape is punished under Article 354 in the 1982 Penal Code. Sexual mutilation, as a form of violence against women, is punished under Law 98–757, dated December 23, 1998. Sexual harassment, forced labor and early or forced union are punished under Law 98–756, dated November 23, 1998.
Exclusion	The United Nations Convention on the Elimination of all kinds of Discrimination Against Women (CEDAW) was ratified in 1995.

[a] However, in the new Constitution, which was adopted by referendum in 2000, the death penalty was abolished.

Such efforts to actualize legislation are crucial in countries such as Côte d'Ivoire, where, as in many other developing countries, there is a wide gap between the law and actual practices.

The objectives of the National Reproductive Health and Family Planning Program are liable to some criticism. A major one is that not all of those objectives are realistic (Table 116–5). For example, a 25% reduction in the maternal mortality rate meant that the rate should drop down to 450 for every 100,000 live births in 2003. This would have been possible only if an extremely strong program had been enforced. Another example is that the desired 20% reduction in the rate of infertility prevalence among women seemed exceedingly ambitious, considering the poor health infrastructure available and the behavior of the populations.

According to a recent study conducted in the region of Aboisso (Anoh, 2001), few women go to health centers for infertility reasons, and the few cases identified were referred to health centers that have a gynecologist, particularly to the regional hospital of Aboisso. The fact that few people go to a health center does not mean that few people are affected with this problem; the fact is that infertility is negatively perceived by society and that those persons who have this problem prefer to go to traditional healers for help. Moreover, the notion of a "sexuality-related mortality rate" (Executive Direction/ National Reproductive Health and Family Planning Program, 1999) is not usually used, and measures of it are not available often enough for any precise value of the desired reduction to be given. Those remarks show that the process of objectives development was not built on any in-depth knowledge of the facts on which the program was to focus.

The strategy for the implementation of the program was based on three lines for action: increased availability and accessibility of services, stimulated demand, and improved quality of services.

As far as availability is concerned, in 2002 the program was managing 250 reproductive health/ family planning centers, which meant a coverage rate of 33%. Pilot programs were implemented in the health districts of Bondoukou, Bouaflé, and Tanda. A Reproductive Health in the Armed Forces project is being developed, and awareness campaigns for youth are being organized by the Network for the Promotion of Sexual Health among youth and adolescents (known as RIPS/AJ).

Accessibility of services is increased by means of integration of activities into primary health care; the activities are served by the strategy of the minimum package of activities (MPA). A study of the situation in the districts of Bondoukou and Bouaflé has shown that

> the MPA as defined by the Ministry of Health is not available in the health centers, nor it is open on all working days. Geographical accessibility is limited and infrastructure is insufficient. Reproductive health services are not always available, due to products stock shortages; except for the poor quality of reception, the patients were satisfied with the quality of the care they received. (Ministry for Planning and Development, 2002)

The government's commitment to reproductive health and family planning has been proved through

TABLE 116–5 Specific Impact and Result Objectives for the 1999–2003 NRHFPP (from January 1999 to late 2003)

Components	Impact objectives	Result objectives
Mothers' and women's health	Reduce the maternal mortality rate by 25%	Ensure that 60% of pregnant women go through at least three prenatal medical visits Raise the rate of deliveries attended by qualified health workers from the present 45% to 55% Ensure that 30% of women who have just had a child go for a postnatal visit Raise the present 5.7% contraceptive prevalence among women of procreating age up to 14% Bring the present 50% proportion of men in favor of family planning practice up to 80%
	Reduce the sexuality-related maternal mortality rate by 50%	Contribute to ensuring that 50% of women who go to a health center for asexually transmitted infection (STI) are examined and treated appropriately Contribute to keeping down HIV prevalence among women at the present rate of 12% Reduce the prevalence of infertility among women by 20%
Adolescents' and young people's health	Reduce the sexuality-related mortality rate among adolescents and young people by 15% Reduce the sexuality-related morbidity rate among adolescents and young people by 30%	Bring the rate of early pregnancy among girls aged under 18 from 37% down to 20% Reduce the rate of induced abortion among adolescent girls and young women by 25% Contribute to ensuring that 25% of young people who come to a health center for an STI are examined and treated appropriately Reduce the practice of female genital mutilation among adolescent girls and young women by 10%
Men's health	Reduce the sexuality-related mortality rate among men by 10%	Bring the proportion of men who know about the different means of preventing STI/AIDS up to 95% Bring the proportion of men who use condoms for sexual intercourse with occasional partners up to 25%
	Reduce the sexuality-related morbidity rate among men by 40%	Increase the rate of treatment of infertility cases among men by 25% Contribute to ensure that 25% of the men who come to a health center for an STI are examined and treated appropriately

Source: EDNRHFPP, 1999.

changed parameters concerning family planning supply and use of contraceptive methods.

3. The Evolution of Contraceptive Practice and Choice of Methods

People must know about the methods and how to have access to them before they can use them. It is therefore necessary to deal with such knowledge before assessing the progress and evolution of choices.

a. Contraceptive Knowledge

Even though the level of contraceptive knowledge is usually higher among nonmarried people, it is the knowledge among married individuals that matters most as far as fertility control is concerned. Three categories of methods can be distinguished with respect to how familiar they are to the population.

The first category, that of the best known methods, includes male condoms, the pill, injections, and periodic abstinence. Already in 1994 male condoms were well known (to 61% of married women and 82% of married men), but by 1998–1999 virtually everybody knew about them (83% and 93% respectively). Such a result is directly related to the social marketing strategy implemented by Population Services International (PSI) and Economie et Développement (ECODEV) as part of the program on AIDS control. PSI is currently developing such social marketing activities in collaboration with a recently created NGO, the Côte d'Ivoire Agency for Social Marketing (AIMAS). The pill, injections, and periodic abstinence were also fairly well known in 1994, and knowledge of those methods has expanded since then: the proportion of women who have heard about them rose from 58% to 74% (pill), from 53% to 67% (injections), and from 40% to 58% (abstinence) between 1994 and 1998–1999. Among men, the proportion rose from 56% to 66%, from 52% to 61%, and from 61% to 62% respectively.

The second category, that of not so well known methods, includes female sterilization and male withdrawal. From 1994 to 1998–1999 the proportion of those who had heard about female sterilization rose from 37% to 44% among women; the proportion of men who knew about withdrawal rose from 46% to 53%.

The third category includes those methods that are definitely less familiar: intrauterine devices (IUDs), traditional methods, and, still less familiar, vaginal methods, male sterilization, and Norplant. In 1998–1999 the proportion of people who knew about those methods ranged from 4% (Norplant) to 34% (vaginal methods) among married women, and from 6% (Norplant) to 34% (IUDs) among men.

The fact that some of the methods are persistently unfamiliar is partly due to the policy itself. There are gaps where such methods as sterilization are restricted by law, or others, such as IUDs and Norplant, require qualified staff to prescribe them. Moreover, such methods as withdrawal or traditional methods are not included in the range of methods suggested in the program.

b. The Progress of Contraceptive Practice

The use of contraceptive methods has been progressing at a variable pace since the early 1980s. Among married women, the proportion of women who were using a modern method was under 1% in 1980–1981, but rose to 4.3% in 1994 and 7.3% in 1998–1999 (Table 116–6). Since the early 1980s, contraceptive prevalence among married women has then risen by 6.8% (in absolute terms), which means an average annual growth rate of 16%.

Considering all women, regardless of their marital status, the growth pace has been only slightly faster: 16.8% compared with 16.1% per year over the period from 1980–1981 to 1994, and 12.8% compared with 12.5% per year during the more recent period from 1994 to 1998–1999. However, there is one similarity: in both cases, the progress in modern methods use has been slower since 1994 than it used to be. The average annual growth rate dropped from 18.1% between 1980–1981 and 1994 to 12.8% between 1994 and 1998–1999 for all women, and from 17.3% down to 12.5% respectively among married women.

Observations made at Aboisso, the seat for a pilot family planning program, show that the level of contraceptive methods use there has risen faster than at the national level. For all women at Aboisso, the proportion of those who use any method rose from 19.1% in 1993 to 27.5% in 1999—an average annual growth rate of 6.3%—whereas the proportion of those who were using a modern method rose from 6.8% to 14.7%—an average growth rate of 13.7%. At the national level, the average annual growth rate for the recent period 1994 to 1998–1999 was only 5.2% for all methods and 12.8% for modern ones. However, the

TABLE 116–6 The Progress of the Level of Contraceptive Methods Use since the Early 1980s in Côte d'Ivoire (in %)

Contraceptive methods	All women			Married women		
	CIFS 1980–1981	DHS 1994	DHS 1998–1999	CIFS 1980–1981	DHS 1994	DHS 1998–1999
Côte-d'Ivoire						
All modern methods	0.6	5.7	9.8	0.5	4.3	7.3
Average annual growth for the preceding period	—	18.1	12.8	—	17.3	12.5
All methods	1.5	16.5	20.7	2.9	11.4	15.0
Average annual growth for the preceding period	—	19.4	5.2	—	10.6	6.3
Aboisso						
Modern method	—	6.8	14.7	—	4.1	13.1
Average annual growth 1993–1999	—	—	13.7	—	—	21.4
Any method	—	19.1	27.5	—	14.0	21.5
Average annual growth 1993–1999	—	—	6.3	—	—	7.4

The annual growth rates were calculated over exactly 13.5 years for the period 1980–1981 to 1994, and 4.5 years for the period from 1994 to 1998–1999, i.e., a total 18 years for the periods under study.

CIFS = Côte d'Ivoire Fertility Survey; DHS = Demographic and Health Survey.

Sources: Direction de la statistique, 1984, p. 285; N'Cho *et al.*, 1995, p. 56; National Statistics Institute (INS) and Macro International, 2001, p. 9; National Research Council, 1993, p. 31; Koffi *et al.*, 2000; Anoh, 2001.

period under observation was slightly shorter, and the real difference was consequently rather small.

The situation shows a larger contrast when only married women are taken into consideration. At Aboisso, the average annual growth rate was 7.4% for all methods and 21.4% for modern methods during the period 1993–1999; at the national level, those rates were 6.3% and 12.5%, respectively, between the two Demographic and Health Surveys.

The various strata of the population of Côte d'Ivoire are at different levels of contraceptive practice. For all of them, the use of contraceptive methods, particularly modern ones, has grown more frequent. The progress, however, much depends on social and demographical characteristics. For example, the use of methods has increased faster among women with a primary education level than among others—from 20% up to 25% (+5%), as against 5% up to 8% (+3%) among illiterate women and from 39% to 40% (2%) among women of secondary education level and beyond; increased use of methods is to be observed among all the different generations (Anoh, 2001). Those differences between social classes suggest that the progress of contraceptive practice does not solely result from the effect of family planning programs.

c. Changes about Method Choice

Contraceptive practice has progressed, but it is still limited and mostly relies on natural and traditional methods. Over 50% of users use natural and traditional methods (Figure 116–10). However, the proportion of conceptive users of modern methods grew between 1994 and 1998–1999. The modern methods most used are the pill and condoms, and to a much lesser degree, injections. The latter involve high leaving rates and do not contribute much to increasing contraceptive prevalence, when compared with other methods such as IUDs or sterilization, for which a small number of new cases a year is enough to induce significant boost in the prevalence. The level of contraceptive prevalence results both from the history of

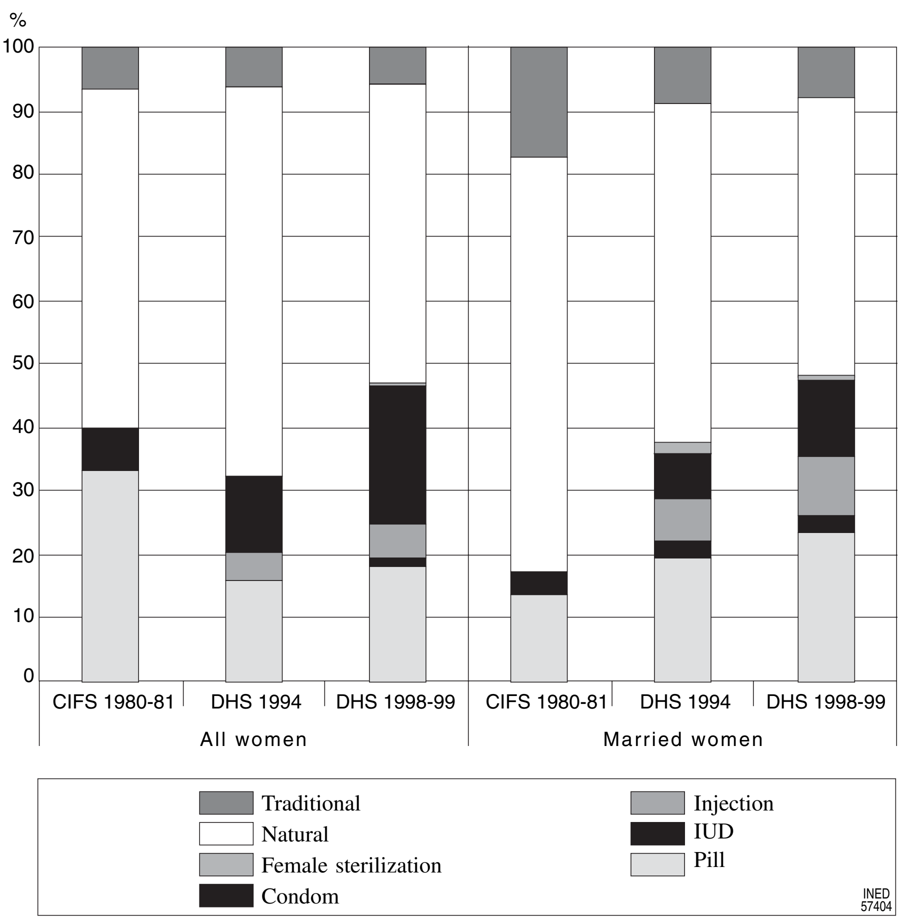


FIGURE 116–10 The evolution of the kinds of methods used. (Sources: Direction de la Statistique, 1984; N'Cho *et al*., 1995; INS and ORC Macro, 2001.)

how the methods were taken up and of their continued use, which in turn depends on the nature of the methods being used (Ross *et al*., 1999).

Suppliers for pills are very diversified in Côte d'Ivoire. In particular, small shopkeepers are playing an increasingly significant part (3% and 12% of women in 1994 and in 1998–1999 respectively), which shows that not all the pills in use require a medical prescription. This raises the question of compatibility between a woman's health condition and the method she is using, and also of the possible side effects that should be monitored.

Condoms, the second most used method, could be bought mostly from the nonmedical private sector in 1998–1999, as a direct consequence of the social marketing organized by PSI and ECODEV.

CONCLUSION

Côte d'Ivoire became aware at very early stages of the relations between population and development. The trend of its economic policy and certain ideological motivations, however, caused the country to seek to increase its population rather than slow down its growth. Such a populationist option was kept up until the late 1980s, implying an ultraliberal immigration policy and pronatalist measures. The change of position and the adoption of a National Population Policy in the 1990s are part of a regional dynamic that affected most African countries, particularly the Francophone ones, which had been reluctant up until then. The turnaround is the direct consequence of the economic constraints that resulted from the economic crisis and recession, on the one hand, and from the political constraints related to the structural adjustment programs, on the other hand.

The armed conflict that broke out in Côte d'Ivoire on September 19, 2002 has slowed down the implementation of the first program of action on population. However, certain priority actions and investments scheduled in the program have begun to be implemented because they had been defined out of operations already ongoing as part of programs that government departments were implementing in different sectors, due to the lack of centralized planning. In that respect, the reproductive health and family planning program, which can be considered part of the Control of Natural Growth subprogram, is somewhat more advanced than others.

This shows that, although the approach is one in which population programs are supposed to be integrated into social and economic development strategies, the population policy in Côte d'Ivoire, as is often the case, may very well focus on the family planning component at the expense of those components based on the economic, social, and environmental context, which in fact require much more effort in terms of investment.

Lastly, it must be emphasized how important political stability and sustained support from development partners are for population programs to be properly implemented. In addition, the process of implementing the programs must be based on in-depth analyses of demographic behaviors, so that the objectives that are set are not unrealistic.

Acknowledgment

This chapter was translated from the French by Mireille Rabenoro.

References

Adé Adiaffi Jean-Marie, 1998. Ivoirité, ethnicité et transethnicité, *Racines, revue culturelle africaine*, vol. 3(1), p. 77–91.

Aggrey Albert, 1999. *Quel est le statut des étrangers en Côte-d'Ivoire? Guides droits des personnes et de la famille.* Abidjan, Juris-Editions, 40 p.

Ahonzo Etienne, Barrère Bernard, and Kopylov Pierre, 1984. *Population de la Côte-d'Ivoire. Analyse des données démographiques disponibles.* Abidjan (Côte-d'Ivoire), Ministère de l'économie et des finances, Direction de la statistique, 324 p.

Aka-Anghui Hortense, 1989. *Discours d'ouverture.* Abidjan, Ministère de la promotion de la femme and Société Intermanagement, 4 p. (Speech made at the Approaches to Birth Spacing Programs seminar, Abidjan, January 23, 1989).

Antoine Philippe, 1985. Évolution de la mortalité et de la fécondité en Côte-d'Ivoire, *Espace, populations et société*, no. 3, p. 567–580.

Anoh Amoakon, 2001. *Émergence de la planification familiale en Côte d'Ivoire.* Nanterre, Université de Paris X-Nanterre, 439 + 36 p. (Doctoral thesis in Demography).

Anoh Amoakon, Fassassi Raïmi, and Vimard Patrice, 2002. *Politique de population et planification familiale en Côte d'Ivoire.* Paris, Centre français sur la population et le développement, 45 p. (Les Dossiers du CEPED, no. 68).

Any-Gbayere Sahou, 1992. *L'éducation en Côte-d'Ivoire: des programmes d'ajustement structurel au programme de valorisation des ressources humaines.* Abidjan, GIDIS/ORSTOM Petit-Bassam, 14 p. (Paper read at the round table on Crisis and Structural Adjustment, Bingerville, November 30–December 2, 1992).

Bamsié Roger and Kouamé Koffi, 1992. *La planification familiale en Côte-d'Ivoire: bilan et perspectives.* Abidjan, ENSEA. (Paper presented at the national seminar on Depographic Prospects in Côte-d'Ivoire 1988–2028, Abidjan, November 18–20, 1992).

Barbieri Magali and Vallin Jacques, 1996. Les conséquences de la crise économique africaine sur l'évolution de la mortalité, *in:* Jean Coussy and Jacques Vallin (eds.), *Crise et population en Afrique. Crises économiques, programmes d'ajustement et dynamiques démographiques*, p. 319–343. Paris, CEPED, 580 p. (Les Études du CEPED, no. 13).

Brou Kouadio and Charbit Yves, 1994. La politique migratoire de la Côte-d'Ivoire, *Revue européenne des migrations internationales*, vol. 10(3), p. 33–59.

BUNAP, 2002. *Programme national d'actions en matière de population 2002–2006.* Abidjan, The Prime Minister's office, Ministère de la planification et du développement, BUNAP, 221 p.

Caldwell John, 1993. Les tendances démographiques et leurs causes, le cas de l'Afrique est-il différent? *in:* Jean-Claude Chasteland, Jacques Véron, and Magali Barbieri (eds.), *Politiques de développement et croissance démographique rapide en Afrique*, p. 21–45. Paris, INED and PUF, 314 p. (Congrès et Colloques, no. 13).

Chamie Joseph, 1988. Les positions et politiques gouvernementales en matière de fécondité et de planification familiale, *in:* Dominique Tabutin (ed.), *Population et sociétés en Afrique au sud du Sahara*, p. 167–189. Paris, L'Harmattan, 551 p.

Curtis Sian L. and Neitzel Katherine, 1996. *Contraceptive knowledge, use and source.* Calverton (Maryland), Macro International Inc., 92 p. (DHS Comparative Studies, no. 19).

CVRH, 1991. *Programme de Valorisation des ressources Humaines, Politique de Population pour le Développement.* Abidjan, République de Côte-d'Ivoire, Comité de Valorisation des Ressources Humaines, 17 p. + annexe.

DEPNSR/PF, 1999. *Programme national de la santé de la reproduction / planification familiale Côte-d'Ivoire 1999–2003.* Abidjan, République de Côte-d'Ivoire, Ministère de la santé publique, Direction exécutive du programme national de la santé de la reproduction et de la planification familiale, 108 p.

Direction de la Statistique, 1984. *Enquête ivoirienne sur la fécondité 1980–81, Rapport principal, vol. 2: Tableau statistique.* Abidjan, Ministère de l'économie et des finances, 434 p.

DCGT, 1991. *Recensement général de la population et de l'habitat année 1988, exploitations et comparaisons 1965/75/88.* Abidjan, Ministère du plan et de la comptabilité nationale, Direction de la statistique, Direction et Contrôle des grands travaux, 155 p.

Doudou Seck, 2001. *Urbanisation, Recensement Général de la population et de l'habitation de 1998 (RGPH-98), vol. 4: Analyse des résultats, Tome 13.* Abidjan (Côte-d'Ivoire), Institut national de la statistique, 62 p.

Dozon Jean-Pierre, 2000. La Côte-d'Ivoire entre démocratie, nationalisme et ethnonationalisme, *Politique africaine*, no. 78, p. 45–64.

Dozon Jean-Pierre, 1997. L'étranger et l'allochtone en Côte-d'Ivoire, *in:* Bernard Constamin and Harris Memel-Fotê (eds.), *Le modèle ivoirien en questions*, p. 779–798. Paris, Karthala and ORSTOM, 802 p.

Dureau Françoise, 1985. *Migration et urbanisation. Le cas de la Côte-d'Ivoire.* Paris, Université Paris I, Institut de démographie de Paris, 567 + 84 p. (Doctoral thesis in Demography).

Garenne Michel, Madison Maria, Tarantola Daniel, Zanou Benjamin, and Dogoré Raymond, 1995. *Conséquences démographiques du sida en Abidjan, 1986–1992.* Paris, Centre français sur la population et le développement, 194 p. (Les Études du CEPED, no. 10).

INS and Macro International, 2001. *Enquête démographique et de santé, Côte-d'Ivoire 1998–99.* Abidjan (Côte d'Ivoire) and Calverton (Maryland), Institut National de la Statistique and Macro International Inc, 197 + 98 p.

Jarret Marie-France and Mahieu François-Régis, 2002. *La Côte-d'Ivoire. De la déstabilisation à la refondation.* Paris, L'Harmattan, 144 p.

Koffi N'Guessan, Sika Lazare and Anoh Amoakon, 2000. Une expérience pilote de planification familiale dans la région d'Aboisso (Côte d'Ivoire), *in:* Patrice Vimard and Benjamin Zanou (eds.), *Politiques démographiques et transition de la fécondité en Afrique*, p. 111–130. Paris, L'Harmattan, 297 p. (Populations).

Le Bris Émile, 1998. Urbanisation et politiques urbaines dans les pays en développement, *in:* Michel Loriaux (ed.), *Populations et développements : une approche globale et systémique*, p. 297–356. Louvain-la-Neuve and Paris, Academia-Bruylant and L'Harmattan, 582 p. (Population et développement, no. 5).

Leridon Henri and Toulemon Laurent, 2003. La régulation des naissances se généralise, *in:* Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux*, p. 477–495. Paris, INED, 768 p. (Les Cahiers de l'INED, no. 149, 2nd ed., rev. and exp.).

Locoh Thérèse and Makdessi Yara, 2000. Politique de population et baisse de la fécondité en Afrique sub-saharienne, *in:* Patrice Vimard and Benjamin Zanou (eds.), *Politiques démographiques et transition de la fécondité en Afrique*, p. 263–296. Paris, L'Harmattan, 297 p. (Populations).

Loriaux Michel, 1990. *La fin des certitudes. Modèles démo-économiques à réformer, systèmes informationnels à inventer.* Louvain-la-Neuve, CIDEP, 53 p. (Cahiers du CIDEP, no. 4).

MDPDI, 1997. *Déclaration de politique nationale de population.* Abidjan, République de Côte-d'Ivoire, Ministère délégué auprès du premier ministre chargé du plan et du développement industriel, 40 p.

Makinwa Adebusoye Paulina, 1993. Content of population policies: African countries, *in*: United Nations, *Population Policies and Programmes*, Proceedings of United Nations expert group meeting on Population Policies and Programmes, Cairo, Egypt, April 12–16, 1992, p. 42–52. New York, United Nations, 268 p.

Ministère du Plan, 1980. *Plan quinquennal de développement économique, social et culturel 1976–1980*, vol. 3. Abidjan, République de Côte d'Ivoire, Ministère du plan.

MLS, 2003. *Sida Magazine, the news bulletin published by the Ministry for AIDS Control*, no. 1, p. 1–22.

MPDI, 1997.—*Déclaration de politique nationale de population.* Abidjan, République de Côte-d'Ivoire, Ministère du plan et du développement industriel, 40 p.

N'cho Sombo, Kouassi Lucien, Koffi Kouame Albert, Schoemaker Juan, Barrère Monique, Barrère Bernard, and Poukouta Prosper, 1995. *Demographic and Health Survey, Côte d'Ivoire 1994.* Abidjan (Côte d'Ivoire) and Calverton (Maryland), Institut national de la statistique and Macro International Inc, 294 + 97 p.

Ncho Sombo, Yapo Eugène and Zanou Benjamin, 2001. *Migrations et Société en Côte-d'Ivoire.* Abidjan (Côte-d'Ivoire), Institut National de la Statistique, 215 p.

NRC (National Research Council), 1993. *Factors affecting contraceptive use in sub-Saharan Africa.* Washington D.C., National Academy Press, 252 p.

Raulin Henri, 1968. Le droit des personnes et de la famille en Côte-d'Ivoire, *in:* M Baye Kéba (ed.), *Le droit de la famille en Afrique noire et à Madagascar*, p. 221–241. Paris, Maisonneuve, 295 p.

Ross John, Stover John, and Willard Amy, 1999. *Profiles for family planning and reproductive health programs, 116 countries.* Glastonbury (Connecticut), The Futures Group International, 87 + 81 p.

Roussel Louis, 1967. *Côte-d'Ivoire 1965. Population: études régionales 1962–1965, synthèse.* Abidjan, Ministère du Plan, 192 p.

Sala-Diakanda Daniel Mpembele, 2000. Les positions des gouvernements africains vis-à-vis des politiques en matière de fécondité, *in:* Patrice Vimard and Benjamin Zanou (eds.), *Politiques démographiques et transition de la fécondité en Afrique*, p. 13–45. Paris, L'Harmattan, 297 p. (Populations).

Sembajwe Israël, 2000. Planification familiale et facteurs de la pratique contraceptive au Kenya et en Ouganda, *in:* Patrice Vimard and Benjamin Zanou (eds.), *Politiques démographiques et transition de la fécondité en Afrique*, p. 73–89. Paris, L'Harmattan, 297 p. (Populations).

TAPINOS Georges P., 2002. La population étrangère, *in:* George P. Tapinos, Philippe Hugon, and Patrice Vimard (eds.), *La Côte-d'Ivoire à l'aube du XXI^e^ siècle*, p. 387–408. Paris, Karthala, 498 p.

UNITED NATIONS, 2001. *World population prospect. The 2000 revision*, vol. 1. New York, United Nations, Department for Economic and Social Affairs, Population Division, 745 p. (ST/ESA/Ser.A/198).

UNITED NATIONS, 2003. *World population prospect. The 2002 revision*, 2 vol. New York, United Nations, Department for Economic and Social Affairs, Population Division, 781 p., 939 p., and CD-Rom. (ST/ESA/Ser.A/225).

VIMARD Patrice, ANOH Amoakon, and GUILLAUME Agnès, 2002. La croissance démographique, *in:* George P. Tapinos, Philippe Hugon, and Patrice Vimard (eds.), *La Côte-d'Ivoire à l'aube du XXI^e^ siècle*, p. 15–87. Paris, Karthala, 498 p.

ZANOU Benjamin, 2001. *Migrations. Recensement Général de la population et de l'habitation de 1998 (RGPH-98), vol. 4: Analyse des résultats, tome 2.* Abidjan (Côte-d'Ivoire), Institut national de la statistique, 122 p.

ZOBO Paulin N., 2003. Immigration, *Fraternité Matin*, 13–14 Dec. 2003.

APPENDIX 116–1 Population Densities by Department

Department	1965	1975	1988	1998	Department	1965	1975	1988	1998
Abidjan	65	149	285	1474.5	Séguéla	6	7	11	15.3
Aboisso	12	23	36	53.9	Sinfra	16	42	74	103
Adzopé	19	33	45	54	Vavoua	4	11	27	46.4
Agboville	19	39	53	61.5	Zuénoula	25	35	41	50.8
Divo	11	25	49	65.9	Odienné	5	6	8	10.6
Grand Lahou	5	9	22	37.7	Boundiali	11	14	17	20.9
Tiassalé	8	23	38	53	Ferkessédougou	4	5	10	14.4
Abengourou	11	24	42	56	Korhogo	19	22	31	36.2
Agnibilekrou	14	33	52	62.3	Tingrela	17	18	25	28.2
Bondoukou	8	13	17	29.2	Bangolo	21	28	39	61.8
Bouna	2	4	6	8	Biankouma	12	16	20	24.3
Tanda	12	27	32	35.2	Danané	27	38	49	67.9
Béoumi	32	37	34	43.1	Duékoué	10	18	33	68.5
Bongouanou	21	40	41	43.8	Guiglo	6	7	15	25
Bouaflé	15	25	39	57.2	Man	35	46	59	75.7
Bouaké	42	56	94	153.6	Touba	8	9	13	16.1
Dabakala	5	6	8	10.4	San Pédro	1	7	24	57.8
Daoukro	14	26	23	29.2	Sassandra	6	9	20	40.9
Dimbokro	25	33	29	52.7	Soubré	4	9	37	73.2
Katiola	7	9	14	17.6	Tabou	3	4	10	26.3
M'bahiakro	11	18	19	21.3	Alépé				31.7
Sakassou	26	31	32	45.8	Dabou				87.3
Toumodi	20	29	31	39.1	Jacqueville				77.8
Yamoussoukro	22	31	47	79.8	Tiébissou				29.8
Issia	14	26	53	75.8	Bocanda				25.7
Daloa	18	37	67	97.8	Adiaké				56.4
Gagnoa	26	38	60	79.3	Grand_Bassam				142.7
Lakota	22	27	42	53.9	Toulepleu				61
Mankono	7	7	11	18.9	Côte-d'Ivoire	13	21	33	47.8
Oumé	15	37	60	77.9					

Sources: For 1965, 1975, and 1988, DCGTX, 1991: for 1998, INS.

APPENDIX 116–2 The Evolution of Fertility Rates by Age in Côte d'Ivoire since 1958 (per 1,000)

Age groups	1958[a]	1962–1964[a]	MRS 1978–1979[a]	CIFS 1980–1981[b]	RGPH 1988[b]	DHS 1994[b]	DHS 1998–1999[c]
15–19	217	192	224	207	193	151	129
20–24	319	289	296	310	289	245	223
25–29	289	264	278	298	289	240	225
30–34	209	226	217	245	252	227	198
35–39	167	158	158	189	193	172	151
40–44	66	102	86	126	98	82	80
45–49	36	44	35	57	38	26	37
TFR	6.52	6.38	6.54	7.16	6.76	5.72	5.21

Sources: [a] Roussel, 1967, qtd in Antoine, 1985.
[b] N'cho *et al*., 1995.
[c] INS and Macro Int., 2001.

CHAPTER

117

Population Policy in the United States

CARL HAUB

Population Reference Bureau, Washington, D.C., United States

Officially, the United States has no population policy. There have never been numerical goals or targets for a particular population size or rate of growth. Indeed, the country has always considered itself to be a "country of immigrants," a distinction that sometimes ignores that the populations of other developed countries of the "New World," such as Australia, Canada, and New Zealand are themselves largely products of immigrants from distant lands. There has never been any sort of demographic target or goal regarding birthrates, although concern has been raised in the past when birthrates declined to low levels, as happened in the 1930s. In responding to the United Nations Population Division's biennial survey on population policies, the United States stated that population growth was satisfactory and that no intervention to alter the rate was needed (United Nations, 2002).

As of this writing, U.S. population increases at about 3 million per year, 1.6 million through natural increase of births minus deaths and 1.2 million through net immigration, both legal and illegal (see Figure 117–1). This is the largest amount of growth in absolute terms in the country's history and is nearly the equivalent of adding the population of Sweden every 36 months.

The closest the country has come to considering some sort of national population policy was in the early 1970s at the urging of Republican President Richard Nixon with the establishment of the Commission on Population and the American Future, also known as the Rockefeller Commission. The Commission consisted of a large number of sociologists, demographers, and environmentalists, giving it by far the largest assembly of expertise ever focused on the population issue. At the time, population had begun an acceleration spurred by the period of high fertility during the baby boom and by rising immigration. Over 50 million had been added to U.S. population since 1950, partly due to the effects of the landmark immigration act of 1965. The sweeping 1965 immigration reform opened U.S. borders to immigration from the nontraditional regions of Africa, Asia, and Latin America, a part of Democratic President Lyndon Johnson's Great Society, itself an outgrowth of the civil rights movement of the 1960s. Prior to the 1960s, immigration had been at very low levels throughout the depression years of the 1930s through the 1950s as immigration from traditional European countries declined. Ironically, the Commission also began its work on the eve of the steepest decline in fertility in U.S. history, that of the 1970s.

The Commission's findings, presented in 1974, were a controversial departure from the "growth is good" mentality, to say the least. "The health of our country does not depend on population growth, nor does the vitality of business, nor the welfare of the average person" (Negative Population Growth website, *Footnotes*, March 1998, accessed May 14, 2003). The Commission, whose report filled six volumes, called for the decriminalization of abortion, removal of any remaining restrictions on contraceptives (the pill had been approved by the Food and Drug Administration in 1960, but individual states had their own laws), the wide availability of sex education, freezing immigration at 400,000 per year and stopping illegal

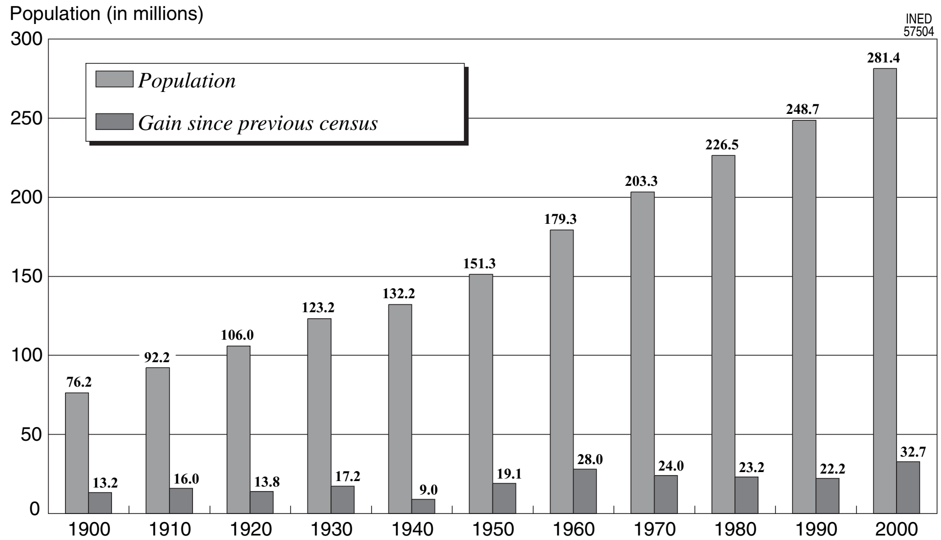


FIGURE 117–1 Population of the United States at each census and gain since previous census, 1900–2000. (Source: USCB, 2002.)

immigration, as well as support for bans on any type of sex discrimination. President Nixon ultimately rejected the report, possibly due to both the controversial nature of the recommendations and the distraction of the Watergate scandal, which led to Nixon's resignation in June 1974.

If population growth was one of the Commission's fears, then many of their concerns have come to pass. Population, numbering 203 million at the 1970 census, has grown to 291 million in 2003 and will reach the 300-million mark in 2006. Still, many of the Commission's wishes were realized without Presidential approval. In 1973, the historic *Roe v. Wade* Supreme Court decision legalized abortion on demand, families voluntarily chose the two-child family it sought (the total fertility rate [TFR] actually dropped to 1.7 in 1976, oddly the year of the country's Bicentennial celebrations), and there are now no legal restrictions on contraceptives whatsoever. The sole area where the Commission's goal has not been realized is immigration.

Although debate over population growth has waned since the 1970s, individual groups do keep the notion alive. Negative Population Growth in Alexandria, Virginia, Californians for Population Stabilization in Santa Barbara, California, the Center for Immigration Studies in Washington, DC, Population Environment Balance in Washington, DC, and Population Connection (formerly Zero Population Growth) in Washington, DC are just a few examples of organizations that lobby or otherwise promote slower population growth, primarily by advocating reduced immigration.

But, in some ways, all countries have a population policy, be it clearly stated or inferred indirectly. Still, by not having an expressly stated policy, the United States has implicitly stated that it has no position on population size. Significantly, while the 2000 census enumerated nearly 7 million more residents than anticipated, increasing the annual estimated rate of population growth, there has been no particular public concern or media attention generated as a result. The single area in which the United States does impose numerical limits to a component of demographic change is that of immigration law, yet even here there is no direct connection with any sort of population size or growth rate target.

For much of its history, immigration was not only uncontrolled but was openly encouraged. During the 19th century, immigrants were welcomed to help develop and occupy the largely empty continent, stretching over 4,000 kilometers from the Atlantic to the Pacific. Viewing the United States today, it is easy to ignore just how new the country really is. During the 1990s, many cities and towns in the central United States, or the "Midwest," celebrated just the 100th anniversary of their founding. New residents arrived by the millions and all were welcome, although arrivals from new areas, such as Eastern or Southern Europe were initially subjected to various forms of discrimination. But virtually all of these immigrants were from Europe, the traditional sending countries of new Americans.

Responding to the UN policy survey, the United States stated that immigration levels were satisfactory and that current immigration of permanent migrants,

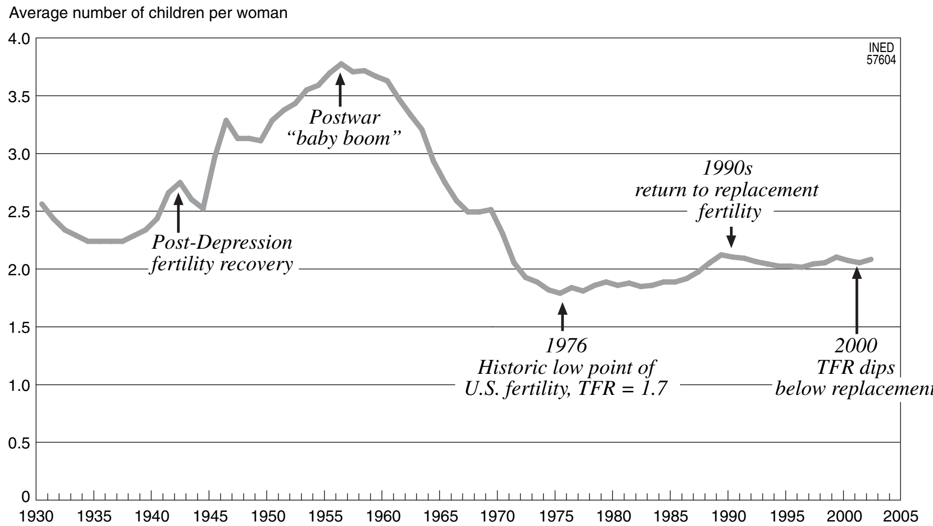


FIGURE 117–2 U.S. total fertility rate, 1931–2003. (Source: U.S. National Center for Health Statistics, Vital Statistics of the United States, Natality, various issues; National Vital Statistics Reports, various issues.)

migrant workers, and dependents should be maintained as well as the policy toward integration on non-nationals. U.S. policies on growth and migration can be compared to that of France, insofar as that country also said that current population growth was satisfactory, but indicated that the level of immigration was too high and that the official policy was to lower it. Similar sentiments were expressed by Germany and the United Kingdom, sharp contrasts to U.S. stated contentment with the level.

Fluctuations in fertility rates (see Figure 117–2) appear to be far more a product of the economic background than of any government action to provide support for families The United States today has the highest TFR of any industrialized country (*The Economist*, 2002). In the 20th century, there were two periods of significant decline in the total fertility rate, both of which reflected similar trends in Europe. The TFR declined to 2.1 during the global Depression years of the 1930s, recovering as World War II approached and then reached its modern zenith in 1957 at 3.8, the peak of the well-known and much-celebrated baby boom. A second plunge took place in the early 1970s, perhaps partly attributable to rising feminism, but quickly fueled by the oil crises of the same decade and resulting inflation as well as high unemployment. Fertility responded to a favorable economic outlook in the late 1980s, rising to the replacement level by 1990, a level generally maintained until the end of the decade. Following the larger than expected 2000 census count, there has been a downward revision in the U.S. TFR for both 2000 and 2001 of approximately 0.1, to 2.056 and 2.034, respectively. Replying to the United Nations, the U.S. considered that their fertility level was satisfactory and in no need of intervention, but labeled the adolescent fertility rate a "major concern" with policies and programs to influence it. While fertility has certainly fluctuated, it is quite difficult to identify any specific public policy that has had a measurable effect on the TFR.

In terms of mortality, the current level of expectation of life was reported as unacceptable to the United Nations. With respect to mortality, the United States does have an implicit policy to improve overall health conditions, although no specific targets for measures such as life expectancy are mentioned. This overall goal or policy can be contrasted with that France and Germany, who stated to the United Nations that current mortality levels were acceptable. The United States does have any number of programs to improve health conditions such as Medicare, Medicaid, and public health programs and clinics for low-income populations, which can certainly be seen as intended to improve medical care and overall health, ostensibly increasing life expectancy. Government programs as well as media attention focuses on racial differentials in rates such as infant mortality and the prevalence of specific diseases, but references to overall life expectancy are infrequent.

I. IMMIGRATION

Half of U.S. population growth now stems from immigration, despite the fact that the level of natural increase is quite high by industrialized country standards. Currently, there is little public debate on the level of immigration, although illegal immigration was widely discussed in the late 1980s and early 1990s. Since the terrorist attacks of September 11, 2001, public sentiment has changed insofar as steps to control the entrance of those who would do the country harm are generally accepted, albeit with some misgivings that such steps are now necessary. It is interesting to note that the issue of "urban sprawl," the rapid spread of suburbanization around cities into former farmland, receives much attention, but it is viewed more as an issue of population distribution policies, rather than as a result of the country's growth of 3 million per year.

As welcoming a country as the United States has been, its past immigration history has been filled with friction between more established immigrant groups and those arrivals from newer source countries. From the country's earliest days, there were concerns about immigrants from countries seen as nontraditional. Even Benjamin Franklin fussed about the arrival of Germans in Pennsylvania:

> Why should the Palatine boors be suffered to swarm into our settlements and, by herding together, establish their language and manners to the exclusion of ours? Why should Pennsylvania, founded by the English, become a colony of aliens, who will shortly be so numerous as to Germanize us, instead of our Anglifying them?' (Esmond, 1986; p. 83)

The "Know Nothing" political party worried about the arrival of Catholics from Ireland in the 1840s and were elected to Congress on sentiments of anti-immigration of such groups, but failed to sponsor any successful legislation.[1] Following the Civil War in the 1860s, some laws were passed prohibiting convicts, prostitutes, paupers, and "mental defectives." In 1882, the Chinese Exclusion Act was a landmark law aimed at barring a specific group. Many Chinese had been recruited to build the railroads across the West, but after the work was complete, fears of a flood of cheap Chinese labor prompted the ban, which continued in force as late as 1943. Subsequently, other laws tried to ensure that immigrants could read and write and were intended to limit immigration to countries of Northern and Western Europe, due to the perceived undesirable qualities of migrants from other areas. It was in 1921 that the first numerical limit was set on the number of immigrants, and, in 1924, a limit of 150,000 was imposed, with quotas favoring traditional countries. As a result, the vast majority of immigrants continued to be from Northern and Western Europe until well after World War II.

Following World War II, Presidents Truman and Kennedy favored eliminating national quotas, but it was not until the landmark amendments to the Immigration and Nationality Act, passed in 1965, that U.S. immigration was truly opened to all countries. The amendments of 1965 were, in all likelihood, the single most defining pieces of legislation in U.S. history, forever changing and defining the ethnic structure population of U.S population. After 1965, the countries of origin of immigrants shifted dramatically. Symbolically, the new legislation was signed by President Johnson at the base of the Statue of Liberty in New York harbor. During the 1950s, 53% of immigrants were from Europe; by the 1980s, only 17% were. The full impact of the 1965 law has been such that, by about the year 2050, the country will likely have no majority ethnic group at all. Such a development would most certainly have horrified Franklin, but seems to disturb most contemporary modern Americans little if at all.

United States' policy on immigration is not specified in any official statements but remains implicit in the laws that govern it. The U.S. population is composed of a large number of ethnic and national groups, each arriving in different waves over the course of history. The notion of a "country of immigrants" seems to be a well-accepted aspect of American social change, although such change is never without its frictions. Nonetheless, it is probably fair to say that significant restrictions, let alone a total ban on new immigrants would be seen by most as unthinkable. Concern over illegal immigrants surfaces from time to time, although even that may be viewed as beneficial in that many illegal immigrants take jobs that many Americans no longer want. During the 1980s, there was considerable debate over illegal immigration and the means that could be used to stop it. Most recently, concern has been universal over immigrants whose sole purpose is to do the country harm, with many feeling that some immigrant rights may be abused in the name of security. Beyond considerations of a certain perceived moral obligation toward immigrants, serious restrictions on immigration would cause a political firestorm, as individual national groups would raise a loud protest.

Nonetheless, it is a clear, if tacit, policy that immigration will continue although the United States reserves the right to regulate who and how many come to its shores, as does any sovereign nation. The policy is to favor immigrants who can benefit the country via needed occupational skills and special expertise, allow a generous program of family reunification, and accept a reasonable number of refugees and others seeking asylum.

II. CURRENT IMMIGRATION LAW

The laws defining who comes in and who does not are quite complex. Immigration is currently governed

[1] Much of this discussion is drawn from Martin and Midgely, 1999.

by the Immigration Act of 1990 (Public Law 101–649), although the law is often amended as Congress deems appropriate, typically to include new groups, such as special categories of refugees. This law, passed at the height of the debate over illegal immigrants, actually raised the amount of allowable immigration.

At present, over 30 million people are known to enter the United States each year. The vast majority of these are tourists or people on temporary business who leave in a short period of time. A restriction on those visitors is that they do not generally have the right to work, a right restricted to migrants who receive permanent resident visas or specific work-related visas under other provisions of law.

Current immigration law specifies two broad categories of immigration, permanent and temporary, but only the former are designated with the legal term of *immigrant*. Only immigrants receive permanent resident visas, popularly called "green cards," although the card itself is no longer green. When used in the discussion below, the word *immigrant* refers only to those granted lawful permanent residence and, subsequently, the right to apply for citizenship after 5 years of continuous residence. It is most important to take note that immigrants are not necessarily new residents in the demographic sense in that many have been living in the country for some time, possibly quite a long period. An example of the latter might be an employee of an international organization who reaches retirement age and only then applies for permanent residence.

III. IMMIGRANTS

The legal use of the term *immigrant* causes some confusion in interpretation of immigration data with some users naturally assuming that 1,064,318 immigrants physically arrived on U.S. shores for permanent residence in Fiscal Year (FY) 2001, as shown in the *2001 Statistical Yearbook of the Immigration and Naturalization Service*.[2] In FY 2001, there were 1,064,318 immigrants, but only 411,059 were new arrivals in that year; the balance were previous residents whose status was adjusted from temporary to permanent. A second caveat is the use of immigration data as a times series. Annual numbers fluctuate with the ability of the INS to process visa applications. In 1998, Congress transferred the processing of undocumented aliens from the State Department to the INS without increasing staff. As a result, processing fell behind. The high figure cited above of just over 1 million can be compared to the previous two fiscal years. In FY 1999, there were but 646,568 visas processed and, in FY 2000, 849,807. The apparent rise in numbers from FY 1999 to 2001 was a result of "catching up" the processing of forms. There is also a per-country limit on visa allocations for a particular year to 7% of the annual total.

An additional category of immigrants are the so-called diversity immigrants, a special group of 55,000 added under the 1990 law to encourage immigration adversely affected by the 1965 law. These are potential immigrants from "old-seed" sending countries before the 1965 law, chiefly those of European origins and especially those from Ireland. This addition was partly sponsored by Senator Edward Kennedy of Massachusetts.

IV. IMMIGRATION LIMITS PRESCRIBED BY LAW

As described in Table 117–1, different categories of immigrants are assigned numerical limits. There are two broad categories of immigrants, those granted permanent residence for employment purposes (employment-based preferences) and those admitted for the general purpose of family reunification. Employment-based preference immigrants are often those most often thought of as new immigrants to the United States in that they do not necessarily have any previous connection to the United States, but are admitted for needed job skills or may be investors who are likely to create new jobs. Their number is "capped" at 144,000 per year. Family-sponsored immigrants are relatives of legal residents who are admitted for family reunification. Their number is set at from 226,000 to 480,000, the exact limit being equal to the larger of 226,000 or a calculation of 480,000 minus the previous year's total of immediate relatives. There is also a per-country limit on visa allocations for a particular year to 7% of the annual total.

V. NONIMMIGRANTS

Temporary migrants, the vast bulk of whom are tourists and visitors for business purposes totaled 27.8 million in FY 2003, are actually referred to by law as "nonimmigrants," since they do not receive green cards for permanent residence. Significantly, this number was nearly 6 million less than in FY 2000, a

[2] The Immigration and Naturalization Service (INS), as such, ceased to exist with the establishment of the new Department of Homeland Security (DHS) in the wake of the terrorist attacks of September 11, 2001. A new agency, the U.S. Citizenship and Naturalization Services (UCIS), has been created as its replacement.

TABLE 117–1 Immigrant Categories Included in Worldwide Annual Limit Specified in Section 201 of the Immigration and Nationality Act: Fiscal Year 2003 Limits

Categories		
Total		**701,532**
Employment-based categories		**171,532**
First preference	Priority workers	49,058
Second preference	Professionals with advanced degrees or aliens of exceptional ability	49,058
Third preference	Skilled workers, professionals, needed unskilled workers	49,058
Fourth preference	Special immigrants (ministers, religious workers, etc.)	12,179
Fifth preference	Employment creation (“investors”)	12,179
Family-sponsored immigrants		**480,000**
Family-sponsored preference		**226,000**
First preference	Unmarried sons and daughters of U.S. citizens and their children	23,400
Second preference	Spouses, children, and unmarried sons and daughters of permanent resident aliens	114,200
Third preference	Married sons and daughters of U.S. citizens	23,400
Fourth preference	Brothers and sisters of U.S. citizens (at least 21 years of age)	65,000
Fifth preference	Immediate relatives of adult U.S. citizens (spouses, children, and parents) and children born abroad to alien residents	**Not limited**
Diversity immigrants		50,000

Note: The annual limits are adjusted based on visa usage in the previous year.
Source: UCIS, 2004.

result of the September 11 terrorist attacks. Beyond the 87% of this group arriving in FY 2001 who were simply tourists or businesspeople, 3.5 million were accommodated under a variety of programs, ranging from students to foreign ambassadors and their families. There is no ceiling on the total number of this group who may be admitted in any one year.

The largest group of nonimmigrants, outside of visitors for business or pleasure, are temporary workers and trainees who, along with their spouses and children totaled 726,613. This group contains two visa classifications that are numerically limited, H1B and H2B visas. H1B visas are issued to specialty workers who must have a university degree or the equivalent and a license to practice their profession, if one is required. The typical H1B recipient is, as the INS describes it, a 29-year-old Indian working in the computer field and earning $55,000 per year. The next largest group, 624,917, was composed of students.

Temporary workers who are employer-sponsored employees may remain in the country a maximum of 6 years (the H1-B program). They may, of course, apply for permanent residence at some point. In FY 2001, there were 331,206 such nonimmigrants, about half of whom were from India. The provisions of the various nonimmigrant programs specify persons with college degrees and specific training and are almost always in some type of professional occupation or needed occupation. Other categories include people performing services not available in the United States, athletes and entertainers, religious workers, and the spouses and children of such workers.

There are additional groups such as refugees and asylees, and special agricultural workers to fill labor force needs, particularly in the Southwest. An additional group comprises “diversity” immigrants, a special provision to reverse what was seen as disadvantages to immigrant countries from “traditional” sending countries, particularly Europe. This group numbered 42,015 and did receive green cards.

Are limits imposed on the annual number of immigrants? While there are numerical limits imposed on many categories, in actual practice, application of the law is rather complex.

VI. FERTILITY

It is probably true that no program of family support has ever been enacted in the United States with the intent of directly influencing the level of the TFR. Rather they have been designed to ensure that Americans can have the number of children they wish to have, more as a right than as a demographic goal. Increasingly, the provision of equal rights for women in employment has emerged as a primary concern.

The TFR in the United States stands at 2.0, a level that has remained relatively constant since about 1990. It is often thought that higher fertility among ethnic groups accounts for this comparatively high rate, given that Hispanic women, for example, have a TFR just below three children. But the TFR of “majority” women, defined as White non-Hispanic, stands at 1.9, well above most countries of Europe and Australia, Canada, and Japan. Why is American fertility so high? While it is not possible to quantify an answer, it is reasonable to speculate that U.S. society is simply more oriented to the expectation to have children, particularly in certain geographic areas. Church attendance is still relatively common and the expectation that one will raise a family is almost certainly more persistent. Or, perhaps such a perception is misplaced, that U.S. fertility in the 1990s is more a result of a robust, optimistic economy. Perhaps U.S. fertility is not truly "high," but only appears to be in comparison to very low European fertility. It is still true, however, that, in the 1995 National Survey of Family Growth, 14.8% of women interviewed said that their last birth was mistimed and 6.5% said that it was unwanted (NCHS, 1997).

While programs benefiting parents of children may have had an unplanned effect on fertility, the intent of those programs has always been to provide financial relief to families and to help support the two-earner family, thus facilitating careers for women. The latter has been a response to the rapid entry of women into the full-time career labor force as well as a genuine need for two earners as the cost of living rose substantially from the 1970s. High living costs can be illustrated by some examples. In the early 1970s, a single-family house in the Washington, DC area cost about $25,000 or less, but by even the 1980s, the cost had risen to $250,000 and continues to increase. Families with two full-time earners have also served to make the expense of the two-car family commonplace, particularly given the great geographic expansion of the suburbs and the lack of practical public transportation. Such economic factors could certainly be seen as having a depressing effect on fertility, as has arguably happened in Europe, but it is likely that support for families with children has played some role in maintain the two-child family, however inadvertent.

Along with Australia, the United States is alone among industrialized countries in that there is no government-funded program for maternity leave, placing the burden of providing such leave on employers. In many ways, this is in keeping with the general national attitude that individuals, and not the government, are responsible for arranging and financing their own affairs. It should be kept in mind that Americans do not look to the government as the ultimate form of socialized support and even find "government meddling" unattractive. At the same time, income taxes are considerably lower than the social taxes paid by many Europeans. Americans, much more so than in Europe, are expected to provide for themselves.

In the 1960s, pregnancy often meant that a working woman would leave her job to become a full-time mother, and an employer who granted any sort of maternity leave was rare indeed (USCB, 2001). Since that time, however, women have entered the world of work in increasing numbers. In 1970, 43.8% of women 16 years of age and over were in the labor force (79.7% of men) but, by 2001 the proportion had risen to 60.1% (74.4 of men). In 1978, the Pregnancy Discrimination Act was passed by Congress, forbidding discrimination in employment for pregnancy and childbirth. Along with this development, families were able to take an income tax credit for child care expenses from 1976. Subsequently, the ability to work flexible hours, child care centers at the workplace, and a general notion that working women must be accommodated became frequently debated national issues. These began to be offered by employers as incentives to obtain and keep trained workers.

For those parents who must pay out-of-pocket fees for daytime child care while the parents work, the income tax credit allowed by federal law amounts to a significant subsidy. Expenses up to $2,400 for one child or $4,800 for two or more children up to age 13 may be claimed, and 20% to 30% of the claimed amount may be taken as credit against the amount of income tax, not simply deducted from taxable income. The percentage that may be taken as a credit varies with income, with lower income parents eligible for the full 30% credit.

In 1993, the Family and Medical Leave Act mandated up to 12 weeks of unpaid leave for employees of companies of 50 or more employees and who worked at least 1,250 hours in the past year for their current employer. There is, however, no provision for *paid* leave in the federal law. Paid leave remains an option for an employer, although in some states, such as New Jersey, women receive 75% of their weekly salary up to a maximum of $364 (National Center for Policy Analysis, accessed April 4, 2003, from www.ncpa.org).

While the rise in maternity leave has certainly been dramatic, paid leave of some type is still enjoyed by barely half of all women (see Table 117–2). "Paid leave" also refers to a combination of different options that may or may not be available to a women, such as the use of accumulated sick and vacation time in addition to actual paid maternity leave that may or may not be available. Thus, maternity leave remains something of an ingenious assembly of available options to a family, not a matter of public policy. Overall, about three-fourths of all women who worked during pregnancy in 1991–1994 returned to work in the first year, 42% returning within 3 months.

TABLE 117–2 Leave Arrangements Used by Women Who Worked during Pregnancy: 1961–1965 and 1991–1995 (Percentage)

Type of leave	Year of first birth	
	1961–1965	1991–1995
Total	100.0	100.0
Quit job	62.8	26.9
Paid leave[a]	16.0	42.7
Unpaid leave	14.1	40.3
Disability leave	(—)	11.2
Dismissed from job	4.2	5.0

(—) not applicable.
[a] includes paid maternity, use of sick and vacation leave, and other paid leave.
Source: USCB, 2001.

While any effects that maternity/paternity leave may have had on keeping the United States TFR at a modestly high level cannot be directly quantified, there has been a clear effort on the part of government and the private sector to accommodate working mothers and fathers. Day care centers at the workplace have become commonplace, the use of sick leave to care for a sick child is widely accepted, and some mothers even bring their young infants to work. In addition to programs and benefits, it is almost certainly a major factor that attitudes toward work and motherhood have undergone a sea change in the United States. Certainly, this has arisen as a result of the movement for equal rights for women, but it can also be seen as pragmatic in that employers are less likely to lose trained, valuable employees.

In 1996, the first significant change in public assistance to poverty populations, popularly referred to as "welfare reform," was passed in the form of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA). This act placed new limits on the length of time public assistance benefits could be received and encouraged disadvantaged mothers with children to seek gainful employment. Due to the latter provision, the act is popularly referred to as "workfare." Prior to the passage of PRWORA, there had been much public debate on welfare payments discouraging women from remaining on the rolls at public expense. PRWORA, then, is a landmark experiment and a major restructuring of public assistance programs and the philosophy behind them, a shift from long-term support to a program that helps poor mothers adjust to the need for self-support. And, it can certainly be argued that welfare reform was a matter of public policy, designed to encourage those on welfare to seek gainful employment as well as discouraging out-of-wedlock childbearing supported by public funds.

Will PRWORA affect fertility levels among poor women? This question takes on added significance when we consider that 22% of the 35.5 million mothers in 1996 received some form of public assistance (USCB, 2002a).[3] Overall, mothers who participated in one or more public assistance programs had given birth to 2.361 children in their lifetime, compared to 2.071 who did receive benefits. Of the mothers who received benefits, the first birth of 58.5% was outside of marriage, compared to 24.7% of mothers not receiving benefits. Clearly, the question is: does welfare encourage childbearing among poor women, particularly out of wedlock births?

During the 1990s, the only significant change in the TFR by ethnic group was that of Black women, who recorded nearly a 12% decrease. The decrease in the birthrate for the 15–19 age group was particularly sharp, leading to speculation that welfare reform was the cause. However, it is evident that fertility decline was under way well before the 1996 act became law. While it is possible that pre-passage publicity may have had an effect, the timing of age-specific fertility decline in the 1990s suggests that there were a multiplicity of causes, such as antipregnancy publicity campaigns in the schools.

Looking at birthrates for unmarried young women (see Figure 117–3), a primary target group for welfare reform, the effect of PRWORA may be somewhat more evident. The decline in birthrates for unmarried Black women appears to have accelerated in the years leading up to passage and then continued in a clear downward path. For Hispanic women, the increasing trend reversed itself, leveling off in the second half of the 1990s. For White women, there is no apparent trend, although poor mothers receiving public assistance are a small minority of White women overall.

Support for parents is available from a variety of other sources, notably financial aid for university attendance. In the United States, even state-sponsored colleges and universities are not free, with tuition, room, and board expenses falling to the parents. Many parents aspire for their children to attend college, which has become a virtual necessity for a well-paying position. The cost of even the less expensive state schools easily runs $10,000 per year and that of the highly desired private institutions at least two times that figure. Saving for tuition is something young parents strive to do early on, and parents may even wish to space births so as not to have two or more

[3] The cohort for these data are women ages 15–44 who had had at least one birth.

TABLE 117–3 Age-specific Birthrates and Total Fertility Rate, White non-Hispanic, Black non-Hispanic, and Hispanic Women, 1990, 1996, and 2003

Categories	Rate by age group (per 1,000)						TFR
	15–19	20–24	25–29	30–34	35–39	40–44	
White non-Hispanic							
1990	42.5	97.5	115.3	79.4	30.0	4.7	1.851
1996	37.6	90.1	104.9	82.8	33.9	6.2	1.781
2003	27.5	83.5	110.9	97.7	43.2	8.1	1.858
Percentage change 1990–2003	–35.3	–14.4	–3.8	23.0	44.0	72.3	0.4
Black non-Hispanic							
1990	116.2	165.1	118.4	70.2	28.7	5.6	2.548
1996	91.9	137.0	96.7	63.2	29.1	6.2	2.140
2003	64.8	128.2	102.2	67.5	33.5	7.7	2.030
Percentage change 1990–2003	–44.2	–22.4	–13.7	–3.8	16.7	37.5	–20.3
Hispanic							
1990	100.3	181.0	153.0	98.3	45.3	10.9	2.944
1996	94.6	170.2	140.7	91.3	43.9	10.7	2.757
2003	82.2	163.4	144.4	102.0	50.8	12.2	2.785
Percentage change 1990–2003	–18.0	–9.7	–5.6	3.8	12.1	11.9	–5.4

TFR = total fertility rate. Sources: NCHS, 2003, 2004.

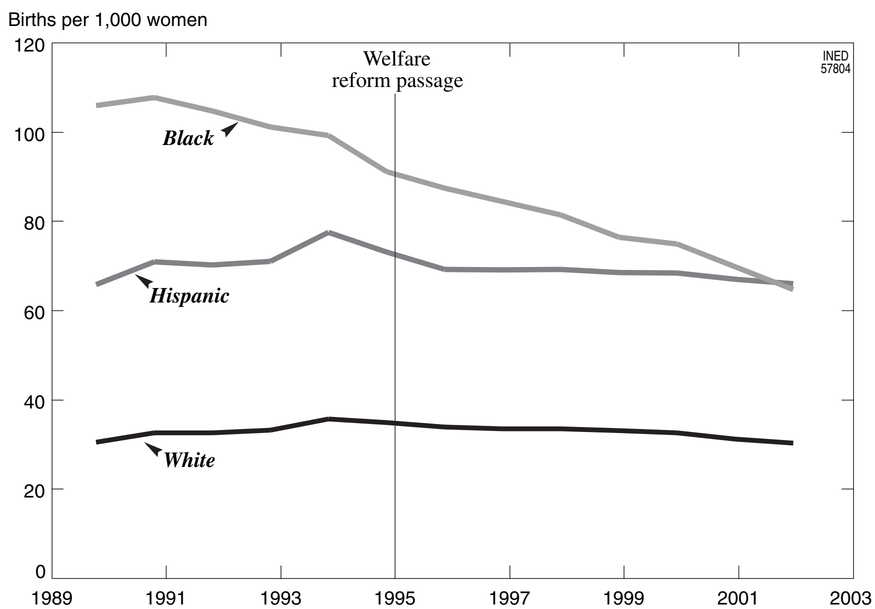


FIGURE 117–3 Birthrates for unmarried women ages 15–19, by ethnic group, 1990–2002. (Source: Martin *et al.*, 2002.)

children in college at the same time, depending on how strong their motivation for college educations might be. In part, parents are assisted in their dilemma by a variety of state and federal student aid programs (for which a full financial statement must be completed) and scholarships that provide some relief, particularly for low-income families, although the chief burden is borne by the parents.

VII. DISTRIBUTION

While there may be no national policy on population growth, many local governments have taken steps to limit growth, particularly in the vicinity of cities, a development now receiving wide media publicity as urban sprawl. Urban sprawl refers to the rapid expansion of suburban settlement circling cities into areas once considered rural. A typical scenario is a county near a city that has historically been lightly populated and with an attractive, bucolic setting. At some point, housing developers begin construction of new units, which often attract younger families due to the lower cost compared to expensive houses "in town." The nature of the rural setting is itself an attraction, but newcomers soon find themselves joined by increasing numbers of even newer arrivals and soon the bucolic nature of the area disappears. A result is heavier traffic and a common expression is that one left the city but "the city followed me here." A major contributor to sprawl is the fact that many jobs leave the city and relocate to suburban areas, allowing commuters to live further and further from the city itself.

Generally, sprawl itself is seen as a result of *distribution*, not overall national population growth itself. This tends to ignore the fact that the country grows by 3 million per year and that those new residents must clearly live and work somewhere. As a result, population growth tends to be concentrated in two general areas: the large suburban areas of larger, established cities, such as Atlanta and Washington, DC, and smaller cities and their suburbs which have gained attractiveness in recent decades, such as Charlotte, North Carolina and Boise, Idaho.

The geographic spread of settlement has often been exacerbated by what has been called "the American dream," a single-family house on a tract of land ranging from one-quarter acre to an acre in size. Local jurisdictions often attempt to limit growth with moratoriums on sewer hookups, limits on how many houses may be built on lots, and the requirement that developers provide funds for needed roads before housing permits are issued. There are often, however, conflicting goals between those who wish to limit growth and county officials who may wish to attract more tax dollars.

VIII. MORTALITY

Particularly since the 1960s, there has been a concerted effort on the part of local, state, and the federal governments to become involved in providing greater access to full health care to those unable to afford it. In fact, few Americans could afford to pay for health care. Employed persons nearly always benefit from some type of group insurance plan, but those without regular employment and the elderly were typically left to their own resources, which often meant little or no health care at all. In a very real sense, public policy has become interventionist in health care, following a largely laissez-faire policy for most of the country's history. This was in part due to the rising cost of health care which, in turn, resulted from an explosion of diagnostic methods and the expensive equipment they require. Health care shifted from home visits by a local doctor to a vast array of costly in-patient and out-patient services. In short, the nature of health care itself underwent a sea change after the 1950s.

1. Medicare and Medicaid

Medicare, established in 1965 along with Medicaid as Title XIX of the Social Security Act, is a government program that covers Americans age 65 and over as well as some persons below age 65 with disabilities. The law was a reaction to the fact that rapidly rising medical costs were increasing much faster than the cost of living. About 1960, it became evident that private insurance companies could no longer provide coverage to senior citizens and pensioners at premiums they would be able to pay. It was estimated that only one in four had adequate medical coverage (considered to be insurance that would pay 75% of a hospital bill).

Premiums for Medicare are normally paid throughout one's working life through payroll deduction. Upon reaching age 65, full hospital coverage is provided under Part A. Medical insurance, covering doctor's fees, out-patient services, and the like are included in Part B, but the beneficiary must pay a monthly fee of $58.70, an amount that can also rise. Individual states vary in how procedures are judged to be medically necessary, what procedures are covered, and the amount they pay, but there is nearly always a deductible or percentage that must be paid by the patient. In 1997, Congress passed the Medicare+Choice legislation, a move championed by those who wish to "privatize" government programs, transferring more cost to the consumer. This program gives more options in that patients are free to choose private Health Maintenance Organizations (HMOs), often at a higher cost. About 15% of Americans have opted for the new plan.

Medicare does not cover prescription medicines, an issue of debate at present. The elderly are particularly hard hit, given their disproportionate use of such medications. Television ads promoting prescription insurance show Americans driving to Canada to buy drugs at lower prices. A proposal submitted in 2003 by President Bush would extend drug insurance to participants in new Medicare, and discount cards as well as an unspecified cap on annual drug expense. For many, particularly Democratic lawmakers who generally favor the expansion of government programs, the proposal does not go far enough and it appears to be

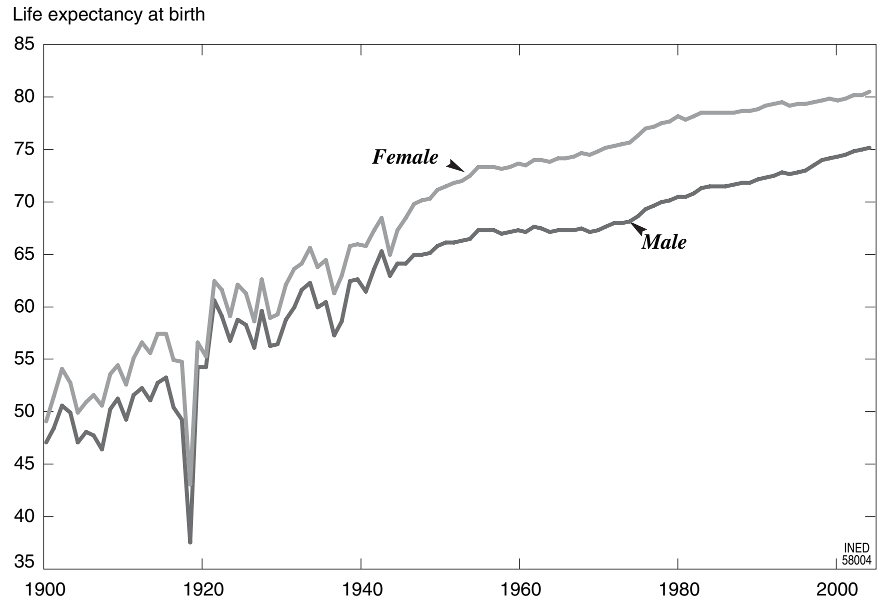


FIGURE 117–4 Life expectancy at birth, United States, 1900–2003. (Source: USCB, 1976; NCHS, 2004a, 2004b).

crafted to satisfy both sides of Congress. It is likely some version of the proposal will become law given that drug expense today has risen in much the same way hospital expense did in the 1960s.

Given that Medicare is the largest government health program in the nation's history, it is logical to inquire if there was an improvement in a possible outcome, such as life expectancy (see Figure 117–4). In 1954, life expectancy at birth was 66.7 and 72.8 for men and women, respectively. In 1965, the year Medicare was passed, life expectancy for men was unchanged and that of women had increased only one year, to 73.8. Twenty years after passage, life expectancy for men had jumped to 71.1 and, for women, to 78.2. The gain has, however, been much less since 1985, rising to 74.4 for men and 79.8 for women. A 1997 paper by Morris Davis at the University of Pennsylvania estimated that the elimination of Medicare and near elimination of Medicaid would reduce the life expectancy of 70-year-olds by .37 years but not affect that of wealthier people. It may be that these life expectancy figures indicate that any improvement that Medicare might have contributed may have run its course. It will be interesting to see if a comprehensive prescription drug plan causes a second burst.

Medicaid generally provides health care benefits for low-income persons. Originally, eligibility was extended only to those who were eligible for Aid to families with Dependent Children (AFDC) (recently modified as Temporary Assistance for Needy Families in line with the current policy to limit the time spent receiving welfare payments). The program is a federal-state matching program, with each state setting an income limit on who can receive benefits, with great variation among the states. The primary beneficiaries are poor families headed by a single parent. Since its inception, Medicaid has been expanded to cover uninsured pregnant women, low-income families not receiving welfare, the elderly and disabled, and other groups. Medicaid also funds full-time nursing home care after determining that the individual is unable to pay, as most are, greatly relieving the expense and time involvement of caring for the infirm elderly on their children.

Any effect of these programs, which primarily benefit the poor, can be evaluated by examining life expectancy at birth trends among minorities. In 1955, life expectancy for the non-White population was 61.4 for males and 66.1 for females. Ten years later, that of males actually declined slightly to 61.1, while that of females rose to 67.4. By 1996, life expectancy for non-White males has risen to 68.9 and, for females, to 76.1.

2. Health Insurance

Given the high cost of medical care and prescriptions and the lack of a system of socialized medicine, Americans depend heavily on health insurance to defray costs, particularly catastrophic costs. In 2001, 41.2 million Americans, 14.6% of the population, were without health insurance of any kind, including government health insurance for the elderly and those below the poverty line (USCB, 2002b). Minorities are the hardest hit by the lack of coverage, with 19.0% of Blacks and 33.2% of Hispanics without coverage.

The largest source of coverage is that provided by an employer through a group plan in which the cost is shared by the employer and the employee. In 2001, this represented 62.6% of the population, although there was a 1% drop from the previous year. Government insurance, through Medicare, Medicaid, and military health care, covered an additional 25.3%.

CONCLUSION

While the United States has no population policy to directly affect the size of the population or influence the growth rate, its many laws and programs do affect its demographic parameters. Legislation controlling immigration is hardly unique to the United States, and it can probably be said that the United States is among the more liberal countries in the area, particularly since most recent legislation has only served to increase the amount. Certain changes in visa processing policies since September 11, 2001 are reducing the number of visas issued to a degree, but, again, the goal here is not demographic but security related.

Policies toward the family have given assistance to families with children, but the motivation has been one of equality for working and poor women, certainly not demographic. The many provisions for health care for the poor and elderly have had an undisputed effect on life expectancy and perhaps that was an unspoken goal. But, in terms of numbers, the U.S. attitude has always been "come what may."

References

Davis Morris A., 1997. *Medicare, Medicaid, Medigap, and the Life Expectancy of the Elderly*. Philadelphia, PA, University of Pennsylvania, 1997.

Martin Philip and Midgely Elizabeth, 1999. Immigration to the United States, *Population Bulletin (Washington, DC)*, vol. 54(2), p. 1–44.

Martin Joyce A., Hamilton Brady E., Ventura Stephanie J., Menacker Fay, and Park Melissa M., 2002. Births: final data for 2000, *National Vital Statistics Reports*, vol. 50(5), p. 1–104.

NCHS (National Center for Health Statistics), 1967. *Vital statistics of the United States, 1965, Vol. II: Mortality, Part A.* Washington, DC, National Center for Health Statistics, 416 p.

NCHS, 1997. *1995 National Survey of Family Growth*. Hyattsville, MD, National Center for Health Statistics.

NCHS, 2003. Births: final data for 2002. *National Vital Statistics Reports*, vol. 52(13), p. 43–45.

NCHS, 2003b. Deaths: preliminary data for 2002. *National Vital Statistics Reports*, vol. 52(13), p. 25–26.

NCHS, 2004. Births: preliminary data for 2003. *National Vital Statistics Reports*, vol. 52(13), p. 8–9

NCHS, 2004. Deaths: preliminary data for 2002, *National Vital Statistics Report*, vol. 52(13), p. 25–26.

The Economist, 2002, August 24. A tale of two bellies, *The Economist*, vol. 364(8287), p. 11.

United Nations, 2002. *National population policies 2001.* New York, United Nations/Nations unies, Population Division/Division de la population, 425 p.

United Nations, 2004. *World urbanisation prospect. The 2003 revision.* New York, United Nations, Department for Economic and Social Affairs, Population Division, 186 p. (ESA/P/WP.190).

USCB, 1976. *Historical statistics of the United States, colonial times to 1970*, 2 vol. Washington, DC, United States Census Bureau, 1200 p.

USCB, 2001. *Maternity leave and employment patterns 1961–1995.* Washington, DC, United States Census Bureau. (Series P-70–79, December).

USCB, 2002a. *Fertility and program participation in the United States: 1996.* Washington DC, United States Census Bureau, 85 p. (SeriesP70–82, May).

USCB, 2002b. *Health insurance coverage: 2001.* Washington, DC, United States Census Bureau, 85 p. (Series P60–220, September).

USCB, 2002c. *Statistical abstract of the United States.* Washington, DC, United States Census Bureau, 1000 p.

USCB, 2003. *Statistical abstract of the United States.* Washington, DC, United States Census Bureau, 1030 p.

U.S. Citizenship and Naturalization Services (UCIS), 2004. *2003 Yearbook of Immigration Statistics*. Washington, DC, Immigration and Naturalization Service, 260 p.

Wright Esmond, 1986. *Franklin of Philadelphia.* Cambridge, MA, Harvard University Press, 404 p.

CHAPTER

118

Population Policy in Russia

SERGUEY IVANOV,* ANATOLY VICHNEVSKY, AND SERGEI ZAKHAROV

*Population Division, United Nations, New York, United States

Center of Demography and Human Ecology, Russian Academy of Sciences, Moscow, Russia

During the Soviet era (1917–1991), the State was much more actively involved in all spheres of society in Russia than in countries with more liberal regimes. Demographic processes were no exception and were strongly influenced by the State. This is no way implies, however, that the State aimed to meet specifically demographic goals. Quite the reverse—it was unable to appreciate demographic reality as a phenomenon in itself, relatively independent of the life of Soviet society, and any influence it had had on the issue was incidental, the consequence of government decisions concerning economic and social policy, the regime's foundations, or even external affairs, and the like.

At the start of the 20th century, Russia was still a high-fertility country. Its population, in expansion within an immense territory, was growing relatively rapidly, offering no serious reason to intervene in demographic processes. The very high level of mortality was the only population problem the authorities were aware of at that time; consequently, shortly after the revolution, great efforts were made to reduce mortality through the development of a real public health policy whose main objectives and results will be mentioned here.

The internal migration created by the country's commitment to forced industrialization became the second demographic problem soon facing the regime. In very little time, tens of millions of peasants left the country for cities or went to exploit the new regions, and the State quickly recognized the need to control the massive migration flows it had itself instigated. As early as this era, it is possible to talk of setting up a migration policy, although the expression was never officially used.

Half a century later, at the start of the 1970s, Russians became aware of a third major population issue, low fertility—a topic that was completely new at that time. The question aroused animated discussions about the relevance of pronatalist family policy, and led to the creation of the first measures to this end.

Finally, in the 1990s, after the break-up of the USSR and with the natural population balance in Russia showing signs of growing increasingly negative, Russian society had to face yet another new problem, that of external migration: strong immigration from other republics of the ex-USSR at first, followed by a substantial wave of emigration. Policies to meet this new challenge needed to be developed.

These four main population policy themes do not completely exhaust the topic. Soviet and post-Soviet policies also affected many other aspects of demographic behavior. Apart from these four broad areas, however, policies were rarely elaborated according to strictly demographic objectives. While not responding to any demographic logic, for example, the *iron curtain* ideology and policy, characterizing almost the entire Soviet era, involved severe restrictions on external migration. The same is true for many of the often-contradictory family policy measures. Loosening or tightening laws on marriage, divorce and the family, permitting or forbidding abortion and such like, these waves undoubtedly influenced the course of demographic processes; however, they were more a reflection of fluctuations in internal politics, one minute permitting greater freedom and the next tightening the

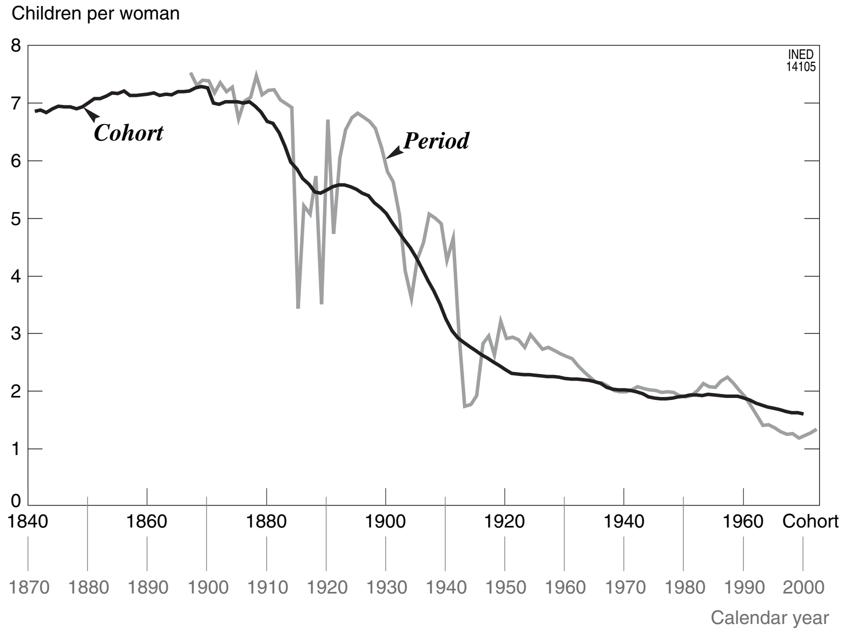


FIGURE 118–1 Completed fertility for the 1841–1970 cohorts and total fertility rates from 1897–2002 (children per woman). (Source: Vishnevsky, forthcoming.)

screws, and vice versa, than of the desire to meet demographic challenges, however real.

I. FAMILY POLICY

Today, Russia is among the countries most affected by low fertility. During the first half of the 20th century, the rapid and often brutal modernization of an essentially peasant society transformed reproductive behavior patterns. Starting much later than in Western nations, and from a higher level, Russia's fertility transition was also much faster. Completed fertility declined regularly from one cohort to the next, despite the many catastrophes (World Wars I and II, the 1918–1920 civil war, massive repressions during the 1930s, famines in 1932–1933 and 1947), which triggered sudden interruptions in fertility timing and strong swings in total fertility rates around the general downward trend; these fluctuations were mainly ironed out during the second half of the 20th century (Figure 118–1).

The completed fertility of all generations born after 1931 is below replacement level; the total fertility rate (TFR) crossed this threshold in 1966. In the two following decades, the TFR remained close to the threshold, oscillating between 1.9 and 2.0 children per woman; then, a brief and gentle rebound in the mid-1980s was followed by an abrupt drop during the 1990s. At 1.2 to 1.3 children per woman, Russia now has one of the lowest TFRs in the world. Despite similarities with Western countries in the level and tendencies of these summary fertility indicators, reproductive behavior in Russia had retained several distinctive features, an earlier entry into reproductive life among them. A change in fertility timing during the 1990s sent the TFR (period indicator) below completed fertility (cohort indicator).

1. From the October Revolution of 1917 to World War II

Within communist doctrine, family policy was viewed in turn as a means to found the new social order, a lever for economic projects, an instrument to reduce social inequalities or, finally, as a tool to reach demographic objectives, depending on the different perspectives taken. Despite these various viewpoints, family policy always occurred within a general context of massive invasion of privacy with the aim of creating an original socialist lifestyle conforming to the superior interests of society as a whole.

According to communist doctrine, society should help women exercise their right to motherhood. Among the first measures taken after the 1917 Revolution, the Soviet regime instituted maternity leave,[1]

[1] Decrees and the Labor Code of 1918–1922 instituted prenatal and postnatal leave at full pay. During this period, work was forbidden and there were legal sanctions that could be applied to employers who contravened these clauses.

introduced family allowances,[2] and started the free distribution of food to children. Female emancipation was also a priority on the political agenda. A decree of December 1917 made divorce easier to obtain—from then on granted on request of one or other spouse, which broke completely with the prohibitive laws of Imperial Russia—and created equality in the rights of children, irrespective of whether they were born within or outside marriage. A 1927 law allowed tribunals to recognize legal rights for cohabiting couples. Abortion on demand, forbidden until then, was legalized in 1920.

At the same time, the attitude toward the family as an institution was ambivalent. In the 1920s, the debates centered on the acceptability of various union types and on individual reproductive rights. According to classic Marxism, the *bourgeois family* and the patriarchal rural household were incompatible with socialism. However, the Bolsheviks were hesitant about implementing such principles as abolishing the institution of marriage or socialization of children. Instead, the regime ventured into less radical experiments, introducing pilot housing projects where individual apartments had no kitchens and in which residents were supposed to take turns preparing meals for everybody, and where children were cared for in nurseries or kindergartens.

The public debate on new union types and collective households quickly lost relevance. During the 1930s, the aim to build a solid socialist fatherland gradually replaced the doctrine of the disappearance of the nation-state under socialism, and the State embraced the precommunist principle within which the family is the basic unit of society. At the same time, mass industrialization provided a practical reason for public family-assistance programs. Vast numbers of women entering the labor force flew into industry, construction, and offices. Work was proclaimed women's duty and basic right. This led to conflict between women's dual role as mother and homemaker and as worker. This tension appeared quite independently of fertility levels, even if it emerged at a time of fairly rapid fertility decline. This decrease was not as yet perceived as a threat for State objectives.

Simultaneously, the idea that population growth would benefit the communist cause emerged. Since population was indeed expanding at that point, the Communist Party stated that high fertility and rapid population growth were inherent to socialism—a myth that prevented discussion about population issues for several decades. At the end of the 1930s, population growth slowed. The regime preferred to blame declining fertility than to accept responsibility for massive losses during the Great Famine of 1932–1933 and the results of repressions during the 1930s. In keeping with its predilection for coercion, its main response took the form of prohibitions. In 1936, abortion was banned (except when the mother's health was directly threatened). In 1944, the liberal law on marriage introduced in 1927 was repealed, and divorce became more complicated and beyond the reach of most couples.

The State also introduced economic support for motherhood. In 1936–1937, it doubled the expenditures for maternity wards and preschool child care. The government introduced monthly allowances for mothers with seven children or more (1936); however, this covered only a small fraction (6% on average) of children's consumption needs (Kharaeva, 2001). The allowance for single mothers of children under 16 years (introduced in 1944) was higher, equal to 15% of the average wage. In addition, allowances for third or higher order births, and for children aged 1 to 4 years, were created. Funding for these new financial awards was met by reducing payments on state bonds. Maternity leave was increased from 9 to 11 weeks. A series of six medals was created (1944), going from the first level, the "Medal for Motherhood" to the title of "Mother-Heroine" (10 children, all living at the last child's first birthday). These medals conferred certain material advantages on their holders, such as preferential positions on waiting lists for kindergartens, summer camps, and housing, as well as priority in queues for shops.

2. The 1950s to the End of the Soviet Regime

During the 1950s, official interest given to demographic aspects of motherhood was low, mainly because, although falling, fertility remained comfortably above replacement level (Figure 118–1). Abortion on demand was once again legalized in 1955, and the divorce procedure was steadily simplified in 1965–1969. The State's involvement in support for family well-being strengthened, notably at the turn of the decade. Huge housing construction enabled many families to move from tiny communal apartments and factory barracks into small individual apartments equipped with a minimum of modern amenities and virtually rent-free. Preschool services were improved and extended.

At the same time, in Russia as in the other USSR European republics, fertility dropped below replacement and population growth slowed considerably. The

[2] A birth allowance equal to the average wage, and a child-maintenance allowance equal to a quarter of the average wage for a period of nine months after the birth.

political climate at that time permitted relatively open debate on the determinants of the less-than-ideal population growth and on the ways to reverse the trend. Discussion was limited in two ways, however. First, of the three demographic components of population growth, only fertility was considered. Immigration from foreign countries, all perceived as too foreign or as plainly hostile, was not even mentioned. The authorities also preferred to maintain their taboo on the question of high mortality rather than look for ways of reducing its level. On the other hand, nobody dared question the assumption that the State had the ability to reshape any type of individual behavior.

In the late 1960s and during the 1970s, the theoretical foundations for fertility policy were formulated and a group of practical measures, aimed to encourage childbearing, were developed. They dealt as much with the obstacles to the satisfaction of fertility preferences as with the preferences themselves. Despite the obvious theoretical difference that ultimately separates the two approaches, in practice this difference mattered little, as most of the concrete recommendations were compatible with both.

The recommendations were widely debated at academic conferences, in professional journals, and in the press, and ended up reaching the ears of policy decision makers. The 24th Party Congress (1971) indirectly recognized that the demographic situation was unsatisfactory, and the 25th Congress (1976) called for the development of an "effective population policy," meaning a fertility policy. The 26th Congress (1981) identified specific areas for political intervention: financial and in-kind assistance to young married couples and families with children, improvements of housing conditions and preschool establishments, expansion of part-time and home-based employment for women, and introduction of partially paid child care leave. This policy canvas included no quantified demographic targets, however. This contrast with the well-established practice of matching a quantified goal to every political action, however illusory or distant it might be, reveals doubts about the real demographic effect of population policy.

These policy decisions were followed by legislation. Maternity leave on full pay was extended from 77 to 112 days (then to 126 days in 1990). The fringe benefits for pregnant women and mothers of young children in the workplace, in terms of shifts, overtime, or types of tasks, were reinforced and extended to lone fathers and legal guardians. Three new child care options were introduced: (1) 18 months of leave on full pay to take care of a sick child, (2) leave with 20% of the average wage between the end of maternity leave and the day the child reached 18 months, and (3) unpaid leave to raise a child from 18 months to 3 years. In 1989–1990, these options were offered to fathers, grandparents, and even to more distant relatives. Moreover, in addition to the relevant allowances, the caregiver was granted the right to part-time or home-based employment. Businesses were obliged to offer part-time or full-time work or, under certain conditions, flexible hours, to women with children under 14 years, if they requested it.

A monthly income-related allowance was introduced in 1974 for children under 12 years. Until 1991, a flat housing allowance of 12 rubles per month was paid for each child if the family's per capita income fell below the unofficial but implicitly used low-income threshold of 50 rubles per month. In 1991, both the condition of eligibility and the amount of this poverty allowance were linked to the minimum wage.

The government also imposed a special tax on men aged 20–50 years, and women aged 20–45 years who were either single or married, but with fewer than three children. The tax rate was 6% of income for childless individuals, 1% for families with one child, and 0.5% for families with two children. The money from this tax went toward financing the monthly allowances for single mothers or for women with large families.

In the early 1980s, the government once again reinforced its population policy. The proportion of all family allowances in the federal budget rose from 0.2% in 1980 to 0.3% in 1985, whereas, in the budgets of the constituent republics, this proportion increased from 0.1% to 0.2% (Kharaeva, 2001). Funding for child care establishments increased. The number of establishments rose from 75,000 in 1980 to 88,000 in 1990, and their staff increased from 638,000 to 994,000. The growth of enrolment ratio was accompanied by an improvement in the quality of services offered to children, as the number of children per employee fell from 13 to less than 9.

Family allowances were diversified and increased. Birth allowance for mothers who were working or studying was raised to 50 rubles (the equivalent of a quarter of the average monthly wage) for the first child, 100 rubles for the second, and the same for the third. The allowance for higher-parity births (introduced in 1944) was also increased. Also created was an allowance of 30 rubles for working fathers with a wife at home. In 1990, these allowances were merged into a single allowance independent of parity, and indexed on the minimum wage.

Although the allowance for single mothers of a child under 16 years quadrupled relative to its starting level, this still did not restore its value relative to the average income. Certain categories of parents became eligible for similar, or even larger, benefits, such as widowed or single mothers who had themselves been orphans, widows not receiving reversion pensions, conscripted soldiers, and ex-spouses of persons not paying child support or paying only at a level below a given threshold.

The number of children a woman needed to quality for the special status of "mother of many children" (*mnogodetnaya*, which literally means "having many children"), which gave access to certain additional benefits, was reduced from seven to five, four or three, depending on the benefit. For instance, a woman who reared five or more children to the age of 8 years became eligible for a pension at 55, as long as she had worked at least 5 years, or at 50, if she had worked for 15 years. For parents with four or more children, the fee in preschool establishments was reduced by half in 1990. Mothers with three or more children received free summer camp vouchers. Probably, these multiple definitions of *mnogodetnayat* not only resulted from budgetary constraints, but they also reflected the confusion reigning around what could have been reasonable policy goals with respect to fertility.

On the surface, this new set of policy measures seemed to have some degree of success, as the total fertility rate rose from 1.9 in the late 1970s to 2.1–2.2 in the mid-1980s. The authorities congratulated themselves on the "positive sociodemographic impact" of strengthening their family support policy (Ryzhkov, 1986). However, this increase in TFR resulted essentially from a tempo effect; many women gave birth to their first child earlier than they would have done without the population policy effort of the early 1980s, but did not increase their completed fertility. The impact of these new measures on period fertility indicators thus disappeared rapidly (Zakharov and Ivanova, 1996), and the new trend reversed from 1988 onward.

Fundamentally, the government was extremely suspicious of family planning, which it saw as a direct cause of low fertility. In 20th-century Russia, abortion was by far the most common birth control method. Although the incidence of abortions fell during the 1960s, it was still among the highest in the world in 1970, with a rate of 135 abortions per 1,000 women aged 15–49 years; two-thirds of pregnancies ended in an abortion (Vichnevski, 1994). Although officially abortion was condemned, no theoretical or practical alternative was offered. Very few contraceptives were available, not even the condom. Abortion, on the other hand, was virtually free of charge and on demand, and the number of clinics and doctors specializing exclusively in abortion increased constantly throughout the period.

As the 1970s drew to a close, a new policy approach was adopted, aiming to replace abortion with modern contraceptives. A few years later, the Ministry of Health purchased huge quantities of intrauterine devices (IUDs) and birth control pills, and launched a vast information campaign. However, although it constituted an important ideological shift, practically speaking, this turnaround only inspired an increase in traditional contraceptive methods, in the early days, and a modest decline in abortion (Avdeev, 1994).

Only in the late 1980s and early 1990s did the authorities begin to encourage the development of networks of family planning centers and authorize the creation of the National Association for Family Planning. The new policy's positive effect on abortion thus expanded. In 1995, the abortion rate fell to 73 per 1,000 women aged 15–49 years, and then to 50 in 2002. However, plummeting fertility levels mean that, today, almost four out of five pregnancies end in abortion, and age-specific abortion rates put the total abortion rate at two per woman (Visnevsky, 2004). The abortion rate is estimated in the same way as the TFR, that is, by adding period age-specific rates. This means that at the current abortion rate, a woman will have an average of two abortions in the course of her reproductive life.

3. The Transition Years

The TFR fell to 1.75 children per woman in 1991 and, since 1996, has plunged to 1.3. The demographic situation thus became the subject of bitter debate. Since 1993, the government's official line on fertility is that it is "too low" and, since 1998, it has been proclaiming its desire to raise the level (United Nations, forthcoming). Given that this fall coincided with the country's political and economic disruptions, a popular explanation holds the abruptness of the transition responsible for the demographic crisis (Caldwell and Schindlmayr, 2003). Although the new deprivations may well have amplified the reduction in fertility intentions, the evidence is inconclusive. Another explanation focuses on the similarity between most developed countries, in terms of the recent evolution of fertility, and on the behavioral changes underlying this evolution (Ivanov, 2000; United Nations, 2004; Vishnevski, forthcoming). These two approaches, however, are not mutually exclusive.

Contrary to widespread assertions, family support policies have often survived the disruption in the country's economic system and political regime. The government approved a National Plan of Action for Children (1995) and a Family Policy Framework (1996), which focus on both family autonomy (as opposed to the Soviet era's subordination to the

State) and the principle of partnership with social institutions.

The swift and radical drop in the number of preschool establishments, personnel, and enrollment (Figure 118–2) is often interpreted as the destruction of an essential social service. The majority of nurseries and kindergartens did indeed undergo severe local budget cuts. At the same time, however, the number of preschool-aged children fell even more quickly than the number of children registered in these establishments (Figure 118–2); thus, the staff-to-child ratio increased, and the deficit in number of places available became a surplus. This evolution should be interpreted as the rational adaptation of the means to the real needs of the population rather than as a breakdown of the system.

This does not mean that the transition went without a hitch. The enrollment ratio fell by 17 percentage points during the first half of the 1990s, due to the combined effect of two factors. First, the evolution of property rights and budgetary constraints disrupted normal functioning of many nurseries and kindergartens and obliged their managers to collect registration fees. Although these fees were often optional, paid only by those who wished, in 1992 the State limited the proportion of costs for which parents were responsible to 20%, or 10% in the case of parents with at least three children. Second, the huge reduction in employment transformed many working mothers into housewives who no longer needed daily child care services. At the end of the 1990s, the recovery of the job market reversed the trend in enrollment ratio. The development of private nurseries for the middle and privileged classes also contributed to this growth.

The State rightly attempted to maintain the system of monetary allowances for children, notably in the mid-1990s, when the national economy and government finances were at their lowest. No benefit was abolished, and some were, on the contrary, even expanded. Thus, paid maternity leave, at 100% of the lost wage, was increased to 140 days (in 1992), and applied to the period during pregnancy after stopping work (1993). The monthly allowance for children aged 1.5 to 5 years was extended to younger children (in 1994), and a monthly allowance for children aged 6–16 years was also created. In 1996, these two types of child allowance were merged into a single allocation for children under 16 years. In the same way, parents on parental leave became eligible both for child care allowances and for child allowances. In the attempt to alleviate the massive rise in poverty, the income criterion for child allowance was abolished in 1992. The proportion of combined expenses from different sources (consolidated State Budget and social insurance funds) allocated to family allowances rose from 1.6% in 1993 to 3.1% in 1996, before falling to 1.3% in 2000, gaining 0.1% in 2001, and increasing again in 2002.

The family allowance system had only a mitigated effect, however. It was effective neither in redistributing income to those in need, nor in easing the financial burden of motherhood for all income groups. Short-term and one-time benefits largely predominated to the detriment of long-term allowances. Benefits were not adequately indexed to inflation. The budgets allo-

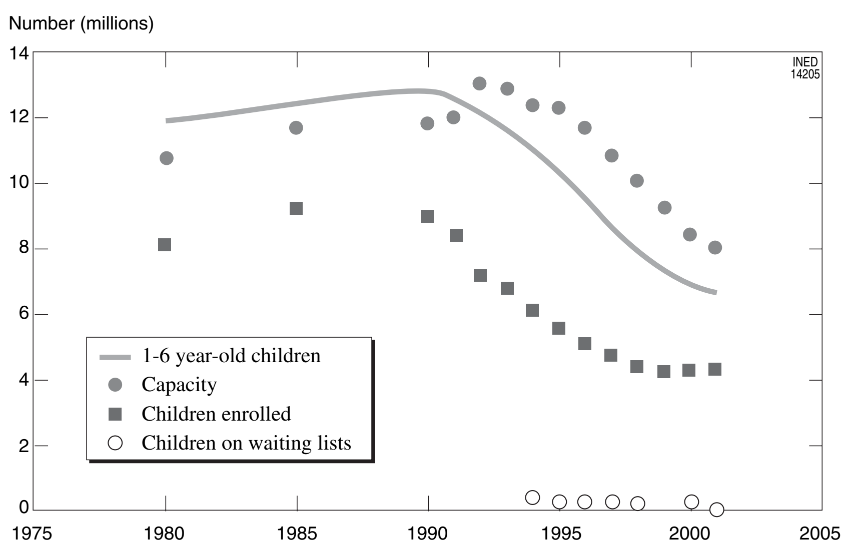


FIGURE 118–2 Children of preschool age and enrollment in preschool establishments, 1980–2001 (in millions). (Sources: estimated from data published in Goskomstat, 1994–2002).

cated were lower than the entitlements they were supposed to fund.

The indexing mechanism did not stop depreciation in the real value of allowances. The minimum salary instituted in 1990 as a base for indexation was not itself indexed to the cost of living. Consequently, the ratio of the minimum wage to the subsistence level fell to 7% in 2000. The multiplier for allowances was increased several times to counterbalance the gap between the cost of living and the minimum wage. For example, the birth allowance was raised from three times the minimum wage in 1990–1992 to 15 times the minimum wage in 1996–2000; in 1995, child care allowance increased from an amount equal to the minimum wage to twice the minimum wage. Nonetheless, the ratio of family allowances to the level of subsistence, or to the average wage, fluctuated enormously (Figure 118–3).

In fact, the birth allowance was the only substantial amount given. The child care allowance was at only 13% of the subsistence level. Child allowance (less than 5% of the average wage) represented a negligible amount, whatever the income level, and many parents did not even go to the trouble of applying for it. Despite continuing economic hardships, the proportion of actual applicants for child allowance among those eligible for it fell from 82% to 70% in 2000.

Other changes made to the system during the 1990s contradicted one another. Discarding the link between birth order and birth allowance reflected pronatalist leanings. At the same time, repealing the income criterion of eligibility weakened the redistribution role played by family support. The child allowance (the most costly element for the State budget) was not large enough to make any significant difference to its beneficiaries. The federal government transferred the responsibility for funding family allowances to the local level, without providing them with the necessary resources.

Funding for family allowances comes from two sources. Maternity, birth, and child care (for children aged under 1.5 years) allowances are financed by the Social Insurance Fund and employer contributions. The consolidated State budget (federal, regional, and municipal funds, to be more precise) finances child allowances (for children under 16 years). The Social Insurance Fund has generally met its obligations. In contrast, child benefits weighed heavily on the State budget, and the different funds were often insufficient to meet the obligations. Furthermore, the federal budget withdrew from funding monthly child allowances in 1994. The difficult economic conditions in many regions strongly affected the availability of the funds (Elizarov *et al*., 1997), and regional budgets built up considerable delays in child allowance payments.

The prolonged fertility decline over the last 15 years has provoked heated debate. The most vociferous pro-

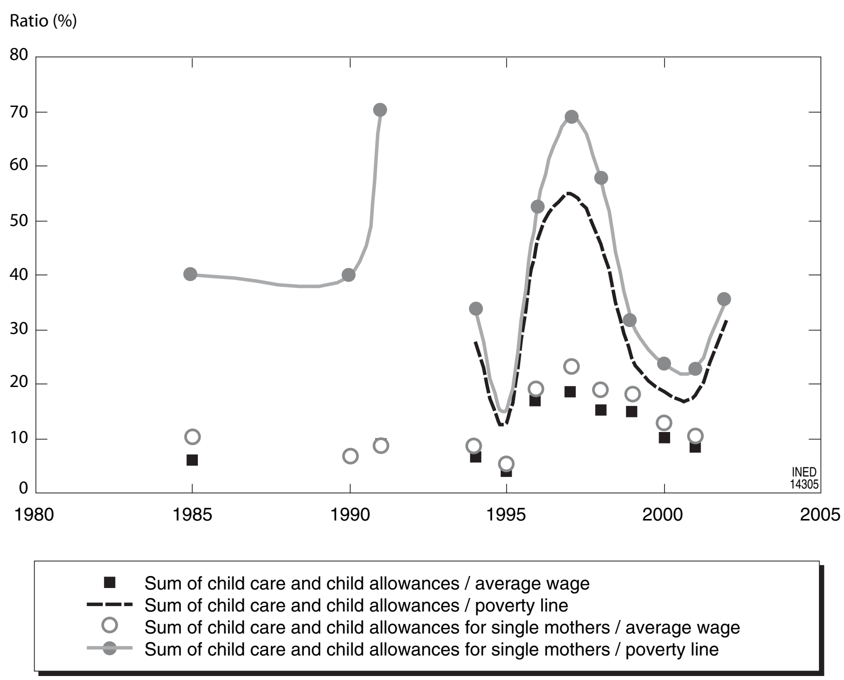


FIGURE 118–3 Trends in the ratios of family allowances to selected indicators, 1985–2002 (per 100). (Sources: estimated from data published in Goskomstat; 1996–2002; Valentei, 1985.)

tagonists are not simply demanding a return to replacement-level fertility. On the contrary, they readily endorse such policy goals as an average family size of "three or four children" (Russia, 2004). The government's white paper on the country's demographic situation (2001) repeats this visionary objective. Several participants in the debate suggested introducing generous birth allowances such as, for instance, an interest-free 20-year mortgage to reward a first birth (written off if the parents raise their child), a free car for a second birth, and a plot of land for a third (Popov, 2004). Others recommended that only biological parents should obtain public housing, but added, paradoxically but in keeping with the old Marxist doctrine, that the government and not parents were responsible for providing for children (Russia, 2004). Equally audible were calls for the return to the traditional family based on the father as breadwinner and mother as homemaker (Antonov and Arkhangelskii, 2002).

Few wealthy regions had the means and desire to establish their own family support programs with the stated intention of improving the region's demographic situation. Family allowances created by Moscow's municipal government in 2004 are several times higher than the payments offered at the national level. The city's residents aged under 25 years (27 for students and 30 if they are married) qualify for an allowance of 15,000 rubles (around 500 dollars) for a first birth, 21,000 for a second, and 30,000 for each higher-order birth. These payments are equivalent to 5, 7, and 10 times respectively the official subsistence level; this compares with 1.5 times for the federal allowance, which is independent of birth order. If no child was born within the first three years of marriage, the couple no longer qualified for these benefits. The official line on these measures is that their goal is to "improve significantly the city's demographic situation." The city of St. Petersburg created a similar program, although on a more modest scale.

Thus, raising fertility has become a high priority for a substantial part of Russian society. At the same time, little attention is given either to other options concerning demographic evolution, or to socioeconomic adaptations to the demographic situation. The pronatalist focus could undermine rather than reinforce the existing family support system. One danger is that even where ambitious programs have been adopted, they will not be sustainable; in other cases, the suggestion to introduce exotic benefits could jeopardize existing programs in the eyes of policy decision makers. Another weakness of this debate is that participants tend to move directly from the assumption that fertility is too low, to formulating solutions. Generally speaking, findings related to the effectiveness of government policy on reproductive behavior in low-fertility countries are weak and unconvincing. In any case, public debate on the issue in Russia scarcely refers to whatever knowledge has been gained from international experience. As a result, the solutions offered combine hyperambitious goals and instruments whose effectiveness has proved very weak, or remedies which might perhaps cure the illness but would certainly kill the patient.

This does not mean that it would be better to leave current family support policies in their present state. These policies, which form a major part of the social safety net and whose justification is therefore not limited to purely fertility objectives (Vishnevski, 1994), need to progress in order to better fulfill their role.

II. HEALTH POLICY

Russia entered the 20th century with very high mortality levels, characteristic of a backward rural country where traditional norms govern the natural movement of the huge majority of the population. Traditional mortality factors, especially the most dangerous infectious and parasitic diseases, were still far from eradicated; for instance, cholera epidemics were still frequent, notably those of the 1890s and of 1905–1911. Mortality rates from scarlet fever, typhoid, and smallpox were still double the maximum levels observed in other European countries (Brochaus and Efron 1898, 1898). At the local and even national levels, famine remained a threat and, over and above widespread chronic malnutrition, bad harvests, exacerbated by outbreaks of infectious diseases, lead from time to time to major crises with exceptional peaks in mortality (as in 1892, e.g.). Thus, life expectancy at birth in Russia was 10 to 12 years lower than in developed countries (Table 118–1). The infant mortality rate, over 300 per 1,000, was the main component of this high mortality. Deaths among children under 5 years accounted for 60% of all deaths; 52% of boys and 40% of girls died before the age of 15 years.

The life expectancies given in Table 118–1 come from recent reestimations based on data adjusted for underregistration of deaths among infants and among the very old from the 1950s onward (Meslé *et al.*, 2003), as well as on a more general correction for earlier periods (Andreev, 1998). Figure 118–4 illustrates the difference with Goskomstat's official estimates. It also shows in detail annual fluctuations in life expectancy for the periods where reconstitutions had been carried out (1927–1938 and since 1946).

TABLE 118–1 Life Expectancy at Birth in Russia Compared with the Average for a Group of Developed Countries, 1900–2000

Year	Russia		Average for 27 developed countries[a]		Difference	
	Men	Women	Men	Women	Men	Women
1900	29.4	31.4	42.4	43.4	13.0	12.0
1930	34.6	38.7	53.4	56.6	18.8	17.9
1938	35.7	41.9	58.1	61.8	22.4	19.9
1950	53.4	60.9	64.0	68.2	10.6	7.3
1960	62.9	70.6	67.4	72.5	4.5	1.9
1970	62.6	73.1	68.4	74.7	5.8	1.6
1980	60.9	72.5	69.9	76.8	9.0	4.3
1990	63.4	74.0	71.9	78.6	8.5	4.6
1995	57.9	71.4	73.1	79.5	15.2	8.1
2000	58.8	72.1	74.0	80.1	15.2	8.0

[a] Australia, Austria, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Finland, France, Germany (FRG, 1946–1988), Greece, Hungary, Iceland, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom, United States.

Sources: Andreev *et al.*, 1998; Meslé *et al.*, 2003.

Mortality had undoubtedly started to decline by the end of the 19th century as a result of social, economic and political reforms undertaken since the early 1860s.

From then on, the history of Russian mortality was composed of extreme contrasts, with periods of progress interrupted by major crises. In reality, Russia experienced only three relatively short periods of genuine progress during the 20th century (before 1914, from 1923 to 1928, and from 1947 to 1964), separated by two disastrous episodes (World War I and the civil war, on the one hand, and the Great Famine and repressions of the 1930s, followed by World War II, on the other) and ending in a long period of slow deterioration of the situation, after 1965, notwithstanding the fluctuations of the 1980s and 1990s. Setting aside past crises and recent fluctuations, one could say that in the course of a century Russia, trying to make up lost ground first experienced a period of convergence with western countries; on the contrary, in the mid-1960s the country entered a phase of radical divergence due to a decline in life expectancy unconnected to any crisis and as such unprecedented in the history of developed countries. At the start of the 21st century, the gap between Russia and industrialized countries was 15 years for men and 8 years for women—as great, on average, as it had been 100 years earlier, at 13 and

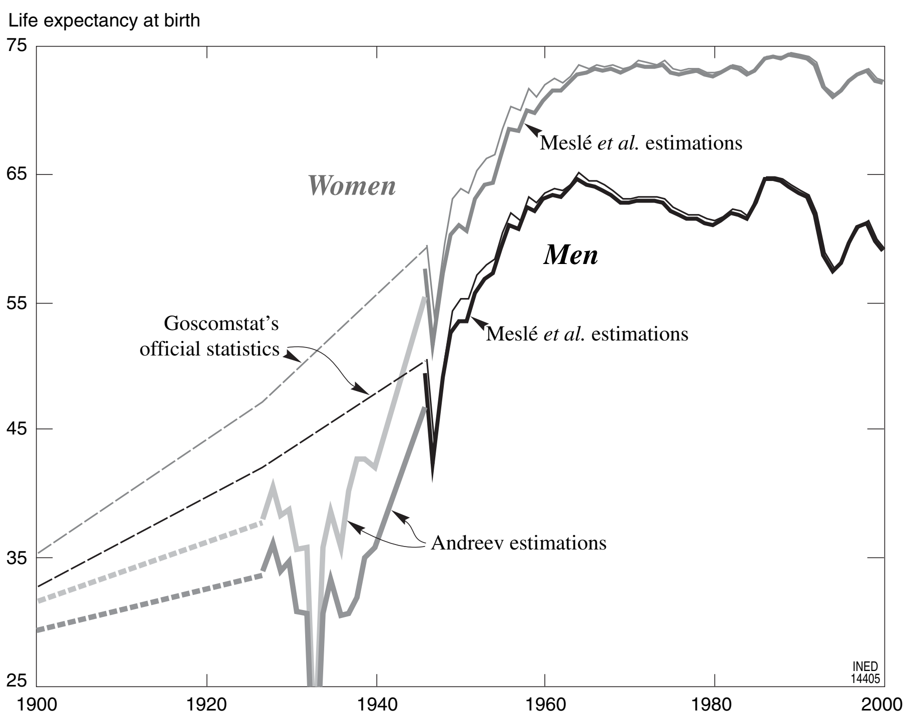


FIGURE 118–4 Trends in life expectancy in Russia since 1900. (Sources: Goscomstat, 1988; Andreev, 1998; 1900–1950, Meslé *et al*. 2003.)

12 years respectively (Table 118–1). This development is obviously not unrelated to health policy and, more generally, to the country's political, economic and social evolution. We will demonstrate this by focusing on four different issues, namely (1) the successes and contradictions of the state monopoly on health care (1918–1965), (2) the public health care crisis and the State's abortive attempts to check it (1965–1990), (3) the ups and downs of the campaign against alcohol, and (4) the return to negative trends.

1. State-Controlled Health Care: Achievements and Contradictions (1918–1965)

World War I, the 1917 Revolution, and the 1918–1920 civil war halted the process of modernization and the first reduction in mortality. However, the latter resumed in a while under new conditions and despite the catastrophes of the 1930s and 1940s, by the 1960s life expectancy attained its highest level ever. The main determinant of this chaotic evolution was State policy in all areas of social life, without significant modernization of health-related behaviors.

As early as 1919, health protection had been declared a priority for the new government. From the moment they took power, the Bolsheviks proclaimed their wish to provide the whole population with free access to high quality medical care, to create a more secure health environment, to establish scientific means to prevent the development and spread of infectious diseases, and especially to fight social diseases such as tuberculosis, sexually transmitted diseases, and alcoholism.

This statement of intent later became one basis for the myth of the unprecedented success of the Soviet socialist system in terms of health protection. Only a few decades later, however, it became apparent that this success was only relative. There may well have been significant positive changes, but they were not extraordinary ones. Mortality was falling in many countries, and the fall was often faster and more profound than in the USSR. Moreover, many of the medical achievements were built on pre-revolutionary developments, as a result, for instance, of the culture and actions of professors and physicians of the *zemstvo* (local assembly that functioned as a body of provincial self-government in Russia from 1864 to 1917). The Soviet regime's only real innovation was the principle of a State monopoly on health care (Mirsky, 1996).

The lack of resources was the first problem facing the founders of the Soviet health system (known in the West as the Semashko system). Having abolished the *zemstvo*, the State had to recreate a whole network to manage the prevention and treatment of diseases, and provide the necessary resources for its functioning, in terms of materials, technical instruments, medicine and personnel. To obtain the funds needed to fulfill its aspirations, the State had diverted into its own budget the profits generated in the commercial sector, and had artificially devalued wages. The substantial funds collected in this way were first invested in heavy industry and transportation, with particular attention given to the military-industrial complex. The social sector as a whole, including public health, received only the meager remnants. The acute lack of resources and the notorious depreciation of medical personnel's wages became the Soviet public health system's chronic ailment in all subsequent stages of its development.

During the early decades following the Revolution, mortality decline in Russia appears to have been fairly rapid (due to a general change in lifestyle, fast rising education levels and the spread of information, as well as to relatively cheap mass measures within health and hygiene, such as improvements in the urban sanitation, vaccination campaigns, quarantine measures, etc.). The many studies that found an association between these latter factors and morbidity or mortality reinforced the optimism.

Indeed, a substantial movement toward modernization was developing in the country. Education at all levels was spreading rapidly; science, including medicine, and mass culture were progressing. Significant investments were made in health education as part of the fight against illiteracy. At the same time, medical institutions were expanding, an accelerated training for medical personnel had been organized, and the pharmaceutical industry was developing (albeit too slowly), and so on.

On the other hand, during the entire period that followed, the Soviet government became obsessed with maintaining the growth of quantitative, formal health progress indicators (number of beds, doctors, medical check-ups, etc.) that had appeared in the 1920s and 1930s. Public health's centralized development entailed target-based budgeting. The priority given to such investment parameters as numbers of beds, doctors or hospital nurses, or the proportion of the population receiving medical check-ups, was certainly justified at the start. Later, however, by preventing the use of other parameters to track progress during the health transition, this approach ossified the system and blocked any new directions for investment (and thus advances in health).

In particular, too little attention was paid to qualitative parameters—the provision of modern equipment and decentralized services to hospitals, the introduction of new treatment methods, the development of new drugs and progress within the pharmaceutical and medical industries. Decades later, in the mid-1980s, 15% of hospitals were still without running water, 49% had no hot water on tap, 24% had no sewer systems, and 45% no baths or showers (Goskomstat of the USSR, 1990). Already problematic early on, health investment efficiency has become completely inadequate since the 1970s. In addition, the health authorities rarely took regional, economic and demographic particularities into account. The central administration gradually lost control over the functioning of the vast network of medical institutions, and management efficiency deteriorated; meanwhile, the inflexible rules and rigid hierarchies left regional authorities little room for maneuver in responding to evolving structures of population, its epidemiological profile, and medical and health needs.

Undoubtedly, partly because of these deficiencies of the health system put under the complete control of the Ministry of Health, alternative health institutions developed in parallel within other ministries or organizations, with their own standards, structures and personnel. Those in the highest echelons were not the only ones to have their own medical services; it was also the case for the Ministries of Defense, the Interior, the Defense Industry, Nuclear Industry, as well as the different Ministries responsible for transport. After the Ministry of Health, the Ministry of Railways had the largest network of medical services. In addition, establishments such as the Ministry of Culture or the Academy of Sciences also had their own dispensaries or clinics. All these alternative medical institutions were better equipped and had more highly qualified personnel than those controlled by the Ministry of Health. Evidently, the multiplication of these parallel, and often overlapping, institutions enabled many individuals to register themselves and their families with several medical establishments at the same time: at their place of residence, through employment in a particular sector of industry, owing to their privileged status. The Soviet system, supposed to provide all citizens with equal access to high-quality medical services, thus made this access closely dependent on profession or social status, to the detriment of the health of the great majority of the population.

Health statistics became more difficult to compile; the sector-related medical services did not divulge indicators of their activities and were not therefore taken into account by the Ministry of Health, making any objective evaluation of the situation difficult.

Despite high ratios of medical personnel and hospital beds per capita (Table 118–2), long waiting lists for consultations became one of the thorns in the side of the medical system, particularly in the network of territorial institutions managed by the Ministry of Health. The overload in the health services was not only due to high morbidity levels. The doctor providing the first line of care to patients also issued documents authorizing sick leave; in the post-Stalin period, such documents could be supplied on false pretences, often just at the patient's request, without any medical check-up. Real diagnoses and the appropriate care were virtually unobtainable outside hospitals, while prevention and primary health care became increasingly formal and ineffective.

Until the 1960s, infectious diseases were the main focus for public health policies. There is no denying the success of the battle against these diseases, that the hypercentralized health service fought, even though it was not total (Table 118–3). Starting in the period between the two world wars, the frequency of diseases such as diphtheria, typhoid, and especially smallpox

TABLE 118–2 Growth in Numbers of Hospital Beds and Doctors (per 100,000 persons), Russia, 1928–1990

Year	Number of beds	Number of doctors
1928	186	43
1932	289	48
1937	411	64
1940	500	81
1950	592	145
1960	821	193
1970	1125	266
1980	1298	383
1990	1375	450

Sources: Goskomstat, 2000, p. 211, 214; Vishnevsky, in press

TABLE 118–3 Incidence of Selected Epidemic Diseases (per 1,000 persons per year)

Disease	1913	1926	1928	1933	1936
Typhoid	24.2	9.1	6.7	12.0	9.2
Rickets	6.6	4.2	2.2	50.4	6.5
Smallpox	5.2	1.3	0.9	2.0	0.0
Measles	31.0	29.5	44.8	28.5	48.4
Scarlet fever	29.8	21.3	20.8	6.0	30.2
Whooping cough	34.2	32.2	32.8	20.7	7.3
Diphtheria	31.0	4.6	5.0	10.7	6.2
Dysentery	30.7	17.2	10.0	12.6	39.8
Malaria	253.0	319.7	206.5	372.4	235.7

Source: Vishnevsky, in press.

declined significantly. However, in the late 1930s it was still too early to talk of a radical decrease of the incidence of malaria, dysentery, measles, scarlet fever, and rickets.

The contradictory nature of the Soviet type of modernization must be taken into consideration. Rapid industrialization had been accompanied by a formidable increase in population density around the main construction sites and new industrial centers, which were very often built out of nothing. In the old cities, the infrastructure was inadequate to accommodate the huge waves of immigrants. The forced collectivization of agriculture and its consequences, such as the great famine of 1932–1933, pushed millions to migrate, and provoked a resurgence of infectious diseases both in famine zones and where refugees and deportees congregated, and in what have been called "special settlements" (different categories of forced migrants within the context of the repressions: dispossessed peasants, ethnic groups, transgressors of the regime of interior passports, the unemployed, etc.) World War II and the massive evacuation of tens of millions of people led to the reemergence of infectious diseases exacerbated by malnutrition and famine.

Although the advantages of a centralized health care system were used to the full for anti-epidemics operations, it was not up to dealing with the challenge created by a continuous chain of social disruptions.

With the end of mass repressions, and with the country's economic and political climate relaxing a little, the mid-1960s marked a turning point with respect to morbidity and the structure of mortality by causes of death. Lifestyles improved somewhat. Mass urban housing construction enabled millions of families to leave the factory barracks, shantytowns, or overcrowded municipal dwellings. The quality of drinking water, the development of central heating, and, finally, medical progress made it possible to drive back many infectious diseases more quickly.

The planned large-scale introduction of essential medical supplies, such as sulpha drugs and antibiotics, played a key role in the general fall of mortality rates and in transforming their age structure. In the early 1950s, the USSR organized wide-scale production and dissemination of antibiotics, and for a decade and a half mortality plummeted. As in other countries, infant mortality and deaths from infectious and respiratory diseases declined most, enabling Russia to complete the first phase of its health transition (see Chapter 57). However, this was the Soviet regime's final victory in the health domain.

The state monopoly on health services lasted several decades longer (until the early 1990s), but with no significant positive outcomes and no fresh mortality reduction. Then the calcified Soviet health system had only a limited success in the area for which it had been created initially, that of controlling the main infectious and parasitic diseases, where the essential task had already been accomplished; in addition, it proved totally incapable of meeting the challenge of modern ailments, such as cardiovascular diseases, chronic old age-related diseases, accidents and injuries of all sorts. In contrast with Western countries, it failed to enter the second phase of the health transition (see Chapters 57, Volume II, and 69, Volume III, as well as Chapter 108, Volume IV).

2. Crisis in Health and Ineffective Remedies (1965–1990)

During the second half of the 1960s and the 1970s life expectancy declined throughout the USSR, but the deterioration was most serious in Russia. Infant mortality slightly fell and old-age mortality stood still; their respective levels had not yet fully caught up with those of developed countries. At the same time, working-age mortality rose substantially. Moreover, Russia's exceptionally large gap in life expectancy between men and women grew even wider (Meslé and Vallin, 1998; see also Chapter 54, Volume II). In addition, the differentials between rural and urban populations deepened. In 1977–1978, the life expectancy of rural men was only 57.7 years, whereas that of urban women reached 73.3 years. In this period, the "northeastern" mortality gradient became prominent: the further north and east one moved through the country, the more the population's health status degenerated (Vasin and Costello, 1997).

One of the most glaring aspects of the mortality crisis is the spectacular rise in violent deaths: accidents, accidental poisonings, injuries, homicides, and suicides. The crisis, nonetheless, resulted just as much from premature deaths through cardiovascular disease. According to the life tables by causes of death, in 1965, the average age at death by cardiovascular disease was 74.3 years in Russia; by 1980 it declined to 70.7 years, and by 1995, to 67.5 years (compared with 77 years in Western countries at the start of the 1990s).

On the other hand, the crisis did not affect infectious, respiratory or digestive tract diseases, nor cancer. Deaths from tuberculosis, gastro-intestinal disease, and viral infections declined. Pneumonia, chronic bronchitis and influenza also evolved in the right direction, although more modestly. Apart from some Russian particularities, tumor-related deaths followed approximately the same evolution as that witnessed in Western countries (Meslé *et al.*, 1996).

The government and Communist Party leadership were informed of the unfavorable trends in health and mortality early on. Many major documents of the 1970s (minutes of the Communist Party meetings, Council of Ministers' decrees, etc.) reflect the authorities' concerns and generally identify correctly problems such as excess working-age mortality or excess male mortality. Civil society, however, had no complete information on what was happening in the USSR in this area.

In the first half of the 1970s, as in Stalin's time, the USSR once again banned the public use, publication and transfer to international organizations (such as the United Nations or the World Health Organization) of basic information on the evolution of health and mortality (life tables, morbidity indicators, data on causes of death). In particular, causes-of-death statistics were classified, and for certain causes access restrictions were even stricter. The publication of demographic and epidemiological research findings was strictly controlled. These severe limits on the dissemination of information lasted until the mid-1980s, when Mikhail Gorbachev's arrival in power opened the way for *glasnost*.

This embargo on information also prevented any serious reform that might improve the health system's efficiency. This does not mean that no effort at reform was made, however. During the 1970s, in particular, rural health services were reorganized as part and parcel of the restructuring of local rural communities. The aim was to increase the population's concentration in settlements "with a future," and gradually empty settlements "with no future"—those in which emigration had already significantly reduced the number of residents. As a result, most small hospitals, outpatient offices, primary health care centers, and obstetric clinics were closed. Instead, larger multifunction hospitals with outpatient services were created in district centers. The organization and management system for rural health was simplified, and the quality of diagnosis and treatment improved. However, this recentralization of medical services often made access more difficult because of low density in the means of transport and communication. The selective effects of migration that skimmed off the rural population's most active and healthy members, as well as the destruction of traditional rural social structures made major contributions to the deterioration of the demographic and health situation in the rural areas in the 1970s.

During this decade, the USSR Ministry of Health continued its policy of establishing a system of specialized polyclinics, hospitals, research centers, coupled with outpatient departments. Centers for cardiology, oncology, pediatrics, surgery, and other specialties also appeared, draining the best specialists, the best diagnostic equipment, and the best pharmaceutical and medical supplies. In the context of weak growth in public spending in the health and social spheres, however, this policy created a situation when the regional health services (the lowest level in the public health care structure) increasingly lagged behind. Thus, the basic level of medical intervention (where mass prevention programs are implemented) started to deteriorate not only in rural areas but also in the large cities.

This structural imbalance in the health system should have been corrected by a population-wide program of "general mandatory medical check-ups" aiming to improve early diagnosis. The vast territorial network of polyclinics and the number of doctors per person (higher than anywhere else in the world) should have made it possible to achieve this goal. However, the program's implementation was blocked by financial constraints, lack of medical, technical and computer equipment, and the medical personnel's insufficient training in screening methods and in the use of diagnostic equipment. People themselves did not see the point in attending competent services without a serious health problem. The attempt to introduce a "medical passport," required when entering employment, also failed. Generally speaking, this was the weakest and most ineffective health policy and attempt at health system reform in the country's entire history. No positive results were gained during this period; on the contrary, mortality trends were unfavorable, particularly in the area of cardiovascular and chronic diseases, the policy's principal targets.

3. Ups and Downs of the Anti-Alcohol Campaign

In May 1985, the USSR launched a campaign against alcohol. From the start, it was presented as a means to reduce violent deaths. At the same time, Soviet specialists had gathered sufficient proof that excessive and expanding alcohol consumption was largely the cause of increasing working-age mortality, excess male mortality, large rural-urban mortality differentials, and regional disparities in mortality. It was just as evident that alcoholism was not only responsible for heavy demographic losses but also was at the root of many social and economic problems, whether at the macrosocial or family level.

The campaign was implemented with the aim of limiting the production, sale and consumption of alcohol. It was incisive and involved all possible political, administrative and economic levers. As a result, in the space of a few years, per capita alcohol consumption dropped. The effect on mortality curves was immediate: according to official statistics, in only 2 years, life expectancy at birth rose by 2.6 years for men and 1.2 years for women. The fall in mortality rates was even greater in republics of the ex-USSR and regions of the Russian Federation that had experienced the steepest rises in mortality during the previous period. Working-age groups benefited from maximum changes. This unique social experiment fully confirmed the importance of alcohol consumption's impact on Russian mortality (Shkolnikov and Nemtsov, 1997). The anti-alcohol campaign not only reduced deaths from external causes, but it also helped reduce cardiovascular mortality (Meslé *et al*., 1996).

This campaign did not produce durable results, however; rather than providing the means to eradicate the cause of alcoholism, the authorities only introduced prohibitions and restrictions on the production and sale of spirits. Once the surprise effect had worn off, these measures rapidly elicited a return to a soon-thriving domestic production of traditional alcohols, the consumption of nonalimentary alcohol products (medicinal tinctures, *eaux de Cologne*, etc.) and other substitutes (solvents, industrial alcohol, etc.), not to mention the growth in drug consumption and the dependency phenomena that accompanied it. From 1988 onward, mortality stopped falling and soon afterward started to climb again to pre-1985 levels. The losses are thus in complete symmetry with the gains produced by the campaign (Meslé *et al*., 1998). Mortality regained ground essentially in working-age population groups; violent deaths (homicides, accidents, suicides) and deaths form accidental poisoning (which, in the Russian classification of causes of death, includes alcoholism) played the dominant role, with cardiovascular diseases in distant second place. The most affected regions were those that had benefited the most from the anti-alcohol campaign's positive effects. The evolution of cancer-related mortality, in contrast, was scarcely affected either by the drop in mortality at the start of the campaign, or by its subsequent rise. Similarly, mortality among children, the elderly and the highly educated adults proved less sensitive to the fluctuations. Like certain cohort analyses, these facts suggest that those spared during the campaign against alcohol did in fact die a few years later. In all cohorts affected by this fluctuation, the excess deaths observed in the late 1980s and early 1990s correspond almost exactly to deaths avoided during the preceding years thanks to the campaign against alcohol; in the end, the campaign had virtually no effect on the number of years lived by these cohorts (Avdeev *et al*., 1997; Zakharov, 1999; Andreev, 2000).

Nevertheless, mortality increased even more rapidly between 1992 and 1994, this time as a result of the economic and social crisis following the brutal transition to the market economy. In 3 years, men's life expectancy fell by 6 years, and women's by 3 years. Life expectancy was thus lower than before the campaign against alcohol in both rural and urban areas—even lower than the level that might have been predicted from a simple return to past long-term negative trends (Meslé *et al*., 1998; Vallin and Meslé, 2001). Violent deaths and cardiovascular diseases certainly continued to play a particularly important role in exacerbating this decline, but all the other causes also contributed to it; even though working-age groups remained central to the problem, mortality among children and the elderly was also affected. When the initial shock was over, mortality fell again and in 1996–1998 reached the level it would have been at had it followed the long-term negative trend uninterrupted by those specific events.

Thus, responsibility for the first fluctuation (1985–1991) can be assigned to the disruptions of the alcohol market which affected behavior, first by reducing consumption, then by directing consumers toward alternative sources of supplies (Leon *et al*., 1997). The general economic crisis brought on by political reforms, with its huge decline in living standards for the majority, and, in particular, the deterioration of health services, is responsible for the second, inverse fluctuation, in 1992–1998. Very probably, the sociocultural shock was accompanied by a collective psychological stress, related to the transition to the new model of society, which also helped aggravate mortality's resurgence in 1992–1994 that brought life expectancy to an unimaginably low level in 1993–1994 (Shapiro, 1995; Cornia and Paniccià, 2000).

4. Return to Long-Term Negative Trends and Problems with Reform

With the fall of the communist regime, the very principles on which the Soviet system of funding and management in the social sphere rested were obviously questioned, including in matters of health. The fundamental principle of free access to medical care was among the first issues to be dealt with in the context of transition to a market economy.

At the same time, it was impossible to ignore the fact that the principle of free access to medical services had already been significantly eroded during the period of "advanced socialism" (official label introduced by Brejnevian ideologues), that is in the 1970s and 1980s. For the great majority of people, the supposed ease of access to medical services was already largely vitiated by the mediocre quality of the services actually provided. For rapid, high quality medical assistance, it was normally necessary to call on parallel informal services paid in kind or in cash, unless one had access to special services reserved for certain social categories. For instance, highly qualified gynecologists were very wealthy but they owed it to the bonuses, monetary or in kind, received illegally from their patients. Corruption developed very quickly in the health sector.

In the final years of the Soviet regime, the principles of fee-free and easily accessible medical services, as well as the preventive orientation of health policies, were thus undermined by the deterioration of the quality and efficiency of the services provided, and became the mere echo of an ideological rhetoric unrelated to reality (Krichagin, 1989). These problems, more or less concealed under the socialist guise of the Soviet regime, came explosively to light when the regime fell.

Fifteen years after the reforms began, however, health policy goals are still undefined, and priorities remain vague. In the post-Soviet period, health policies have often changed course, reflecting the instability in the balances of power in the public health sector. Russian society is clearly far from ready to face the struggle needed to reverse the unfavorable trends plaguing the population's health.

At the present moment, the main problem affecting public health in Russia is the rift separating the State's guarantee of broad access to free medical care from the inadequate funds allocated to it (for a detailed analysis of the situation, see Bogatova *et al.*, 2002). The health sector was already chronically underfinanced during the Soviet era, but budgetary constraints reached a new dramatic threshold in the early years of the transition to the market economy. Employees in the social sectors, medical practitioners included, were no longer paid their salary. For months, the maintenance of medical buildings and equipment which escaped looting was neglected.

It appeared that the target-oriented approach for public financing of priority goals was partly justified by the fact that no investment mechanism had been set up that could combine the principles of territorial-industrial funding with those of target-oriented funding. Consequently, the federal target programs adopted over the years (Prevention through Vaccination, Insulin-Dependent Diabetes, Children of Russia, Family Planning, Motherhood without Risks, Urgent Measures against Tuberculosis, Anti-AIDS) were constantly underfinanced.

The attempts made to transform the state health system to one based on insurance (Law on the Medical Insurance of the Citizens of the Russian Federation adopted in 1991) were largely formal, and have failed on the whole. The financing of medical institutions hardly differs from the budgetary system set up during the Soviet regime. In reality, the funds generated by the new compulsory insurance merely redistribute tax revenue and fulfill none of the functions of insurance institutions that should evaluate the amount, type, quality, and validity of services offered to people. The compulsory medical insurance premiums paid by the employed persons, defined as a percentage of the salary received (3.6% in 1993), bear no relationship to the real financial needs of medical care. Those insured are excluded from the negotiation process and cannot influence either the timing or the quality of the medical services made available.

Since 1998, the Government has examined many proposals for a free health care program guaranteed by the State based on providing a minimum of well-defined benefits limited by allocated resources. However, none of these numerous proposals was ever adopted. The management of public health institutions operates on the basis of temporary, and constantly changing, standards. Their financing depends directly on the region's economic conditions (efficiency of its businesses, governor's political influence, and so on). As a consequence, basic medical services offered to the population differ enormously throughout the country in terms of both quantity and quality.

Alongside free medical services, the state medical institutions also provide paying services. Regional authorities and autonomous local institutions establish the list, the order and, in some cases, the price list of paid medical services provided in public medical institutions (belonging to State or to municipalities). Furthermore, since the mid-1980s a private market of medical services grew rapidly. Next to institutions providing medical services as such, a large number of businesses have surfaced in the last two decades, offering traditional or alternative medical treatments and paying services for diagnosis, prevention, and treatment for the whole range of known diseases. Simultaneously, optional medical insurance has emerged. As a consequence, the population's global medical bill has soared.

From 1994 to 1999, health expenses (including payments for medical care and drugs) from non-state sources as a proportion of total health expenditure thus climbed from 13% to 36.5% (Belyaeva, 2001). During the decade starting in 1993, private expenses for paid medical services and optional insurance premiums increased several times, by a factor of 100 in constant prices. Experts have concluded that in the late 1990s the ratio of the State's global expenses on health to those of the population was 55 to 45 (Bogatova *et al*., 2002, p. 15–16). According to *Russia's Longitudinal Monitoring Survey* (RLMS), the proportion of the average Russian family's total expenditure allocated for health care and drugs rose from 0.9 to 4.6% between 1992 and 2001. In 2001, 10% of patients paid for outpatient treatment, and 15% of patients paid for hospital inpatient treatment. Among those who paid for health services, approximately half did so officially, at the medical establishment's cashier office, whereas the others paid under the table (Bogatova *et al*., 2002), p. 17).

The extent of "informal" payments varies considerably from one region to the next; it also differs according to the type of medical institution and the nature of the services provided (Shishkin *et al*., 2004). Even in the same regional hospital, the proportion of patients paying on the side varies from 10% to 70% depending on service and treatment. It is practically impossible to be operated free of charge in such fields as neurosurgery or vascular and thoracic surgery (Shishkin, 2003). Among families of modest means, official and non-official medical expenses for theoretically fee-free services represent a higher proportion of income than among wealthier families. In poor regions, this social discrimination is stronger than in the richer ones (Shishkin *et al*., 2004). Informal payments are higher and more frequent in the poorest regions, implying that, in fact, people alleviate through these payments the regional disparities in state financing.

Thus, the black market is thriving alongside the official market for paid services. In 1990–2001, this informal health care market pocketed 600 million dollars per year (Satarov, 2002), whereas the State's total expenses on health barely reached 7 billion dollars.

Thus, the Russian health care system appears as a patchwork of state activities (managed by the Ministry of Health), sector activities (independent of the Ministry of Health) and private activities providing services legally, semi-legally or illegally; services are provided either completely or partially free of charge (funded at the federal or local level, or through compulsory health insurance funds), or fully payable (with the possibility of subscribing to voluntary insurance programs). Having lost its monopoly on health care planning, the Russian Federation's Ministry of Health no longer has the power to make substantial improvements. The main group of reformers for the health services (and for the social sector in general) is made up of health professionals representing the government's macro-economic bloc (the members of which are constantly replaced) and of experts from various research centers with government instructions to work on particular programs. Federal and regional legislative authorities then slightly edit government proposals and programs, often in an unprofessional manner, with the clear aim of satisfying the political and ideological preferences of the majority of deputies.

During the 1990s, the international agencies that financed the research projects on how best to reform the health system were very active. Thus, the project *Zdravreform* (Public Health Reform) was set up with the help of the World Bank. Within the TASIS program, the European Union sponsored the following projects: Program of Health System Development, System of Prophylactic Measures and Population's Health, Reform of Health Care Financing, and Insurance-based Health Care. It is difficult to assess the effectiveness of these and other lesser known projects. At the same time, they unquestionably played an educational role with respect to health system organizers and in terms of international experience in solving similar problems and setting up modern health systems.

The current Governmental Health Development Plan up to the year 2010 (see http://www.med-pravo.ru/common/conception/concepmed-1.htm) proposes a new transformation in the supply structure of medical care by reducing the range of treatments requiring hospitalization, transferring them to the outpatient sector. At present, financial resources are distributed in the following manner: hospitalization 60%, outpatient consultations 30%, emergency services 10%. The plan intends to invert these proportions, thereby reducing the number of hospital beds needed (by 17% between 1991 and 2002). The newly created service of family doctors (currently comprising less than 3,000 of the 2 million doctors working in the health system) would occupy the central position in the primary health care system. Improving the efficiency of emergency services has also been given priority.

The main objectives toward which the Ministry of Health is channeling most of its energies include the following (in the order in which they appear, and as formulated, in The Concept of Public Health Development) (see http://www.med-pravo.ru/common/conception/concmz2001-1.htm): (1) to reduce the level of premature mortality (from cardiovascular diseases, accidents, poisoning, traumas, and cancers); (2) to fight diseases that are particularly relevant for the country's demographic situation (diseases threatening reproductive health, mother's and child's health, and middle-age and old-age diseases); (3) to wage war on diseases posing a particular threat for the whole nation's health (tuberculosis, AIDS, alcoholism, sexually transmitted diseases); (4) to set up specific mechanisms giving people access to the resources necessary to achieve these targets; (5) to develop preventive measures, and set up ways of maintaining and rehabilitating people in good health; (6) to become involved in activities designed to give people a better understanding of the value of good health and encourage them to adopt behaviors to help safeguard it; (7) to develop medical sciences and advanced medical technologies, in every way possible, including information technologies; (8) to meet the population's need for medications and other medical products; (9) to set up a system of sociomedical insurance; (10) to fix a price policy in the market for medical services; (11) to introduce professional liability insurance, and improve the social protection for health sector personnel; (12) to create a system of medical rehabilitation, and develop preventoriums and sanatoriums; (13) to extend the organizational and legal framework of public health institutions, and optimize the nomenclature of medical specialties.

On the whole, this document reads more like an ideological text, in the manner of numerous writings of the Soviet period, than a practical guide for action. It reflects the ideas and interests of the Ministry of Health's main leaders before its reorganization in 2004; with little danger of being proved wrong, it can be said that this document will go through major adjustments in the coming years.

It is not surprising that the transition period's blurred policy did not lead to any major success. Apart from a reduction in the number of abortions, declines in maternal and child mortality, and gaining control over a certain number of infectious diseases, such as diphtheria, whooping cough, hepatitis A, typhoid, brucellosis, rabies, or malaria, the population's health status has shown no positive change, nor any promising tendency. Incredible as this may sound, Russians have a lower life expectancy at birth at the start of the 21st century than they had 40 years ago.

It was easy to identify the key heath challenges, and certain policies could have met these challenges. Looking first at the rapid worsening of various infectious diseases (tuberculosis, STDs, AIDS); for all these diseases, morbidity rates increased at an unprecedented rate in the last decade. Although the rising incidence has slowed down in the last two or three years, the future evolution of the situation remains uncertain and worrying. Second, exogenous mortality (accidents, accidental poisoning, homicides) shows no sign of letting up. On the contrary, it has begun to rise again, probably as a result of increasing alcohol consumption. Third, the risks of premature death, essentially through cardiovascular and chronic diseases, have grown substantially. In the current state of affairs, it is impossible to foresee how society and the State could curb a trend that started more than 40 years ago.

III. MIGRATION POLICY

1. Internal Migrations

During the Soviet period, the size and nature of internal migratory flows were essentially determined by two State-driven processes: forced industrialization, which accelerated the process of urbanization, and the development of peripheral regions, which triggered off massive flows toward the east and north of the country. The aim was to transform an agrarian country into an industrial power as quickly as possible by transferring resources, human and others, from the countryside to the cities, or toward regions to be developed. The immense population movements engendered by urbanization and regional development were completely consistent with the policy of the State. In so doing, however, the authorities rightly feared losing control over massive migrations that risked ruining the system, a danger they sought to avoid.

Deep regional inequalities in living conditions were present throughout the entire Soviet era. An immense urban-industrial periphery living in poverty existed alongside small enclaves where living standards were relatively high and artificially maintained (capital cities, several regional centers, or privileged industrial centers). Residents of the periphery were constantly on the lookout for the slightest opportunity to move toward more prosperous areas, thereby undermining the very idea of privileged enclaves. The authorities also feared too swift an exodus from the countryside, which played the role of an internal colony whose exploitation had for a long time made it possible for them to achieve their ambitious plans. In a word, massive migrations were necessary for development, but their strict control was just as necessary to anchor and maintain the totalitarian political regime then asserting itself. These contradictory aims spawned a hypocritical and paradoxical migratory policy. On the one hand, the State did its utmost to bolster the factors forcing people to leave the countryside and small towns, and to ensure an uninterrupted flow of undemanding human resources toward industry and other urban sectors, the army, and remote regions. On the other hand, however, the State implemented a system of restrictions and barriers to ensure its total power over migration by submitting every individual move to bureaucratic controls.

The key instrument of this policy was the internal passport system. Making it possible to control all citizens' movements, this system had been widely used in Western Europe in the 19th century but had more or less disappeared everywhere by the beginning of the

20th century. It had been kept in place in Russia (at least for those who left their permanent place of residence), however, providing Lenin with the opportunity of declaring, in 1903:

> Social-democrats demand complete freedom of movement and enterprise for the people. But what does "freedom of movement" mean? [. . .] It means that passports must also be eliminated in Russia (in other states, passports have long since disappeared) [. . .] that no policemen, no zemstvo civil servant will have the power to stop people from living and working wherever they see fit. [. . .] Isn't this a case of feudal serfdom? Isn't it insulting people? (Lenin, [1966]).

Once in power the Bolsheviks eliminated the internal passport and, despite the fact that serious obstacles to the movement of citizens resurfaced in the years following the revolution, the general policy of abolishing "feudal serfdom" remained in force for some time. In 1922–1923, a whole series of legislative documents were adopted to abolish the system of internal passports and all institutions related to it, as well as all other constraints on the right of Russian Federation citizens to move or reside anywhere on the Russian territory, by demanding that they have a residence permit or other such document. Thus, *The Small Soviet Encyclopedia* could write that "Soviet law does not recognize the system of internal passports [which] had been an important tool of police intervention and fiscal policy in a so-called police State" (*Encyclopedia*, 1930).

Less than 10 years later, however, a new article in *The Great Soviet Encyclopedia* tells of the Bolshevik State's sudden policy turnaround:

> Passport system: rules of administrative registration, of control and regulations over population movement through the introduction of passports for the said population. The Soviet legislation [. . .] never concealed that its passport system was class-based, using it according to the conditions of class struggle and the aims of the working class's dictatorship at different stages in the construction of socialism. (*Encyclopedia*, 1939)

These words referred to the new system of internal passports already established by a decree of the Central Executive Committee and the Council of People's Commissars on December 27, 1932. According to this decree, all people living in cities, industrial towns, sovkhozes, and new cities had to have an internal passport stamped by the authorities testifying to their place of domicile (the *propiska*). Ipso facto, this inhibited the right to move since the choice of residence depended on the possibility of obtaining the *propiska*. Very quickly, in cities such as Moscow, Leningrad, the federal republic capitals, and other important cities, the *propiska* was restricted, and even citizens with a passport could obtain the *propiska* only with special authorization. As for the peasants—the majority of the country's population—without passports, they were unable to move to the city and were effectively prevented from leaving their village or kolkhoze.

Admittedly, the State did not intend to ban all migrations by introducing such restrictions. On the contrary, its economic policy demanded a radical upheaval in the population's geographical distribution and, as a consequence, vast migration. The aim was to define and control these flows from the center, regardless of its citizens' individual aspirations. As Stalin declared in 1939, "at present it is only a matter of proposing to the kholkozes to accept our request and allow at least one and a half million young members of kholkozes to depart each year to satisfy industry's increasing needs" (Stalin, 1952). Although rather hypocritical, Stalin's formulation clearly reveals the cynical order thus created: the individual could only move if "he was allowed to leave" by being granted a passport authorizing him to go and live elsewhere. From this point on, migration flows were thus strictly regulated.

By the late 1930s, the already severe controls placed on population movement proved insufficient. As a result, in 1940, a new law added penal sanctions against any individual leaving his work or changing his place of work without authorization. Conversely, engineers, technicians and qualified workers who refused to accept a change of residence dictated by the State could be penalized. Although this law was at first followed to the letter, it was undoubtedly applied less rigidly over the years; nonetheless, the 1940 law was abrogated only in 1956.

These regulations aimed at controlling spontaneous migrations, or those considered as such, even though the desire to migrate often resulted from very pressing economic or political constraints. In counterpoint, forced migration also existed, involving mass displacements and even repressive population transfers.

The Soviet regime frequently resorted to political repression to solve the most diverse problems, with economic issues first, or almost first, among them. As early as 1929, on the basis of a proposal put forth by a special commission of the *Politburo*, the Sovnarkom adopted a secret decree on "the use of prisoners' labor," which specified that people condemned to 3 years or more of "deprivation of freedom" by common law tribunals had to serve their sentence in labor camps by working for OGPU. And, to welcome those deprived of their freedom, the OGPU had in turn to enlarge or build new labor camps in remote areas "in order to colonize them and exploit their natural resources by using the labor of those deprived of freedom" (Materik, 2000, p. 64).

Very quickly, deprivation of freedom for criminal or political offences became an important instrument for the "colonization of outlying regions." These regions soon became the sites of a vast network of camps, colonies, and special settlements (*spetsposelenie*) where millions of people newly deprived of their freedom were sent: *dekoulakised* peasants, Gulag prisoners, ethnic deportees, and so on. In 1953, the total number of prisoners and *specposelency* rose to 5.4 million people, while Siberia, the Far-East and Kazakhstan combined, where most of these prisoners were concentrated, barely numbered 30 million inhabitants in total.

Stalin's death ended these massive deportations, but the system of migration regulation initiated under Stalin did not die with him.

The discrimination imposed by the system of internal passports went on for a long time. It took more than 20 years before new passport regulations in the USSR were introduced by a USSR's Council of Ministers Decree on August 28, 1974; according to this new decree, passports had to be issued to all USSR citizens aged 16 years and over. For the first time, this decree created the same system for rural and urban dwellers. "The passport campaign began on the first of January 1976 and was completed on 31st December 1981. In 6 years, 50 million passports were delivered in rural localities." (Lioubarski Kronid, *Régime de passeport et régime de propiska en Russie*, http://www.hrights.ru/text/b2/Chapter5.htm).

The *propiska* lasted even longer, until the USSR collapsed, and has still not completely disappeared. In principle, any citizen holding a passport could move freely throughout the entire country; in reality, however, he could not settle in any place where the *propiska* was restricted as, without it, he had access to neither work nor housing. As the demand for labor was in general very high in these regions, the State's need for labor in key economic sectors (industry, science, etc.) was in constant contradiction with a migration policy that paralyzed the labor market. In practice, the creation of non-viable artificial social statuses made it possible to bypass these contradictions. Thus appeared millions of *limitchik*—people whom businesses and state organizations were entitled to employ by issuing them with a temporary *propiska* for the duration of their contract of employment. This authorization to live in large cities was precarious and did not extend to other members of the family (even marriage to a *limitchik* did not entitle them to a provisional *propiska*).

Only on October 26, 1990, did the USSR's Supreme Soviet Committee for Constitutional Supervision acknowledge that the *propiska*'s authorization procedure in force for so many years "prevents citizens from enjoying their fundamental right to freedom of movement, work, and education"; legally, this paved the way for abolishing the *propiska*. Furthermore, on June 25, 1993, President Boris Yeltsin approved the law on "the right of citizens of the Russian Federation to freedom of movement and to the choice of their place of residence within the limits of the Russian Federation," which should have put a term to the *propiska*. In reality, this did not happen. In many places, the attempt to eradicate the *propiska*'s authorization procedure met with resistance from local authorities, and it still survives today. Moreover, many bureaucrats at all levels still openly favor a return to the *propiska*.

Although prohibitive and repressive measures were the main tool of Soviet migration policy, this was seldom referred to. Instead, the focus was constantly placed on the incentives that had also always existed. A whole system of privileges had been created at the time the State needed to increase flows of migrants toward certain cities or regions. A most typical example is the priority accorded to regions of the extreme North, rich in natural resources but suffering from harsh climatic conditions. The list and particular status of these territories were defined in 1967 by a decree of the Soviet government. In 1990, the list was extended to 70% of Russia's territory (11.9 million sq. km.), without any real reason according to some authors. Those volunteering to come and work in these regions enjoyed a higher salary, more days off, subsidies for temporary work stoppage, higher pensions, and so on. These measures met with relative success. The benefits attracted many people to the northern regions; in 20 years, from 1970 to 1989, the population in the far eastern regions increased noticeably, from 7.2% to 8.6% of Russia's total population (12.6 million people). In subsequent years, these regions lost some of their economic attraction and their population size, in absolute terms and as the percentage of the country's total population, has started decreasing.

Another type of incentive was used to encourage migrations toward developing regions. It consisted of combining economic interests with young people's enthusiasm referred to as the "*komsomol* appeals." In the decades following World War II and until the mid-1980s, it was common practice to call on volunteers to work on the large sites of natural resource development, construction of railways, power stations, and so on. In this way was ensured, for instance, a flow of young people to Kazakhstan in the 1950s, to exploit virgin lands or build the Baikal-Amour railroad during the 1970s.

During World War II, people and industries were evacuated from the occupied territories, which played a considerable role in increasing the population and developing the economy of the Asian part of Russia.

The population census of 1989 makes it possible to assess the results of internal migration during the post-war period and, as a consequence, the effect of Soviet migration policy.

According to the first Soviet census of 1926, 82% of the population within the borders of the USSR as well as those of Russia were rural (Goskomstat, 1988, p. 8). At the end of the 1920s, right from the start of forced industrialization, a rural exodus took place on an unprecedented scale. Between the 1926 and 1939 censuses, the urban population of the USSR (within the pre-1939 frontiers) increased by 30 million people (from 26.3 to 56.1 million) and its share of the total population rose from 18 to 33% (Goskomstat, 1988). Over the next 50 years, the USSR's total population (within its new borders) increased by 94.5 million but, during the same period, the urban population grew by 127.5 million whereas the rural population fell by 33 million despite its fertility being higher than that of the urban population. Between 1950 and 1990, the USSR's urban population increased, on average, by around 30 million people every 10 years (Goskomstat, 1997, p. 7). At the end of the 1980s, the urban population represented two-thirds of the USSR's total population and close to three-quarters of Russia's population. Of the 29.8 million–strong increase in the urban population of the USSR between 1926 and 1939, 18.7 million, or 63%, were migrants from rural areas. Between 1939 and 1969, within the USSR's post-1939 borders, the urban population rose by 75.6 million people, 61 million of whom were migrants (Urlanis, Borisov, 1984, p. 414).

The migrations taking place during the Soviet period also modified the regional distribution of Russia's population. These changes, however, were less important than those brought about by urbanization. The USSR had inherited from the Russian empire an immense territory far from uniformly populated and developed. During several decades after the end of the 1920s, occurred a substantial shift of population eastwards, as a result of migrations from the European to the Asian parts of Russia. The total population east of the Ural increased from 13 to 22 million people between 1926 and 1939 and reached 32.1 million in 1989 (thus multiplied by a factor of 2.5, whereas Russia's total population increased only 1.7 times). However, population movements toward the East hardly changed the overall picture (Figure 118–5). As in the past, Siberia and the Far East have remained very sparsely populated and the concentration of population in the Central region that includes Moscow, has only slightly decreased (Figure 118–6).

During the period following the USSR's collapse, former restrictions on freedom of movement were either abolished or considerably weakened, and the State's once active policy in the realm of internal migrations disappeared. Migrations then took place against a backdrop of depopulation, economic crisis, and political conflicts (especially in the Caucasus), and have often flowed in directions contrary to those imposed for decades by previous policies. Noticeable since the late 1980s, this migratory ebb has gathered momentum, mostly from the eastern and northern regions that the State once sought to populate.

This is not really because the State, with its more liberal political regime, no longer allows itself to resort to methods used during the Soviet regime. More fundamentally, this occurs because the potential of internal migration became severely limited by the level of urbanization already reached and by the trend toward depopulation. In Russia, internal migration can no longer play the role it enjoyed from 1930 to 1980. Today, the migration issues are related much more to external migration, and more precisely to immigration, which completely alters the objectives that policies may be aiming at.

2. External Migration

a. Emigration

The Russian empire contributed much less than other European countries to the trans-Atlantic migrations of the late 19th and early 20th centuries. This can be explained by Siberia's vast and vacant lands, and by the enormous potential for internal agrarian colonization, encouraged by the government. Nevertheless, the country produced emigration flows, mostly consisting of non-Russians. Estimates put the number of people who emigrated from Russia in the period 1861–1915 at 4.3 million, of whom 2.6 million left during the first fifteen years of the 20th century. Two-thirds went to the United States (close to 80% during this period) (Obolensky-Ossinsky, 1928). Among those who left, predominated the Jews (approximately 2 million between 1881 and 1914), and other non-Russians. From 1860 until the beginning of World War I, the majority of migrants to North and South America were non-Russians: Jews (44%), Poles (25%), Lithuanians (8%), Finns (7%), and Germans (6%) (Obolenski, 1928, p. 24).

Although World War I abruptly halted these emigration flows, the Revolution and the civil war generated new, and very different, streams. In fact, these political upheavals incited a massive and definitive emigration. Although estimates vary greatly from one author to the next, it is now estimated that the USSR,

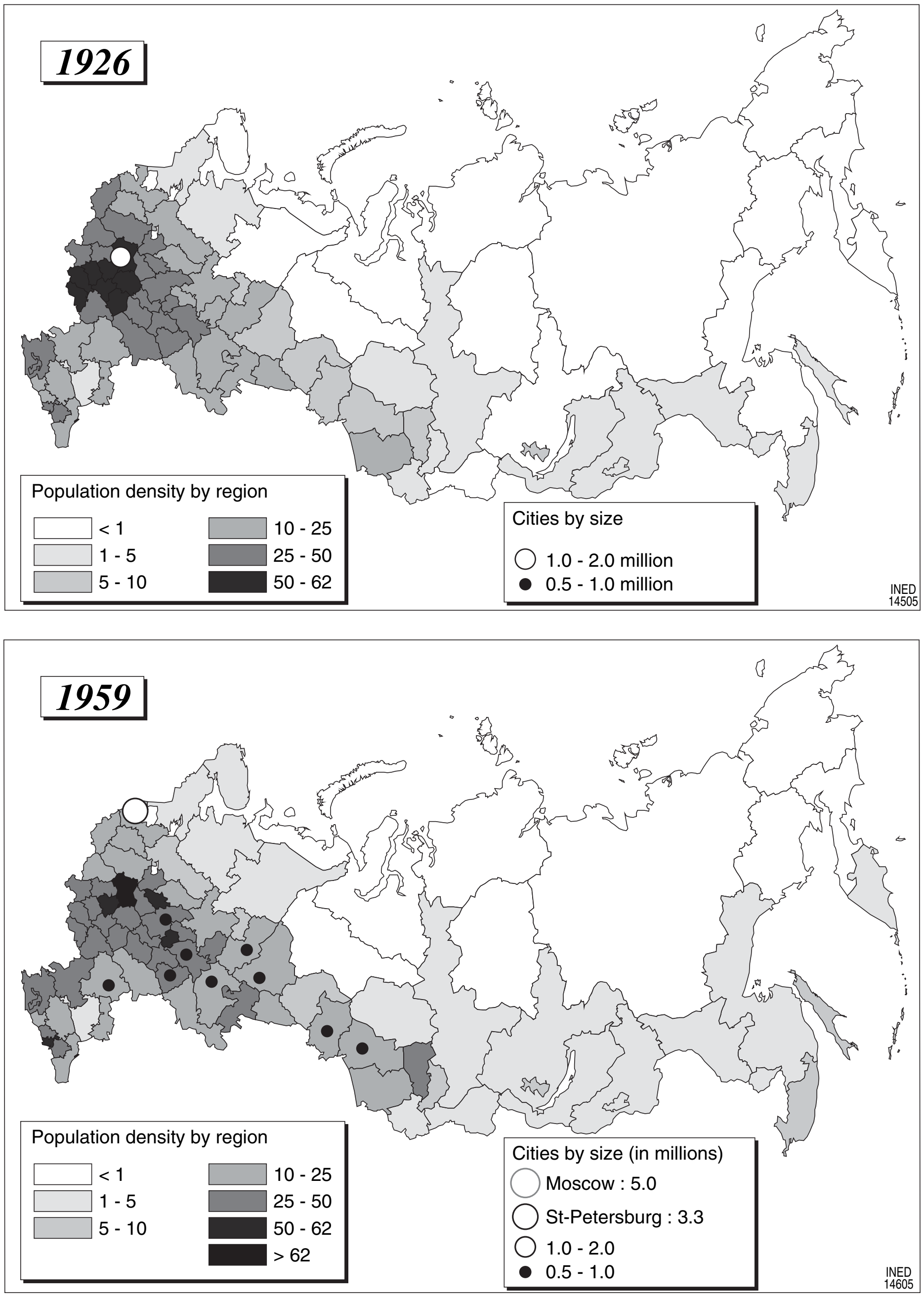


FIGURE 118–5 Maps showing the evolution of the Russian population's geographical distribution during the 20th century.

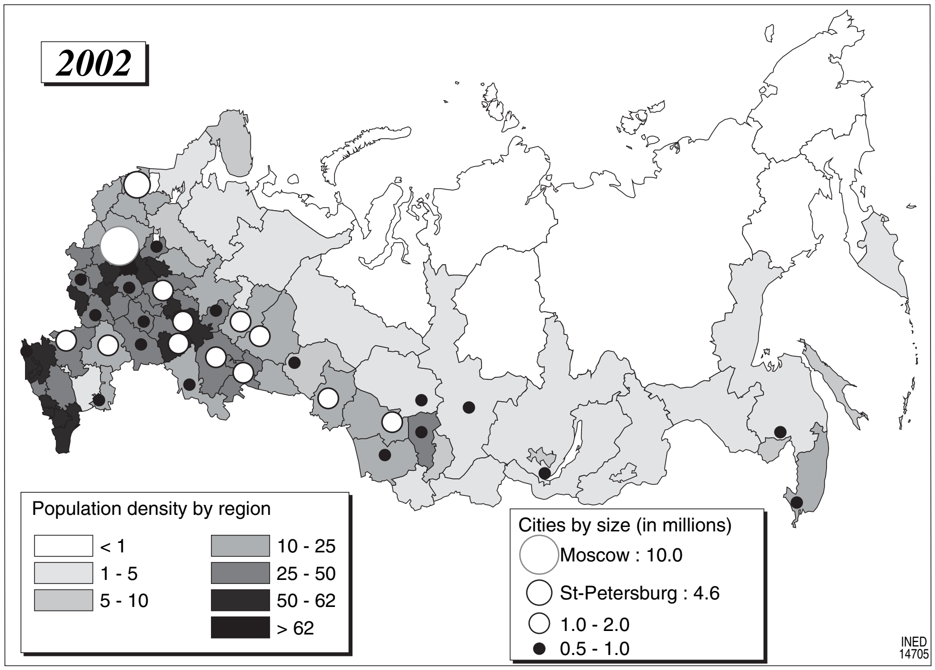


FIGURE 118–5 (*Continued*)

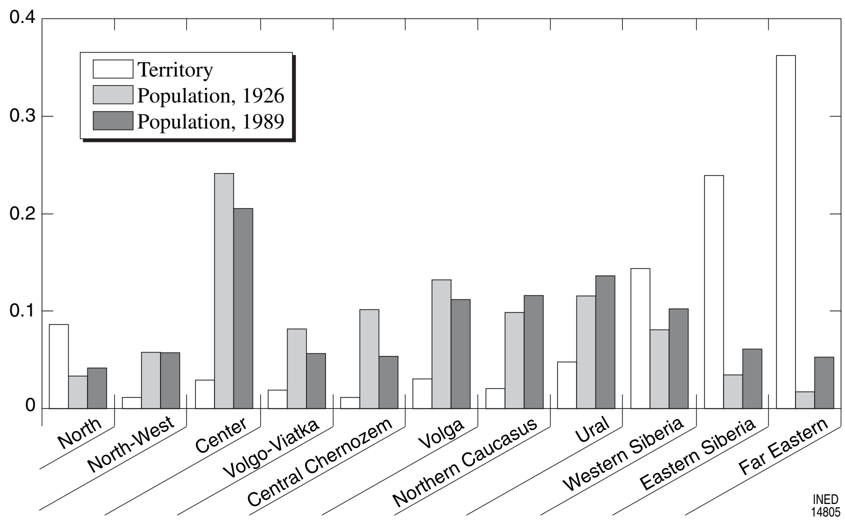


FIGURE 118–6 Large economic regions as a percentage of the entire Russian territory and total population. (Source: Goskomstat, 1998, p. 50–53.)

within its 1922 borders, established at the time of its foundation, lost then approximately 2 million inhabitants (Jiromskaia, 2000, p. 134–139; Adamets, 2003, p. 284–285).

When the civil war ended and the Soviet regime imposed its power over the entire country free emigration outside the USSR stopped. It became exceptionally rare for individuals to receive authorization to leave the country (an exit visa), except for official duties abroad. A few waves of uncontrolled emigration took place nonetheless. In 1931–1933, according to some estimates, some 200,000 Kazakhs definitely left for China, Mongolia, Afghanistan, Iran, or Turkey, to escape from a raging famine (Kozybaev *et al*., 1991, p. 226).

World War II spurred new, and very substantial, waves of emigration, some more or less uncontrolled and others, on the contrary, forced and severely controlled by the authorities. On the one hand, a vast number of refugees and displaced persons refused to return to the USSR, despite the fact that the Soviet authorities did their utmost to get them back. Various estimates have been made of their numbers. According to official sources, on January 1, 1951, 452,000 Soviet citizens remained abroad (Polian, 2002, p. 573). However, some of those who refused repatriation after 1945 had died by 1951, so that the total number of permanent departures was greater. Other estimates provide much higher figures: 1.2 to 1.5 millions, or even 2 million (Heitman, 1987, p. 10), a figure considered exaggerated, however (Polian, 2002, p. 576). On the other hand, the State decided to expel nearly 900,000 Germans (approximately 400,000 in accordance with the 1939 Soviet-German Pact, and 500,000 after Russia annexed Eastern Prussia) from the USSR territory within the borders established after the war. In addition, close to 1.5 million Poles from Western Ukraine, Western Bielorussia, Lithuania, and other regions of the USSR were moved to Poland, and nearly 400,000 Finns were moved from Karelia to Finland, and so on (Heitman, 1987; Piesowicz, 1988).

Once the waves of migration prompted by World War II were over, the iron curtain fell again, putting a sudden end to all emigration from of the USSR.

During the 1960s, departures out of the USSR were minimal, and were more than counterbalanced by arrivals, themselves very few (Armenians returning to their historical homeland, refugees from China or students from Asian or African countries); as a result, before 1970, net migration was slightly positive.

In the two following decades, the number of entries fell and, despite the rarity of departures, net migration was negative. In the 1970s, net emigration fluctuated between 10,000 and 15,000 people, only exceptionally reaching 30,000 to 40,000. During the 1980s, emigration had slowed even further.

A major change took place in 1988, as Jews, Germans and Greeks were allowed to emigrate with relative ease, and trips abroad on personal invitation were authorized. The reaction to this breath of freedom was immediate. Departures from the USSR almost trebled from 1987 to 1988 (108,000, as opposed to 39,000), and doubled again in 1989 (235,000), and once again in 1990 (452,000). As immigration remained a rare event, net emigration skyrocketed.

In May 1991, the USSR adopted a new law on the entry and exit modalities, guaranteeing that international law in matters of freedom of movement would be respected; this law was supposed to be implemented throughout the whole USSR territory from January 1st, 1993. The USSR ceased to exist at the end of 1991, but the law was nonetheless implemented in the Russian Federation and remained in force until a federal law was adopted, on August 15, 1996, on "the modalities of entry to and exit from the Federation of Russia." This law stated: "Every citizen of the Russian Federation can freely circulate outside the borders of the Russian Federation and freely return to the Russian Federation." Thus the final restrictions on leaving the country, in place for decades, disappeared; as a result, emigration increased substantially.

According to published government statistics, Russia's net emigration from the ex-USSR's borders rose to 1,071,000 people during the years 1990–2000 (Goskomstat, 1997, 2001).

b. Immigration

Whether taking its present frontiers, those of the Russian empire or those of the USSR, Russia never experienced large-scale immigration waves in the past, despite the fact that the government occasionally accepted relatively numerous groups of migrants belonging to specific categories. Thus, for instance, German settlers were made welcome in the 18th century, and Armenians in the 19th and 20th centuries. On the whole, however, within its present borders, Russia has been a sending rather than a receiving country. This is even more the case in the European part of Russia, which for a long time was an important source of migration toward the Asian part of Russia as such or toward other republics of the USSR which became independent states after the USSR's collapse.

In the days of the Empire and under the Soviet regime, these migrations were always considered internal and, fitting neatly in its strategy to develop peripheral regions, were encouraged by the central government. However, for demographic or sociopolitical reasons, migration gradually lost its role and Russia became host country to in-migrants as early as the second half of the 1970s (Figure 118–7). In addition, the USSR and RSFSR (Russian Soviet Federal Socialist Republic) leaders became aware of the need to open their doors to immigrants. Experts and policymakers hotly debated the question of how to supplement a scanty demographic tank of Russia with the help of migration from regions with excess manpower, such as the central Asian republics; their attempt at measures of this nature met with no great success.

As the USSR was dismantled and the Russian Federation became an independent state, the migration context and interpretations have radically changed. On the one hand migrations between republics, hitherto considered internal, became international migrations overnight. On the other hand, the population's natural increase became negative since 1992 and immigration suddenly became an important counterbalanc-

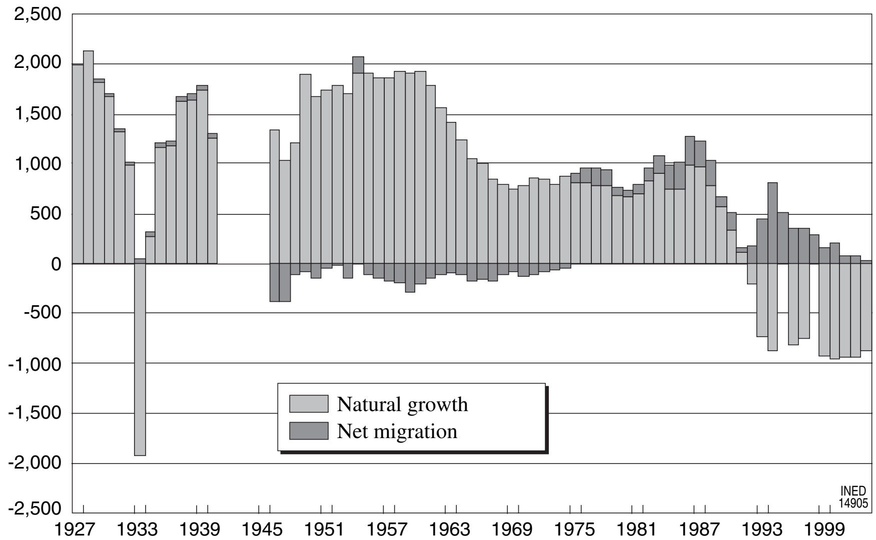


FIGURE 118–7 Components of population growth in Russia, 1927–2003 (per thousand). (Source: Vichnevski, 2004, p. 18).

ing factor: even if it is not sufficient to offset the population's natural decrease, it nonetheless considerably curbs the population decline.

In the early 1990s, Russia, which, at its level, had no experience managing migration flows, created institutions and legislation essential for this management. In 1992 the Federal Migration Service was established as an independent federal agency that subsequently was reshuffled many times. In 2002, this service lost its independence and its functions were passed on to the Ministry for Matters of the Federation, Nationalities and Migration Policy. This Ministry was in turn dismantled and migrations became the responsibility of the Ministry of the Interior where, the Federal Migration Service was reconstituted in 2002 as a ministerial department.

Present laws concerning the treatment of refugees and displaced persons were adopted in 1993, together with a presidential decree aiming to introduce migration controls. However, more general federal laws on Russian Federation citizenship and on the status of foreigners, replacing regulations dating back to the Soviet regime, were only passed in 2002. Without going into detail on the content of these documents, it is nonetheless worth noting that they aroused strong criticisms within society. Recognizing its weaknesses, even the President of the Russian Federation criticized the law on citizenship and, in 2003, some amendments were made.

The unfairly restrictive nature of these two laws does indeed open them up to criticism. They create too rigid a boundary between Russian Federation citizens and those of former constituent republics of the USSR who shared, until recently, the same citizenship as the inhabitants of Russia. By putting the inhabitants of Ukraine, Armenia, or Kazakhstan on the same legal footing as those from China, Vietnam, or an African country, they groundlessly sever kinship ties, prevent the return to Russia of persons of Russian culture, who either have Russian as their mother tongue or are fluent in Russian, and thereby deprives Russia of an extremely valuable source of immigrants.

Despite these unfair restrictions, it is not true that Russian migration laws completely close the door to immigrants. They acknowledge that foreigners can legitimately remain on Russian soil provided they have a residence permit, temporary residence permit, visa, or any other type of document accepted by federal law or international agreements confirming the right of foreign citizens to stay or reside on Russian Federation territory.

A 5-year residence permit can be issued to any foreigner residing in Russia and in possession of a temporary visa; it can also be extended many times. Anyone holding such a residence permit is considered a permanent resident of Russia, whereas all other legal immigrants are defined as temporary residents (the temporary residence permit is issued either according to quotas defined yearly by the government of the Russian Federation or, in certain legally defined cases, outside quota), or as persons staying temporarily in the Federation of Russia (with visa, or without visa when they come from countries for which the visa is not required). Every foreigner who has resided in Russia for more than 5 years can apply for Russian citizenship.

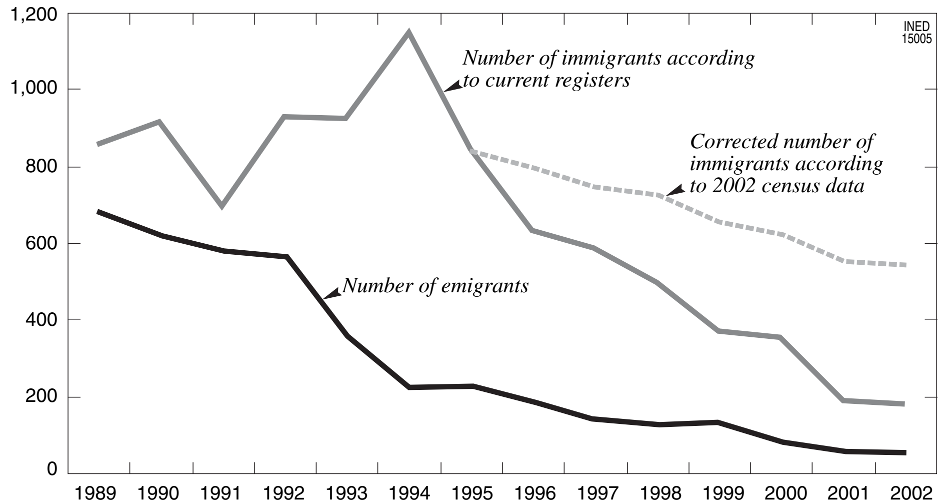


FIGURE 118–8 Trends in external migration in Russia, 1989–2002, in thousands of persons. (Source: Vichnevski, 2004, p. 140.)

Therefore, the current legislation provides a legal framework that is relatively favorable for immigration and for gradual integration of immigrants into Russian society. Nonetheless, the presence of a large number of illegal immigrants, often mentioned by the authorities and the media, shows that the system provided by this legislation functions poorly, that it was created without taking into consideration actual operative policies, and that it fails to regulate spontaneous, weakly controlled migration flows about which no reliable information exists.

The official statistics only provide data on registered migrants. According to these statistics, the exchange of migrants between the ex-republics of USSR changed radically with the Union's collapse. Entries into Russia suddenly outnumbered departures, resulting in an abrupt increase in Russia's net migration. This increase was short-lived and by the end of the 1990s, net migration had greatly diminished. However, it is quite possible that the real evolution of immigration flows was different, as a result of increasing illegal immigration.

This may to some extent stem from a change in the nature of migration movements. In the early years following the USSR's demise, repatriated Russians, returning to live permanently in Russia, made up the bulk of immigrants; with time, however, immigration has increasingly included migrants in search of temporary work. Mounting difficulties in obtaining permits for work, residence, or temporary stay, were soon felt.

After the 2002 population census, Russia's Goskomstat (which became the Federal Service of State Statistics in 2004) adjusted the data provided by current migration registration after 1995, raising immigration estimates (Figure 118–8). This correction is supposed to include illegal immigration as well, but this is questionable because census data on illegal immigrants are probably incomplete.

In the absence of reliable information, public opinion tends to overestimate the number of illegal immigrants, and increasingly resents it, which explains the government's ambivalent attitude and migration policy. On the one hand, the authorities are starting to appreciate immigration's new role as the main means to counterbalance natural depopulation and shortage of manpower. On the other hand, central authorities are more preoccupied by public opinion than by the active development of migration policy. They prefer to leave the responsibility of organizing and implementing this policy to the lower (departmental) echelons and regional authorities; the latter, however, tend to manage migratory processes according to their own interests, which are sometimes at odds with those of the country as a whole.

As a result, the President of the Russian Federation declares, "What we need [. . .] are not prohibitions and obstacles, but an efficient migration policy that can benefit the country while being acceptable to people" (message from Russia's President to the Federal Assembly, dated May 16, 2003). At the same time, the President acknowledges that "we [. . .] have not yet succeeded in developing a civilized mechanism to recruit manpower from other countries" (comment made by Russia's President on the radio, on December 18, 2003). A migration policy tailored to Russia's present needs remains to be invented.

Acknowlegment

This chapter was translated from the French by Heather Juby.

References

Adamets Serguei, 2003. *Guerre civile et famine en Russie. Le pouvoir bolchevique et la population face à la catastrophe démographique 1917–1923*. Paris, Institut d'études slaves, 376 p.

Andreev Evgeni, 2005. forthcoming. *in:* Anatoli Vishnevsky (ed.), *Demographic modernization of Russia 1900–2000 (Демографическая модернизация России).*

Andreev Evgeni M., 2000. Health and longevity, *in:* Vassily Kolesov and Terry McKinley (eds.), *Human development*, p. 193–196. Moscow, UNDP, Moscow State University and Human Rights Publishers, 383 p.

Andreev Evgeni, Darskii Leonid, and Kharkova Tatiana, 1998. Демографическая история России: 1927–1959 *(Demographic history of Russia: 1927–1959)*. Moscow, Информатика (Informatica), 187 p.

Antonov Anatolii I. and Arkhangelskii Vladimir M., 2002. *Демографические процессы в России XXI века (Demographic developments in 21st century Russia)*. Moscow, Graal, 168 p.

Avdeev Alexandre, 1994. Contraception and abortions: trends and prospects for the 1990s, *in:* Wolfgang Lutz, Sergei Scherbov, and Andrei Volkov (eds.), *Demographic trends and patterns in the Soviet Union before 1991*, p. 141–144. London and New York, Routledge, 218 p.

Avdeev Alexandre, Blum Alain, Zakharov Serguei, and Andreev Evgeni, 1997. Réaction d'une population hétérogéne à une perturbation. Un modèle d'interprétation des évolutions de mortalité en Russie, *Population*, vol. 52(1), p. 7–44. Translated into English in 1998: The reactions of a heterogeneous population to perturbation. An interpretative model of mortality trends in Russia, *Population: An English Selection*, 10(2), p. 267–302.

Belyaeva Nina, 2001. Источники финансирования в системе счетов здравоохранения России, *Экономика здравоохранения. Счета России*, vol. 59(7).

Bogatova Tatiana V., Potapchik Elena, Tchernetz Vladimir, Tchirikova Alla, and Shishkin Sergey, 2002. Бесплатное здравоохранение: реальность и перспективы. Независимый институт социальной политики., *Пробел -2000*, 216 p.

Brochaus and Efron, 1898. Россия, *in:* Энциклопедический словарь (*Dictionnaire encyclopédique*), vol. 54. St. Petersburg, Brockhaus & Efron, 1890–1907 , 86 vol.

Caldwell John C. and Schindlmayr Thomas, 2003. Explanations of the fertility crisis in modern societies: A search for commonalities, *Population Studies*, vol. 57(3), p. 241–264.

Cornia Giovanni A. and Paniccià Renato, 2000. The transition mortality crisis: evidence, interpretation and policy responses, *in:* Giovanni A. Cornia and Renato Paniccià (edss.), *The mortality crisis in transitional economies*, p. 3–37. New York, Oxford University Press, 456 p.

Elizarov Valerii V., Djanaeva Natalia G., Zatonskikh Natalia and Feoktistova Elena, 1997. Семейная политика в СССР и России (Family policy in the USSR and Russia), *in:* Valeri V. Elizarov and Natalia V. Zvereva (eds.), Домохозяйство, семья и семейная политика *(Household, family, and family policy)*, p. 142–148. Moscow, Dialog-MGU, 190 p.

Encyclopedia of the Soviet Union (Советская энциклопедия), 1930. Паспорт (Passport), *in:* ISE (ed.), *Малая Советская энциклопедия*, vol. 6. Moscow, Советская энциклопедия, 10 vol.

Encyclopedia of the Soviet Union (Советская энциклопедия), 1939. Паспортная система, *in:* ISE (ed.), *Большая Советская энциклопедия*, vol. 44, p. 322–323. Moscow, Советская энциклопедия, 718 p.

Goskomstat, 1988. *Naselenie Rossii 1987. Statistitcheski sbornik (Population of Russia 1987. Statistical yearbook).* Moscow, Goskomstat, 440 p.

Goskomstat, 1990. *Demograficheski ejegodnik SSSR (USSR demographic yearbook).* Moscow, Goskomstat, 439 p.

Goskomstat, 1992–2002. *Social situation and standard of living of the population of Russia (1996–2002).* Moscow, Goskomstat (Russian Federation State Committee on Statistics), one volume annually.

Goskomstat, 1994–2002. *Russian statistical yearbook (1994–2002).* Moscow, Goskomstat (Russian Federation State Committee on Statistics), one volume annually.

Goskomstat, 1997. *Demograficheski ejegodnik Rossii 1997 (Russian demographic yearbook 1997).* Moscow, Goscomstat, 498 p.

Goskomstat, 1998. *Naselenie Rossii za 100 let (Population of Russia: 1897–1997).* Moscow, Goskomstat, 221 p.

Goskomstat, 2001. *Demograficheski ejegodnik Rossii 2001 (Russian Demographic Yearbook 2001).* Moscow, Goscomstat, 315 p.

Goskomstatde of the USSR (Госкомстат СССР), 1990. Социальное развитие СССР *(Social development of the USSR).* Moscow, Статистический сборник, 258 p. (Collection of statistics).

Government Bulletin *(Правительственный Вестник*), December 1991.

Heitman Sidney, 1987. *The third Soviet emigration: Jewish, German and Armenian emigration from the USSR since World War II.* Cologne, Berichte des bundesinstituts für ostwissenschaftlische und internationale Studien, 108 p.

Ivanov Serguey, 2000. Новое лицо брака в развитых странах (New face of marriage in developed countries), *Demoscope Weekly*, no. 63. Online publication, available at http://www.demoscope.ru.

Jiromskaia Valentina (ed.), 2000. Население России в XX веке (*The population of Russia in the 20th century*), vol. 1. Moscow, ROSSPEN, 462 p.

Kharaeva Olga A., 2001. (*Особенности проведения и финансирования мер демографической политики в СССР и РФ* (Features of implementation and financing of the demographic policy in the USSR and the Russian Federation). PhD thesis. Moscow, Moscow State University.

Kozybaev Manash K., Abylkhojine Jouldouzbek and Tatimov Makash B., 1991. La tragédie du Kazakhstan, *in:* Manach Kozybaev (ed.), История и современность. Alma-Ata, Gylym, 254 p.

Krichagin V., 1989. Public health: the time of decisions (Народное здравие: время решений), *in:* Anatoli Vichnevski (Вишневский) (ed.), *Human dimension (человеческом измерении)*, p. 247–267. Moscow, Progress (Прогресс), 487 p.

Lenin Vladimir I., 1966. Aux paysans pauvres, *in:* Vladimir I. Lénine, *Oeuvres, Tome 6*, Moscow, Ed. du Progrès, 45 vol.

Leon David A., Chenet Laurent, Shkolnikov Vladimir, Zakharov Serguei Shapiro Judith, Rakhmanova Galina, Vassin Serguei and Mckee Martin, 1997. Huge variation in Russian mortality rates 1984–1994: artifact, alcohol, or what? *The Lancet*, vol. 350(9075), p. 383–388.

Lioubarski Kronid, 2002. Паспортная система и система прописки в России *(Passport and propiska in Russia)*. Available online at http.www.hrights.ru/text/b2/Chapter5.htm.

Materik, 2000. Гулаг *(Главное управление лагерей) 1917–1960 (Goulag Administration, 1917–1960: documents*). Moscow, Materik, 888 p.

Meslé France, Shkolnikov Vladimir, Hertrich Véronique, and Vallin Jacques, 1996. *Tendances récentes de la mortalité par cause en Russie, 1965–1994*. Paris, INED, 140 p. + 2 disquettes p. (Données statistiques no.2).

Meslé France and Vallin Jacques, 1998. Évolution et variations géographiques de la surmortalité masculine: du paradoxe français à la logique russe, *Population*, vol. 53(4–5), p. 1079–1102.

Meslé France, Vallin Jacques, Hertrich Véronique, Andreev Evgeni and Shkolnikov Vladimir, 2003. Causes of death in Russia: assessing trends since the 50s, *in:* Irena E. Kotowska and Janina Józwiak (eds.), *Population of Central and Eastern Europe: challenges and opportunities*, p. 389–414. Warsaw, Statistical Publishing Establishment, 724 p. + CD-Rom (Book published for the European Population Conference 2003, Warsaw, Poland, August 26–30, 2003).

Meslé France, Vallin Jacques, and Shkolnikov Vladimir, 1998. Reversal of mortality decline: the case of contemporary Russia, *World Health Statistics Quarterly*, vol. 51(2–4), p. 191–206. (Historical epidemiology: mortality decline, and old and new transitions in health, special issue edited by Odile Frank).

Mirsky Marc, 1996. *Medicine in XVI–XIX century Russia (Мирский М.Б. Медицина России XVI–XIX* веков*).* Moscow, ROSSPEN (РОССПЭН), 313 p.

Obolensky-Ossinsky Valerian V., 1931. Emigrations from and immigration into Russia, *in:* Walter F. Willcox (éd.), *International migrations*, vol. 2, p. 521–580. New York, National Bureau of Economic Research.

Piesowicz Kazimerz, 1988. *Studia demografichne*, vol. 94(4), p. 51–90.

Polian Pavel M., 2002. *Les victimes des deux dictatures.* Moscow, ROSSPEN, 573 p.

Popov Gavriil Kh., 2004. Размышления социал демократа о народно-демократической программе (Reflections of a social-democrat on the popular-democratic program), *Moskovskii Komsomolets*, no. 14, January.

Ryzhkov Nikolai I., 1986. *Доклад Председателя Совета Министров СССР на XXVII* съезде КПСС *(Report of the Chairman of the Council of ministers to the 27th Congress of CPSU).* Moscow, Poliyizdat, 42 p.

Satarov G., 2002. *A diagnosis of Russian corruption: a sociological analysis (A short summary of the report)* (Диагностика российской *Диагностика коррупции: социологический анализ [Краткое резюме доклада]).* Moscow, INDEM (Фонд ИНДЕМ). Available online at http://www.anticorr.ru/awbreport/index.htm.

Shapiro Judith, 1995. The Russian mortality crisis and its causes, *in:* Anders Aslund (ed.), *Russian economic reform at risk*, p. 149–178. London, Pinter, 198 p.

Shishkin Sergey, Besstremyannaya Galina, Krasilnikova Marina, Ovtcharova Lilia, *et al*. (Ш ишкин С.В.. (Рук) Бесстремянная Г Е. Красильникова М. Д Ов арова Л. Н. и др.), 2004. *Public health in Russia: payments in cash (Российское здравоохранения: оплата за наличный расчет).* Moscow, State University Higher School of Economics (Издательский дом ГУ ВШЭ), 208 p.

Shishkin Sergey (Шишкин С.В.) (ed.), 2003. *Informal Payments for Medical Care in Russia (Неформальные платежи за медицинскую помощь в России).* Moscow, Moscow Public Science Foundation, Independent Institute for Social Policy (Московский общественный научный фонд Независимый, институт социальной политики), 46 p.

Shkolnikov Vladimir and Nemtsov Alexander, 1997. The anti-alcohol campaign and variations in Russian mortality, *in:* Jose Luis Bobadilla, Christine A. Costello and Faith Mitchell (ed.), *Premature death in the New Independent States*, p. 239–261. Washington D.C, National Academy Press, 504 p.

Stalin Joseph, 1952. Report presented at the XVIII[e] Congress of the All-Union Communist Party (Bolshevik) on the activities of the Party Central Committee, 10th March 1939, *in:* Joseph STALIN, *Questions du Léninisme*, p. 625–626. Moscow, 651 p.

United Nations, Population Policy Databank, maintained by the Population Division, Department of Economic and Social Affairs. New York, United Nations.

United Nations, 2004. *Partnership and reproductive behaviour in low-fertility countries.* New York, United Nations, 115 p. (Sales No.: 03.XIII.5).

United Nations, forthcoming. *World Fertility Report 2003.* New York, United Nations, 504 p. (ESA/P/WP.189)

Urlanis Boris and Borisov Vladimir, 1984. *Народонаселение стран мира* (*World Population*). Moscow, 447 p. (Reference Manual).

Valentei Dmitri I. (ed.), 1985. *Демографический энциклопедический словарь (Demographic Encyclopedic Dictionary).* Moscow, Sovetskaia Encyclopedia, (*Правительственный Вестник* (Government Bulletin), December 1991).

Vallin Jacques and Meslé France, 2001. Trends in mortality in Europe since 1950: age-, sex- and cause-specific mortality, *in: Trends in mortality and differential mortality*, p. 31–186. Strasbourg, Council of Europe Publishing, 334 p. (Population Studies no. 36).

Vassin Serguei A. and Costello Christine A., 1997. Spatial, age, and cause-of-death patterns of mortality in Russia, 1988–1989, *in:* José L. Bobadilla, Christine A. Costello, and Faith Mitchell (eds.), *Premature death in the New Independent State*, p. 66–119. Washington, D.C., National Academy Press, 404 p.

Vichnevski Anatoli G. (ed.), 1994. *Население России. Второй ежегодный демографический доклад (Population of Russia in 2002. The second demographic report).* Moscow, Evrazia, 165 p.

Vichnevski Anatoli, 1997. La démographie soviétique après la seconde guerre mondiale, *in:* Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 157–187. Paris, INED and PUF, 434 p. (Conference proceedings 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995).

Vichnevski Anatoli (ed.), 2004. *Население России. Десятый ежегодный демографический доклад (Population de la Russie en 2002. Le dixième rapport démographique).* Moscow, KDU, 223 p.

Vichnevski Anatoli (ed.), 2005, forthcoming. *Demographic modernization of Russia: 1900–2000 (Демографическая модернизация России).*

Vichnevski Anatoli and Zayontchkovskaïa Jeanne, 1991. L'émigration de l'ex-Union soviétique: prémices et inconnues, *Revue européenne des migrations internationales*, vol. 7(3), p. 5–30.

Zakharov Sergei V., 1999. Cohort analysis of Russian mortality: long-term and short-term effects of inequality of generations in mortality, *Studies on Russian Economic Development*, vol. 10(2), p. 199–211.

Zakharov Sergei V. and Ivanova Elena I., 1996. Fertility decline and recent changes in Russia: on the threshold of the second demographic transition, *in:* RAND (ed.), *Russia's demographic "crisis"* p. 36–83. Santa Monica, RAND.

CHAPTER

119

Population Policy in France, 1896–2003

JEAN-CLAUDE CHESNAIS

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION: AN UNUSUAL DEMOGRAPHIC DESTINY

France, a country of smallholders, has for a long time been highly populated, and was even readily compared in the 18th century to the Chinese world (Braudel, 1986). It was here, from 1760 onward, that revolutionary ideals, such as emancipation from divine authority, abolition of royalty, and a quest for liberty and equality, were born.

The secular decline in fertility was in keeping with this context and reinforced by the reform of the civil code, which abolished primogeniture in relation to succession and introduced equality of treatment among the siblings of the same family by partition of inheritance in equal shares. It was from this that arose the risk of fragmenting plots of land, which was later highly disliked for its direct contribution to impoverishment.

It was in this way, with a good century in advance of its neighbors in Western Europe, that France started on its secular decline in fertility. In certain regions, such as the Southwest, Normandy, and the mountainous areas, the movement arose even earlier and was more marked. It was therefore a rural society that was the first to initiate this movement. As early as 1850, before the change in fertility had arisen in the other countries of Western Europe, the fertility rate in France was no more than an average of 3.5 children per woman, and had indeed already covered half the distance separating tradition (5 or more children) and modern times (2 children or less). Associated with the rural exodus, this weakness in fertility (well below the replacement threshold on account of the high early death rate of the period) did not take long to create pockets of depopulation and even global depopulation by the end of the 19th century.

An overview of the long-term tendencies supports this diagnosis. Finally, the reconstitution of the completed fertilities of the cohorts shows that, since the Napoleonic era and with the exception of the 1919–1946 cohorts that participated in the post-war baby boom, all of the female cohorts have had (taking into account the mortality of their epoch), a fertility below the replacement level. Consequently, ever since the mid-19th century, the economy has been undermanned, in both agriculture and industry, and has had to call on foreign immigration. Here again, the situation was contrary to that prevailing in the neighboring countries, which were still in a phase of high fertility and decreasing mortality, and sending an ever-greater demographic surplus toward the New Worlds. Napoleon wanted soldiers, irrespective of their nationality. A half-century later it was workers who were to arrive. Immigration to France was to become a structural feature. In spite of the intensity of the naturalization policy adopted in 1851 and reinforced at the end of the 19th century, foreigners numbered 7% of the total population by 1931. Because of the low fertility situation, France only participated marginally in the great wave of population that peopled the Americas.

There was another lesser-known but important feature specific to France: its relative excess mortality (reflection of a poorer state of health), which was to endure until the 1950s. Disturbed by the losses in

World War I, frightened at the prospect of a German revenge, demoralized by its economic backwardness, subject to immobility (lack of construction, rigidity of institutions and regulations), crushed by the sugar beet growers and drink salespeople, France remained a victim of the serious social scourges of tuberculosis and alcoholism. Most of the dwelling places were old, dilapidated, devoid of comfort, and particularly lacking in sanitary installations. The smallest village could count several taverns. The numerous and powerful home distillers were not only protected, but privileged. This health situation contributed to obstructing the population momentum of the country, and to increasing absenteeism from work and lowering the productivity of the labor force. It recovered in a spectacular way after the 1950s, when France became one of the countries where average life expectancy was the highest.

Until the 1940s, subfertility, immigration, and excess mortality made France a special case. The turnaround linked to the reconstruction and family policy of the war period lent its baby boom a more powerful and durable character. As for the phase of subfertility introduced into the industrial world of the 1960s, it has, up to the present, affected France less than its chief partners (except for the United States), and this preservation appears to be linked both to a collective mentality (see the Euro barometer reviews on the ideal family), and to a population policy that, although admittedly inadequate, is better adapted than that within the neighboring countries. This historic reversal is clearly insufficient to curb the decline in the French share of the world population: 3.4% in 1750; 2.5% in 1900; 1.7% in 1950; 1.0% in 2000. Recent decades have certainly been marked by an acceleration in population growth in the less developed countries (see Chapter 69), but it is on the European scale that the results are most significant. Estimated at 28 million inhabitants in 1800, the French population on the present territory occupied second place in Europe after Russia (40 million), and represented just over 15% of that of the continent (180 million). In 1939, with its 42 million inhabitants on the eve of the war, the weight of France halved (7.9%). Due to the early decline in fertility (a good century before other Western countries) and the fall in mortality, the population increased by only 50%, whereas it doubled or tripled in the neighboring countries of the North. It must also be stressed that from 1870 to 1940, in the absence of foreign immigration, the French population would have diminished. It is appropriate therefore to analyze here the three large components of the policy that provoked this distinctive drive toward lesser growth (until 1945), and then the stronger growth than other European countries since 1945, with its policies relating to childhood, fertility, public health, and internal and international migration.

I. THE POLICY OF ENCOURAGING FERTILITY

1. Demographic Movement Compared

The singularity of the French fertility policy can only be understood by reference to the historical context of a relative decline in France and an increase in the power of France's larger neighbors, at a time of tension when the number of men still played a strategic role in military confrontations.

Between 1850 and 1950, with a stable territory, the population of France increased by only 16%. This was despite the fall in the death rate, the relative initial youthfulness of the population pyramid, and the migratory contribution that Alfred Sauvy (1945) estimated at 5 million people around 1940. The relative growth in a century had been two times lower than during the following period of scarcely 40 years (1950–1988). All the other great countries in the industrial world had evolved in the opposite direction. Between 1850 and 1950, the other European countries (except Russia), taken as a whole, saw their overall population growth exceed 100%, whereas, between 1891 and 1946, France experienced a long stagnation in its total numbers, because, within the framework of its present borders, although it had turned 40 million since 1891, the population in the 1946 census was still 40 million! France had therefore experienced a cessation of population growth just at the time when it was at its peak in most of the other European countries. As a result, the changes in ranking have been profound and irremediable. It should be recalled that in 1850, France was more highly populated than Japan or the United States.

In 1945, France was in the forefront of demographic stagnation and aging, and in its secular decline in fertility a good century in advance (Figure 119–1). The death rate was higher than that of England. This was abnormal for adults, especially for males , and was due to two main causes: tuberculosis and alcoholism. Compared with other countries at the same level of socioeconomic development, food in France was satisfactory, but living accommodations were frankly insufficient and drinking was excessive. The number of insane people in asylums had risen due to alcohol dependence, and yet 2 years of restrictions on alcohol during the war sufficed to halve the number of confinements. The reduction in the numbers of internees in psychi-

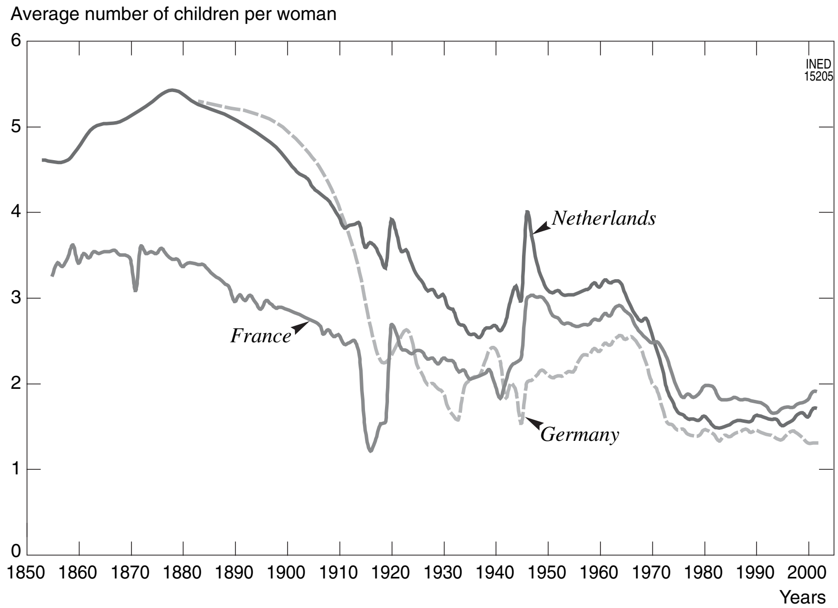


FIGURE 119–1 Total fertility rates in France, Germany, and the Netherlands, 1855–2002.

atric hospitals was not solely due, however, to the war restrictions. The policy of the Vichy government with regard to the insane also largely contributed to it (Meslé and Vallin, 1981). Although, as shown by the recent study by Isabelle Bueltzingsloewen, the charge of genocide is doubtlessly exaggerated (Guyotat, 2003), there was certainly, as written by Max Lafont (1987), a policy of "soft extermination." In Paris, where the alcohol restrictions were better observed than elsewhere, the decrease in the number of people confined to asylums was 90% (Sauvy, 1945).

TABLE 119–1 Birthrates in France and Germany (per 1,000 Inhabitants)

Period	France	Germany
1841–1850	27.4	36.1
1851–1860	26.3	35.3
1861–1870	26.3	37.2
1871–1880	25.4	39.1
1881–1890	23.9	36.8
1891–1900	22.2	36.1

a. The German Eruption

The contrast between France and Germany, and the known military consequences, are clearly visible in the birthrates of the two countries during the period 1841–1900 (Table 119–1).

Already weak during the 1840s, the French birthrate continued to fall until the end of the century (–20%), moving toward a previously never observed value of 20. It was quite the reverse in Germany where the birthrate was maintained at a clearly higher level to the order of 35 to 40 per thousand. This difference was for a long time to be an influence in the power struggle. The German age structure still remained young for many decades, and only fell after 1900. This fall was drastic, however, and the total fertility rate decreased by two-thirds in 33 years, passing from an average of 5 children per woman in 1900 to 1.6 in 1933. But in 1913, disequilibria in the mobilizable forces (to the order of 2 to 1) was such that the National Assembly had to lengthen the period of military service.

b. The English Takeoff

In pre-revolutionary France, fertility was hesitant and relatively high, with on average more than five children per woman. The revolutionary period and the first half of the 19th century (1790–1850) marked a profound change, in that the total fertility rate fell from 5 to 3.5, and the change in fertility was therefore halfway between the pretransitional and the cohort replacement levels. The epoch of the Second Empire (industrialization and the birth of modern capitalism) was characterized by a halt in the fall, and even a slight rise. The decline resumed at the end of the 19th century to the point where the total fertility rate became less than an average of 3 children per woman—a level without parallel (and at a time it should be recalled when fer-

TABLE 119–2 Total Fertility Rates (Average Number of Children per Woman) in France and England, 1760–1900

Period	France	England	Difference	Period	France	England	Difference
1760–1769	5.21	4.88	+0.33	1830–1839	3.76	5.25	–1.49
1770–1779	5.06	5.21	–0.15	1840–1849	3.57	5.00	–1.43
1780–1789	5.13	5.24	–0.11	1851–1860	3.43	4.97	–1.54
1790–1799	4.91	5.67	–0.76	1861–1870	3.50	5.22	–1.72
1800–1809	4.43	5.76	–1.33	1871–1880	3.43	4.91	–1.48
1810–1819	4.37	6.07	–1.70	1881–1890	3.25	4.35	–1.10
1820–1829	4.18	5.98	–1.80	1891–1900	2.93	3.76	–0.83

tility in Germany was still more than 5 children per woman).

English development was quite different. The earlier industrial revolution created thousands of jobs and people found employment at early ages. They were then able to get married more easily and earlier, and have larger families (Kindleberger, 1964; Deane and Cole, 1967). An extraordinary growth in fertility resulted, which attained its historical maximum toward 1820 with an average of 6 children per woman, whereas the long-term tendency before the industrial revolution had only been between 4 and 4.5 (see also Chapter 68).

Table 119–2 compares the course of total fertility rates (the average number of children per woman) in France and England, in 10-year periods between 1760 and 1900.

The effect of this long swell in English fertility was impressive, accounting for a fivefold increase in the English population between 1750 and 1900, and for the massive emigration and supremacy of the English language, especially in the new worlds growing in population. At the end of the 19th century, 440 million people altogether, that is, almost one-third of the world's population, had English as their official language. For 80 years (1800–1880), as is shown in Table 119–2, the difference in fertility to the advantage of England relative to France was of the average order of 1.5 to 1.8 children per woman. It was the combination of an exceptionally high English fertility rate (5 to 6 children per woman) and an abnormally low French fertility rate that produced the contrast in their destinies—rise of the Anglo-Saxon world and rapid relative decline of France and its language (see also Chapter 4, Volume I for a comparative analysis of the evolution of the French and English population pyramids). Because England, having turned exclusively toward the high seas, no longer had any continental ambition in Europe, its demographic domination had only indirect and negligible implications for the continent. But the repercussions of this excess fertility soon became universal, above all thanks to the supremacy of the United States.

2. Birth of a Consensus

The situation was quite different in Germany, the military rival with continental ambitions, where nationalism had become virulent and where, according to the geopoliticians such as Friedrich Ratzel and especially Karl Haushofer[1], there had to be a conquest of the living space (*Lebensraum*) that the German people required. This was to be at the expense of Denmark, and later of France, with the loss of Alsace-Lorraine at the close of the Franco-Prussian war of 1870. In his desire to build a "thousand-year Reich," Adolf Hitler was to carry the application of this doctrine to its final conclusion, when in 1942–1943, his troops occupied almost all of Europe. In only 70 years, the French territory was invaded on three occasions: 1870–1871, 1914–1918, and 1940–1945. Long accustomed to dominating the European continent, France was living with the nostalgia of its super power. From then on, France was weakened (in the 1930s, France was at the forefront of demographic stagnation and aging), and had lost confidence. Meanwhile, in 1940 France did not rearm and refused to fight—it was the "phony war." During the years of occupation France was to favor collaboration with Germany rather than with the Resistance.

But the Anglo-Saxon allies and the "resistants" were able to galvanize the awareness of national unity and to create the feeling of a victory and of entry into a new era. It was essentially during the period at the end of the war that a whole series of institutions such as Social Security and the *Commissariat Général au Plan* (National Planning Body) came into being. A contem-

[1] Heir to this tradition, general and professor of geopolitics in Munich, Karl Haushofer initiated the journal *Zeitschrift für Geopolitik* in 1924. One of his leading ideas was to move borders (by war). He was to be an adviser to Adolf Hitler.

porary or retrospective consensus existed among the ruling elite, between such different individuals as Philippe Pétain, Charles de Gaulle, and François Mitterrand, which held that the defeat and occupation were linked to demographic weakness, and that national reconstruction had to come about through a revival of fertility. In the polls subsequent to the war, public opinion was unanimous in attributing the recovery of fertility to the payment of family allowances, and it is true that at that time they constituted a novelty and their relative amount was high.

The shock of the defeat in 1940 was present in all minds. Henceforth, the linkage to demographic anemia, stressed by the military general staff, was recognized by all classes of the population.

The first indicator of the actual demographic state of the country came with the unexpected defeat by Prussia in 1870. This reverse (that followed the Prussian victory against the Austro-Hungarian Empire at Sadowa in 1866) was seen as a humiliation. One of the reasons put forward by contemporaries—apart from the deficiencies in training of the French troops—was mainly the greater numerical superiority of the enemy.

This type of analysis was then shared abroad, both by English specialists and their German counterparts; demographic differences were everywhere perceived as the main factor in changing the military equation. Population development was primarily viewed according to its external implications, in terms of shifting strategic equilibria. From this point of view, France was described as the seriously sick man (or woman) of Europe.

Nevertheless, right up to the German occupation, save for certain restricted circles, there were relatively few among the civil population who were aware of the real factors in the relative decline of the country. An association, however, was to play a leading role in the evolution of minds: the *Alliance nationale pour l'accroissement de la population française* (National Alliance for the Growth of the French Population). Founded in 1896 on the initiative of Jacques Bertillon, doctor and statistician, its aim was to alert the population and the public authorities to the risks that "depopulation" involved for French society. Its most celebrated recruit was the writer Émile Zola (1899) who, at the time when the Dreyfus affair was in full swing, wrote the famous novel *Fécondité* (*Fertility*), in which he vaunted the delights of family life. (During the Dreyfus affair, Émile Zola introduced a manifesto in the *L'Aurore* newspaper titled "J'accuse"[(I accuse], undertaking the defense of Captain Dreyfus the condemned person, a Jewish army officer accused in 1894 of having delivered some military documents to foreigners. Dreyfus was stripped of his rank and sentenced to transportation, but was in fact a victim of intrigue. His belated rehabilitation, after some 15 years of violent controversy concerning the Jewish question, generated a state of deep division.) Supporters of a pronatalist policy included influential Ministers, such as Adolphe Landry, author of *La Révolution démographique* (1934) and Paul Reynaud, inspirer of the CDF (*Code de la famille* [statutory family law] of July 29, 1939). The CDF, which is seen as the originator of the new fertility policy, laid the foundations for a coherent and pronatalist family legislation.

But in 1940, it was too late. Aware of their weakness and traumatized by the blood bath of 1914–1918 (almost every family had lost a relative), the French people and its leaders were in fear of a well-trained and fanatic Germany. This was the "phony war." Invaded and occupied, France was subjugated, stripped, and robbed of its freedom. Yet 20 years or so earlier, Georges Clémenceau had already issued a warning by declaring that the first of the clauses that should have figured in the *Treaty of Versailles*, was the "need to have more children." Otherwise, whatever the clauses in the Treaty, "France will have lost, since there will no longer be any French people" (Ceccaldi, 1957).

The best expression of national consensus occurred in late 1945, when the dependants' allowance, the taxation principle based on ability to pay and faithful to Article 13 of the *Declaration of the Rights of Man and of the Citizen*,[2] was adopted unanimously by Parliament. Despite some misunderstandings, this taxation device marrying concern for vertical equity (progressive tax rates according to income level) with concern for horizontal equity (taking into consideration the number of children in the household) has been maintained up to present times. This taxation principle is called "quotient familial" (family splitting); it has a moderating influence on the rate of contribution for families with dependent children; it arose from the demands of the Revolution.

3. Reason of State, Social Motivations, and Individualistic Ideals

The population policy adopted in the developed countries of Europe from the 1930s onward has for long been confused with that of the dictators of the extreme right: Nazi (Hitler) and fascist (Mussolini and Franco). Often tainted with racism (especially in

[2] Article 13 of the *Declaration of the Rights of Man and of the Citizen* (1789) stipulates, "The maintenance of the police force and administration expenses require public contributions. These contributions are to be borne by the citizens equally according to their 'contributive capacity' " (combination of income and household size).

Germany), this policy carried out a veritable policing of bodies and considered the individual as being an instrument to serve the requirements of the State (the "thousand-year Aryan Reich" of Adolf Hitler, the reconstruction of the Roman Empire by Benito Mussolini, the conscience control by allying Church and State under Franco). After the war, it was the dictatorships of the extreme left that were to follow, with the deplorable precedent set by Stalin in 1936 (brutal prohibition of abortion, legalized in 1920). The most tragic case was that of Romania under the iron rule of Ceaucescu (1966), where abortion—the main regulator of births—was suddenly prohibited, provoking a wave of births of unwanted, and frequently abandoned, children. There, yet again, the aim was that of power and completely unrealistic. The dictator wanted a population of 30 million people by the year 2000; the figure actually achieved was 22 million.

There is a tendency, nevertheless, to forget that present-day family policies were initiated by the pluralist democracies from the 1930s onward, as a response to fertility deficiency. Thus, the *Beveridge Report* (Beveridge, 1943) in England, although primarily concerned with the struggle against poverty and unemployment, eventually reached conclusions similar to those of Adolphe Landry (1934) in France: that since family size, alongside unemployment, is the greatest generator of inequalities, it is necessary, in order to overcome the dangers of depopulation, to introduce a system that compensates for dependants.

It was, however, the work by Alva Myrdal, *Nation and Family* (1945), that formed the focal point in the reasoning of advanced urban societies. Alva Myrdal and Gunnar Myrdal (1934) raised the question of freedom of choice: to have children (monetary allowances, equipment) or not to have children (freedom of contraception and abortion); and pondered about the need to reconcile family and professional life. Alva Myrdal was to become Minister of Social Affairs and a Nobel Prize winner. Already, in 1934, she and her husband had considered the challenge to freedom within a democracy facing a demographic crisis (Myrdal and Myrdal, 1934), and more particularly within a social democracy.

Many ideas are common to England, France, and Sweden. The French policy, often considered exemplary for the degree of protection it ensures for mother and child, is not solely the outcome of demographic preoccupations (number, continuity of institutions, and reason of State). It is based primarily on social motivations, in particular the concern for equality that consists in guaranteeing a satisfactory level of well-being for all households, irrespective of size. It must be conceded, however, that the full exercise of freedom to choose the number of children is not independent of the conditions of daily life of the families (the cost of a child has become very high, and the contribution to this by public financing is only partial, thus causing budgetary problems for large families). Under these conditions, the two possibilities of choice are unequal—it is easier to defer the setting up of a family than to extend it.

Nevertheless, for some 50 years (discovery of the contraceptive pill, which only became legal in France with the *Neuwirth Law* of December 28, 1967, and abortion, which was only authorized in 1975 by the *Weil Law*), the family policy in France has complied with a new ethical principle that requires the exercise of a new freedom, quickly to become fundamental: that of having only the children you want and having them when you want them. It is probably because it associates some complex individualistic and sometimes contradictory ideals (such as the emancipation of women) with collective requirements for historical continuity and social solidarity that this policy has successfully resisted the hazards of political life and found itself consolidated under the Vichy regime, as well as under the presidencies of Charles de Gaulle and François Mitterrand.

Thus, since the 1960s, a new dimension of individual freedom has gradually asserted itself: the freedom to control the number of one's children and their moment of arrival. This happening concerns the couple, and even more the woman, especially where she has the possibility of contraception, abortion, or sterilization. The right of a woman to prevent pregnancy has materialized with the perfection of safe antinatal techniques that operate durably, if not permanently. Thanks to this medical progress, women have seen a thousand-year dream become reality—that of having access to sex that is free from the fear of an unwanted pregnancy. Sexual pleasure has become an explicit component of individual happiness; it has opened the way to hedonism. The contraceptive revolution has ultimately been the matrix of the sexual revolution.

4. The Important Dates

The first conspicuous measure aimed at encouraging fertility was one that favored the families of public servants, SFT (*Supplément Familial de Traitement* [Family Supplementary Income]), adopted during World War I. Up until that time, there had only been selective and local measures linked to the private initiative of company directors, but these varied considerably from one branch or province to another.

a. 1920: Repressive Law against Contraception and Abortion

At the close of the war and in spite of the reduction in population, State intervention only expressed itself by adopting repressive legislation.

The intention was to remedy the hemorrhage of the war years, when almost 1.5 million were killed, all about 20 years old (Huber, 1931; Vallin, 1973), and the number of men mobilized rose to 8 million, thus causing a serious shortage of births and having the cumulative effect of provoking an increased aging of the population. But the law of July 31, 1920, which suppressed incitement to abortion and forbade antinatalist propaganda and the distribution of contraceptives, was without appreciable effect on the birth curve. This, following a short recovery in 1920 and 1921 linked to the regrouping of families, resumed its downward slope, and from 1926–1928 the total fertility rate was appreciably lower than its pre-war level (an average of 2.3 per woman, instead of 2.5 as in 1910–1913). Abortion had already been assimilated to a crime since 1791 and forbidden by the penal code since 1808. The 1920 law did no more than reiterate and confirm these penal provisions. It was to be reinforced in 1923 by the so-called *Barthou Law* of March 23, which sent cases of abortion before a court of summary jurisdiction, in order to avoid the habitual indulgence of assize court juries. What had previously been a crime became an offense, but the number of adjudged cases doubled.

The failure of the 1920 law (whose objective had been to check the decline in fertility) could be understood in the following way: (1) the confidence of the population had not been restored, and remained in a state of shock; (2) French people have a secular know-how concerning the limitation of births, such as coitus interruptus, clandestine abortion, adoption of separate beds by couples, and condoms. (Bergues *et al*., 1960; Himes, 1963).

b. 1932: Extension of Family Allowances to Wage Earners in the Private Sector

In 1932, there was an important breakthrough in the initiative for social compensation for dependants. It was in that year that the State took over from the private initiative by making membership of family allowance funds for wage earners in industry and commerce compulsory for businesses. Family allowances came about mainly through an initiative by Christian employers. Observing the living conditions of their workers, they noted that those who were single lived comfortably, those who were married without children were uncomfortable, those with one child had some difficulty in living, and those with two children had to appeal for help (often in kind) from their parents who had remained in the countryside. As for families with three or more children, these sank into poverty, since the same wages had to provide for a growing number of mouths. From this arose the idea of paying a supplementary wage to compensate for the cumulative decline in living standards (Ceccaldi, 1957).

But such a measure incurred the risk of distorting competition. Employers paying family allowances then found themselves penalized by higher labor costs. Thus, the 1932 law made the compulsive power of the State available to progressive employers by enabling the contribution to dependants' compensation to be extended to all employers. It should be noted, however, that universal coverage was only attained after the war, with the extension of the system to farmers and independent workers. On the other hand, from then on family allowances could no longer be viewed as employer generosity, but as an absolute right for the wage earner and a legal addition to the salary. Frequently perceived, until then, as a paternalist measure, and also denounced as a method of preventing wage rises, they saw their image transformed and soon to figure (retrospectively) under the headings of social progress and legitimate claims of the working class. In 1936, family allowance benefits were extended to agricultural wage earners. At the end of 1938, more than 90% of the families from the business, industrial, and liberal professions envisaged by the law were effectively receiving family allowances. As for those in the public sector, they continued to benefit from a special indemnity regime for dependants that was more favorable than the general regime of family allowances.

Large families, who most frequently lived in poverty or even destitution, were those, however, who ensured the major responsibility for the reproduction of the society (Vincent, 1946).

The above-mentioned proposals, intended to improve the lot of families with children, were insufficient to prevent the fall in fertility during the Great Depression; at most they were able to limit its extent. The scales of family allowances at that time still varied considerably from one fund to another, but the amount paid out remained small and without a common measure of the needs arising from the dependants. At the end of the 1930s, deaths again exceeded births, and in 1938 the deficit reached 35,000.

c. 1938–1945: The Turning Point (Statutory Order of November 12, 1938, Adoption of the Code de la Famille [Family Law], Improvement in the Family Allowance Scales, Inception of the Social Security Plan)

It was in 1938 precisely, on the eve of the new invasion, that everybody felt that positive (incitement) measures of considerable extent should be taken to counter depopulation. The statutory order of November 12, 1938 included, within a succession of statutory orders, some provisions relative to the compensation for dependent children. Prepared by Alfred Sauvy, then a member of Paul Reynaud's (Minister of Finance) cabinet, it was this order that gave the benefits paid up to then the true nature of family allowances and conferred on them a daring demographic character. The benefits could no longer be identified as a supplementary wage allied to the trade carried on and the uncertainties of wage negotiations. They became independent of the salary and of the company. Their amount, on the other hand, varied according to the ranking among the siblings (it was progressive in order to take account of the additive nature of the upkeep costs of children). Paid monthly, their rate was fixed according to the average salary of the legislative department. It was established at 5% of this salary for the first child, 10% for the second, and 15% for each of the following dependent children, with a guarantee of annual revision (updating), according to the change in wages noted in the interval.

At the same time as it achieved an important uniformity of the rates, the 1938 reform immediately caused a substantial increase, in the order of 50%, in the amount of benefits paid out. It envisaged, moreover, an increase to compensate for the loss of salary suffered by the mother in giving up a professional activity in order to devote herself to her family (this was the future Single Salary Allowance, adopted in 1941). This increase corresponded to 2% of the total payroll and increased the overall cost of this reform by 3% or 4% of it. In 2 years, between the beginning of 1937 and the beginning of 1939, family allowances increased two or three times more than they had done over the preceding 20 years.

By the end of 1938, the broad outlines of the CDF, which marked another major innovation, had already been drawn. This statutory law was the work of the HCP (*Haut comité de la population* [High Committee of the Population], where Adolphe Landry was based), created some months earlier. It was adopted by a Chamber stemming from the Popular Front. The text of the law sought to discourage voluntary infertility—frequent in France in the period 1890–1939—and to encourage more particularly those families whose composition would contribute to an increase in population, that is, families with at least three children (Sauvy, 1954). In practice, however, the rate of family allowances was scarcely restructured except for very large families (for families with three children, the rate remained unchanged at 30% of the reference salary). For the family with two dependant children, the rate was revised downward (10% instead of 15%), and the allowance for one child was abolished. It was replaced by a high birth premium designed to encourage newlyweds to have their first child quickly. In 1941, under the Vichy government, the allowance for the mother in the home was increased by the single salary allowance, which quickly occupied a dominant position in the system for mitigating the cost of dependants—since 1943, indeed, the payment of this allowance has accounted for more than one-half of the total allowances paid out. It is appropriate to resituate the context of the period, at a time when most of the mothers did not have salaried employment and devoted themselves to raising their children. The payment of the single salary allowance came about after marriage.

5. Priority of the Child and Peak of Family Policy

The war crystallized demographic awareness at the same time as it gave rise to a new, less individualistic and calculating state of mind. After the poverty, trials, privations, and food rationing, the merits of solidarity were recognized by all as being evident. The family was in fashion, and it increasingly personified the symbol of happiness—the privileged place for personal development. Against a background of destruction and ruins, history was being refashioned. It was the atmosphere of reconstruction, the pioneer spirit.

With the 1945–1946 decrees, family allowances became part of the general system of Social Security managed by Pierre Laroque, himself a former member of the Adolphe Landry cabinet. The principle of compensation for dependants was established:"Family allowance is justified for the same reason as retirement; it represents in short a salary advance that society grants to the apprentice, who will in turn repay it" (Sauvy, 1954).

Sickness and maternity insurance was also set up, and development of family policy continued with an improvement in the contributions and alignment of rates in the provinces with those of the capital.

But the contribution rates deducted from the salary (subject to a ceiling) for the benefit of the family branch increased from 5% in 1938 to 12%–13% in 1947–1948, and was soon to culminate in the 1950s at around 16%–17%. The 1947 *Charte des prestations familiales* (Law on Family Benefits), which envisaged an indexation of productivity increases, was not applied. The amount of the family allowances was to follow the simple progress of the consumer price index, but the frequent delays in so doing were to account for the enormous gap that arose in the course of the decades between the volume of benefits contributed to young parents by the family branch and the amount of retirement and sickness insurance expenditures.

6. Implications

The real recovery in fertility did not begin (as is frequently written) in 1942—it was then only the matter of a minor change (+10%), linked partly to the movement of returning prisoners—but rather took place at the end of the conflict (+30%), after the return of the soldiers and the sudden increase in the rate of family benefits.[3] This baby boom was not limited to France, but affected all of the countries of the Western allies. It was not experienced, however, in Germany, Italy, or Spain, or in the countries of Central and Eastern Europe that had fallen under the Soviet yoke. This baby boom, however, had certain features that were peculiar to France.

1. The recovery there of the post-war years was more sudden and powerful than in the homologous countries. The relative position of France in the international natality hierarchy between the pre- and post-war periods was inverted. From 1945 to 2002, the average rate of growth of the French population was higher than that of Germany, Italy, and the United Kingdom.
2. The demographic characteristics of the baby boom corresponded to the raising of the family allowance rate. Actually, it was with the emphasis on the second child that the expenditure effort was by far the most marked (sixfold). With two dependent children, a family received almost the equivalent of a second wage (in the aftermath of the war, three-quarters of the mothers with two children remained in the home and received a single salary allowance). This change in the rate created a very strong incentive to abandon the family model so widespread in the pre-war France, that is, that of the single child. It was in agreement, moreover, with the wishes of the legislator. There was thus a concordance between the variations in the legislation scale (differential incitements) and the increase in fertility following the pre- and post-war birthrates—couples without children or with a single child became rarer, to the advantage of limited or average families (those with two children especially, and a few with three), who were the main beneficiaries of the single salary allowance).

   It is also appropriate to bear in mind that, at that time, the payment of monetary benefits was a complete innovation in a country that was coming out of a barter economy.
3. The revival assumed greater importance than elsewhere because it generated 28 cohorts (1946–1973) with numbers that were one-third higher than those of the previous cohorts. It was the weighting factor of these cohorts, moreover, that was to cushion the fall in the number of births from the 1980s onward, and it was this momentum that made it evident that the country would be immunized against a further fall in fertility.

   Under the influence of Georges Pompidou, family policy became of secondary importance. At the time when productivity took off (this was the heart of the "Trente Glorieuses" ["Glorious Thirty," i.e., the 30-year period from 1945 to 1975], according to the Jean Fourastié formula, 1979), financial transfers for the benefit of young parents scarcely followed prices (disindexing had in fact started before 1950, in violation of the CDF). Above all, there came about a distortion in family policy. Its essential purpose had been to correct horizontal inequalities of income, that is, between families whose only difference was the number of children, whereas there was a gradual increase in benefits conditional on resources that henceforth transformed horizontal transfers into vertical transfers, thus adopting the form of redistribution from rich to poor. The application of the same reasoning to sickness insurance or state pension schemes was to arouse fierce opposition, but properly speaking there was no existing family lobby, and even less for children. Owing to the increasing numbers of exclusion thresholds, family policy declined in importance (it is easy to lower inflation-adjusted ceilings) and especially in universality. It became opaque, incomprehensible,

[3] Many analysts, among them Gérard Calot (see his Chapter 105 in this volume) date the recovery to 1942, but with 2.1 children per woman, fertility was then appreciably lower than what it had been in the previous period (on average 2.18 for 1927–1939). In 1946, on the other hand, there was a phenomenon without precedent since 1890–1893, when fertility reached 3 children per woman and maintained this for the following four years (1947–1950) and subsequently remained above 2.5 until 1969.

> and remote from the initial concern about demographic equilibrium. Family policy therefore became an aid policy for specific groups—the attribution rates vary from one case to another, and the number of benefit categories in 2002 reached 26 or 28, instead of only 6 in 1946.

This decline in family policy is illustrated by the evolution of contribution rates (deducted from earnings) intended for family allowances. First of all there was an increase, especially during the wartime, expressed by a rise of more than double (5% in 1938, 12% in 1947), followed by a new increase (16.75% in 1951–1958). The following period has been one of decline, as the rate has now returned to 8%, its level during the war.

7. Financial Drift of Social Regimes in the European Union

What happened in France has often taken place in an even more marked fashion among other European Union (EU) countries. Controlling social services (retirement, health) is an old sea serpent. The fall in fertility and disappearance of large families enabled substantial economies to be made with reference to the family sector, and was therefore redeployed in disguised intergenerational transfers from the rising generations (children) to the retiring generations of aged people. This applies to all EU member States, whose budgetary preferences clearly favor the old rather than the young.

Although France has achieved highly effective results in the retirement area by combining the contradictory qualities of early retirement and generosity and universality of pensions (occupying 6th position among the 15 EU members in terms of average pension measured according to purchasing power), in 1999 the country devoted 13.4% of its gross national product (GNP) to the payment of retirement pensions. This is one of the highest values in the world, and actually exceeds that of Germany, and even of Sweden. Only Italy attains a higher level of expenditure, due to the exceptional frequency of very early retirement and the rates of salary replacement by the pensions.

If the benefits paid in the form of pensions and health care are added together, we now come to countries such as Sweden, Germany, Italy, or France, where it is nearly 25% of the GNP. Taking into consideration the inversion of the population pyramid, which is often only at a halfway point, the viability of such devices in the long term can be questioned, and the tax burden and increase in public debt are regularly invoked by international financial institutions for their effects in terms of the loss of competitiveness in highly aging populations.

There is a striking contrast in the expenses allocated to young parents and their children, for the observed maximum reaches only 3.8% and 3.5% of the GNP (Denmark and Sweden respectively). France, where child policy is reputedly pronatalist, no longer merits this qualifier (at least on the financial level), as it occupies a similar position to that of Germany (3%). But it is especially the position of countries most sparing in public expenses in favor of their children that merit the most attention. There are, in fact, two EU countries where the financial distribution in favor of children is less than 1% of the GNP: these are Spain and Italy, countries where the fascist dictatorships (Franco and Mussolini) had set up family allowances, but where these have suffered a continual monetary erosion to the point of becoming derisory and no longer of benefit to even the most humble households. It should be noted in passing that these countries are precisely those where fertility is lowest (Eurostat, 2002, 2003).

Even in France, it is a long time since family allowances occupied the forefront of social expenses. The negative correlation between the erosion of allowances and the decline in fertility is shown by the spectacular case of the Saar (French, then German) in the immediate post-war period.

8. Balance of Family Policy: Strong and Weak Points

a. Strong Points

France's fertility policy has the unique advantage (that it shares only with Belgium) of pre-elementary schooling that is both public and early: from the age of 2 years, almost 50% of French children are already attending nursery school. At the following ages (3, 4, and 5 years), all children are enrolled. The generalization and nonpayment for nursery schooling are the envy of neighboring countries, where parents have to undertake the care of their children up to the age of 5, 6, or even 7 years. From the age of 2–3 years, the network of nursery schools is well thought out and available to the greatest number of children. It fulfils the needs of both parents and children and is a powerful socializing instrument.

It only remains to close the gap between the end of maternity leave and entry into the nursery school, which involves the need for the development of solutions, such as recourse to maternal assistants, parental leave, or day-care centers.

Another original French feature is the system of direct taxation, which is, in fact, a moderator of progression adjusted according to the number of dependent children. The system was set up at the time of the Liberation, following centuries of debate. It is based on an idea of fiscal justice. Instead of applying the taxation rate to the income itself, but by income share, taxation is applied to the living standard of the household. This idea of family support, which is nowadays sometimes incorrectly interpreted and considered to originate from the right, dates back in fact to the period of Enlightenment in the 18th century. It fulfills a republican and revolutionary ideal inscribed in the *Declaration of the Rights of Man and of the Citizen*. Whenever taxation is evoked in this fundamental text, there is mention of the need to take into account the contributive ability of the person paying (i.e., the number of dependants).

b. Weak Points

Since the accumulation of benefits conditional on resources (late 1960s), an increasing number of beneficiaries have been excluded from benefits under family policy, thus marginalizing it. The family policy was clearly of high priority in 1946 (concern was for national reconstruction based on the revival of fertility) with more than 40% of the national social budget (instead of hardly 10% at present). Even more symptomatic is its wretched aspect, which has appeared and grown over recent decades. It is the Family Allowance Funds that manage the allowances for adult handicapped persons, and above all carry out the payment of the RMI (*Revenu minimum d'insertion* [income support: unemployment pay]). Children are thus relegated to the same ranking as the poor, infirm, and unemployed. Here, there is clearly a diversion of aim. Just as the policy of sickness insurance aims to ensure a financial transfer between the healthy and the sick, and the retirement policy to pay replacement incomes from the present active population to the former active members, the function of the family policy is to ensure compensation between small families and those without children and those of greater size. But this principle has often been ignored to the benefit of a more visible and political consideration, such as the struggle against poverty, which itself entails employing fiscal means.

The underindexing of family allowances also goes a long way to explaining the fall in the share of benefits within the GNP. It has enabled the appearance of the famous CNAF (*Caisse nationale des allocations familiales* [National Agency for Family Benefits]) surpluses, which, regardless of the principle of separating various branches of social protection from the budget, have regularly been employed to make good (without repayment) the deficits from other sectors. Families with dependent children have found themselves impoverished through this, and all the more seriously as the number of their children is greater.

Although other countries also have child support mechanisms in mind, it does nevertheless appear that in France the strong points appreciably outweigh the weak. The family policy, however, still presents many shortcomings. The most serious concern early childhood and post-adolescence (children around 18–20 years-of-age enter a period of maximum cost, whereas the family allowances cease all payment).

II. PUBLIC HEALTH POLICY

Defining the role of health policy in lowering mortality requires much greater subtlety, for the determinants of the main causes of death are multiple and intertwined.

1. Initial Delay

Toward 1930, life expectancy in France was 10 years lower than in the Netherlands (54 and 64 years, respectively), and France occupied the place in Western Europe, with a net disadvantage for men.

But after having long occupied a mediocre ranking in terms of the struggle against death, France now finds itself well placed among the advanced countries. Although the leading position has been occupied by Japan since the 1970s (a model of individual and collective providence), the second place is held, almost equally, by the large countries of Western Europe (France, Germany, Italy, Spain, and the United Kingdom) (Figure 119–2). The convergence of the health performances contrasts with the differences in fertility. Thus, toward the year 2000, Spain and Italy presented total stabilized fertility rates to the order of 1.1 to 1.2 children per woman; Germany around 1.3; the United Kingdom, in slight regression, around 1.6–1.7: and lastly France around 1.8–1.9. These relative divergences in fertility are important and may be due to differences in collective psychology, early childhood institutions, territorial and national policies (explicit or not), and fertility support. Progress concerning the struggle against death form a consensus and may be related to the general socioeconomic level of development and to technical means (vaccines, medicaments, equipment, etc.), which are more easily transferable from one country to another. This development is the result of voluntarist actions carried out

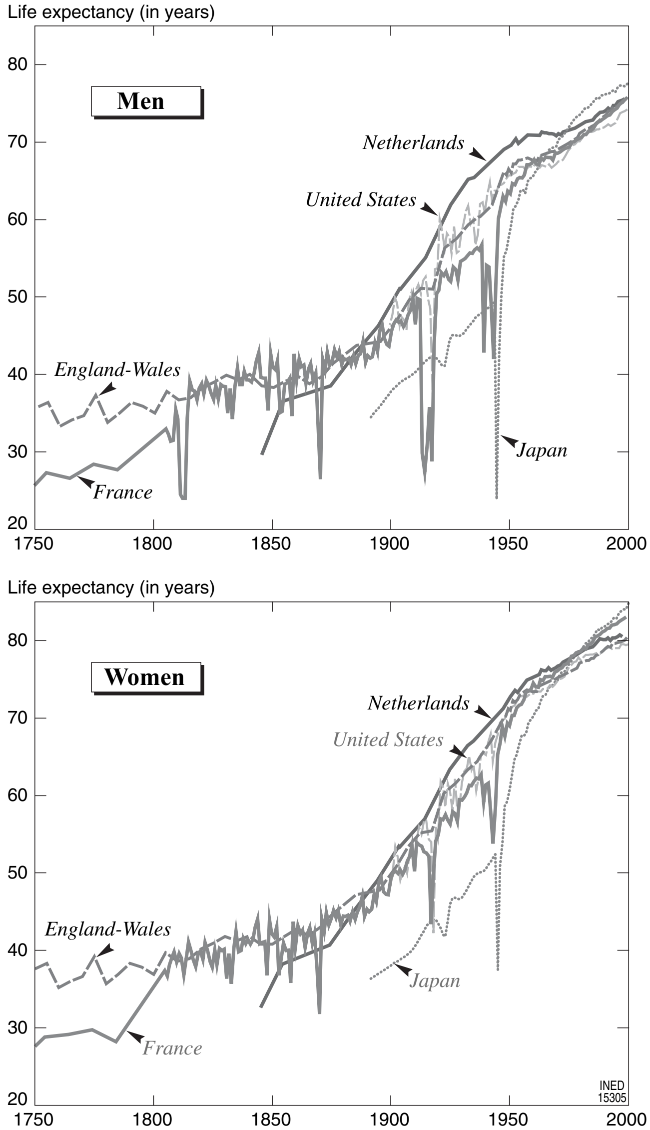


FIGURE 119–2 Trends in life expectancy at birth in France and in other selected industrial countries, since 1800. (Sources: Vallin and Meslé, 2001, 2004.)

by public authorities. In the first place, there was the effort to eliminate tuberculosis, which was strengthened by the extraordinary revival (construction and renovation) of the housing policy in the 1950s.

Although alcoholism, the traditional scourge of France, has reduced under the IVth Republic phase of society modernization and urbanization, the success has been relative. Admittedly, the number of deaths from this cause has fallen since the 1940s due to prevention, but excess mortality allied to alcoholic poisoning among men sets France apart from other EU countries and even the OECD (Organization for Economic Cooperation and Development). Alcohol consumption has certainly diminished among the young generations, but the struggle against alcoholism remains insufficient in relation to that prevailing in the northern countries.

2. A Prevention Model: Road Accidents

Road deaths increased up until the first oil crisis, placing France in the worst international rank category. Annual deaths rose from 3,354 in 1950 to 16,617 in 1972, not because roads were becoming more dangerous (the number killed per kilometer traveled was, on the contrary, already greatly decreasing[4]) but because the expansion of motor vehicle traffic had been very rapid. Arising from ecological awareness, it was at this time in the seventies that thought was given to the cost and advantages of choosing the motorcar. The oil crisis had had a major psychological effect on the way the automobile was viewed in relation to collective transport or the bicycle. A discussion arose about the damage caused to the urban atmosphere, to congestion, environment, and road insecurity. In successive stages, the government adopted stern measures, such as mandating the wearing of a seat belt, limiting speeds, reducing and monitoring drinking and driving, technical checking of vehicles, and so forth. Above all, the public became aware of the annual toll of the roads, carrying off thousands of human lives in the prime of life, and even more slyly leaving tens of thousands of young people handicapped for life.

France formed a group apart, with an atypical road death rate double of that of England, for example (or of Japan). Over recent years, the scourge has been designated by the name "road rage," and has become the object of priority action by the government. The results are impressive: death on the road is declining regularly. By 2002, road deaths were half as many as in 1973 in absolute figures, and the rate reduction was even more marked, for the population had in the meantime increased. Sanctions are becoming ever more severe against reckless or drunk drivers, and these people are now brought before the courts and frequently awarded severe terms of compulsory imprisonment.

The same cannot be said for death from cancer, which continues to worsen (about 150,000 deaths per year, or 28% of the total mortality). It is highly resistant, doubtlessly linked in part to the aging of the population and also to the delay in prevention. It is in France that the frequency of deaths from cancer is highest, due especially to the importance of lung cancer among both women and men (more women

[4] In fact, the number killed per 100 million passengers/km fell from 10.9 in 1953 to 5.4 in 1968, and to 3.5 in 1976 (Vallin and Chesnais, 1975; Chesnais and Vallin, 1977).

smoke than men, especially at young ages). Nevertheless, the decline in deaths from lung cancer among men has started.

3. Resistance to Alcohol and Nicotine Addiction

The *Évin Law of* January 10, 1991 marked a first in the struggle against nicotine addiction. Although the law aims to combat alcohol and nicotine addiction, public opinion has only remembered the basically more innovative effort to fight against nicotine. It involves the prohibition of smoking in public places (except those specially reserved for it), in order to protect both smokers and non-smokers from inhaling toxic substances. France is thus in line with the Anglo-Saxon countries (United States and England), which were the first to undertake a firm and effective struggle against the ravages of tobacco. The results are still inadequate.

4. Records of Infant Mortality

Through the total number of institutional and socioeconomic factors that it depends on, infant mortality is one of the most robust indicators of development. It is no surprise that Japan heads the world classification, followed quite closely by the advanced Western countries such as Germany, Italy, France and the United Kingdom, which present fairly similar values. The setting up of maternal and infant protection (post-war), followed by the program for preventing prenatal mortality (1970s), comprising regular monitoring of the pregnancy and of the infant's development, and lastly the systematic application of vaccinations have had the effect of provoking a spectacular fall in infant mortality and eliminating the French lag. Between 1930 and 1932, the French infant mortality rate reached 77 deaths per 1,000 live births, thus largely exceeding that of Norway (46), the Netherlands (48), Switzerland (50), Sweden (54), and England (64), among others. Between 2000 and 2002, with a rate close to 4 deaths per 1,000 births, France had, for some 20 years, descended beneath the 10 per thousand floor, long considered to be impassable. It has become one of the front-runners among the countries with the best health record (Figure 119–3).

Premature death has fallen, but not only just before the age of 1 year; death before 5 years has in fact almost disappeared, and reference is now increasingly made to death before 60 years. For the newborn, this is increasingly due to technical progress (especially in the area of neonatology) and to prevention (medicalization of childbirth, generalization of prenatal medical visits, struggle against prematurity, echography, electronic monitoring of childbirths, etc.).

Although clearly declining in the long term, however, the regional disparities are far from disappearing. This indicates that differences in mortality are rooted in lifestyles, eating habits, differences in industrial pollution, and in the consumption of alcohol and tobacco.

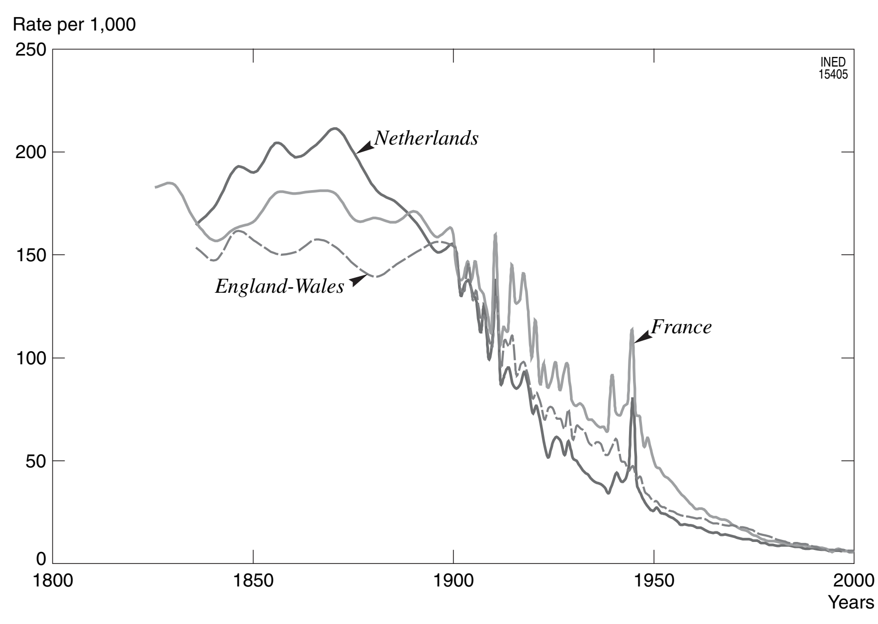


FIGURE 119–3 Trends in infant mortality rates in France, England-Wales, and the Netherlands, 1830–2000.

5. Disappearance of Infectious and Parasitic Diseases

The most striking feature of 20th-century mortality is the collapse of infectious and parasitic diseases (Vallin, 1988; Vallin and Meslé, 1988), and that henceforth it is the progress in chronic diseases (principally cardiovascular diseases, but also cancer) that dominates pathology (Meslé and Vallin, 1988). Following the law of October 30, 1935, the reporting of contagious diseases by families became obligatory in France (obligatory for doctors and midwives since the end of the 19th century). In a few years, between 1938 and 1950, the requirement for vaccination was extended against diphtheria, tetanus, typhoid and paratyphoid, and tuberculosis (1950), and also in 1964 for poliomyelitis. This generalization of vaccinations was to cause a fall in infectious diseases. It should be borne in mind that in 1950, tuberculosis accounted for the deaths of 24,364 persons (i.e., 4.6% of total deaths).

6. A Constellation of Progress Factors

But among the factors in the victory over death (fully described in Section I of Volume II), it is important to recall here the progress in hygiene, disinfection, and asepsis; advances in surgery made during World War I; the birth of laboratories; the availability of drinking water and elimination of waste; the modernization of hospitals; perfection of effective medication; globalization of medical research; discovery of the first sulphamides in 1935 (providing treatment for those with bacterial, pulmonary, and digestive infections); and finally, the decline in a certain millennial fatalism with regard to disease and death (Biraben, 1988), and the post-war discovery of antibiotics that accelerated the eradication of infectious diseases.

A turning point appeared to have been reached, but it was only a stage. The advances against death by infectious diseases appeared to be partly nullified by the increase in the so-called diseases of civilization (transport accidents, pollution, drug abuse, alcoholism, tobacco addiction, etc.). Sometimes, this resistance was also imputed to the powerlessness of struggling against diseases due to the biological aging processes. Increase in the average length of life has therefore proceeded in stages that were sometimes rapid and sometimes slow, according to the advances in medicine.

7. Institution of Social Security (1945)

But the most fundamental factor in the French recovery took place without doubt in 1945 with the introduction of social security, which very soon facilitated access to medical coverage. Although the French post-war population growth exceeded that of neighboring countries, it was not solely due to an advantage in fertility but also to a more rapid mortality reduction than most of the European countries, especially among females. France made up its delay, and then even moved ahead. This fact is related to the extension of social security coverage (health and retirement), restricted until the 1930s to the limited category of salaried workers. Since 1970, the faster progress in the means and personnel of the health sector (induced by the soaring health insurance expenses) is probably the reason for the renewed fall in mortality. Similarly, the important decline in deaths among elderly people (60 years and over) during the same period is probably attributable to the intense effort in redeploying social finances in their favor (massive increase of the minimum old age pension, high adjustment of purchasing power relative to retirement pensions), added to strong income growth and improvement in living conditions. At the beginning of the 21st century, the female life expectancy at birth was 83 years, only 2 years from the world record of 85 years held by Japan. The male excess mortality, continually increasing since the 18th century, had begun to diminish—the average life expectancy of men exceeded 75 years.

It is indeed possible that due to a number of new factors (emergence of new diseases, increase in obesity, environmental neglect), we are moving toward a slowing down of the recent progress, obtained moreover at an ever higher cost (growing deficit of sickness insurance), but up to the present, the pessimistic forecasts have been denied by the facts.

Added to the economic anxieties concerning health are those relating to territory and to difficulties, in times of high unemployment, of integrating foreign populations. The latter largely condition migration policies both through national and regional development and through immigration policy.

III. NATIONAL AND REGIONAL DEVELOPMENT

1. National and Regional Development Policy

a. Paris, a "Primaspatial"[5] City

The Parisian agglomeration occupies a disproportionate place in relation to the other large French towns

[5] According to the expression used by Denise Pumain and Daniel Courgeau (1988).

in that its population has always remained six or seven times higher than the second city, Lyons or Marseilles according to the epoch.

France is a macrocephalic country, with a hypertrophic center, congested and with many aged provinces that are often suffering depopulation. Considerable disparities (demographic among others) characterize the national area and hence the necessity for national and regional development.

Already invoked after the post-war period by Jean-François Gravier in the founding work *Paris et le désert français* (*Paris and the French Desert*, 1947), the wish for decentralization in the service of national and regional development has been ignored since the 1970s (decline in planning and the DATAR[6]), but has now become a major preoccupation with a central ecological dimension—rural people are no longer considered simply as farmers, but henceforth as guardians of the environment. In addition, the principle of equality of access to the public services (post offices, schools, hospitals, and infrastructure), which was for a time called into question, has been preserved. Guidelines for action are regional balance, the creation of employment, and correct utilization of European structural funds.

The integration of Europe and its expansion toward central Europe have redrawn a vision that is directed eastward.

Despite a lower polarization of internal migrations, however, serious disequilibria remain to be reduced between the overpopulation of the Ile-de-France and the desertification of the disadvantaged rural zones: the differences in density and potential are extreme. Tourism and mobility (secondary residences) can only change these marginally, for in the end it is only the presence of employment that dictates the setting up of new households. Different formulae have been attempted, such as the creation of regional centers and establishment of new towns, a law on decentralization, and even the setting up of high-speed trains (TGV), but the Parisian region, situated at the heart of the country, continues to drain manpower and wealth. Lastly, as in the United States, in a context of deficit State expenditure, the question arises as to whether decentralization does not also pursue the aim of transferring expenditures to the regions and municipalities.

b. Space and Demographic Transition

The secular fall in fertility began in the countryside, thus accounting for the relatively low migration toward the cities and delay in urbanization. The long period of demographic atony contributed to the disequilibria between Paris and the provinces. This migratory movement took place to the detriment of the rest, and rural preponderance was also allied to the low agricultural productivity and the predominance of small concerns. On the eve of World War I, more than half the French population was still living in the country, in rural districts of less than 2000 inhabitants clustered around the chief town. It was not until 1982 that France became three-quarters urban, and it was only after World War II, with the end of the demographic freeze, that urbanization became explosive. The population abandoned its lethargic state of secular quasi-stagnation in order to grow until the 1970s at a rate of about 1% per annum, accompanied by a rapid industrialization of the economy and attraction of its labor force from the country districts.

The expansion of the urban system frequently took on two forms: the birth of new towns and the extension of the existing fabric (see Chapter 61). There was then periurbanization rather than disurbanization (Pumain and Courgeau, 1988). Many rural districts belonged to the ZPIU (Zones de peuplement industriel et urbain [Industrial and Urban Population Zones]). The secular reversal of the migration between country and towns succeeded in giving rise to talk about "the end of towns" (Chombart de Lauwe, 1982); the faltering of the industrialization and urbanization process at the end of the 1960s gave way to tertiarization during the 1970s. The industrial crisis then reached the North and East, creating gigantic zones of temporarily inactive industry and unemployment, and inducing the developers to act without delay by reconversion aid and the setting up of new businesses. A large number of départements (akin to counties) were continuing to lose population through net emigration, but the essentially industrial and urban areas (Paris basin, Mediterranean South, Lyons, and Marseilles) continued to grow.

Departures from or nonentry into the agricultural world, however, led to considerable migration to the towns, where the possibilities of industrial or tertiary employment were concentrated. Nevertheless, mobility away from the rural communities was largely intraregional: two-thirds of the migrants remained within the same region.

In the early 1980s, there was a reversal in the migratory current. Up until then, industrial and urban centralization functioned as a suction pump that emptied most of the peripheral regions of the country to the advantage of a small number of regions. In the 21st century, a movement in the opposite direction is taking place, where old industrial and urban regions are being emptied to the advantage of other regions, previously considered to be disadvantaged but

[6] Delegation à l'Aménagement du territoire et de l'action régionale (Agency for Land Planning and Regional Development).

better equipped for the implantation of services and for attracting the working population of the tertiary sector. England is still a more striking case, in that the industrial sector has almost disappeared, following privatizations that quickly led to delocalizations.

c. Centralizing Tradition

Its central position in the heart of the country gives Paris a natural geographical advantage. But this asset has been reinforced by some political choices according to a very ancient centralizing tradition. Since 1842, the choice of railway plan put Paris in the center of the network, thus favoring the administrative conception of a network radiating from a center in Paris. The servicing of the industrial regions was not considered to be a priority, and the command functions were therefore concentrated in Paris, which monopolized the centers of political, economic, and financial decisions, and even that of research (Figure 119–4). The power of Paris has, however, been gradually relativized by the reversal of the above-mentioned migratory flows, especially those between Paris and the provincial areas.

d. Some Success in National and Regional Development

The DATAR was created in 1963 by the interministerial structure allied to the planning entrusted to the *Commissariat général du plan* (General Planning Commission). It was provided with ample means and played, at least until the 1970s industrial crisis, an important voluntary role. It constituted an essential piloting instrument, which weakened with the mythical rallying call for the correction of regional inequalities. It succeeded, however, in opening up certain poor regions, with the 1968 *breton plan*, the Massif Central development program in 1975, and the restructuring of the urban web. Less successfully, it set up *métropoles d'équilibre* (regional centers) and important development schemes. Lastly, it created nine new towns: five in the Paris region and four in the provinces.

IV. IMMIGRATION POLICY

1. 1945: The Desire to Integrate Immigration into the Population Policy

In the aftermath of the war, the planners aimed to include immigration among the objectives for relaunching the population of the country. It was the time when Alfred Sauvy (1946) evaluated the immigration needs of France, but the avowed aims were never to be followed, for employers adopted various channels, legal or otherwise, and sometimes even went to recruit workers from abroad. In addition, the robustness of the baby boom lessened the preoccupation with population, creating, however, deficiencies of labor force in certain sectors.

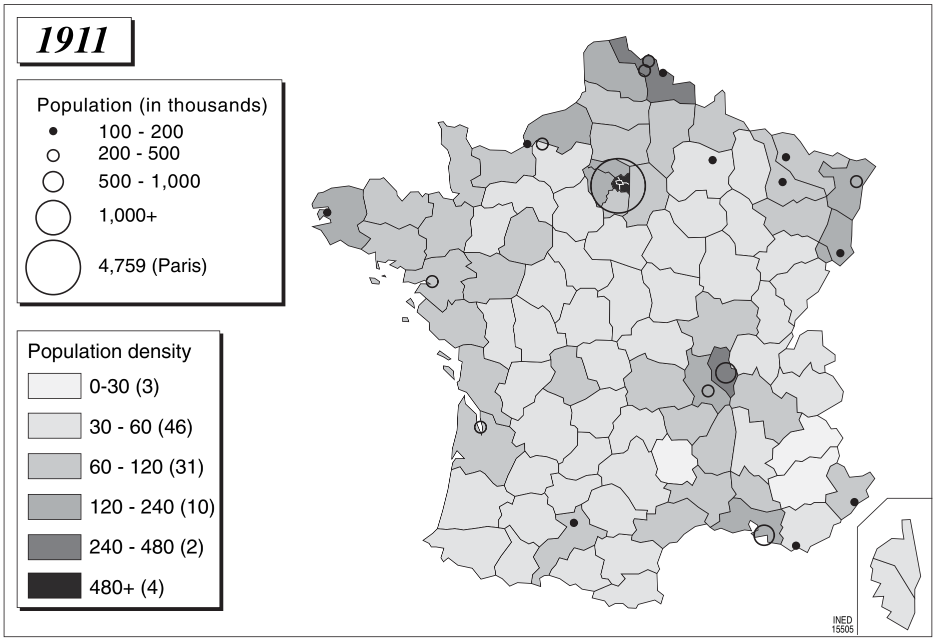
With the first oil crisis in 1973, the government altered its position and aligned itself with most of the European OECD members in closing the borders to the inflow of new workers, encouraging return, and integrating the numbers of the foreign population legally installed within the territory. It was the so-called policy of zero immigration. We now know that this policy failed everywhere (in Germany, Switzerland, France, etc.). By its very nature, foreign immigration is a life project that does not concern just a single individual, and the granting of a bonus is therefore without effect as an incentive to return. (Zero illegal immigration is mentioned later in the chapter.) There again, this is to disregard a powerful reality that imposes itself on all rulers: international disequilibria, political crises, massive poverty in the countries of the South, the role of families already in place on arrival, activities of smuggling rings, corruption around the borders (illicit payments, falsification of documents etc.), everything that combines to sustain migratory pressures.

In addition, from the internal point of view and despite the persistence of a high rate of unemployment, there were indeed important sectoral needs among the workforce (domestic services, construction, hospitals, and personal assistance, etc.). In France, as elsewhere, unemployment was no longer experienced as an indignity or social shame, and an increasing number of jobs were considered to be degrading by reason of what had to be endured (excessive working hours, low salary, dirty work).

This distortion, in fact, between rhetoric (cessation of immigration) and reality (sustaining entry flows) is at the origin of misunderstandings, and to a large extent doubtlessly of pressure from the extreme right.

2. Freedom of Movement (19th Century)

France is traditionally an immigration country. According to the 1851 census, there were already 379,000 foreigners (more than 1% of its total population). This is the threshold statistic of 1% that for European countries corresponds to the end of secular emigration and beginning of the contemporary era of immigration. Thus France had a good century of advance over its European partners with regard to immigration: Among its neighbors in Western Europe, this threshold was only crossed between the 1960s and



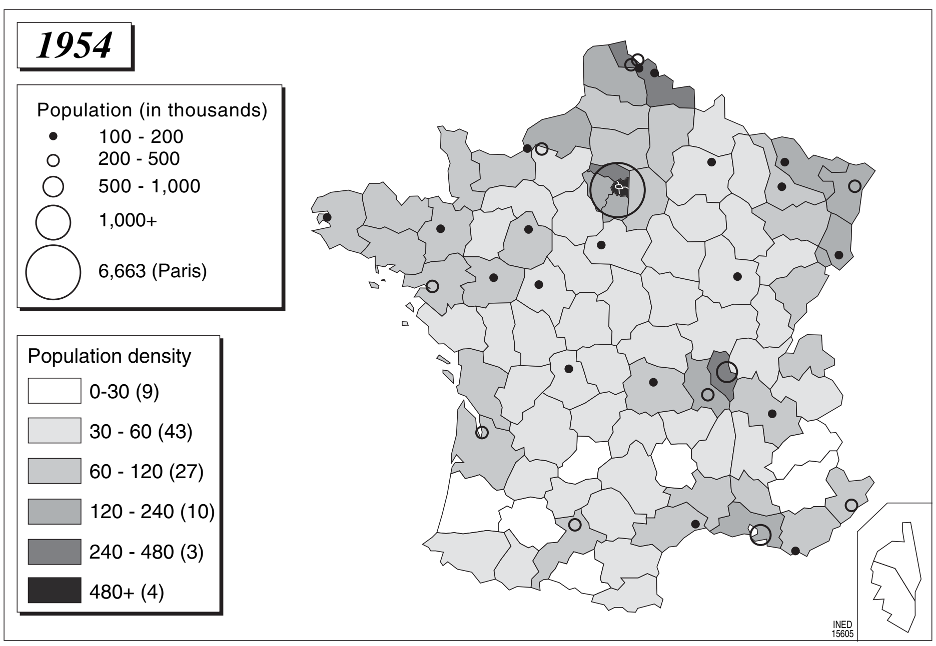


FIGURE 119–4 Population densities by département in 1876 and 1990, and population of the urban agglomerations of more than 50,000 inhabitants within the present French territorial limits. (Sources: Brunet and Aurillac, 1995; Noin and Chauviré, 2002, p. 447.)

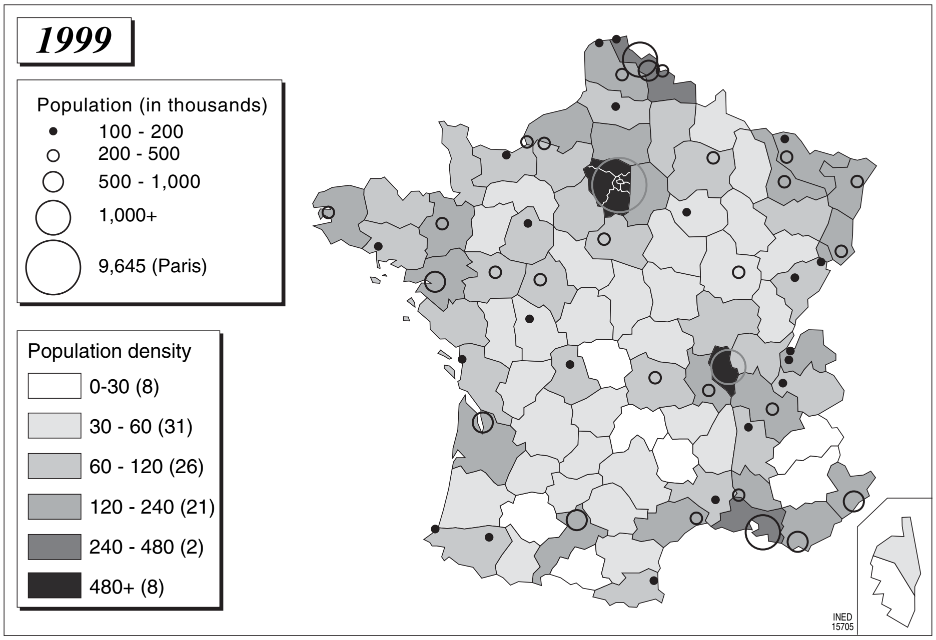


FIGURE 119–4 (*Continued*)

1980s. This fact is just a reflection of the century of advance in the secular decline of fertility, and the lack of fertility was not long in finding its expression in the sectoral shortages of the labor force.

Immigration into France was completely free until the end of the 19th century. The settlement of immigrants has subsequently been progressively controlled, and their employment was limited in 1932, following the economic crisis (discussed later).

From the mid-19th century, foreign immigration increased in importance. In the 1886 census there were more than 1 million foreigners (i.e., 3% of the population). The gross contribution of this immigration to population movement, even if limited solely to the migratory balance in the total population increase, was decisive. Thus, during the whole 1851–1896 period, the arrival of foreigners contributed to 35% of the total growth in the French population, and the direct contribution was even greater in a period of demographic stagnation (such as between 1881 and 1911) when foreign immigration then accounted for half the population growth.

The number of foreign nationals acquiring French citizenship served to mask the slowing down or even decline in the national population. The law of June 26, 1889 eased the acquisition of French nationality to the extent that, in spite of the inflow of immigrants, the number of foreigners remained stationary almost until 1911, thus largely compensating for the decline in French population.

3. Interwar Period: From the Peak of Immigration to Rejection

After the blood bath of World War I, France lacked a young labor force, and immigration increased to the extent of placing France in the front rank of immigration countries. It is also true that the United States had, by their restrictive laws in 1921 and 1924, dried up the flow to their country. In 10 years, despite the many naturalizations following the 1927 law on nationality, the number of foreign nationals in France increased by more than 1.2 million (in addition to Poles and Belgians, this wave included colonial workers), thus reaching the figure of 2.7 million in the 1931 census (i.e., 6.6% of the French population). If the population born outside France is considered (which excludes naturalized neo-French), the situation is even clearer. The number thus increased from 866,000 in 1891 to 2,942,000 in 1931—an increase of more than 2 million.

It is well-known that the Great Depression and unemployment led to xenophobic tensions, and the government decided to limit the employment of foreigners and encourage their return to their countries of origin; this affected arrivals. The law of August 10, 1931 aimed at protecting the national work force by a decree that fixed the proportion of foreign nationals having the right to employment. The war also stopped immigration and motivated return to such an extent that in 1946 the number of foreigners did not exceed 1.7 million, and was therefore below its 1931 peak.

All in all, the interwar epoch was most marked as a time of high immigration (1918–1931), allied to the replacement of those killed in combat. Between 1920 and 1939 the influx of foreigners accounted greatly for the growth in French population—a most important migratory contribution was added to natural growth.

4. Post-war: From Opening to Restriction

In the aftermath of World War II, the national leaders felt the necessity for massive foreign immigration in response to both demographic needs and reconstruction requirements. The legally regulated ordinance of November 2, 1945 laid down the conditions for entry and residence of foreigners in France. The award of a residence permit was subject to obtaining a work permit, and bilateral agreements were signed with the countries of origin. Family regrouping was seen as a demographic aim, and was encouraged. The new *Code de la nationalité* (Nationality Law), promulgated by the Ordinance of October 19, 1945, was relaxed for the youngest generations. The child became French if either of the parents were French or if it was born in France of foreign parents, and lastly the foreign bride of a Frenchman also became French (the reverse case was to be adopted in 1973).

Labor immigration became more frequent after 1955 and quickly found itself swollen by the return of repatriates from the former colonies (1.3 million as a whole, 710,000 in 1962 alone).

Three waves of permanent workers followed during 1955–1973 period: Italians, Spanish, and Portuguese. Their entry took place under the control of the ONI (*Office national d'immigration* [National Immigration Office]), since renamed OMI (*Office des migrations internationales* [International Immigration Office]). As for Algerian immigration (the oldest), this came about according to an evaluation by the Ministry of the Interior (*Direction des renseignements généraux* [Central Department for General Information]).

Similarly to the Great Depression of the 1930s, the oil crisis was to close down the entry of workers and encourage their return to their countries of origin. But the anticipated effects did not come about. There was a rise in family regroupings (from North Africa for the most part) and a development of requests for asylum and clandestine entries (and the precariousness with which they were associated).

TABLE 119–3 Absorption Indicator of Foreign Population

Country	Indicator
Australia	319
Canada	249
Switzerland	170
United States	138
Germany	119
Sweden	94
France	89

5. A Statistical Balance: France within the Main Currents of World Immigration, 1950–2000

In the absence of adequate statistics, especially on the population born abroad, one can assume as a first approximation that the difference between the total increase in population and the natural balance will measure the flow in migration that is applied to a sole national population. In accordance with this rule, the migratory balance of the main Western countries over the whole period of 1950–2000 can be evaluated. As total population is very different from one case to another (285 million in the United States, 7 million in Switzerland, e.g.), we have divided the volume of net immigration obtained by the corresponding population of each country in the middle of the period (1975).

The result provides an absorption indicator of the foreign population. Table 119–3 shows the indices obtained for the seven main Western immigration countries, per 1,000 inhabitants:

The gross data are thus relativized. Despite having a migratory balance of 30 million people, the United States is situated far behind Australia and Canada. As for France, it is in last position.

CONCLUSION

Comparisons within the EU (CEC, 1990; Testa, 2003) regarding the ideal family bring out the unique quality of the French situation. As everywhere else (except in the German-speaking countries in 2001), the family size preferred by young couples is two children, but the second choice is for three children, and this accounts for almost one-quarter of the answers, and is clearly more than in other countries. Doubtlessly, it is also appropriate to attribute the lesser fertility deficit in France to this attitude. For some 30 years, fertility has, according to the situation, been from 10% to 20% below the cohort replacement threshold. A recentering of family policy in favor of the third child could therefore be envisaged, and all the more so since the demographic effectiveness of public transfers to families has proved its effectiveness on the European scale (Chesnais, 1982).

The room for maneuver of the public authorities with regard to population has been largely underestimated. This also applies to fertility, mortality, and international migration. Although very probable, the European population implosion is not entirely inevitable.

Every population wishes to ensure its own replacement (United Nations, 1999). There is nothing to be ashamed of in this objective, and it is even praiseworthy in the name of social equilibrium and biodiversity. In France, it has been regularly repeated by the Supervisory Ministers in prefaces of the reports from the Institut national d'études démographiques (INED) (National Institute of Demographic Studies) to Parliament. It is, moreover, by no means beyond the reach of countries that devote adequate means for preserving the relative position of young couples and their children in society.

As for immigration, it is not only inescapable but also necessary (IFRI, 2003). Where, however, fertility has not collapsed, the gaps to be filled will be minimal and more easily managed. France has a century and a half of, often tentative, immigration policy behind it and two-thirds of a century of political support for fertility (admittedly declining). Public opinion is receptive to population concerns, and a return to cohort replacement is not impossible if unemployment among young people falls and if the government slackens the pressure on pension expenditures and redeploys it to the young people. This time, France could avoid the pitfall of depopulation, which henceforth threatens, often in the short term, most European countries.

In France, indeed, as elsewhere, the philosophy implicit in the redistribution system has been split in two: a universal and comfortable sector for the sick and retired and a partial and weak sector for the poor and the children. Everything takes place as if it is a matter of insurance for the former and assistance for the latter, thus recalling the fate reserved in other times to the destitute. We are far indeed from social investment.

Acknowledgment

This chapter was translated from the French by Ronald Wilkens.

References

Babeau André, 1985. *La fin des retraites?* Paris, Hachette, 432 p.

Bergues Hélène, Ariès Philippe, Hélin Étienne, Henry Louis, Riquet Michel, Sauvy Alfred, and Sutter Jean (ed.), 1960. *La prévention des naissances dans la famille. Ses origines dans les temps modernes.* Paris, INED, PUF, 500 p. (Travaux et Documents, Cahier no. 35).

Beveridge William H., 1943. *Social insurance and allied services.* London, Allen and Unwin, 300 p.

Biraben Jean-Noël, 1988. Le progrès de la médecine, *in*: Jacques Dupâquier (dir.), *Histoire de la population française*, vol. 4, p. 263–277. Paris, PUF, 590 p.

Braudel Fernand, 1986. *L'identité de la France. Espace et histoire.* Paris, Arthaud-Flammarion, 368 p.

Brunet Roger and Aurillac Franc, 1995. *Atlas de la France*, vol. 2. Paris, La documentation française, 128 p. (RECLUS, Population).

Calot Gérard, 1980. Niveau de vie et nombre d'enfants, *Population*, vol. 35(1), p. 9–55.

Calot Gérard, 1991. La fécondité en Europe. Évolutions passées, perspectives d'avenir. Les actions possibles et leur efficacité, *in*: Eurostat *(ed.), Le capital humain à l'aube du XXI^e siècle*, p. 1–34, Luxembourg, Eurostat (Conference on Human Resources in Europe at the Dawn of the 21st Century, held in Luxembourg November 27–29, 1991).

Ceccaldi Daniel, 1957. *Histoire des prestations familiales.* Paris, UNCAF (Union nationale des Caisses d'allocations familiales), 166 p.

Chesnais Jean-Claude, 1982. L'État et la fécondité, *in*: Conseil de l'Europe (ed.), *Conférence démographique européenne, Strasbourg, 1982*, p. 371–391. Strasbourg, Council of Europe, 481 p.

Chesnais Jean-Claude, 1985a. Les conditions d'efficacité d'une politique nataliste: examen théorique et exemples historiques, *in*: IUSSP (ed.), *Proceedings of the IUSSP international population conference, Florence, 1985*, vol. 3, p. 413–425. Liège, IUSSP, 525 p.

Chesnais Jean-Claude, 1985b, July–Sept. Sécurité Sociale et population, *Revue Française des Affaires Sociales*, vol. 39(special issue), p. 45–59. ("40 ans de Sécurité Sociale," Pierre Laroque [ed.]).

Chesnais Jean-Claude, 1999. Vers une récession démographique planétaire?, *in*: Institut Français des Relations Internationales (ed.), *Rapport annuel mondial sur les systèmes économiques et stratégiques 2000*, p. 51–63. Paris, IFRI, 424 p.

Chesnais Jean-Claude and Vallin Jacques, 1977. Évolution récente de la mortalité et de la morbidité dues aux accidents de la route, *Population*, vol. 32(6), p. 1239–1265.

Chombart de Lauwe Paul-Henri, 1982. *La fin des villes: mythe ou réalité.* Paris, Calman-Levy, 249 p.

CEC (Commission of the European Communities), 1990. *The family and the desire for children.* Bruxelles, CEC, 119 p. (Eurobarometer 32).

Deane Phyllis and Cole William A., 1967. *British economic growth, 1688–1959: trends and structures.* London and Cambridge, Cambridge University Press, 350 p. (University of Cambridge, Department of Applied Economics Monographs, 8).

Desplanques Guy, Vallin Jacques, and Biraben Jean-Noël, 1995. La belle époque de la santé, *in*: Jacques Dupâquier (ed.), *Histoire de la population française*, vol. 4, p. 233–285. Paris, PUF, 586 p. (Quadrige rééd.).

Dupâquier Jacques, 1988. Préface, *in*: Jacques Dupâquier (ed.), *Histoire de la population française*, vol. 4, p. 1–11. Paris, PUF, 590 p.

Ekert Olivia, 1986. Effets et limites des aides financières aux familles: une expérience et un modèle, *Population*, vol. 41(2), p. 327–348.

Eurostat, 2002. *Annuaire 2002. Le guide statistique de l'Europe, données 1990–2000.* Luxembourg, Eurostat, 455 p.

Eurostat, 2003. *Annuaire 2003. Le guide statistique de l'Europe, données 1991–2001.* Luxembourg, Eurostat, 317 p.

Fourastié Jean, 1979. *Les Trente Glorieuses ou la révolution invisible de 1946 à 1975.* Paris, Fayard, 288 p.

Gauthier Anne Hélène, 1996. *The State and the Family: a comparative analysis of Family Policies in Industrialized Countries.* Oxford, Clarendon Press, 231 p.

Gravier Jean-François, 1947. *Paris et le désert français. Décentralisation.* Paris, Le Portulan, 416 p.

Guyotat Régis, 2003. Le drame des asiles de Vichy, *Le Monde*, vol. 59(18265), p. 6.

Himes Normann, 1963. *Medical history of contraception.* New York, Gamut Press, 521 p.

Huber Michel, 1931. *La guerre et la vie sociale. La population de la France pendant la guerre.* Paris, PUF, 1025 p.

IFRI, 2003. *Les caractéristiques du commerce mondial au XXIe siècle. Scénarios pour l'avenir de l'Europe.* Paris, Institut Français des Relations Internationales, 332 p. (plus annexes).

Kindleberger Charles P., 1964. *Economic growth in France and Britain 1851–1950.* Boston, Harvard, 378 p.

Lafont Max, 1987. *Extermination douce.* Paris, Éditions Le bord de l'eau, 272 p.

Landry Adolphe, 1934. *La révolution démographique. Étude et essai sur les problèmes de la population.* Paris, Sirey, 231 p.

Ledermann Sully, 1956. *Alcool, alcoolisme, alcoolisation*, vol. 1. Paris, INED, PUF, 314 p. (Travaux et Documents, no. 29).

Le Play Frédéric, 1871. *L'organisation de la famille suivant le vrai modèle signalé par l'histoire de toutes les races et de tous les temps.* Tours et Paris, Mame et Tequi, 318 p.

Mathieu Mathilde, 2003. Le gouvernement propose plusieurs mesures pour aider les jeunes parents, *Le Monde*, no. 18121, April 29, p. 11.

Meslé France et Vallin Jacques, 1981. La population des établissements psychiatriques: évolution de la morbidité ou changement de stratégie médicale? *Population*, vol. 36(6, p. 1035–1068.

Meslé France and Vallin Jacques, 1988. Les composantes de la mortalité cardio-vasculaire en France depuis 1925: résultats d'une reconstitution historique, *Population*, vol. 43(2), p. 391–425.

Myrdal Alva, 1945. *Nation and family.* London, Kegan Paul, 441 p.

Myrdal Alva and Myrdal Gunnar, 1934. *Kris i befolkning-frågen (The demographic crisis).* Stockholm, A. Bonnier, 332 p.

Nizard Alfred, 1974, June. Politique et législation démographique, *Population*, vol. 29(special issue), p. 285–326.

Noin Daniel, 2001. *Le nouvel espace français.* Paris, A. Colin, 256 p.

Noin Daniel and Chauviré Yvan, 1986. *La population de la France.* Paris, Masson, 203 p.

OCDE/OECD, 2001. *Les tendances des migrations internationales.* Paris, Service d'Observation Permanente des Migrations (SOPEMI), 287 p. (Annual Report, 2001 and preceding years)

Pressat Roland, 1974. Évolution générale de la population française, *Population*, vol. 29(special issue), p. 11–29.

Prost Antoine, 1985. L'évolution de la politique familiale en France de 1938 à 1981, *in*: Commissariat général au plan (ed.), *La politique familiale en France depuis 1945*, p. 7–37. Paris, Commissariat général au plan, 387 p.

Pumain Denise and Courgeau Daniel, 1988. L'évolution du peuplement de 1911 à 1990, *in*: Jacques Dupâquier (ed.), *Histoire de la population française*, vol. 4, p. 387–420. Paris, PUF, 590 p.

Sauvy Alfred, 1945. *Bien-être et population*. Paris, Payot, 230 p.

Sauvy Alfred, 1946. Évaluation des besoins d'immigration française, *Population*, vol. 2(1), p. 91–98.

Sauvy Alfred, 1954. *Théorie générale de la population*, vol. 2. Paris, PUF, 375 p.

Sauvy Alfred, 1967. *Histoire économique de la France entre les deux guerres*, vol. 2. Paris, Fayard, 627 p.

Testa Maria-Rita, 2003. *Attitudes of Europeans towards fertility: ideals, desires, intentions and realizations.* Brussels, European Commission, 2003, 68 p. (Eurobarometer 56.2).

Thélot Claude and Villac Michel, 1998. *Politique familiale: bilan et perspectives.* Paris, Documentation Française, 252 p.

Tugault Yves, 1974. Les migrations internationales, *Population*, vol. 29(special issue), p. 115–123.

United Nations, 1999. Below replacement fertility, *Population Bulletin*, no. 40–41, p. 1–348.

Vallin Jacques, 1973. *La mortalité par génération en France depuis 1899.* Paris, INED, PUF, 484 p. (Travaux et Documents, Cahier no. 63).

Vallin Jacques, 1988. Les causes de décès en France de 1925 à 1978, *in*: Jacques Dupâquier (ed.), *Histoire de la population française*, vol. 4, p. 244–262. Paris, PUF, 590 p.

Vallin Jacques and Chesnais Jean-Claude, 1975. Les accidents de la route en France. Mortalité et morbidité depuis 1953, *Population*, vol. 30(3), p. 443–478.

Vallin Jacques and Meslé France, 2001. *Tables de mortalité françaises pour les XIX^e et XX^e siècles et projections pour le XXI^e siècle.* Paris, INED, 102 p. + CD-rom. (Données statistiques, no. 4—2001).

Vallin Jacques and Meslé France, 2004. Convergences and divergences in mortality. A new approach to health transition, *Demographic Research*, Special Collections S2 "Determinants of diverging trends in mortality," p. 1–23.

Vincent Paul, 1946. Le rôle des familles nombreuses dans le renouvellement des générations, *Population*, vol. 2(1), p. 148–154.

Weil Patrick, 1995. *La France et ses étrangers: l'aventure d'une politique d'immigration de 1938 à nos jours.* Paris, Gallimard, 592 p.

Zola Émile, 1899. *Fécondité. Les quatre évangiles.* Paris, Bibliothèque Charpentier E. Fasquelle, 751 p.

CHAPTER

120

World Population Growth and the International Community from 1950 to the Present Day

JEAN-CLAUDE CHASTELAND

Institut national d'études démographiques (INED), Paris, France

INTRODUCTION

Without a benign "invisible hand" to guide individual demographic behavior toward the interests of the group, there was little option but to adopt a voluntarist approach that few accepted, or even envisaged, at the end of World War II, as much because of the religious interdictions still controlling the private life of many societies as because of the era's cold war ideological and political context. Today, this voluntarism has become quasi-consensual, as is all the more apparent given the extraordinary convergence of behavioral norms observed in the last few years, concerning population, perhaps, more than any other field. Five decades, however, is a short time for global attitudes toward an issue as sensitive as population to evolve. In 1959, the Belgian delegate, a conservative Catholic, could still declare that the United Nations Population Commission "should not attempt to propose a solution to what was a serious problem, and least of all suggest that it advocated curbing or stopping demographic expansion" (Symonds and Carder, 1973; p. 87). As for the USSR delegate at the same Population Commission session, he was opposed to "any suggestion that the Commission should recommend action which would lower birth rates" (Symonds and Carder, 1973; p. 88).

To reach this point, during the 1940s influential American researchers like Frank Notestein had first to undergo a type of intellectual *aggiornamento* which led them to conclude that the demographic transition's natural mechanisms were not sufficient in themselves to trigger fertility decline in time and that it would be necessary to intervene in the process (Szreter, 1993). The original small group of researchers and decision makers who had joined them had subsequently to show great persistence and persuasiveness in order to mobilize around this project their country's resources, and later those of the international community within which the United Nations played a central role.[1] Americans for the most part, but also Scandinavians, these were the pioneers of the Population Movement. This group of scientific and political figures was to have a decisive influence on the international stage. Throughout the period, they were to be found in different halls of power: the foundations, American universities and government, the United Nations, international NGOs, and the like. Among those creating this *Population Movement*, or holding influential positions within it, however, the same small set of sociologists, demographers, physicians, and other scien-

[1] The international community refers here to a collection of players on the world stage, grouped within and around the United Nations and its member states, and includes academic institutions, nongovernmental organizations (NGOs), foundations and the media, whose respective influence can vary over time.

tists are particularly noticeable. A shared conviction brought them together: to place their expertise at the service of a political endeavor. They constituted what is referred to by Anglo-Saxon political scientists as "epistemic communities," meaning, "a network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue area" (Haas, 1992, p. 3). These networks tended to develop rapidly into transnational networks, as was certainly the case in the field of population.

Attitudes toward this issue have not followed a linear evolution; as discussions in Mexico in 1984, and even in Cairo in 1994, have shown, an about-turn is always possible. These difficulties come as no surprise when the population and resources (or development, to use today's term) debate is placed within a historical perspective that extends back far beyond the last decades of the 20th century. The first World Population Plan of Action adopted at Bucharest echoes concerns rooted in the 18th century, when the appearance of what Michel Foucault (1976) termed "bio-politics" made possible the better administration of the human resources necessary for the development of Europe's future industrial states.

Today, the population and development debate continues its arduous course, still associated nowadays as it was during the 18th and 19th centuries with three frequently irreconcilable elements: ethics, ideology, and science. This uneasy combination has not succeeded in holding back voluntarism, as the adoption of three successive world population action plans has shown; it contains within it, nonetheless, sources of tension that can either stimulate voluntarism or, on the contrary, paralyze it, according to the circumstances.

This chapter is organized around five foundational events that marked the period between the end of World War II and today. The creation of the Population Movement in the United States after World War II is the first; the second, the adoption by the United Nations in 1966 of a resolution authorizing the United Nations to support family planning programs in countries requesting them. Next is the organization of a series of three large intergovernmental meetings that were crucial, as we will see, to legitimize and generalize antinatalist policies in the Third World. The Bucharest Conference opened a decennial cycle of world population conferences; the Mexico conference followed in 1984; Cairo, in 1994, was the last in the series. A final prospective section explores the future of population policy at the international level.

I. 1946–1965: CREATION OF THE POPULATION MOVEMENT IN THE UNITED STATES AND THE EARLY MOBILIZATION OF THE INTERNATIONAL COMMUNITY

From the late 1940s to late 1960s, the *Population Movement* in the United States developed primarily around researchers and representatives of the large foundations and NGOs. These individuals later created links with the American administration, which did not itself become involved until the mid-1960s (Donaldson, 1990).

Two concerns awakened people to the problems raised by present and future Third World population growth: one was geopolitical in nature, and stemmed from the discovery, as early as the 1940s, of very rapid population increase in Latin America and Asia, two regions constituting a crucial part of the United States' strategic horizon. This growth was seen as a threat to political stability in regions that were considered essential for United States' security and for world peace, whose main guarantor this country perceived itself to be. William Vogt's book, *Road to Survival* (1948), in which the author describes the apocalyptic situation in which humanity would find itself if it did not take rapid steps to limit population growth, perfectly illustrates the fear this issue instilled in public opinion as soon as it was recognized. The book created quite a stir in the United States and in the rest of the world. The other concern was a humanitarian one. It focused on the well-being of the populations involved, and on their access to family planning, considered to be a universal right. With policy as a principle thus justified at the strategic and humanitarian levels, only an economic rationalization needed to be found in order to gain the support of Congress and the administration. It was here that American researchers, whose studies had the advantage of being highly oriented toward world problems, played a significant role. In the late 1940s they were already alerting funding agencies and government officials to the speed of population growth; all that remained was to demonstrate that such an increase was a major obstacle to development, and that development would, in fact, be impossible unless they got it under control.

At that time, economic theories of growth were dominated almost exclusively by considerations regarding the formation of physical capital, believed to be the unique key to development. Ansley Coale and Edgar Hoover's (1958) work, largely inspired by India's experience but with a more general application, established a rigid and negative association between population growth and the creation of savings and investment. Falling easily within the framework of this issue, it provided the economic argument to justify the adoption of birth control policies. The same line of reasoning reappeared later from the lips of the Democratic President, Lyndon Johnson, when he declared before the United Nations in 1965, "*Five dollars invested in population control is worth 100 dollars invested in economic growth*" (Green, 1993).

Demographers supplied additional arguments in favor of family planning (thanks particularly to findings from the KAP [Knowledge, Attitude, and Practice] surveys). By demonstrating that demands in this sphere were not met and that one could therefore legitimately respond to them, on the one hand, and by clarifying the determinants of fertility, on the other, they had cleared the way for possible intervention on these determinants (Davis and Blake, 1956).

Nothing could have been achieved, however, without key developments within biomedical research. Certainly, militants within the Population Movement believed that the diffusion of birth control practices depended essentially on making effective and cheap contraceptives available. In 1954, therefore, the discovery of the contraceptive pill by physicians Gregory Picus, John Rock, and Min Cheuh Chang was viewed as a major breakthrough; their work, on the initiative of Margaret Sanger, had been funded up to $150,000 by Mrs. MacCormick, the wealthy heiress of the company International Harvester. Several years later, the development of a much-improved version of the Population Council's intrauterine device (IUD), followed by that of Norplant, was another important step in extending access to contraceptive methods in terms of their efficacy and ease of use in poor countries (McLaren, 1990). Later still, even more effective contraceptives appeared, not to mention improved sterilization and abortion techniques.

The American administration was not particularly enthusiastic at first. In 1959, with reference to family planning, the republican president Dwight Eisenhower stated, "Birth control is not our business. [. . .] I cannot imagine anything more emphatically a subject that is not a proper political or governmental activity or function or responsibility" (Green, 1993). He later came to regret the statement and had to admit that he had been mistaken at the time. In the meantime, the foundations and American NGOs were left to spread the word and finance the first programs, and especially research developing efficient contraceptives. Personalities like John D. Rockefeller III, one of the Population Movement's outstanding figures, General William H. Draper Jr., a frequent participant in negotiations with the United Nations, Dr. Clarence Gamble, very wealthy heir of the pharmaceutical firm Procter & Gamble, Frederick Osborne, former vice president of the American Eugenic Society, President of the PAA (1949–1950) and of the Population Council (1957–1959) as well as the Ford, Mellon, MacArthur, and other foundations, all had prominent roles in creating or financing such organizations as the Milbank Memorial Fund (1944), the Population Council (1952), the Pathfinder Fund (1957), and the like. A special mention should be made of John D. Rockefeller III, who devoted to population issues an important part of his considerable political influence within the American administration and United Nations, of his financial power, and of his prestige among scientists (among other things, the Population Council in 1952 was his creation) (Harr and Johnson, 1991).

The US Agency for International Development (USAID), whose expertise also included population programs, only really got going later on. In 1965, it provided no more than 2 million dollars of aid compared with the Ford Foundation's 11-million-dollar contribution. It was only after the first private initiatives, encouraged mainly by Democrat presidents like Lyndon Johnson (Richard Nixon, through his support, was an exception among the Republicans) that the USAID Population Bureau (responsible to the State Department) became one of the main funding sources for population programs in many Third World countries and at the United Nations. From its creation in 1966, the Bureau was directed by a family planning militant, Dr. Reinert T. Ravenholt, whose extreme activism ended up alienating not only the academic world but also the administration, who demanded his resignation in 1979. During the 1960s, private and public American funds accounted for almost all international funding destined for Third World family planning projects; during the 1970s, it still represented three-quarters of it. Only after 1981, under Reagan's administration, did the relative weight of this funding drop significantly, due to the arrival of other donor countries (Germany, Norway, the Netherlands, Japan, etc.); its contribution remained substantial nonetheless, representing 40% to 50% of all international financing in the 1990s (Donaldson, 1990). Also since 1981, the administration's policy position on family planning program funding has been challenged on several occasions, due to growing disagreement between Democrats and Republicans about contraception and abortion, with the Republican Party's right wing dominated by Christian fundamentalists for close to 15 years.

Outside the United States, countries like Sweden had made a commitment to support for Third World family planning programs early on. This was the case for Sri Lanka from 1958 and Pakistan in 1961, and personalities like Alva Myrdal, ambassador to India at the time, played key roles (Wahren, 1986). A pioneer in many fields, Sweden was quickly joined by other Nordic countries, such as Denmark, Norway, and, more hesitantly, Finland. Owing to their moral authority and dynamism, these countries formed a highly influential pressure group on all questions relating to population at the United Nations, as well as in other organizations or international NGOs, like the Organization for Economic Cooperation and Development (OECD) or the International Planned Parenthood Federation (IPPF). What was lacking, perhaps, was a certain autonomy in relation to the American scientific domination. Furthermore, their bilateral aid strategy, structured around public aid organizations in the Third World (DANIDA for Denmark, NORAD for Norway, and SIDA for Sweden) and well endowed financially through their government's strong commitment to public assistance for development, proved to be of high quality.

Some developing countries also did not remain idle, however. In 1952, at the close of a conference organized in Bombay on the Family Planning Association of India's (founded in 1949) initiative, the participants decided to create the IPPF. Twelve years later, in 1964, this association finally obtained consultative status at the United Nations Economic and Social Council (ECOSOC). At that time, the IPPF already brought together several dozen national associations. Today it unites 141 associations in 180 countries and its donations reached 59 million dollars in 2001 (www.ippf.com). The scope of its activities has also diversified considerably; in addition to its traditional family planning interests, the IPPF has adopted most of the Cairo conference's vast array of reproductive health objectives, to be examined further on.

It is not by chance that the IPPF was created in postwar India. This developing country was, in fact, the first to officially declare its resolve to take control of population growth. The roots of this commitment go even further back, however: as early as 1937, the Indian Congress Party had examined the prerequisites for establishing a family planning policy. In 1952, India launched a family planning program as part of its first five-year plan (see Chapter 113, this volume, by S. Irudaya Rajan and Jacques Véron). From its English administrators, the country had inherited a solid statistical tradition and, from its researchers, a strongly Malthusian view of the role of population. Its experience would therefore have a great influence on the whole Indian subcontinent (Caldwell, 1998). In 1965, it became the first country to ask the United Nations for technical assistance to develop such a program. As early as 1956, on China's behalf, Chou En Lai had also expressed the need to develop a family planning program, but without asking for outside help. We know that this political direction would be challenged several times, particularly on President Mao's initiative, until the one-child objective was adopted at the end of the 1970s, an objective maintained up to the present day (see Chapter 112, this volume, by Isabelle Attané).

The growing number of countries adhering to the general aim of population growth control, however, should not make us forget the formidable opposition aroused by this endeavor. This opposition grouped a coalition of highly disparate countries. First, the communist bloc (including China, until the Bucharest Conference) perceived family planning policy as a manifestation of a false conception of the relationship between demographic increase and economic development (in their view, population growth stimulated economic growth); it was also seen as a convenient way for the West to protect the advantageous geopolitical status quo, by preventing expansion among populations who were the most deprived and potentially, therefore, the greatest threat to world order. Numerous so-called nonaligned countries of that era (who formed the Group of 77 soon afterward at the United Nations) shared this last point of view. Their hostility was reinforced through reading not only the American media or polemical works like that of William Vogt, but also by the writings of eminent scientific Population Movement members, such as Frank Notestein (who was later called successively to the direction of three organizations that played a crucial role on the international scene: the Office of Population Research at Princeton University, the United Nations Population Division, and the Population Council). By 1944, inspired by the purest geopolitical realism, he unhesitatingly stated, "By launching a program of modernization, the now dominant powers would in effect be creating a future world in which their own peoples would become progressively smaller minorities, and possess a progressively smaller proportion of the world's wealth and power" (Notestein, 1944; p. 156). In 1968, at the Vatican, Pope Paul VI had just published the encyclical *Humanae Vitae*, a virulent condemnation of any form of "artificial" contraception, and several Latin American and European countries joined the opposition as much for religious as geopolitical reasons (as in the case of Brazil, e.g.). France, along with other French-speaking countries, perceived these American endeavors as reflecting their desire for hegemony, though couched at times in humanitarian terms, as illustrated by this quotation from another influential demographer of the period, Dudley Kirk: "We will probably be serving our own political ultimate interests by speeding the social evolution that will bring about slower population growth. Most important of all, we shall have led all humanity to new possibilities of life" (Kirk, 1944; p. 35).

The double talk of the American position, notably at the Population Commission, was an easy target for criticism from the French representatives who defended an opposite agenda. During the 1960s, the French position corresponded to what could be termed a sort of "demographic Gaullism," whose objective was to maintain or if possible to increase the weight of the French speaking countries in the world. France had only adopted the Neuwirth law on contraception in 1967, and the Weil law on abortion in 1975. This delay in the much-needed evolution of its national legislation handicapped France in international debates. It also blocked progress in the legislature of its ex-colonies, who modeled themselves closely on it. In taking a critical stance toward the *Population Movement*, France often played a useful role, but by remaining in this position, was also marginalized. A kind of natural division arose between the intellectual importance of French demographic research and France's weak political influence on the international scene.

In 1946, the Population Division was created at the United Nations, a step that facilitated the process of reflection on world population problems; this body, mandated to carry out studies for the Population Commission (a subsidiary body of ECOSOC), was composed of government representatives chosen from different regions of the world (27 at the present time). The creation of an organization exclusively concerned with world population questions was far from evident as, at the time, population issues were viewed as falling naturally within the sphere of national governments. The Population Commission was to provide governments and the United Nations Secretariat with opinions and recommendations on population questions bearing on the international community. The United Nations, it should be noted, took over and considerably extended activities initiated on a modest scale at the time of the Society of Nations in Geneva.

The Population Division applied itself to the task without delay, under the authority of its director, Frank Notestein, who had also worked for the Society of Nations, notably in the sphere of population projections. From 1951 onward, its publications started to appear, with the first population projections (United Nations, 1951) and, among others, an important report titled: *The Determinants and Consequences of Population Trends* (United Nations, 1953). The discussions surrounding the planning of this report in 1951 provided the opportunity for a thorough review of the different positions held by the Population Commission's member states on the population and development question. Rereading the report today, it is impossible not to notice how painstakingly and carefully it was argued. It concludes by suggesting very cautiously that rapid demographic growth could create difficulties. The report was nonetheless violently attacked by the USSR representative and, to a lesser extent, by the French one, Alfred Sauvy, director of Institut national études démographiques (INED) at the time. A new version had to be drawn up, and this was the one finally published in 1953 (United Nations, 1953). As for the population projections, regularly revised from 1951 onward, they had a significant influence in raising awareness within the international community. Their remarkable reliability over time made their impact all the greater (see Chapter 77).

As this is not the place for a detailed description of the Population Division's achievements, we will mention only that in the early years they had a huge influence on demographic research in developing countries, particularly with the publication of model life tables, manuals for indirect analysis of incomplete data, and stable population studies. In addition, five training centers were created: in the Middle East (Cairo), Africa (Accra and Yaounde), Latin America (Santiago, Chile), and Asia (Bombay); their impact was critical for the education of several generations of Third World demographers. Later, it became more oriented toward the comparative analysis of the growing quantities of available data, from censuses or World Fertility Surveys (WFSs) during the 1980s, and later from Demographic and Health Surveys (DHSs), one of the most massive undertakings for the scientific exploration of human behavior worldwide. The Population Division also pioneered population policy surveys among its member states, and is the holder of the largest comparative data bank on these policies and on international migration. In its many publications, it has always taken care to be open to as broad a demographic field as possible, especially avoiding any extreme polarization on questions of fertility, while endeavoring simultaneously to remain neutral at the politico-ideological level.

During this period, in collaboration with the International Union for the Scientific Study of Population (IUSSP), the Population Division directed preparations for the first world population conference in Rome in 1954, which brought together 455 participants from 74 countries. A conference of specialists with no government mandate to adopt any recommendations, however, its aims were purely scientific. Nonetheless, the organizers hoped that the work it achieved would enlighten government actions. As Milos Macura observed, it helped spread the idea that "all populations of the world should be properly studied and their problems understood in the context of their particular economic, social and cultural conditions" (Macura, 1986). This was a way of responding in advance to criticisms concerning the global population concept.

In Asia, which found itself up against substantial population problems early on, and where American influence on the political, economic, and intellectual elites was strong, the reaction was rapid. Japan, defeated and facing the return of its settlers from countries within the defunct Grand Asia project, passed the so-called Eugenics law in 1948, which made it possible for an ambitious population control program to be set up. India, Sri Lanka, and Pakistan followed suit. When the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) organized a regional conference at the governmental level in New Delhi in 1963, therefore, no one in the region was surprised that the final document supported recommendations regarding the creation of family planning programs. This initiative had been seriously criticized by certain Population Commission members, however, who had no wish to have their hand forced (Symonds and Carder, 1973).

The climate was still not ripe, therefore, to set up a governmental level world conference. When the second world population conference was organized, yet again by the IUSSP, in Belgrade in 1965, its objectives were once more restricted to a scientific analysis, by close to 1,400 experts from 118 countries, of the population problems confronting the international community. Nevertheless, in his opening speech, the Frenchman, Philippe de Seynes, Deputy General Secretary at the United Nations and head of the Department responsible for the Population Division, had the courage and lucidity to declare that "at the opening of this Congress whose declared objectives are technical, why do we have the feeling to be participating in an exceptional event which is taking on a political dimension?" (United Nations, 1965; p. 29). At Belgrade, fertility issues were abundantly debated (more than 100 of the 487 presentations focused on fertility and family planning). For the first time at a conference organized by the United Nations, family planning and the problems of program implementation were discussed, giving them a type of implicit recognition. The evolution since the Rome conference was thus very clear, and reflected the evolution of the countries most involved. Whereas only three countries had family planning programs in 1954, the number had already risen to 23 by 1965.

The collaboration between the Population Division and the IUSSP needs to be briefly touched on as it was a key ingredient of developments in the study of problems raised by global population trends during the two decades following the end of World War II. The experience acquired in jointly organizing the Rome and Belgrade Conferences, in preparing the *Multilingual Demographic Dictionary*, in sharing the preparation of several reference manuals particularly on the subject of fertility, and their participation in the various IUSSP scientific committees in their official capacity as Division members, produced a close collaboration right up to the early 1980s. After that, two developments altered the nature of this collaboration. First, Division members gradually became less involved in these committees, as the need for multidisciplinary expertise had somewhat reduced the importance of demography, and that of the Division with it. Second, the Division's role in the union's general conference preparation and organization dwindled, while, simultaneously, the role of the United Nations Fund for Population Activities (UNFPA) within these same conferences, and in financing its activities, expanded (IUSSP, 1988). Like the IPPF, the IUSSP has developed significantly in the last 50 years and today includes more than 1,900 members from 127 nations.

In 1966, a year after Belgrade and following arduous discussions, the United Nations General Assembly took a decisive step in adopting its famous GA Resolution 2211 authorizing the United Nations to support family planning initiatives in any country expressly requesting it. Sponsored by 15 countries, this resolution was proposed by Sweden, Norway, and Denmark, with the backing of 11 Third World countries and of Yugoslavia, a nonaligned country. The central, though relatively unknown, part played by Philippe de Seynes during preliminary discussions about the adoption of this resolution at the General Assembly's second commission is once again worthy of note. After integrating numerous amendments, it was adopted by consensus on December 17, 1966. The ground had already been prepared by the World Health Organization's (WHO) adoption, on May 20, 1966, of WHA Resolution 19–43, authorizing the organization to "lend its technical support to the development of family planning activities in countries requesting it".

Having gained international legitimacy, from then on the *Population Movement* grew progressively stronger.

II. 1966–1973: THE UNITED NATIONS' GROWING INFLUENCE

Although the United States enjoyed almost total control of the *Population Movement* at all levels—ideological, political, intellectual, and financial—during the period from 1950 to 1966, this position gradually evolved as the Movement grew more international following the United Nations' strong entry into its field of action. Well aware of their problem with direct intervention in numerous Third World countries with relatively hostile governments, the U.S. administration had actually encouraged this internationalism. It entered into a new collaborative relationship with the United Nations, not unlike the one it had experienced and continued to experience with NGOs and American foundations. During this period, the role of the foundations and NGOs remained important, but they encountered an increasingly powerful United Nations' presence in territory they had occupied alone for many years. Some institutions, very active ones, were still being set up at the end of the 1960s, such as the Alan Guttmacher Institute (1968) or the Population Crisis Committee (1968).

The United States' intellectual supremacy scarcely changed during this period. Stimulated by an abundance of funding, numerous studies even helped consolidate their intellectual hegemony. Among those attracting the attention of the era's decision makers, the National Academy of Science's (NAS) 1971 report should be cited, with its pessimistic conclusion on the consequences of population increase for Third World countries (NAS, 1971). Paul Erlich's (1968) work and Garrett Hardin's (1968) article also made waves, and achieved lasting influence by introducing the ecological dimension into the population and development debate. This new line of argument, which in fact resuscitated the old idea of physical limits, had the success it is known to have had and arrived at exactly the right moment, namely when theories of economic growth were beginning to free themselves from considerations related exclusively to the creation of physical capital. Furthermore, empirical observations were becoming available that show the impact of demographic growth on the formation of savings and investment to be less negative than previously expected.

Although American population policy in the Third World was based on an abundant neo-Malthusian literature, there was also a highly dynamic body of literature contesting their arguments, but which the Administration ignored (Hirschman, 1958; Kuznets, 1967); within the American academic community, individuals such as Colin Clark (1967) and Dame Esther Boserup (1965) were also very influential. The strongest critiques of American political involvement did indeed hail from the academic world. The conservative and religious political wings had not yet managed to organize a counter-offensive comparable to the one they succeeded in setting up in the early 1980s.

Outside the United States, the opposition was made up of the same heterogeneous coalition as in the early 1960s. Rubbing shoulders with the Vatican were communist countries (soon to lose one of their influential allies, China), nonaligned countries like Algeria, then at the forefront of the Third World struggle (these countries formed the Group of 77 at the United Nations in 1964), as well as the Catholic countries of Africa, Latin America, and Europe. Still taking the same critical stance held from the start, a French representative at the United Nations Economic and Social Council suggested making a distinction between family planning, on the one hand, and birth control, on the other.

> That process whereby the population unbalance resulting from the fall in the death rate was to be corrected by deliberately reducing the birth rate was called "family planning" although, in fact, it implied birth control measures resulting not from the free decisions of individuals or families, but from the application of a veritable national policy (Symonds and Carder, 1973; p. 179).

He warned the international community of the danger of attempting to encourage birth control on the pretext of supporting family planning. This warning was something of a premonition; the campaigns for sterilization and forced abortions in China, India, and elsewhere, and the not-always equitable methods of dissuasion and encouragement used, proved it to be well founded.

At the United Nations, a number of factors contributed to the rapid creation of technical assistance programs in the area of family planning, forbidden until then: the General Assembly's adoption of Resolution 2211 in 1966; the publication by U Thant, the United Nations Secretary General, of a Declaration in favor of family planning, signed by 12 Heads of State in September of the same year, at John D. Rockefeller III's suggestion; and, in 1968 at Tehran, another publication on human rights containing a paragraph on the right of couples to decide freely the number and spacing of children. These programs developed not only at the United Nations itself, in the New York Secretariat, and in UN regional commissions at Bangkok, Addis Ababa, Beirut and Geneva, but also in the United Nations' large specialized agencies: first at the United Nations Education, Sciences, and Culture Organization (UNESCO), no doubt due to the involvement and personality of it director at the time, Julian Huxley; then at the International Labor Office, executive body of the International Labor Organization (ILO), and the Food and Agriculture Organization (FAO).

On account of its mandate, the WHO had started discussing the possibility of giving support to family planning initiatives very early on, but there was also an effective opposition until 1965, the year preceding the adoption of the WHA Resolution 19–43 which, as mentioned, had opened the way to United Nations involvement.

The International Bank for Reconstruction and Development (IBRD), known today as the World Bank, enjoyed a special status within the United Nations system, as one of the so-called Bretton-Woods institutions, where communist bloc countries were not represented. At first, it was satisfied simply with showing certain countries, like Thailand in 1958, that rapid population increase was leading them into an economic impasse and with helping fund Ansley Coale and Edgar Hoover's celebrated work on India in 1966, within the context of studies on population growth in developing countries (Kasun, 1981). Up to this time, however, the Bank had no wish to be directly involved in financing family planning programs. GA Resolution 2211 and Robert MacNamara's nomination as President in 1968 accelerated considerably its involvement in this issue, all the more so that Robert MacNamara rapidly became simultaneously one of the Population Movement's leaders and the inspiration for the World Bank's antipoverty programs. In Robert MacNamara's mind, however, geopolitical considerations needed to remain at the forefront. Thus, in 1969, in a famous speech at Notre-Dame University, he declared, "population growth creates stress in the body politic which in the end can bring on conflicts among nations" (Crane and Finkle, 1981). Since then, the World Bank has continued to play a decisive role in the population arena, no longer only at a financial level where the volume of its low-interest loans exceeds the volume of grants from UNFPA (more than 500 million dollars in 2003) but also at an ideological and political one, using its political weight widely to influence vacillating governments. At the same time, it broadened its field of action to encompass such varied activities as providing financial support for health, education, and the improvement of women's status programs, and even the promotion of "methods of good governance" for local authorities.

Still at the United Nations, the General Secretary had decided to set up a trust fund in 1967 to collect voluntary contributions from donor countries. This was in response to the many countries opposed to the United Nations using its regular budget to finance family planning programs. Administered for 2 years by the Population Division, and with the United States as the first contributor (a sum of 1 million dollars), in 1969 this fund became the UNFPA, mentioned earlier. Its administrative and especially political management was entrusted to the United Nations Development Program (UNDP), a decision fraught with consequences. It had been instigated by the United States and its representative at the Population Commission, General William H. Draper Jr., who opposed the management of this new program going to the Population Division on the pretext that it would give too academic a twist to future programs. This decision was contrary to the advice of Philippe de Seynes, who complained to the United States' representative at the United Nations, none other that George Bush, future United States president, that the Secretariat had been pressured by General Draper and his assistants (Symonds and Carder, 1973). In fact, the motive was essentially political: the United States has always had great influence at the UNDP, whose director, right up to the present day, has consistently been an American. Their influence is more modest, on the other hand, at the United Nations Secretariat, to whom the Population Division is responsible. Alfred Sauvy, together with the USSR representative, had no difficulty, therefore, calling attention to the fact that this new fund's political management had in fact fallen to an advisory board operating within the UNDP, with neither political representativity nor legitimacy, whereas, legally, this function was among the Population Commission's responsibilities, whose mandate clearly included monitoring population policy and programs. This decision particularly reflected the weakening of ECOSOC, the Population Commission's regulating body, mistrusted by the United States since the Group of 77 and the Soviet Bloc had come to dominate it. This decision would have a durable influence on the fund's operation as, unlike the Population Division, it never attempted to appear politically neutral and took a decidedly Malthusian line throughout.

The first Fund director was nominated by the Secretary General in 1969. He was a high-ranking Philippine official, Rafael Salas, former Secretary General of the President of the Republic during President Ferdinand Marcos' time. Rafael Salas and his successors developed the fund in a remarkable way; starting out as a handful of collaborators, some of whom came from the Population Division, the fund has since opened offices in 112 countries and finances projects in 141 countries. In 2003, it received 398 million dollars entirely from voluntary contributions; the cumulative total since the fund's creation in 1969 has risen to almost 6 billion dollars. It has played an important and often unrecognized advisory role among Third World countries, which were mostly complete novices with respect to the organization and management of family planning programs. The often-archaic national administrations could certainly not serve as models. In fact, the newly created family planning services have often been the only modern sector of their government.

While family planning program funding continues to be central, UNFPA's activities have broadened considerably; in addition to promoting reproductive health, they include financial support for programs against violence to women, the AIDS epidemic, or the promotion of sustained development, for example (www.UNFPA.org).

Since the Belgrade Conference in 1965, an increasing number of nations—more than 30 in 1973—had adopted population policies aiming to reduce their level of fertility. Within the international community, the idea of a conference devoted uniquely to population policy problems had taken root. A small group of five experts, Jean Bourgeois-Pichat among them, was set up in 1969 under the aegis of the Population Division to explore the possible options concerning the organization of the next world population conference. This group recommended that this conference focus uniquely on policy questions, particularly those concerning population programs. The year chosen for the conference, 1974, was also designated World Population Year.

III. 1974–1983 BUCHAREST: THE VOICE OF THE THIRD WORLD

1. The Political and Institutional Context

The decision to organize a governmental level conference on population was neither an unexpected nor an isolated event in United Nations' history. It was not unexpected because it fell in line with the goals the *Population Movement*, very influential at the United Nations Secretariat, had been struggling to achieve. Neither was it an isolated event in the United Nations system. On the contrary, organizing this and other similar conferences during the 1970s fell directly within the context, from the late 1960s onward, of a deliberate UN strategy to regain, at the economic and social level, territory lost at the political one. Cold war clashes meant, effectively, that none of the era's great conflicts (Vietnam, Palestine etc.) could be resolved within the system. The idea of constructing a universal society with shared political values had to be abandoned in favor of identifying common interests in areas such as the environment, energy, oceans, and the like, or in promoting individual rights such as women's status, childhood, health, living conditions, nutrition, and, of course, population. The United Nations was still the only forum in which conflicts arising from positions taken by national governments, in areas where States were interdependent, could be arbitrated globally (Smouts, 1983).

Another interesting question is why the United Nations resorted increasingly to large specialized conferences for the preparation and adoption of policy documents on subjects as varied as nutrition, employment, the environment, women or population, and not, for example, to ad hoc sessions of its permanent deliberative bodies like the General Assembly, ECOSOC, or, in this case, the Population Commission. First, this choice demonstrates the declining status of these political bodies, where even debates on purely technical issues were, in the eyes of the developed world, at risk of being disturbed by the intrusion of political problems (e.g., the Palestine question in a discussion on women's status or population, etc.); it also shows a growing preference among industrialized nations for purely technical, temporary conferences which they can control all the more easily since they have often paid for their organization. Until recently, the 1989 UN Convention on the Rights of the Child (adopted nonetheless in 1990 at a one-day World Summit of Heads of States), and that of 1990 on the Rights of Migrants and their families, were among the rare instruments prepared and adopted outside a large conference. The adoption of the latter, incidentally, passed completely unnoticed by international opinion. Yet, it is precisely this awareness and mobilization of international opinion that are sought increasingly nowadays in areas such as the environment, women's status, population, and so forth. Large conferences, and the accompanying media coverage, are considered the best way to achieve these objectives. The United Nations deserve credit for recognizing early on that these conferences, on relatively well-demarcated issues, offer some kind of catharsis for international problems; by unblocking situations from which there appeared to be no way, it helps contributes to their speedier resolution. More recently, they have also provided public opinion, the media, civil society, and NGOs with opportunities to gather information, voice their opinions, and put pressure on governments (Chasteland, 1994). Finally, they are a stake in the power game for members of the epistemic communities, mentioned above, and of the UN bureaucracy. A successful conference improves the position of its instigators and organizers in terms of obtaining a broader mandate for their institution and the greater resources that go with it. This is a compelling and rarely mentioned motive, which nonetheless has a considerable influence on the decision to organize a conference, and in its preparation.

2. The Players

The UN Secretary General appointed Antonio Carillo-Flores, a former Mexican Foreign Affairs Minister, as the conference's general secretary. Two deputy secretaries assisted him, one of whom was Leon Tabah, director of the Population Division. No UNFPA representative was among this group of three whose role would be a decisive one. The UNFPA's only responsibility was organizing the Population Year.

The conference objectives were ambitious: to explore world population trends, the relationship between development and these trends, and the measures necessary to improve population welfare. In fact, the main aim was to discuss and adopt a World Plan of Action whose application would help reduce fertility and improve well-being particularly among developing nations. Aided by a group of five experts, the Population Division prepared a project for the Plan of Action and submitted it to the Population Commission. The resolution authorizing the conference stipulated, furthermore, that prior to the conference itself, five regional conferences (in Africa, Asia, Latin America, Europe, and the Arab countries) and four expert groups (in population and family; population and development; population, resources and environment; and population and human rights) should be set up, thereby separating scientific discussions and policy deliberations. The conclusions reached by the expert groups were intended to clarify the policy decisions taken during the conference. Rather a vain hope. The next two conferences, at Mexico in 1984 and Cairo in 1994, were organized in a similar way, the main difference concerning the preparation process for the final document: the Plan or Program of Action. By 1984, the Division, whose principal partner at Rome in 1954 and Belgrade in 1965 was the IUSSP, would also have a new partner, the UNFPA. At the Mexico conference, the latter even took over the political and administrative direction of world population conferences.

With the help of its expert group, the Population Division prepared a project for the Plan of Action containing approximately 100 paragraphs and submitted it to Population Commission members, who accepted it without major modifications. It was also presented for comment at the five preparatory regional conferences, and no particular opposition was visible on this occasion. As already mentioned, however, conference participants gave it a highly critical reception, and revisions were made to almost three-quarters of the original document.

One explanation put forward for these numerous alterations was the conference Secretariat's lack of attention to the current international political context, and the ensuing poor communication with the principal players on the international political stage in the early 1970s (Finkle and Crane, 1975). In retrospect, the late 1960s and early 1970s do indeed appear as the apogee of the Third World school of thought that perceived the unequal distribution of global resources as the principal obstacle to development which was, according to this line of thinking, the indispensable condition for subsequent fertility reduction. This was the position skillfully defended by Algerian leaders at the time and supported by Latin American countries such as Argentina, Brazil, Peru, and Cuba. China, which had not yet experienced its ideological *aggiornamento* on population, put its whole weight behind this group of countries. During this era, these nations, members of the Group of 77, had a quasi-hegemonic influence over the international organization. Some were ideologically close to countries from the socialist bloc, and many were crowned with the glory of their anticolonial struggles. At the United Nations they obtained what appears in retrospect to be a short-lived victory: a special session was organized at the General Assembly that adopted a Declaration on the Establishment of a New International Economic Order in May 1974. Accepted just before the Bucharest Conference, this declaration contained no reference to population issues.

In many Western nations, however, the tone was resolutely Malthusian. Public opinion was disturbed by the strongly pessimistic conclusions reached in the Club of Rome's report, published in 1972 (Meadows *et al.*, 1972). The report reintroduced the old obsession with physical limits to economic and population growth. Western, and especially American, demographers continued to be extremely uneasy about the world population's rate of growth. Since 1950, this rhythm had never stopped accelerating and, in the early 1970s, the fact that it was finally starting to slow down was as yet unknown. Based on data that could not take this changing trend into account, the United Nations population projections published on the eve of the conference overestimated population increase (United Nations, 1973).

The small group of experts who helped the Division prepare the Plan of Action project were relatively unaware of the deep ideological currents stirring the United Nations. The same was true for the Population Commission, which was largely composed of specialists of population questions. At that point, also, NGOs still had only a minor role in conference preparations, and in discussions taking place within the Population Commission. The power of American NGOs was undisputable, but at that time it was exercised more over the U.S. Administration than over United Nations organs. From the 1980s onward and with the Republicans' arrival in power, as we will see, they would increasingly exert their influence within the United Nations, which had become unavoidable in matters of population. Once completed, the project would reflect the main preoccupation of most Anglo-Saxon demographers: how, rapidly, to bring about a demographic transition which, left to its own devices, would develop too slowly. While not excluding the role played by economic factors in the process, the project only assigned them a function that might be described as contextual. Finally, the project makes two references to quantitative objectives concerning population growth. These references so much reflect the acceptance of constraint in this ultrasensitive area at that time that they have gained a symbolic value. India, with its sudden conversion to a coercive population policy in the late 1970s, is a typical example. It would be almost 20 years, after the downfall of Indira Gandhi's regime and the end of coercive measures, before India solemnly proclaimed that it was discarding quantitative objectives in the area of population. A more coherent vision of family planning was thus restored, legitimated above all by its voluntary nature.

The conference was to be held in Rumania, at Bucharest. Then a socialist nation, independent in its foreign policy but highly orthodox in its internal politics, Rumania was politically a fascinating country in which to organize the first world conference whose sole aim was to analyze policy issues. Its choice shows that a socialist country could give its backing in a way to the conference's main aim—legitimizing population policy within the international community—but with the inverse objective: to increase the birthrate.

3. The Outcomes

Conference participation was a success. One hundred and thirty-six of the member states were represented, out of the Organization's total membership at that time of 138. Of the NGOs invited to take part in the conference, 109 were present at Bucharest. A gallery had been reserved for them where John D. Rockefeller, president and founder of the Population Council, gave a much talked-about speech in which he admitted holding for too long the belief that family planning programs alone could solve the problems raised by demographic increase. "I now strongly believe," he said, "that the only viable course is to place population policy solidly within the context of general economic and social development" (Rockefeller, 1974). It was virtually a paraphrase of the famous Plan of Action formula on the need to integrate demographic policy into development (United Nations, 1975b, para. 1). Made by one of the era's eminent figures, this statement marked a withdrawal from the most orthodox American positions within the Population Movement. It corresponded relatively well, however, to what American researchers like Bernard Berelson were calling for, in the light of the ineffectiveness of early family planning programs essentially organized around the free supply of contraceptive methods. They were demanding an approach to family planning more oriented toward stimulating the demand for these techniques; an approach that, in addition to access to contraception, included direct measures of persuasion and dissuasion as well as indirect measures, such as improving education, health, women's status, and the like, which could encourage a change in ideal family size (Berelson, 1969).

As with all the large United Nations conferences, the media (more than 100 journalists had been invited) helped turn the conference into a real "happening" by rallying a budding international public opinion on the issue for several days. Although important at the time, this number of journalists would soon appear very modest compared with what would happen at future conferences.

Before examining the key results of the conference, a few words need to be said about a procedure commonly employed within the United Nations, but little understood outside it, regarding policy paper adoption: that of *consensus*. Remember that, after considerable modifications and ten days of heated discussions, the Plan of Action was adopted by consensus—by the 136 participating States with the exception of the Vatican. So why is the rule of consensus so important at the United Nations, compared with straightforward majority rule, for instance? Quite simply because most policy papers accepted by the United Nations are only recommendations and cannot therefore be enforced. In order for these recommendations to have an influence, they need to be universal in nature, hence the rule of consensus. Only under these circumstances will they be able to define new behavioral norms, applicable among others to the governments accepting them. Ignoring them can place these governments in a difficult position in relation to the international community as a whole, especially that these norms, as in the case of population, are promoted and defended by powerful transnational networks; besides scientists from the epistemic communities, these networks include leaders of powerful nations and international NGOs, and influential individuals on the international stage—the so-called transnational moral entrepreneurs, of whom John D. Rockefeller III was one of the most conspicuous representatives (Nadelman, 1990). Much more uncommon at the United Nations are documents, known as conventions, which are compulsory for the countries signing them. The United Nations wields substantial resources for activities in the population sector; allocated only to countries respecting the spirit, if not the letter, of the recommendations, this raises the chances of them being followed through.

To reach consensus, governments negotiate compromises. This was the case at Bucharest, where certain compromises were crucial to the conference's success. First among them was the famous compromise allowing those who believed that development alone could reduce population growth to reach agreement with their opponents, who viewed family planning as the sole route to controlling this increase. The idea of incorporating population programs into development planning was born of the conciliation between these two positions. It proved extremely useful, despite being difficult to put into practice. This first compromise allowed a type of ideological reconciliation between two long-conflicting parties. The second important compromise falls within the political sphere. The Plan of Action recognized national sovereignty in population policy decisions, while simultaneously defining the general orientation of what the international community needed to do in this area. The third compromise was an ethical one. The Plan stated that couples and individuals had the right to choose freely, but in a responsible fashion, the number of their children. It accepted individual liberty as long as it was part of a perspective of moral responsibility toward the community and, thus, did not contradict the sovereign right of nations to determine demographic policy.

Besides these compromises that brought reconciliation between opposing positions, the Plan had the immense merit of legitimizing the vocabulary and action of population policy at a national and international level. It also contained an impressive number of useful recommendations, although they failed to rank in terms of priority. Admittedly, defining priorities is extremely tricky in a world conference where too many different interests are at stake.

The Plan's value lies not only in these compromises, however fertile; at the ideological and political levels, it was also undeniably innovative in character. A number of recommendations concerning women's status and fertility (six paragraphs) are clearly ahead of their time. They even anticipate certain recommendations adopted at the first United Nations Conference on Women that did not take place until 1975. Also ahead of its time is the wording of the principle of right to free choice over family size of couples and *of individuals,* as became clear in Cairo when the same topic came under discussion. Likewise, the insertion of paragraphs regarding both legal abortion and maternal mortality in an international document is, in itself, a first. The Swedish delegation certainly had reservations about the abortion paragraph being placed in the plan of action's mortality section; they would have wished the issue to be debated during discussions on the question of reproduction and the family.

Moreover, the Plan displayed a solid economic realism, particularly in the attention given to the economic and social consequences of population inertia. Whatever the effect of population policy, it pointed out, world population would continue to increase, and this expansion would require the international community's economic mobilization (United Nations, 1975b, para. 13). The Bucharest conference benefited from publications from the United Nations' First Conference on the Environment, held at Stockholm in 1972, and from the population and environment expert group. Taking into account the fact that the topic was new, the policy content of recommendations on the environment was relatively weak, but they nonetheless had the merit of raising the issue 20 years before the Rio de Janeiro conference.

Finally, as would be the case with subsequent plans (or programs), except possibly Mexico, which was adopted during a period of radical ideological and economic mutation, the Bucharest Plan is a faithful enough reflection of the balance of power existing at the United Nations at the time of its adoption. Rereading it nowadays, it is hardly astonishing that it seems a little outdated in its excessive celebration of the virtues of planned development and the pre-eminent role of the State. The document remains, nonetheless, the first universal population policy document that the human race has ever agreed to discuss, accept, and implement.

IV. 1984–1993 MEXICO: THE UNITED STATES AND THE THIRD WORLD AT ODDS

1. The Political and Institutional Context

In the early 1980s, the United Nations envisaged organizing another world population conference. Discussions were tight. In 1981, the United States, who had led and inspired the *Population Movement* until then, elected a Republican president on a highly conservative platform that diverged from that of the *Population Movement* on many points. Associations demonstrating against abortion, sometimes violently, had a great deal of influence over the new Republican Administration. They expressed a wish to return to a moral order that excluded, or barely tolerated, abortion and even contraception. The prevailing spirit was also one of cuts in public expenditure, nationally and internationally. Ideologically, those who believed in state intervention in the social and economic sphere were on the defensive. Previously ignored by the administration, the works of the so-called revisionist economists or demographers, who rejected the neo-Malthusian vulgate, at last found a favorable echo in influential political circles. Their research findings fuelled the case of a Republican Administration reluctant to fund family planning programs for diverse ideological reasons (Simon, 1981; NRC, 1986; Kelley, 1991). The welfare state had become the scapegoat for all the developed world's problems, in the same way as planning had in the Third World, given the scope of its spread. The success of East Asian countries, where capitalism alone ruled, contrasted with the repeated failure of countries with a planned economy, of which Africa offered a sad example. The stagnation in Socialist bloc countries could no longer be concealed, at the same time as the Third World abandoned claims for a reformulation of international North–South economic relations. The Declaration on the Establishment of a New International Economic Order was well and truly buried.

The situation in Third World countries diverged increasingly. Whereas some East Asian countries had entered the world economy with outstanding success, others, especially in Africa and Latin America, were living an unprecedented crisis which, among other things, manifested itself in levels of indebtedness that appeared out of control at this point. These were also the final years of the East–West conflict, which did nothing to ease international cooperation, particularly since the United States suspected the United Nations of a systematic bias in favor of socialist and Third World countries.

In terms of demographics, the old polarization between developed and developing countries was now giving way to a new one, this time between developing nations. Of the 4.4 billion people on the planet in the early 1980s, only 900 million still remained outside the general movement toward fertility decline, which accelerated and spread between the early 1970s and the early 1980s. As mentioned earlier, although at the time of preparations for the Bucharest conference in the early 1970s everyone was still unaware that population growth rate had finally peaked, by the early 1980s the downward trend was no longer in doubt.

At the same time as the slowdown in population growth rates was becoming clear, the perception Third World governments had of their demographic situation was also evolving. In their 1973 government survey, the United Nations observed that 28% of governments, representing 57% of the world population, deemed their country's fertility level too high. In 1983, these proportions stood at 37% and 61% respectively. Furthermore, in 1983, 29% of governments stated that they were implementing fertility reduction policies, compared with 26% in 1973. Sub-Saharan African countries were bringing up the rear in this evolution (United Nations, 1975; United Nations, 1995c).

During the same period, empirical observations and theoretical research led to advances in thinking on the question of economic growth and its determinants. Thus, economic growth theories had cleared new ground, like that relating to the role of human capital in economic growth processes. Just before the Mexico conference in 1984, the World Bank, never slow to adhere to the dominant intellectual models, published its annual report on world development, and devoted it to the relationship between population and development. The report's content has been described as harboring a kind of "retreating" Malthusianism; if so, it is because the classic theme that highlights the negative relationship between population growth and physical capital formation has been replaced by one underlining population growth's same negative role but, this time, on the formation of human capital (World Bank, 1984).

2. The Players

These changes in the international context occurred at a time when one of its principal actors, the new U.S. government, was radically altering many of its predecessor's political choices. The new Republican Administration was quick to demonstrate that its position on abortion at the international level would be no different from the one taken internally. It suppressed all grants to UNFPA, on the pretext that the organization was financing pro-abortion measures as part of family planning programs, notably in China.

Public funding for family planning programs grew rapidly more international as the United States stopped funding a United Nations program to whose creation and promotion it had made such an important contribution (as it had also to the IPPF). This evolution also reflected the rising power within the United Nations, as in other areas, of Asian countries like Japan, and of European ones, like Germany, the Netherlands, Norway, and the United Kingdom, as well as that of the European Union itself, whose financial contributions partly compensated for those of the United States.

Finally, an intense lobbying organized mainly by countries receiving UN assistance (India, Pakistan, Mexico, Egypt, etc.), a few donor countries (the

Netherlands, Scandinavian countries, Japan, etc.), and UNFPA's executive director, won the day. China, who had experienced a remarkable about-turn in the field of birth control since 1974, also offered its support, but more discretely in order not to upset certain Group of 77 countries. In 1981, ECOSOC authorized the organization of a conference, this time named the International Population Conference. Its objectives were as follows: to evaluate the implementation of the Plan of Action adopted 10 years earlier at Bucharest, on the one hand, and, on the other, to make recommendations to improve the conditions of its implementation. Evidently, there were also power objectives surrounding the conference within the United Nations itself. In response to criticisms about waste made to the organization by donor countries, the General Secretary decided that they should economize in its preparation by using the existing personnel resources as far as possible. Thus, he nominated Rafael Salas, Executive Director of UNFPA, as the conference's General Secretary, and Sankar Menon, Director of the Population Division, as deputy. The UNFPA was to take responsibility for the political and financial aspects of conference preparations, the Population Division, for its scientific aspects, for the project to revise the plan and for the organization of four expert groups to meet prior to the conference (fertility and family; mortality and health policy; population distribution, migration and development; population, resources, environment and development). The United Nations' regional commissions were responsible for organizing five regional conferences.

In reality, the Director of UNFPA exercised absolute power throughout the conference preparations. The UNFPA took advantage of its strategic position, as it would do with even greater success at Cairo, to try and press its case for transforming UNFPA into a specialized United Nation agency, totally independent from the UNDP and its governing board (like the FAO, the WHO etc.), with certain member states. The fund was not supported by the majority of member states, some of whom, like France, could not accept a change in the organization's status quo, which would have implied the UNFPA absorbing the Population Division, and probably leading to dismantling the Department of International Economic and Social Affairs on which the Division was dependent and which was itself an integral part of the United Nations Secretariat. Unlike the UNDP, the Secretariat was accountable to all member states and not principally to donor countries, as was the case for the UNPD's board of directors.

Although the conference preparation was similar to Bucharest's, there were some important innovations in addition to the decisive role given to UNFPA. The preparation process for revising the Plan was one of the major changes. Bucharest's experience had shown that the process needed to be opened to the various political influences then dominating the United Nations if the conference Secretariat wished to avoid Bucharest's unpleasant political bombshells and the significant alterations that had to be made to the Plan project. The decision was made, therefore, to submit the Plan revision project to a Population Commission extended from the usual 27 members to include all members of the organization. Although it did not avert all the surprises, this decision made it possible to take better account of the views of all member states and helped reduce considerably the number of amendments made during the conference, whose duration had been reduced to 6 days (from 10 at Bucharest).

3. The Outcomes

Once again there was an excellent level of participation, with 147 countries out of 157 member states represented at the conference, held in a developing country, Mexico. In 1973, on the eve of the conference, this country had adopted a demographic policy that subsequently proved to be relatively effective. Besides the official delegates, 154 NGOs were also present. Finally, over 300 journalists helped enormously to arouse international public opinion. As at Bucharest, scientific activities were separated from policy ones, with the former concentrated in the period before the conference around the four expert groups already mentioned. The conference focused exclusively on discussing and adopting the revised Plan of Action project.

Despite all the precautions taken by the organizers, elements external to the conference disrupted it as they had in Bucharest. First, there were interminable discussions about peace and security, which, in retrospect, appear as the final manifestations of the East–West conflict. Two other external issues also complicated the deliberations, one with regard to the formulation of recommendations concerning Third World debt (a North–South confrontation), and the other, the question of Israeli settlers on the West Bank (involving Arab countries, the United States and Israel).

The debates were disrupted most of all, however, by the United States' change of position before the conference and the attitude of an American Administration that was more interested in gaining support from the Republican electorate (just one year before presidential elections) than from Third World countries (Finkle and Crane, 1985). Thus, the American delegation in Mexico declared that population growth was unrelated to economic development. It sufficed to free market forces, end state intervention, and, stimulated in this way, economic growth would make it possible to meet the needs raised by population growth. The Administration paid lip service to the idea of supporting family planning to the extent that it contributed to family well-being and was voluntary. Evidently, this line of reasoning removed any economic justification for family planning programs. Paul Demeny (1994) described the U.S. Administration's about-turn as a temporary setback. As the U.S. government has still not returned to its pre-1980s position since this "temporary setback," the expression was perhaps a little over-optimistic. As our analysis of the Cairo conference results will show, there definitely appears to exist in this country an enduring opposition toward government intervention in issues of population policy. This was certainly the message passed on by the U.S. Congress, Republican in majority, contradicting that of President Clinton's Democratic Administration, and recently taken up again by the current president, George W. Bush.

To return to the Mexico conference, the 88 recommendations in the Plan of Action's revised project were adopted, as was a declaration, thanks to the skill of its organizers and the unqualified support of Third World countries. Any modifications, it should be noted, left intact the principles and objectives accepted at Bucharest (United Nations, 1984). Given the restrictions put on the organizers, no radical initiatives could be expected from the revised Plan. Many of the recommendations replicated, with clarifications, those adopted at Bucharest. The conference did innovate, however, by extending the number and importance of recommendations related to the environment and, in particular, to women's status, which were treated in a section of their own. This progress was due partly to pressure exerted on the spot by American feminist NGOs, who were well represented in Mexico because of its proximity; it signified their strong entry into the population debate. At the international level, the same remark could be made about environmental questions. The debate on abortion made a public confrontation on the issue possible. As at Bucharest, the problem of abortion was raised in the section on mortality. Sweden once again expressed reservations, as it would have preferred a more liberal presentation of abortion issues. All its attempts were blocked, among others by the American delegation, concerned not to upset an electorate opposed, in majority, to abortion. As a whole, the Mexico program stayed carefully within the lines drawn in Bucharest's Plan of Action, even though, as we have seen, the politico-ideological context within which it was adopted had undergone rapid changes.

Finally in Mexico, as mentioned, through its opposition to the Population Movement's traditional theses, to whose formulation it had nonetheless so powerfully contributed, the American administration freed the Movement, and UNFPA, from its control and accelerated their internationalization (Finkle and Crane, 1985). Experience would show, however, that American NGOs filled the gap this created to some extent, until the return to power of a new Democratic Administration.

V. FROM 1994 TO THE PRESENT DAY: CAIRO, AN ANGLO-SAXON VICTORY

1. The Political and Institutional Context

In the early 1990s, neo-liberalism became the dominant ideology and only a few isolated countries, such as North Korea, Cuba, and Vietnam still proclaimed themselves socialists, China standing alone as a particular case. With the collapse of the Socialist bloc, the Group of 77, whose Third World ideology, as mentioned earlier, often drew its inspiration from the same Marxist sources as the Eastern bloc countries, was reduced to a group without clear political direction. Asian countries in full expansion, Latin American nations emerging from an economic quagmire, and African countries on the verge of economic disaster coexisted within it with ever-increasing difficulty.

The one-time objectives of transforming the economy and giving priority to the struggle against social inequality gave way to new objectives of State disengagement. Leaving market forces to stimulate economic growth, the state withdrew to the role of umpire for the protagonists involved, while at the same time reinforcing its stately functions. In this context, the traditional struggle for collective rights to be recognized in the economic and social spheres, in which the State and unions had a decisive role to play, has shifted toward new demands bearing on rights related more to individuals or group identity, where other players from the civil society (e.g., NGOs) play a key role. The struggle against social inequality was thus somewhat demoted to second place, in favor of the struggle against status inequalities. This was the context within which the Cairo Program of Action was conceived and which, symbolically, ceased to be a plan as at Bucharest.

In practice, it quickly became obvious that creating an open market could marginalize whole groups of the population in many Third World countries; this was soon echoed in international documents, in which were multiplied the number of recommendations aimed at protecting the most vulnerable groups, such as children, the elderly, women, rural populations, or those termed outcasts, and in which, more generally, the theme of the struggle against poverty received systematic treatment. In this struggle, however, when the State was called in to help, any interventions were viewed as humanitarian measures rather than as a return to a global policy of social justice.

At this point in time, most developing countries certainly had no wish to resolve their economic, social, and, evidently, demographic problems on such terms. It was only later that they came round to the neo-liberal policies omnipresent today.

The debt crisis, which surfaced in the late 1970s and continued into the years preceding the conference, provides a striking illustration of the gulf still existing at this time between the Group of 77's point of view and that of the rich countries. For the 77, this crisis blocked any chance of development and thus the ability to solve their population problems. The wealthy nations, headed by the United States, perceived the use of this argument in international debate, particularly when discussing population issues, as a way of avoiding the real demographic problems and, more particularly, of challenging the new global economic order that was emerging. A theme that united the poor countries, these debates can be seen as the weakened Group of 77's final attempt to recover their influence, lost after their short-lived victory with the adoption of the Declaration on a New International Economic Order. The way in which this crisis was resolved, notably by very strict economic restructuration measures imposed by the International Monetary Fund (IMF), confirmed their waning influence on the international stage. The 77's failure also marked the United Nations' ideological alignment with the neo-liberal ideas of the IMF or the World Bank, dominated as they were by wealthy Western nations; until the 1980s, strengthened by its representation of the Third World, the United Nations had asserted its ideological difference.

Worries about the environment added to those raised by the persistence of the 1980s economic crisis. The Bruntland report (United Nations, 1987) crystallized these concerns and succeeded in mobilizing both public opinion and governments, resulting in the adoption, at the 1992 Rio de Janeiro conference, of a program of action, Agenda 21, whose aim was to make it possible to bequeath a preserved environment to future generations. Concepts such as "durable development" and "sustained growth" today appear inescapable for political and economic decision makers, and mark the strong return of a concept at the very roots of Malthusianism—that of physical limits to growth. Cited in abundance in the Program of Action adopted in Cairo, they are also the expression of a new alliance between ecologists and family planning enthusiasts. We will return to this later. This alliance was all the more necessary that, in the United States especially, economists and demographers, charged with revisionism, undermined the very foundations of the economic justification for fertility reduction. This questioning did not pass unnoticed by a Republican Administration, little inclined to support these programs for diverse ideological reasons, as we saw at Mexico. The United States were in any case against funding public assistance programs, believing that this approach was ineffective and that only the free play of market forces would make economic take-off possible in developing countries.

The demographic situation at the start of the 1990s accentuated still further the trend toward greater polarization among Third World countries, already in evidence in the early 1980s. Government leaders and public opinion would take all the longer to become aware of this radical change because, although widespread, fertility decline was still not universal, and especially because the timing of entry into the demographic transition between 1970 and 1990 varied greatly (see Chapter 69, Volume III). This polarization made for an ambivalent interpretation of the observed evolution. For the optimists, the evolution between the early 1980s and early 1990s was extremely encouraging. The proportion of the world population with a total fertility rate (TFR) at replacement level (2.1 children per woman) rose from 1.2 billion to 2.7 billion during the period. The percentage whose fertility was declining significantly (between 2.5 and 4.5 children per woman) reached 2.3 billion in 1990, compared with 2.8 billion in the early 1980s. Finally, although numbers had fallen from 900 million to 750 million in 10 years, at this time there nonetheless remained a sizeable proportion of the world population still delaying their entry into the second phase of the demographic transition, the phase in which fertility starts to fall. A growing proportion of the population in developing areas, therefore, were showing a certain convergence toward an evolution close to that experienced by developed countries. For their part, the pessimists emphasized the fact that countries where fertility had not yet started to fall were among the poorest on the planet, located for the most part in sub-Saharan Africa, a region which had faced enormous problems during the previous two or three decades, or in South Asia, an equally underprivileged region. They added that the United Nations projections medium scenario indicated that fertility in most African countries would not reached cohort replacement level until after 2025, by which time their population would have doubled, reaching 1.1 billion, compared with 500 million in 1990. This growth constituted a considerable challenge, given their present and future economic potential, not to mention their political situation. Furthermore, countries in transition, which included the most populous in the world, still had considerable potential for population growth; this potential was all the greater given that fertility decline was slow in some countries, or had scarcely begun (as in Nigeria, e.g.). Finally, basing themselves on the medium scenario of the United Nations long-term projections, the pessimists added that the level at which the world population would stabilize moved from 10.2 billion to 11.6 billion in 2100, depending on whether the projections referred to were estimated in 1981 or in 1991 (United Nations, 1991). With the United Nations long-term projection cycle before last (1998) putting the probable level of stabilization once again at 10 billion, it seems today that the pessimists were mistaken (United Nations, 2000b). Ecologists added further to the demo-economists' pessimism by stressing that those yet to be born in developing countries, whether or not they had completed the transition, would be additional producers and consumers whose improved living standards would help destabilize further the broad ecological balance. Paradoxically, the march toward progress of ever-growing populations thus became a new source of concern.

Evolving attitudes toward population growth issues among Third World governments further reinforced this pessimistic trend. In 1993, before the Cairo conference, the last survey that the United Nations carried out on government attitudes toward their country's demographic trends, and on measures taken to respond to problems raised by these trends, revealed a significant evolution during the previous decade. Compared with 37% in 1983, 45% of governments then thought that fertility levels in their country were too high. Moreover, in 1993, 41% of governments interviewed stated that they had adopted a fertility reduction policy, up from 32% in 1983 (United Nations, 1995c).

With growing governmental support for population policy, one might question the need to organize a third conference, extremely expensive in both money and manpower. The Bucharest conference was a response to fears raised by two decades of rapid population growth. Any decision to intervene in this required global legitimacy, and the ethical and political rules of this unprecedented intervention needed to be specified. In Mexico, the situation had already changed: the legitimacy of these programs was no longer in question, especially in the Third World. On the other hand, it was once again necessary to overcome reservations surfacing in donor countries about funding programs whose macro-economic justification was beginning to be replaced by a humanitarian one. On this justification also depended the funding of many other programs that thus entered into competition with family planning programs.

The early 1990s brought an end to the cycle opened by the Bucharest conference. Some believed that it was virtually impossible to update further the Plan of Action that already contained, in its two versions, close to 200 recommendations whose innumerable directives were no longer very convincing. Problems with funding surfaced once more, as the United States republican administration continued to refuse funding to the UNFPA; at the same time, it also started to reduce its bilateral aid. Other countries had taken over (notably Norway, Germany, the Netherlands, and Japan), it is true, but some sponsors, such as Canada, Finland, Italy, and Australia, had also reduced their contribution or stopped it altogether. Third World aid was no longer a winner, and population programs had lost their former priority. As in 1984, it was necessary to try and find ideas that could rouse international opinion and stir it into action. One such idea had already been used with some success by the Population Movement: to link population problems to environmental ones. The second idea was to associate population problems with those of improving women's status (McIntosh and Finkle, 1995).

2. The Players

The same bureaucratic forces at work within the United Nations during preparations for the Mexico conference once again intervened in a decisive way to convince member states of the need to organize this conference. For the UNFPA—directed since Rafael Salas' death by Nafis Sadiq, a woman physician of Pakistani nationality—the conference also opened up the possibility of finally changing the institutional status quo in its favor, enabling it to gain independence from the UNDP and be promoted to the rank of specialized United Nations agency. This was not an unrealistic perspective, given on the one hand the weakening of the United Nations Secretariat, funded by compulsory contributions that were paid ever more unwillingly by some member states, such as the United States and, on the other hand, given the opposite trend, namely the rising power of United Nations programs like UNFPA, which were funded by voluntary contributions. For the first time in the history of plans of action concerning population issues, a paragraph was included concerning the reorganization of this sector of activity (United Nations, 1995b, para.

16–26). The growing strength of UNFPA's position emanates from the expansion of its activities and financial power (its resources grew from several tens of millions of dollars in 1973 to more than 370 million in 2001). Consequently, it is also linked to the fact that donor countries, members of the Population Commission, increasingly turned their interest toward the UNDP's governing body, which managed programs financed by voluntary contributions like those of UNFPA, and where their political control could be freely exercised. This was where the true power games were played out in the early 1990s. More symbolically, the *World Population Situation Report,* submitted regularly by the Population Division to the Population Commission, went completely unnoticed, while the *State of the World Population Report,* submitted by UNFPA to the UNPD governing body, drew the attention of the media, and thus of international opinion, thanks to a huge public relations drive.

In 1989, ECOSOC authorized the organization of a population conference for 1994. Other resolutions later defined the modalities of the organization, its objectives, and the conference location, namely Cairo. Its objectives were the following: to evaluate progress made since the preceding conferences; to heighten awareness in the international community of population problems arising in relation to development; to provide guidelines for the way in which population issues should be broached at the national and international levels; and to adopt the new recommendations. The General Secretary nominated Nafis Sadiq, Director of UNFPA, as conference general secretary, and the director of the Population Division as her deputy general secretary.

Responsibilities were divided, and the preparation organized, much as they had been in Mexico. UNFPA kept its political and financial duties, while responsibility for the scientific content fell to the Population Division. The latter consisted essentially of organizing six meetings of experts before the conference on: population, environment and development; population policy and programs; population and women; family planning, health, and family well-being; population growth and demographic structures; and, finally, population distribution and migration (no expert group was to deal with mortality). It also included preparing the project for the Program of Action (the word *plan,* used for the final conference documents in 1974 and 1984, was dropped). Moreover, five regional conferences were to be held before the Cairo conference. The conference itself, titled the *International Conference on Population and Development* (ICPD), was to discuss and adopt the project for the Program of Action. As before, the most important modification in the preparation process concerned the drawing up phase for the project of the final document. In addition to extending the Population Commission to all UN member states, as it had in 1984, the decision was made to open it equally to NGOs, and to organize numerous meetings with individuals and representatives of the most diverse groups, in order to gather as many reactions to the project as possible before the conference (more than 30 meetings in all).

Opening the Population Commission to NGOs other than those generally associated with it, and inviting these same NGOs into the national delegations, was a decisive element in the orientation the conference would take. It effectively meant saying goodbye to the closed circle of government delegations and welcoming civil society and its preoccupations. Increasingly solicited by government leaders who delegated all or part of their responsibilities to them, these NGOs became inescapable players at the national level (Haas, 1992). It remained to establish whether these NGOs truly represented civil society, itself far from a homogeneous entity, and to ascertain from which countries they originated. The influence of some NGOs in international organizations was comparable, relatively speaking, to that enjoyed by large multinationals (Keck and Sikkink, 1998). With regard to American NGOs, who played a decisive role in this conference, two points can be made. In the United States, NGOs within the population sector were, and still are, financed primarily by public funding (United Nations, 1994, 2002). This was not the case for NGOs working to promote women's status or the protection of the environment, and they thus had more room for maneuver.

A substantial discrepancy has always existed in the United States between the progressive position on population issues held by the majority of NGOs and that of the essentially conservative so-called moral majority (representing the broad mass of mid-America). The progressive discourse of the former nonetheless differs markedly from the Third World or Marxist ideological perspective: little concerned with questions of social injustice, Third World debt, or unequal development, it essentially focuses on inequalities whose origins stem from biological differences (e.g., sex), ethnic identity, or sexual preferences. Within international bodies, this is the most strongly represented group and, until now, it has exerted most of its influence outside the United States, although this situation may well change given the country's current political context. Like NGOs dedicated to population questions, feminist NGOs, primarily North American (especially the New Organization for Women [NOW], which played a critical role during the preparation of the Program of Action and during the Cairo conference), seized the long-awaited occasion to present their viewpoint on population policy. Their influence can be analyzed as follows: on the one hand, they criticize the macro-demographic approach of traditional family planning programs that do not adequately take into account the right of women to control their own sexual health, including, if necessary, their right to abortion in suitable conditions; on the other hand, they are also critical of development strategies that do not take environmental constraints into account, particularly for women in the Third World (Dixon-Mueller, 1993; Germain, 1987; Sen and Grown, 1987; Sen and Chen, 1994). This was the crossroads where the two ideological trends dominating the conference—the feminist and ecological ones—could meet (McIntosh and Finkle, 1995). This coalition could have been created at the expense of the traditional family planning trend, represented by the IPPF, had not the latter already associated itself with it at the end of the 1980s. It was not by chance, therefore, that the IPPF adopted objectives similar to those of the Cairo conference.

These feminist and ecological trends emerged from the Anglo-Saxon world, in power for the last half-century, and bear a strong Protestant imprint which, in the face of society's great problems, places individual responsibility above all other, and has absolute faith in instrumental rationality to resolve them (Finnemore, 1996). Islam and the Latin world remain outside this ideological canvas, and of its conception, for instance, of what the relationship between man and nature ought to be, and of women's place in society (Lassonde, 1996). This Protestant imprint is just as strong in the economic domain, especially in the perception of relations, on the one hand, between the individual and the economic laws of the market, and, on the other, between the individual and the State. To find this essentially American hegemony reflected in many spheres is an important aspect of globalization.

The Program of Action project, drawn up by the Population Division under UNFPA's rigorous supervision, was presented to the Population Commission for adoption. Difficult discussions ensued. The document's very conception, the emphasis placed on "reproductive health" and what this implied in terms of individual freedom in sexual matters, especially that of adolescents and women, as well as the importance given to the abortion problem and to new family forms, provoked a severe reaction from the Vatican and from both Catholic and Muslim countries who viewed it as challenging the values at the heart of their vision of society. Recent studies have shown that the adhesion to, or rejection of, these values constitute one of the most important cultural splits existing between Islam and Western nations today (Inglehart and Norris, 2003). For the first time, a religious divide emerged during preparations for an international population conference. This new split replaced the earlier secular ideological divisions (Marx vs. Malthus), which pitted the East and the West, or the North and the South, against each other at the United Nations. The project document, more than 100 pages long (a record compared with the Bucharest and Mexico texts of around 30 pages each), of which more than 80% had been approved within the Commission, went to Cairo with the most contested paragraphs still without approval.

3. The Outcomes

Conference participation reached record numbers: of the 190 United Nations member states, 182 national delegations were present, bringing together 3,500 delegates. Particularly unprecedented, however, was the importance given to NGOs and the media: 1,257 NGOs attended (of which 23% were American and 15% European), as well as 3,800 media representatives—over 40 times more than at Bucharest. This civilian presence played a significant role, constantly exposing the delegations' deliberations to the gaze of an international public opinion largely won over to the Population Group's perspective, reinforced on this occasion by the ecologist and feminist movements. As at the conference on the environment at Rio de Janeiro in 1992, the conference on women in Beijing in 1995, and at Seattle's trade conference in 2000, the forceful arrival on the international stage of those representing global public opinion resulted in a sort of marginalization of governments in the discussions, and it led to a questioning of, if not the principle, at least the current architecture of these large international conferences. The 1996 Social Summit at Copenhagen, a modest attempt by the United Nations to restore their lost priority to social questions, was a semi-failure, despite adopting such innovative programs as the Program 20/20 (20% of donor countries' aid, and 20% of receiving countries' budget to go toward social programs). A semi-failure with regard to the unwillingness of some industrialized nations to put social questions back at the center of economic debates, the conference was also a semi-failure for the United Nations, who had seen it as an opportunity to recover the ideological initiative, in relation to the IMF and the World Bank.

In one sense, the conference discussion held few surprises compared with what had already surfaced during the preparations. The Vatican, certain Latin American and African Catholic countries, and the

Islamic States were still among the nations most opposed to certain of the Program's paragraphs, particularly those concerned with reproductive health, the family, and women's status. The industrialized nations, Asian countries, and a good number of Latin American and African countries were in favor. The views of the former Eastern bloc as such were obviously no longer voiced, and the Group of 77, presided over by Algeria, kept a low profile.

The project was finally accepted after very lengthy discussions around the controversial paragraphs. This was the case in particular for paragraph 8.25 on abortion and paragraph 5.9 on family forms; a compromise in the document made it possible to unlock the situation in both cases (United Nations, 1995b). The wording of paragraph 8.25 provides an excellent example of the application of what Louise Lassonde termed the "principle of juxtaposition," which consists of juxtaposing rather than synthesizing the most antithetical positions. It is hardly astonishing that the phrasing of recommendations drawn up following this principle is rather obscure!

To general surprise, the Program was adopted by consensus, including the Vatican's. This consensus was accompanied by so many qualifications on the part of the Vatican and other member states, however, that one might wonder whether it was not an entirely new form of consensus. The final document, 100 pages long, contained 11 pages of provisos, a record compared with Bucharest and Mexico. More worrisome, however, was the introduction of a new principle making the application of the Program of Action conditional on the civil and religious laws existing in each country; this principle existed only in its secular form in the Bucharest and Mexico documents, and considerably weakened the Program of Action's field of application.

Comparing the Bucharest Plan and the Cairo Program highlights the degree of innovation in the areas of action envisaged, and especially in the conception of the document as a whole. First, as already mentioned, the Cairo document is extremely voluminous: 100 pages compared with 30 at Bucharest. The conference secretariat registered 243 recommended actions (without taking subdivisions into account) as opposed to 108 in Bucharest. Consequently, the first impression given by the Cairo document is one of great depth and detail in its recommendations, whereas the Bucharest text remained at a more general level. Given the above-mentioned references to the context in which the Program of Action was prepared and adopted, it would have been astonishing if it had been no more than a simple update of the preceding plan. At the purely formal level, a report from the conference secretariat identified approximately only 60 recommendations arising directly from the Bucharest plan, the others dealing either with new problems or broaching old ones in a new way (United Nations, 1995a). Just as when Agenda 21 was drawn up during the Rio de Janeiro conference, the desire to produce a self-sufficient document, with all the recommendations justified by a contextual analysis and a detailed presentation of the objectives, also accounts for the document's large size.

At a more fundamental level, the perspective selected to surmount the population problem was very different from that of Bucharest. This time, it involved encouraging the realization of individual aspirations for improved well-being; promoting reproductive health constituted an essential but not unique element of this well-being, as improvements in reproductive health should also lead to a control of fertility that would benefit the whole group. In implementing this policy, women were the strategic group; the main efforts would, therefore, be directed toward them, not only in the reproductive health sphere but also in terms of education and, more generally, with regard to everything involved in improving their situation. Incidentally, although the French representatives remained rather marginal in these discussions, it should be noted that they could have claimed the intellectual paternity of this female-oriented strategy, as the Marquis de Condorcet advocated it as early as 1792 in his *Sketch of the Progress of the Human Spirit*. On the international scene, the concept of reproductive health somehow sanctioned the fact that women were now taking control of the reproductive sphere, as a result of efforts made by feminist movements for a radical redefinition of the traditional family planning perspective. At Bucharest, fertility reduction was expected to ensue both from the diffusion of family planning in its traditional sense and from the beneficial effects of planned economic growth; at Cairo, however, fertility control appeared essentially as a consequence of specific measures promoting individual rights, in particular those of women. No longer central, as in Bucharest, the role of economics had moved to the periphery of the Program of Action.

In the Program of Action adopted in Cairo, however, elements brought to the forefront in Bucharest were reaffirmed, such as the fact that family planning programs will only be effective if integrated into other broader programs, and if, within this context, individual freedom and responsibility are respected. Most noticeably, new elements entered the picture, such as reproductive health, evidently, a concept that encompasses family planning in its traditional sense, as well as other aspects of sexual life; the essential role given to the promotion and education of women as decisive factors in the evolution of fertility; recognizing free access to contraception for all individuals, even adolescents; problems linked to abortion; the existence of diverse family types; men's responsibilities; on quite another level, environmental problems; and, finally, the need to establish quantitative objectives in the realm of aid for population programs. In the panoply of measures adopted, no reference at all was made to measures of persuasion or dissuasion, so popular in the 1970s.

Although these are certainly positive and innovative elements, the perspective taken led necessarily to important omissions. It is impossible not to be struck by the absolute precedence given to individual considerations at the expense of macro-demographic ones, or to purely demographic issues at the expense of those in which population and economy are closely linked, as was the case at Bucharest. Thus, the economic and social consequences of population growth associated with population inertia are scarcely mentioned. Even though deliberations for a detailed program of economic or social measures could not be expected in a conference devoted to population policy, one might have hoped that the conference would at least provide an opportunity to raise awareness of the considerable problems created by consequences of spontaneous or controlled demographic evolutions (Vallin, 1994). In a conference on population and development, never has the necessity of economic mobilization received so little attention, nor poverty so much, with so little discussion about social inequalities, one of poverty's principal causes. Is this the consequence, already mentioned, of giving priority to the fight against group inequalities at the expense of the struggle against social inequalities? Finally, among the other subjects scarcely touched upon, or even evaded, were questions of bioethics (abusive use of prenatal diagnosis of sex, human cloning, etc.), adult mortality (except AIDS), internal migration (primacy of urban lifestyle, growth of megacities), or international migration (multiplication of the number of countries involved), as well as those linked to population aging (including in the so-called developing countries). More generally, it could be said that the conference missed a perfect opportunity to examine the demographic problems of developing and developed countries jointly; this synthesis could have brought their positions closer and underscored future convergences, persisting differences, and the interdependencies and their consequences, especially in matters of international migration. Once again, the global population problem was addressed almost uniquely through that of developing countries, as though developed countries were in no way involved. In any case, it becomes increasingly difficult to justify this approach given the growing homogeneity of demographic situations, and the increasing irrelevance, in this respect, of the "developed countries" and "developing countries" categories.

Finally, as Alison McIntosh and Jason Finkle (1995) noticed, the participants at this conference, as at other conferences of this type, tended to follow the current political trend, to be "politically correct," and therefore focus on themes popular with public opinion. This convergence is not surprising, and, as already mentioned, it reflects the diverse conference players' ever-growing participation in a global/worldwide society that establishes universal behavioral norms in areas such as population, as well as in the environment, women's status, childhood, and so forth (Meyer *et al.*, 1997). Alongside the United Nations and other international organizations where, traditionally, consensual norms have been drawn up, particularly during the last 20 years, NGOs have had a growing and sometimes competing role in the creation and diffusion of these same norms at the national and international levels. As might be expected, the norms promoted by NGOs are not neutral in relation to their countries of origin. They often reflect their country's ideologico-political choices, and the NGOs influence depends directly on the power (at times termed *soft power* in opposition to military or political power, known as *hard power*) and influence enjoyed by these same countries in the international arena (Boli and Thomas, 1997). The media and experts, often selected because they adhere to these norms, then disseminate or elaborate these very norms, each in their own way.

4. After the Cairo Conference: A Preliminary Assessment

It is, of course, impossible to appraise in a definitive way a still unfinished Program of Action; such an evaluation would, in any case, be delicate for a number of reasons. One stems from the qualitative nature of many recommended objectives. It is difficult to assess whether these objectives have been reached entirely, partially or, even, not at all. As for the quantitative objectives, in the absence of evaluation methods like those used for family planning programs, for example, it is impossible to distinguish the effect specific to Cairo programs from that of other measures which would have been taken in any case, even without this program, not to mention the intersecting influence of different recommendations. Moreover, all countries do not define the content of their population programs in the same way. By adopting the concept of reproductive health, which broadens considerably traditional area of application for population programs, it has become even more difficult to obtain a precise and comparable definition of population program content from the different players involved. Generally speaking, the examination and evaluation process was devised more to verify the extent to which the accepted recommendations had or had not been implemented than to assess their intrinsic effectiveness (United Nations, 1999). During the Examination and Evaluation of Bucharest's Plan of Action in Cairo, it was clear that government delegations, little used to the problems of evaluation, were only superficially interested in the exercise (United Nations, 1995c). Finally, the extent to which these international programs were implemented depended on how implicated the political leaders of the countries involved were, and thus on how committed they were to the norms defined in the Cairo program. On this topic, in-depth surveys were carried out in some countries (Bangladesh, Ghana, Jordan, and Senegal); none could clearly indicate whether the commitment displayed by these countries' political leaders stemmed from their own voluntary adhesion to the program, or whether it reflected a pragmatic attitude toward a program whose implementation would imply substantial financing (Luke and Watkins, 2002).

As we have seen, one objective of large international conferences is to convince governments and the other players involved to make additional resources available; evaluating the financial resources unlocked by these conferences, therefore, is one criteria of their success or failure. It is possible to track the financial resources allocated to population programs in the years preceding and following the Cairo conference. From 1993 to 2003, the amount of aid provided by donor countries rose from 1.3 to 3.7 billion dollars (loans not included). As with previous conferences, however, it seems that the mobilization it stimulated may not last. Some authors write of "cycles of attention" toward specific problems that determine the degree to which public opinion and governments' involvement are awakened, and thus the level of financial contributions (Schindlmayr, 2001). Also, the rising anxiety among a growing number of donor countries about a potential decline in their own population seems to weaken their commitment by obliging them to take an almost schizophrenic position at home and abroad, and helps explain their decreasing interest. Whatever the case, the figure of 5.7 billion dollars in 2000 is far from the one the Cairo conference was looking for. It is nonetheless possible that more funding has, in fact, been allocated to population programs as defined at Cairo, given variations in the definition of such programs among the donor countries. In this connection, it can be said that the allocation of funds by type of activity has significantly evolved. Family planning activities received only 461 million dollars in 2003 against 723 in 1995. In contrast, activities related to reproductive health have increased during the same period by 275% and those related to the fight against AIDS by 1300% (United Nations, 2005). Until the Monterrey International Conference on Financing for Development in 2002, a stagnation or a downward trend of Official Development Assistance (ODA) was observed. It seems that this conference and the adoption in 2000 by the Members States of the United Nations of the Objectives of Development for the Millenium have reversed this trend. Thanks to the commitment of a handful of donor countries, assistance in the field of population has passed the threshold of 2% of ODA to reach 4% in 2003. As a final point, it is worth noting that contributions made by private foundations and NGOs rose from 240 million dollars in 1999 to 324 million dollars in 2003, and, in contrast with ODA before Monterrey, was steadily growing (United Nations, 2002). These contributions come principally from well-established American foundations, as well as from more recent ones, such as the MacArthur, William and Flora Hewlett, David and Lucile Packard, Warren Buffett, William Gates (76 million dollars in 2001), and Ted Turner (one billion dollars between 1997 and 2007) foundations. In addition to these contributions, donations made to a new United Nations Fund for International Partners (UNFIP) are partly committed to population programs (United Nations, 2000a).

An assessment of population program funding needs to underline the fact that more than two-thirds of these programs receive funding from the national budgets of developing countries. A recent United Nations study estimated at 11 billion dollars the internal resources allocated by governments of developing countries to population programs in 2003 (United Nations, 2005). To this figure should be added 1.1 billion dollars contributed by private individuals, bringing the total to 8.9 billion (United Nations, 2000a). The majority of these funds are nonetheless concentrated in a few large countries, like India and China, whereas other countries are not involved at all (e.g., countries of sub-Saharan Africa). Finally, public funding predominates at the international and national levels.

Beating all records in terms of its size and cost, the Cairo conference appears to be the last in the decennial cycle of large world population conferences started at Bucharest in 1974. Faithfully reflecting the power struggles and contradictions of their times, these conferences also acted as midwives, at times unexpectedly, for programs that would influence population policy throughout the world. In 2004, a modest cycle of conferences organized by United Nations regional commissions, whose results will be analyzed by the Population Commission (renamed the Population and Development Commission since 1994), ECOSOC, and, lastly, the United Nations General Assembly, was all that was on offer. These activities are only commemorative, rather than seeking to create a new order, as was the case in Cairo in 1994.

Reflecting current priorities, a Global Commission for International Migration was recently created, jointly presided by a Swedish public figure and a South African. This Commission will be modeled on the World Commission on Environment and Development whose efforts were published in 1987 as the *Bruntland Report*. The majority of governments asked about the direction this commission's work should take, and who responded to a United Nations Secretary General questionnaire on the question, do not appear to wish it to work on policy problems as such, and to have resolutions accepted that could put constraints on member states (United Nations, 2003a). This dual movement toward regionalization and sectorization undoubtedly signals the waning influence on the international stage of the Population Movement, whose global political project had somewhat served the interests of environmental and feminist groups; it is also an adaptation, however, to a new demographic reality characterized by an extreme disparity in circumstances between countries with fertility levels that are still very high, and the growing number of countries in which fertility has plummeted.

VI. THE FUTURE OF GLOBAL POPULATION POLICY

One thing is almost certain about the place of population questions on the national and international political stage in the future: they are not about to disappear. The growth, composition, and movement of populations will always raise policy questions, but two fundamental changes will make it compulsory, among others things, to create new solutions. The first concerns the political, economic, and cultural context, transformed by a globalization whose effects we are only just beginning to see. Extending far beyond its most visible economic, technical, and cultural consequences, the latter "from then on inserts social facts [and thus demographic ones] in a causal chain whose point of departure is global, not local"[2] (Laidi, 1998; p. 95). It would be astonishing if globalization did not elicit, nationally or regionally, the same contradictory reaction in the realm of population as has been observed in other areas. The second change involves the complete transformation of demographic perspective offered at the dawn of the 21st century. After decades spent mobilizing national and international forces in the attempt to curb a population growth feared out of control and thus prepare for a population explosion, the international community now finds itself facing a radically different demographic evolution, dominated by the extraordinary prospect, in the not too distant future, of a stable, or even declining, world population and, in particular, by the uncertainties of what the post-transition period may bring (see Chapter 69).

We will distinguish between two time horizons: the near future of a single generation and the more distant one of two generations (2025–2050).

1. A Single Generation's Time Horizon (2000–2025)

Although fertility levels as a whole continue to converge toward ever lower levels more quickly than expected, the period is nonetheless one of enormous variability, where populations whose fertility has not yet fallen or has just begun to, such as some West African countries, will survive for some time alongside a rising number of populations with a TFR close to 2.1 children per woman, and even populations where it barely reaches 1.2 children per woman (like Italy and Russia). Over the last few years a new element of variability has emerged, one which remains a constant source of astonishment: the evolution of mortality. In a growing percentage of world population, mortality has stopped decreasing, or is actually rising (see Chapter 69). This turnabout in a tendency that we believed achieved is all the more paradoxical in that it is occurring during a period of peace, and in countries as diverse as Russia and Botswana.

It is also the ultimate period of enormous growth in absolute numbers, however: according to the United Nations projections' medium scenario, the population will still grow by 2.8 billion between 2000 and 2050 (United Nations 2003b). Increasingly concentrated in part of the Third World, nonetheless, this growth diminishes rapidly, with an increase of 1.7 billion predicted for the first quarter of the century compared

[2] "... enserre désormais les faits sociaux [et donc démographiques] dans une chaîne de causalité dont le point de départ est global et non plus local" (Laidi, 1998).

with 1.1 billion in the second quarter. Future population growth in the former Third World is characterized not only by its changing rhythm, but especially by changes in the growth components which have policy implications.

First, given the future distribution of populations in the process of transition, the relative weight of different growth components (fertility and mortality evolution, inertial effects from age structure) will not be identical to that of the previous century. With the exception of sub-Saharan Africa, where fertility is still very high, the age structure of the population of reproductive age will be the principal component of future Third World growth. Mortality and fertility decline will be in second and third place, respectively.

Among countries where the transition has barely begun, where the TFR still hovers around 5 or 6 children per woman (24% of the global population), the task will inevitably be to jumpstart it, or speed it up. In the 49 least developed countries, most of which are found in sub-Saharan Africa, the medium scenario of the most recent United Nations projections indicates that the TFR of 5.5 children per women in 1995–2000 will still be at 3.7 in 2020–2025, and 2.5 in 2045–2050 (United Nations, 2003b). Together with a few South Asian countries, sub-Saharan Africa is thus family planning's *final frontier* (Caldwell, 2002). For this subcontinent, the question is to know whether the Cairo strategy, with its call for a coherent and dynamic civil society and a State capable of providing a panoply of social services, is well adapted to the region's circumstances, where family planning programs have still not been systematically implemented or, where they have been, have not met with great success. Geoffrey McNicoll (1995) stresses that part of Asia's success with demographic policy can be put down to the presence of strong governments in the region; this is not the case in Africa. Furthermore, even supposing that international aid is available, it cannot totally replace national involvement, as the experience of other countries has shown. It should also not be forgotten that the AIDS epidemic is particularly violent in the region and that it is unlikely it will be curbed within the given time horizon.

In countries where the TFR is midway (23% of world population), the task remains to bring fertility decline to completion. International aid and continuing attention given to their problems may play a decisive role in making it possible for them to reach the reproduction replacement level of 2.1 children per woman toward 2030–2035, as the United Nations projections' medium scenario suggests. As we have seen, however, this aid is more uncertain than in the past.

Third World countries in which the TFR is already close to 2.1 children per woman will continue to grow, given the growth potential accumulated within the age structures in the course of the decades following the end of World War II. In countries viewing this growth as excessive, it will no doubt be interesting to experiment, at the practical level, with measures aiming to restrict growth linked to the structure of their population of reproductive age, such as incentives to increase the average age at motherhood (Bongaarts, 1994). Furthermore, as the example of countries like South Korea shows, we will undoubtedly witness the gradual withdrawal of the State from support for family planning programs, with contraception increasingly entering into the private sphere.

The final growth wave of almost 3 billion will be a population qualitatively very different from that of the previous century. Restructuring certain population characteristics, such as education levels, labor force involvement, type of employment, and place of residence will trigger radical changes, in addition to the evolution of population numbers between 2000 and 2050. The new generations will be composed of individuals who are mostly literate or even more (especially women), living primarily in the city, some in gigantic urban sprawls. Most are likely to have swapped their ancestral rural values for new urban ones. This new population wave will therefore display very different characteristics from those of the earlier waves that swelled world population during the second half of the 20th century, and which included a high percentage of uneducated rural dwellers. In all probability, the demographic behavior of this new wave will reflect that of a population exposed to modern information systems, more open to the world and more likely to accept new norms concerning family and women's status.

A tendency to worry about population decline is gradually resurfacing among wealthy nations today, as we have seen. The medium scenario of the United Nations projections suggests that the European population (including the Eastern European countries and Russia) should decline by only 4% between 2000 and 2025 and by 9% between 2000 and 2050. These estimations may result from the excessively optimistic fertility hypothesis selected, one assuming an increase to 1.85 children per woman by 2050. Skilled observers of European demography like Thomas Frejka and Ron Lestaeghe believe, on the contrary, that fertility decline is neither temporary nor merely the result of a timing effect (Lesthaeghe and Willems, 1999; Ross and Frejka, 2001). Unworried by their own population, the perception of donor countries had been dominated by anxiety over too rapid population growth in the Third

World. Little by little this perception has changed radically; the rich countries are now reassured about Third World demographic evolution but starting to be concerned about their own. The prospect of seeing populations like that of Italy declining by 20% by 2050, or that of Russia by 50%, could not but spur their governments into action. (Incidentally, corresponding to the United Nations medium variant, these estimates are on the optimistic side for these countries.)

To tackle their declining population, three types of intervention can be envisaged. First, actions designed to act on fertility levels are likely to be inspired by the Swedish model's cocktail of measures, with the emphasis placed on taking collective responsibility for a more or less substantial part of women's tasks; also included will be institutional reforms affecting intergenerational relationships, such as taking more account of the number of children in pension calculations than is done at present. It is interesting to note that, both in developed countries where fertility has plummeted and in Third World countries where it is still high, women are at the centre of political strategies for fertility regulation. Liberated from certain economic and social constraints by policies designed to meet a variety of ends, women seem to be expected to bring their reproductive behavior in line with the collective good. In a certain way, they have become the "invisible hand" of the 21st century.

The second type of intervention concerns international migration. While migrations that help regulate the labour force are tolerated relatively well, this is not the case for what is known as replacement migration, even if its only aim is to slow down population decline. The type of migration that could be envisaged as a remedy for population aging is unlikely to be implemented given the size of population flows needed. Effective or not at the demographic level, international migration holds a privileged place on the international political agenda because of the tensions it arouses within and between countries. The Global Commission on International Migration, mentioned above, can play an influential role in waking the international community up to the size of the problems international migrations are already creating and, especially, those they will create with a globalization movement that cannot be limited to the circulation of material or immaterial goods alone.

Finally, in third place, there exists a type of political route to reach desired population size, a kind of "collective demographic insurance system." By this is meant the constitution of large population clusters through the union of independent countries. Forming these supranational entities makes it possible to transcend the demographic shrinkage of national boundaries. With demographic giants like Brazil and China catching up economically and technically, population size confers an advantage equivalent to that offered by strategic depth. This "demographic depth" thus becomes today an indispensable factor in the race for economic and geopolitical power. The current period, with globalisation as its dominant theme, has witnessed the birth of the European Union, whose population, in its present form of 25 countries, ranked below that of China and India but above that of the United States and even above that of NAFTA (the North American Free Trade Association, which includes the United States, Mexico, and Canada). In the long run, the community institution specialists are already working with the vision of a European Union expanded to include 30 members. It would be naïve nevertheless to assume that this changing geography could remain unilateral: NAFTA has the possibility to expand into South America.

For the first time, however, anxiety surrounding population decline is no longer confined to European countries. In former Third World countries like Singapore, we have witnessed an extraordinarily rapid turnabout in the government's position compared with the delayed reaction of European nations who found themselves in the same situation: within the space of a few years, from being antinatalist, the country became pronatalist (see Chapter 103, this volume). A reaction typical of countries in which fertility is considered too low, the growing role of international migration has started to seriously preoccupy countries like South Korea or Malaysia. Japan, the country with the famous Eugenic Protection Law of 1948, is reflecting on the consequences of its aging and future demographic decline. We can thus expect to see these countries take political action similar to that taken by European governments who had to deal with the problems associated with plunging fertility levels some time before them.

Population aging is among the problems linked to the rapid evolution of fertility and, to a lesser degree, mortality (see Chapters 68 and 69). The abrupt drop in fertility (instances of a transition from 6 to 2 children in the space of a single generation are not uncommon among developing countries, whereas the same evolution required two to five generations in industrialized nations) will bring about a structurally different population aging process (primarily through contraction at the base of the pyramid and, second, by extending life expectancy at advanced ages); more particularly, the process will be much faster than in European countries, leaving the economic and social structures of these countries without the time to adapt to it, and their economies unable to release the necessary resources. This evolution will especially affect East Asian countries, the first to enter the transition, especially giants like China, and then the Latin America countries that followed, and contrary to what is indicated in the Cairo Programme of Action (par. 6–16), the population aged 60 and over in these countries is not composed in its great majority of women since, in 2000, men represented 46% of this age group (Knodel and Oldestal, 2003). Aging is a truly universal phenomenon, and we should not forget that it will continue in industrialized countries. The North and South will experience its consequences very unequally, however, because of differences in economic wealth and variations in the speed of demographic change. According to the United Nations, the proportion of the population over the age of 65 years in developing countries will exceed that of children under 5 years in 2020 (United Nation, 2003b; see also Chapter 77). This inversion of the population pyramid's profile in an exceptionally short space of time will have many consequences for Third World societies, for example, with respect to health policy. Among causes of death in Third World countries, the tendency today is still to assign the greatest weight to infectious diseases although, except in Africa, degenerative diseases have already become the most important cause of death. Their weight will increase further as populations age. In countries with few financial and human resources, the battle against these diseases will encounter considerable difficulties. This means that all international aid policies in the spheres of health and population, up to now centred on early child mortality and high maternal fertility, need to be rethought (Raymond, 2003).

2. A Two-Generation Time Horizon

At the 2025 time horizon, predicted demographic events involve a global population, the vast majority of whom are already born today; this is evidently not the case for 2050, however, and wider margins of error need to be taken account of with the extension of projection period (see Chapter 77).

According to the United Nations projections' medium scenario, the global demographic transition should be complete toward the middle of the century, when the average level of fertility equal to that strictly necessary for generation replacement–2.1 children per woman—will be reached (United Nations, 2003b). This average figure, however, is obtained by combining the fertility of 16% of the population with a TFR still superior to 2.1 (mostly that of least developed countries) with that of the 84% of countries with below replacement level fertility. It is unlikely that governments of these countries, or at least those where fertility had dropped the most or for the longest, will resign themselves to see their population disappear without making any attempt to prevent it. In such circumstances, the population question may once again become a cause of international tension. Pronatal policies implemented unilaterally, and not only in today's industrialized nations, could jeopardize the perspective of world population stabilization, which most governments undoubtedly find acceptable.

It is easy to foresee enormous difficulties with the coordination, at the international level, of new probirth policies that are national, or at best regional, in nature (Harbinson and Robinson, 2002). No existing international power has the legal and fiscal power necessary to implement probirth policies, and it is obviously impossible to say whether such a power will exist in the future. Consequently, an essential task for the international community will be to reach a new consensus about a demographic objective acceptable to all, such as that of a world population stabilized at the level reached at the end of this century. A programmed decline would not attract much support because of the contradictory interests it would put at stake. An international laissez-faire, on the other hand, could lead to demographic fluctuations resulting from an anarchic, or inopportune implementation of population policies, or quite simply from fashionable demographic behaviors.

CONCLUSION

Over the last 50 years, demographic evolution has been a part of the broad movement toward global modernization. What differentiates it from other areas also undergoing this modernization process, such as education or health, where from the start there was consensus about their aim of improving services, is that during the two decades following World War II much effort had to be put into creating consensus around a major objective of population policy: fertility reduction. This search for consensus was not obvious, given the mixed memories left by pronatal population policies that were implemented in several countries during the period between the two wars and sometimes smacked of eugenics (see Chapter 100, this volume). Furthermore, during the next three decades, it required great persistence and ingenuity to elaborate family planning programmes, a field in which little or no previous experience existed.

There was, and still is, no shortage of criticisms and doubts about the effectiveness of these programs, whatever form they have taken through time. A substantial literature on the topic exists. Suffice it to say, at this point, that intervention from the United

Nations, then from governments, NGOs, and the media, have aroused an influential current of international opinion that has greatly contributed not only to legitimizing new norms of reproductive behavior, but also to making them desirable.

As already noted, two main actors took the initiative in mobilising the international community in a unique way; first, the Population Movement (born and nourished in the United States and, to a lesser extent, in Scandinavia), and second, the United Nations. To the former goes the credit for conceptualising the need for action, for raising awareness, and for initiating the first programs; to the second, for legitimising it in the eyes of the international community, as both a human right and a condition for development, and for contributing substantially to its implementation, both financially and technically. The United Nations could be criticised, however, for not taking a harder line with countries implementing policies whose coercive nature violated the very principles and rights the United Nations were supposed to be promoting. The organization could also be pulled up for ceding too often to pressure from the Population Movement, who viewed the global population problem as an exclusively Third World issue, ignoring that posed by populations in industrialized countries. Thus was destroyed a unique opportunity to prepare the international community for the truly universal and prospective vision on population questions that would be needed in subsequent decades.

Between Bucharest and Cairo—the greatest period of mobilization within the international community—the ideological, economic, and demographic context changed radically. It was only natural, therefore, that the Cairo document reflect these changes and suggest new strategies. In 20 years, one of the most important changes on the international stage, and one which may be the most influential in the area of population given its secular association with the State, was the challenge to the State's role as sole player in development, particularly in the Third World. The State of today has ceased to enjoy its privileged position and, as in Cairo, has to compete with new actors, individuals, civil society, and the NGOs that represent it, as well as with representatives of religious forces, something it finds more difficult to allow. Part of the social and economic responsibilities of a strong state, as understood during the last decades—to guide society toward economic development and the demographic transition with a firm hand—had been transferred to representatives of civil society (NGOs, local authorities, etc.) or entrusted to market forces. Paradoxically, however, this reduced state continues to be seen as an inescapable provider of social services in the most diverse spheres, with reproductive health, education, promotion of women, and the like among the most important.

The second change, evidently linked to this contraction of the state's role, is the emergence or reinforcement of transnational players, such as international NGOs. Their influence was highlighted especially in the section dealing with the Cairo conference and its preparation. This is only the most visible aspect of their role at the United Nations, however. In fact, by gaining the status of observers within UN deliberative bodies, they are in a position to circulate their own proposals, and intervene in the debates under certain conditions: in a word, to lobby in order to advance their programs, their ideas, and the norms and values they stand for. Their lobbying is all the more effective given that, recently, some government delegations have decided to include representatives of these NGOs. Finally, NGOs possess their own transnational networks, making their presence felt almost everywhere.

The third change bears on the way in which population policy has been justified. Neo-liberal ideology triumphed, as mentioned, just as new players were appearing on the international stage, with different objectives from those at the forefront in the 1970s. As a result of these transformations, earlier macro-economic justifications were more speedily abandoned in favour of humanitarian ones, among which reproductive health is a good example. As it is impossible to bypass macro-economic rationalisations entirely in the political sphere, an attempt was made to replace the old macro-economic justifications with new environmental ones; arising out of thinking on the conditions for durable development or sustained growth, they were not always obvious to operationalise, as we have seen. Despite these efforts, the new programs adopted in Cairo are, above all, humanitarian ones that clearly differentiate them from those adopted at previous conferences. As programs with an essentially humanitarian calling, however, they risk having to compete with numerous other humanitarian programs which, in some cases, could have demographic effects. Unfortunately, this reorientation is occurring at a time when the slowdown in global population growth, and, in particular, the prospect of population decline in a growing number of countries—often donor countries—may modify their priorities with respect to family planning program funding.

As we have seen, the fourth change concerns the increasing speed at which universal social and demographic norms are diffused and adopted. In recent years, countries as different politically, economically, and culturally (especially with regard to women's status) as Brazil, Turkey, and Vietnam, for example, embraced the reduced family size norm (two children) in a remarkably short space of time compared with the experience in Europe; this duration continues to shrink in the most recent countries to enter the demographic transition. The demographic transition seems to have freed itself from certain of the initial model's constraints. It is drawn along by the vast modernization movement within societies, itself accelerated by the recent globalisation phenomenon that fuels the diffusion of types of behavior until then limited to the most developed countries.

This is the new context within which future demographic evolution, and the inevitable political reactions it arouses, will take place. With the end of the transition toward the middle of the century, and world population potentially stabilized by 2075, we will witness a vast demographic transformation that will radically alter the international demographic scene and, more generally, our perception of population issues. Only 10 years ago, such a rapid evolution would have appeared utopian. In 1992, as mentioned, the United Nations did not predict stabilization before 2100 (United Nations, 1992). The transition may end sooner than expected, however, and may also spell the entry into unknown territories that will have to be explored without the precious guide provided by the transition's theoretical schema (see Chapter 69). The post-transition era opens up all kinds of possibilities, such as global population stability at the level achieved before the end of this century, gradual decline, or fluctuations around a point of equilibrium constantly pushed forward (see Chapter 78).

Before reaching that point, the international community is likely to have to manage extremely diverse demographic situations, characterised by fast growth in a minority of countries and by population decline in an increasing number of them. Like any period of rapid change, this may also prove to be a time of great tension, and a return to state intervention in countries concerned by an uncertain demographic future is probable. It will also be a crucial period, allowing the international community to prepare for the post-transition period. As happened in the 1960s and 1970s, a new consensus needs to be created around a universally acceptable demographic goal, and international institutional mechanisms must be designed to reach it. Whether or not this new consensus will be instigated through the efforts of a new *Population Movement*, only the future will tell.

Acknowledgments

This chapter is a revised and extended version of a chapter on the same theme in a collection published by INED (Chasteland, 2003).

This chapter was translated from the French by Heather Juby.

The author would like to thank Jean-Claude Chesnais and Jacques Vallin for their careful rereading of this chapter and for their stimulating comments.

References

Berelson Bernard, 1969. Beyond family planning, *Science*, vol. 163(7), p. 533–543.

Boli John and Thomas George M., 1997. World culture in the world polity: a century of international non-governmental organizations, *American Journal of Sociology*, vol. 62(2), p. 171–190.

Bongaarts John, 1994. *Population policy options in the Third World.* New York, Population Council. (Working Paper, no. 59).

Boserup Ester, 1965. *The conditions of agricultural growth: the economics of agrarian change under population pressure.* London and Chicago, Allen and Unwin and Adline, 124 p.

Caldwell John C., 1998. Malthus and the less developed world: the pivotal role of India, *Population and Development Review*, vol. 24(4), p. 675–696.

Caldwell John C., 2002. The future of family planning programmes, *Studies in Family Planning*, vol. 33(1), p. 1–10.

Chasteland Jean-Claude, 1994. La croissance de la population mondiale devant la communauté et l'opinion internationale, *Revue française des affaires sociales*, vol. 48(4), p. 11–29.

Chasteland Jean-Claude, 2003. De 1950 à 2000: la communauté internationale face au problème de la croissance de la population mondiale, *in*: Jean-Claude Chasteland and Jean-Claude Chesnais (eds.), *La population du monde: géants démographiques et défis internationaux*, p. 717–754. Paris, INED, 768 p. (Les Cahiers de l'INED, no. 149, seconde édition revue et augmentée).

Clark Colin, 1967. *Population Growth and Land Use.* London, Macmillan, 506 p.

Coale Ansley J. and Hoover Edgar M., 1958. *Population growth and economic development in low income countries. A case study of India's Prospects.* Princeton, Princeton University Press, 389 p.

Crane Barbara and Finkle Jason L., 1981. Organizational impediments to development assistance: the World Bank's population program, *World Politics*, vol. 33(4), p. 516–553.

Davis Kingsley and Blake Judith, 1956. Social structure and fertility: an analytic framework, *Economic Development and Cultural Change*, vol. 4(4), p. 211–235.

Demeny Paul, 1994. *Population and Development.* Liège, IUSSP/UIESP, 26 p. (Distinguished Lectures Series on Population and Development).

Dixon-Mueller Ruth, 1993. *Population policy and women's rights: transforming reproductive choice.* Westport (Connecticut), Praeger, 287 p.

Donaldson Peter, 1990. *Nature against us: The United States and the world population crisis 1965–1980.* Chapel Hill and London, University of North Carolina, 207 p.

Ehrlich Paul R., 1968. *The population bomb. Population control or race to oblivion.* New York, Ballantine Press, 201 p.

Finnemore Martha, 1996. Norms, culture and world politics: insights from sociology's institutionalism, *International Organization*, vol. 50(2), p. 325–347.

Finkle Jason L. and Crane Barbara B., 1975. The politics of Bucharest: population, development and the new international economic order, *Population and Development Review*, vol. 1(1), p. 87–114.

Finkle Jason and Crane Barbara, 1985. Ideology and politics at Mexico City: the United States at the international conference on population, *Population and Development Review*, vol. 11(1), p. 1–28.

Foucault Michel, 1976. *Histoire de la sexualité. Volume I: La volonté de savoir.* Paris, Gallimard, 211 p.

Germain Adrienne, 1987. *Reproductive health and dignity: choices by Third World women.* Background paper prepared for the International Conference on Better Health for Women and Children through Family Planning. New York, International Women's Health Coalition, 31 p.

Green Marshall, 1993. The evolution of US international population policy 1965–92: a chronological account, *Population and Development Review*, vol. 19(2), p. 303–321.

Haas Peter, 1992. Introduction: epistemic communities and international policy coordination, *International Organization*, vol. 46(1), p. 1–35.

Harr John Ensor and Johnson Peter J., 1991. *The Rockefeller conscience: an American family in public and private.* New York and Toronto, Maxwell Macmillan and Scribner's Sons, 624 p.

Harbinson Sarah and Robinson Warren C., 2002. Policy implication of the next world demographic transition, *Studies in Family Planning*, vol. 33(1), p. 37–48.

Hardin Garrett, 1968. The tragedy of the commons, *Science*, no. 162, p. 1243–1248.

Hirschman Albert O., 1958. *The strategy of economic development.* New Haven, Yale University Press, 217 p.

Inglehart Ronald and Norris Pippa, 2003. *Rising tide: gender equality and cultural change around the world.* Cambridge (United Kingdom) and New York, Cambridge University Press, 226 p.

IPPF. Financial statement. 2003. (www.ippf.com).

IUSSP (International Union for the Scientific Study of Population), 1988. The Population Commission and IUSSP, *Population Bulletin*, *United Nations*, no. 19–20, p. 115–124. (E.87.XIII.2).

Kasun Jacqueline, 1981. The international politics of contraception, *Policy Review*, vol. 15(Winter), p. 135–152.

Keck Margaret E. and Sikkink Kathryn, 1998. *Activists beyond borders: transactional advocacy networks in international politics.* Ithaca (New York), Cornell University Press, 228 p.

Kelley Allen C., 1991. *Revisionism revisited: an essay on the population debate in historical perspective.* Duke (North Carolina), Duke University Department of Economic Workings, 37 p. (DUDEW Working Paper no. 95–45. Paper presented at the Nobel Symposium of Economics, 5–7th December 1991, Lund, Sweden).

Kirk Dudley, 1944. Population change and the postwar world, *American Sociological Review*, no. 91, p. 28–35.

Knodel John and Olstedal Mary Beth, 2003. Gender and aging in the developing world: where are the men? *Population and Development Review*, vol 29(4), p. 677–698.

Kuznets Simon, 1967. Population and economic growth, *Proceedings of the American Philosophical Society*, vol. 111(3), p. 170–193.

Laidi Zaki, 1998. Les imaginaires de la mondialisation, *Esprit*, no. 10, p. 85–98.

Lassonde Louise, 1996. *Les défis de la démographie*, *Quelle qualité de vie pour le XXIe siècle?* Paris, La Découverte, 225 p.

Lesthaeghe Ron J. and Willems Paul, 1999. Is low fertility a temporary phenomenon in the European Union?, *Population and Development Review*, vol. 25(2), p. 211–228.

Luke Nancy and Watkins Susan Cotts, 2002. Reactions of developing-country elites to international population policy, *Population and Development Review*, vol. 28(4), p. 707–733.

Macura Milos, 1986. L'importance des conférences mondiale des Nations Unies sur la population, *Bulletin démographique des Nations Unies*, no. 19/20, p. 14–26. (Published in 1988: E.87.XIII.2).

Macura Milos, 1986. The significance of the United Nations international population conferences, *Population Bulletin*, no. 19–20, p. 14–26. (Published in 1987).

McIntosh Alison C. and Finkle Jason L., 1995. The Cairo Conference on Population and Development: a new paradigm? *Population and Development Review*, vol. 21(2), p. 223–260.

McLaren Angus, 1990. *History of Contraception.* Oxford, Basil Blackwell, 276 p.

McNicoll Geoffrey, 1995. *Demography in the unmaking of civil society.* New York, The Population Council, 38 p. (Population Council Research Division Working Paper, no. 79).

Meadows Donella H., Meadows Dennis L., Randers Jørgens J. and Behrens III William W., 1972. *Halte à la croissance: rapport sur les limites de la croissance.* Paris, Fayard, 314 p.

Meyer John, Boli John, Thomas George M. and Ramirez Francesco O., 1997. World society and nation state, *The American Journal of Sociology*, vol. 103(1), p. 144–181.

Nadelmann Ethan A., 1990. Global prohibition regimes: the evolution of norms in international society, *International Organizations*, vol. 44(4), p. 479–526.

NAS (National Academy of Sciences) Studies Committee, 1971. *Rapid population growth*, *consequences and policy implications.* Baltimore, The Johns Hopkins University Press, 696 p.

Nations Unies/United Nations, 1951. Étude d'ensemble de l'accroissement de la population mondiale dans le passé et dans l'avenir, *Bulletin démographique des Nations unies*, no. 1, p. 1–12.

Nations Unies/United Nations, 1953. *Causes et conséquences de l'évolution démographique.* New York, Nations Unies/United Nations, 432 p. (Études Démographiques no. 17. E 53.XIII.3).

Nations Unies/United Nations, 1975. *Rapport de la Conférence mondiale des Nations unies sur la population. Bucharest 19–30th August 1974.* New York, Nations unies/United Nations, 169 p. (E.75.XII.3).

Nations Unies/United Nations, 1984. *Rapport de la Conférence internationale de la population.* New York, Nations unies/United Nations, 109 p. (F.84.XIII.8).

Nations Unies, 1997. *Examen et évaluation des progrès réalisés dans l'exécution du Plan d'action mondial sur la population.* New York, Nations unies/United Nations, 160 p. (A/Conf.171/PC.3).

Nations Unies/United Nations, 1998. *Population et développement. Programme d'Action adopté à la Conférence internationale sur la population et le développement. Le Caire*, *5–13 septembre 1994.* New York, Nations unies, Division de la population, 204 p. (E.95.XIII.7).

Nations Unies/United Nations, 2000. *Examen et évaluation des progrès accomplis dans la réalisation et les objectifs de la Conférence internationale sur la population et le développement. Rapport 1999.* New York, Nations unies, Division de la population, 104 p. (Population Studies no. 182, ST/ESA/SER.A/182).

Nations Unies/United Nations, 2005. *Flux de ressources financières devant concourir a l'application du Programme d'action de la Conférence internationale sur la population et le développement.* New York, Nations unies, Division de la population, 21 p. (E/CN.9/2005/4).

Notestein Frank, 1944. Problems of policy in relation to areas of heavy population pressure, *in*: Frank Notestein (ed.), *Demographic studies of selected areas of rapid growth*, p. 138–158. New York, Milbank Memorial Fund, 158 p.

NRC-WGPGED (National Research Council, Working Group on Population Growth and Economic Development), 1986. *Population and economic development: policy questions.*—Washington, National Academy Press, 78 p.

Raymond Susan, 2003. Foreign assistance in an aging world, *Foreign Affairs*, vol. 82(2), p. 91–105.

Rockefeller John D. III, 1974. *Population growth: The role of the developed world.* Liège, IUSSP/UIESP, 412 p. (Lecture series on Population).

Ross John and Frejka Tomas, 2001. Paths to sub-replacement fertility: the empirical evidence, *Population and Development Review*, vol. 27, special supplement: "Global fertility transition," p. 213–254.

Schindlmayr Thomas, 2001. The media, public opinion and population assistance: establishing the link, *International Family Perspectives*, vol. 27(1), p. 42–46.

Sen Gita, Germain Adrienne and Chen Lincoln C. (eds.), 1994. *Population policies reconsidered: health, empowerment and rights.* Boston and Cambridge, Harvard Center for Population and Development, Harvard School of Public Health, International Women's Health Coalition and Harvard University Press, 280 p.

Sen Gita and Grown Caren, 1987. *Development crisis and alternative visions: Third World women's perspectives.* New York, Monthly Review Press, 116 p.

Simon Julian L., 1981. *The ultimate resource.* Princeton, Princeton University Press, 415 p.

Smouts Marie-Claude, 1983. La crise des organisations internationales, *Etudes*, vol. 358(2), p. 165–173.

Symonds Richard and Carder Michael, 1973. *The United Nations and the population question.* New York, McGraw-Hill Book Company, 236 p.

Szreter Simon, 1993. The idea of demographic transition and the study of fertility change: a critical intellectual history, *Population and Development Review*, vol. 19(4), p. 659–701.

UNFPA. About UNFPA funding 2003. (www.UNFPA.org).

United Nations, 1965. *World population: challenge to development. Summary of the highlights of the World Population Conference, Belgrade, 30th August–10th September, 1965*, vol. 1. New York, United Nations, 48 p.

United Nations, 1973. *World population prospects as assessed in 1968.* New York, United Nations, Department of Economic and Social Affairs, 167 p. (Population Studies: ST/SOA/Ser.A/53, 72.XII.4).

United Nations, 1975a. Population Policies and Programmes, *in: The population debate: dimensions and perspectives. Papers of the World Population Conference, Bucharest—1974. Vol. 2*, p. 581–674. New York, United Nations, 726 p. (E/CONF.60/CBP/21).

United Nations, 1975b. Report of the United Nations World Population Conference, 1974. Bucharest 19–30 August 1974. New York, United Nations, 147 p. (E.75.XIII.3).

United Nations, 1984. Report of the International Conference on Population, 1984. Mexico City 6–14 August 1984. New York, United Nations, 101 p. (E.84.XIII.8).

United Nations, 1987. *Report of the World Commission on Environment. Our common future.* New York, United Nations, 374 p. (A/42/427).

United Nations, 1991. *Long range world population projections: two centuries of population growth, 1950–2150.* New York, United Nations, 36 p. (E.92.XIII.3).

United Nations, 1992. *Long-range World population projections: two centuries of population growth 1950–2150.* New York, United Nations, Department of Economic and Social Affairs, 32 p. (Document ST/ESA/SER.A/125).

United Nations, 1994. *Work of intergovernmental and non governmental organizations in the implementation of the World Population Plan of Action.* New York, United Nations, 27 p. (E/CN.9/1994/7).

United Nations, 1995a. *Population consensus at Cairo, Mexico City and Bucharest—An analytical comparison.* New York, United Nations, Department for Economic and Social Information and Policy Analysis, Population Division, 165 p. (E.96.XIII.2).

United Nations, 1995b. *Review and appraisal of the World Population Plan of Action. 1994 report.*—New York, United Nations, 149 p. (ST/ESA/SER.A/152).

United Nations, 2000a. *The flow of financial resources for assisting in the implementation of the Programme of Action of the International Conference on Population and Development.* New York, United Nations, 20 p. (E/CN.9/2000/5).

United Nations, 2000b. *The Long Range World Population Projections, Based on the 1998 Revision.* New York, United Nations, Department of Economic and Social Affairs, Population Division, 92 p. (E.00.XIII.8).

United Nations, 2003a. *International migrations and development report of the Secretary.* New York, United Nations, 18 p. (G.A/58/98).

United Nations, 2003b. *World population prospect. The 2002 revision.* New York, United Nations, Department for Economic and Social Affairs, Population Division, vol. I: Comprehensive tables, 781 p., vol. II: Sex and age distribution of population, +939 p. (ST/ESA/Ser.A/225).

Vallin Jacques, 1986. *La population mondiale.* Paris, La Découverte, 128 p. (Repères, no. 45, première édition).

Vallin Jacques, 1994. Le Caire, une double déception, *La Chronique du CEPED*, no. 15, p. 1–3.

Vogt William, 1948. *The road to survival.* New York, Wm. Sloane Associates, 335 p.

Wahren Carl, 1986. The crucial families, *in*: Pierre Frühling (ed.), *Swedish development aid in perspective: policies, problems and results since 1952*, p. 185–194. Stockholm, Almquist and Wicksell International, 327 p.

World Bank, 1984. *World Development Report.* New York, Oxford University Press, 300 p.

SECTION

II

OBSERVATION, AUXILIARY METHODS, TEACHING AND RESEARCH

Introduction

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

This is the final section of Demography: Analysis and Synthesis. Volume I, Section I looked at the central study of the discipline—the in-built mechanisms of population dynamics. Volumes I, Section II, and Volume II went on to address the determinants of the three big forces of population change—fertility, mortality and migration. Volume III, Section I drew on all this groundwork to paint a broad panorama of the historical and geographical forces behind population processes and sketch out prospective future medium and long term trends. Volume III, Section II then examined the biological, economic, social and political consequences of demographic changes and, more generally, the big issues facing society on the cusp of these changes. Volume IV, Section I, on population thought and policies, considered whether society ought to and can act to shape the course of demographic processes, and with what likelihood of success.

While all the big issues of population dynamics had arguably been touched on over those seven volumes, there was nevertheless thought to be a useful purpose in adding an eighth volume given over to a number of matters which either did not fit into or had been broached only to a very small extent or in passing within the closely-ordered framework of the previous seven.

This section in fact does three quite different things. It first sets out to give a more systematic overview of an overarching issue in demographic research: data sources and quality, and how to maximize the use of defective sources. It was thought that this key topic was best placed at the end of the work where readers might find a greater relevance in it after having explored the issues involved through those of the analyses done in the first 3 volumes. Then, some explanation was also needed of a number of analytical tools borrowed from other disciplines that are either extensively used in demography, or have become sufficiently specific to it to warrant inclusion in a treatise on demography. Nor, obviously, could the final word be said without some specific consideration of the discipline itself: its history, how it is taught, and trends in current research. Although each of these merited a separate small volume to itself, the choice was made to consolidate them in a single volume since, to some extent, they work across boundaries in the discipline, from three different angles: sources, the methodological input of other disciplines, and the historical dynamic of the discipline from the very beginnings to the issues for the 21st century.

Part One, Observation Systems and Indirect Estimation Methods, gives the necessary overviews of two aspects that recurred in each preceding volume, but only ever tangentially and never fully, in relation to discussion of whether a particular analysis could be done using available data, or whether specific data needed to be collected, or indirect estimation techniques used, without ever touching on demographic data collection as such, or the processing of imperfect data. Six chapters are devoted to data collection: the first gives an ordered overview of the different data sources, while the other five each address a more specific aspect that seemed to require a more particular focus here. A further three chapters look at the analysis of imperfect data, with particular reference to the specific requirements of historical demography in one case, and those of developing countries in another.

Part Two, Statistical Analyses, Specific Tools and Qualitative Approaches, brings together a number of methods that are auxiliary or supplementary to demography and are either simply borrowed from other disciplines, or the result of recent developments in particular methods which, while not entirely unconnected with demography, evolved out of (or are being developed from) exposure to approaches prompted by other disciplines. They cover a wide range of different standpoints from the philosophy of science to the most practical statistical methods. There are three main focal points to this section: qualitative approaches, statistical tools and the outputs of individual data analysis.

Part Three, History of Demographic Science, Teaching and Research, offers a panorama of the history of demographic science, from the pioneers who invented it up to those who much later named it and gave it structure. It will also give thought to developments in how the discipline is taught, and in the production of teaching aids, from treatises on demography through learning materials up to computer programs. It concludes with a review of recent and current research, and the small community of demographers as it prepares to confront the scientific challenges of the 21st century.

P A R T

I

OBSERVATION SYSTEMS AND INDIRECT ESTIMATION METHODS

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

As stated in the introduction to this section, this first part aims to give a more systematic overview of an issue that is of overarching importance to demographic research, but tends to deter readers not motivated by having a basic knowledge of how significant the outcomes of an analysis of adequate data can be. Investigating existing sources or having to contrive and develop a complete set of new data presupposes having a research purpose in mind, whence the decision to leave the topic until the very end. The issue of data that can be used or needs to be collected has certainly been touched on in the previous volumes, in particular when it dictated the kind of analysis proposed or could bias the results, but never more than to a minor extent and in passing, dealing only with what was strictly necessary for the purpose of the specific chapter topic, without ever discussing demographic data collection proper. It is now dealt with in six general and specific chapters in this volume. Furthermore, a substantial proportion of demographic data may be incomplete or defective, and this raises particular and often awkward issues for analysts. This is considered in the remaining three chapters.

The section on demographic data collection opens with chapter 121, in which Dominique Tabutin gives an overview of demographic data collection tools, whether from mainly administrative sources, where demographers tend to have little direct involvement in their production, the two most traditional sets being vital registration and population census statistics, or data that more or less directly address demographic research issues, usually obtained through specific surveys, not forgetting all the other forms of collection ranging from simple administrative files to the most complex continuous recording systems, and to the possible linkage of data sources.

The following five chapters then focus on particular aspects of demographic data collection, which may be specific either to the research topic, or to the methods of collection. When evaluating Volume II in Section II, on the determinants of migration, the INED publications editorial committee pressed for a specific, in-depth discussion on data collection for the analysis of migration to be included in this treatise. This valid idea is therefore addressed here in two chapters on observation and concepts in internal migration (chapter 122, by Brigitte Baccaïni) and in international migration (chapter 123, by Corrado Bonifazi and Salvatore Strozza).

Chapter 124 deals with a very specific method of collection devised purely in response to prompting from researchers (demographers, epidemiologists and public health specialists) in a bid to remedy the lack or serious failings of more traditional sources, through the long-term systematic and continuous collection of all the data essential to gaining a knowledge of a comprehensively-observed small population. Gilles Pison examines the specific benefits and limitations of demographic surveillance sites, which, on the restricted and, of course, non-representative, territory that they cover, in fact go well beyond simply making good the gaps in vital registration records and censuses.

Finally, in chapter 125, Viviana Egidi and Patrick Festy address the many facets of comparability of international demographic statistics, the desirable harmonization of which faces problems that may be insurmountable without a compromise between comparability and data quality.

Part One concludes with three other chapters dealing with methods that demographers often need to use as a "get-around" to extract meaningful, not overly-biased, conclusions from data that are incomplete, insufficient or of doubtful quality. Two areas in which specific demographic formulae have been applied—historical demography and the indirect estimation of basic demographic indicators for populations (initially in developing countries) for which only imperfect data are available—have expanded remarkably in recent decades, and their value is not set to decline anytime soon.

In Chapter 126, Lorenzo Del Panta, Rosella Rettaroli and Paul-André Rosental look at specific methods in historical demography where direct investigation of now-defunct populations is impossible, and so any means available must be used to try and reconstruct population statistics from a very wide range of often fairly deficient sources. Some, like the old parish registers of Europe, are not far-removed from our present-day systems of vital registration (although sometimes very incomplete), but were only very seldom or partially investigated in their time as a source of demographic data. Their statistical exploitation must therefore be done ex post, usually by sampling to limit costs. But this kind of source is far from historically or geographically universal. The historical demographer has therefore to take into account the paucity of data and the many pitfalls of all kinds from other entirely discrete sources. Consider, then, the plight of the pre-historical demographer!

The indirect estimation of demographic indicators proper ranges, at least in theory, across a field even wider than that of historical demographers, who may at times be able to make as valid use of it as the specialists in developing country demography, by and for whom it was developed, and may also prove useful in the context of countries with highly developed statistical records to address certain particular issues. On the other hand, it is technically far more specific to a particular problem: can indirect observations be used as basis for making any (or more reliable) estimations of demographic indicators that cannot be calculated at all, or with sufficient reliability, due to the lack or inadequacy of direct observations? Kenneth Hill considers this in chapter 127.

Finally, in chapter 128, Josianne Duchêne looks at the history of model life tables, and examines the different means of constructing model life table networks. This specific tool of demography, which is equally essential to the development of indirect estimation methods and the specific reconstructions of historical demography, has its own more general scope, not least of which is framing population projections.

C H A P T E R

121

Information Systems in Demography

DOMINIQUE TABUTIN

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Every science is based on observation, on collecting new data or assembling available data. Demography is no exception to this rule: there is no demography without data, nor are there data without information systems, data sources or data collection methods.

After some brief general remarks on observation in demography (history, position, role), each of the major existing systems will be presented here: continuous registration (civil registration data, population registers and laboratories), censuses (of which there are currently various types), surveys (of varying natures and procedures), without forgetting to include remote sensing, administrative information systems, and qualitative approaches. The advantages and limitations of each system will be discussed and the latest methodological developments will be mentioned, both in developed countries and developing countries.

However, this chapter will only discuss the observation of population phenomena (structure, fertility and nuptiality, mortality and morbidity, and migration), thus excluding, due to lack of space, purely health and socioeconomic information systems. Moreover, the diversity of methods and the variety of problems are so considerable that we will simplify and summarize them as much as possible in order to make the synthesis as readable as possible.[1]

I. FROM OBSERVATION TO ANALYSIS

At the beginning of any research, researchers find themselves faced with a vague reality, especially if the subject or issue under consideration is new. They perceive it subjectively through a prism of varying degrees of deformity, depending on their *a priori*, culture, ideology, and experience, even though, as privileged and well-informed observers, they try to understand and measure in as objective and neutral a way as possible (Fig. 121–1). To define it, they should gather new data or make a selection from the observations already made by others. The data collection strategy should be well thought out before being operationalized according to precise objectives and according to the hypotheses to be verified. The hypotheses may be imagined or result from the existing theoretical corpus.

These three major phases of collecting raw data, producing and monitoring the indicators (rates, etc.), as well as modeling and causal research, are interdependent within the scientific process, even if they follow one another chronologically at the operational level. Data collection precedes analysis, analysis depends on collected material and it is therefore essential that the information system is set up (a survey for example) or that the selection of existing data be organized according to the hypotheses or types of analysis clearly stated at the beginning.[2] In other

[1] For further detail, readers can easily refer to one of the numerous publications devoted to the subjects discussed here or listed in the bibliography.

[2] Without, of course, forgetting common problems encountered by researchers such as lengthy waiting time for results and financial or human resources.

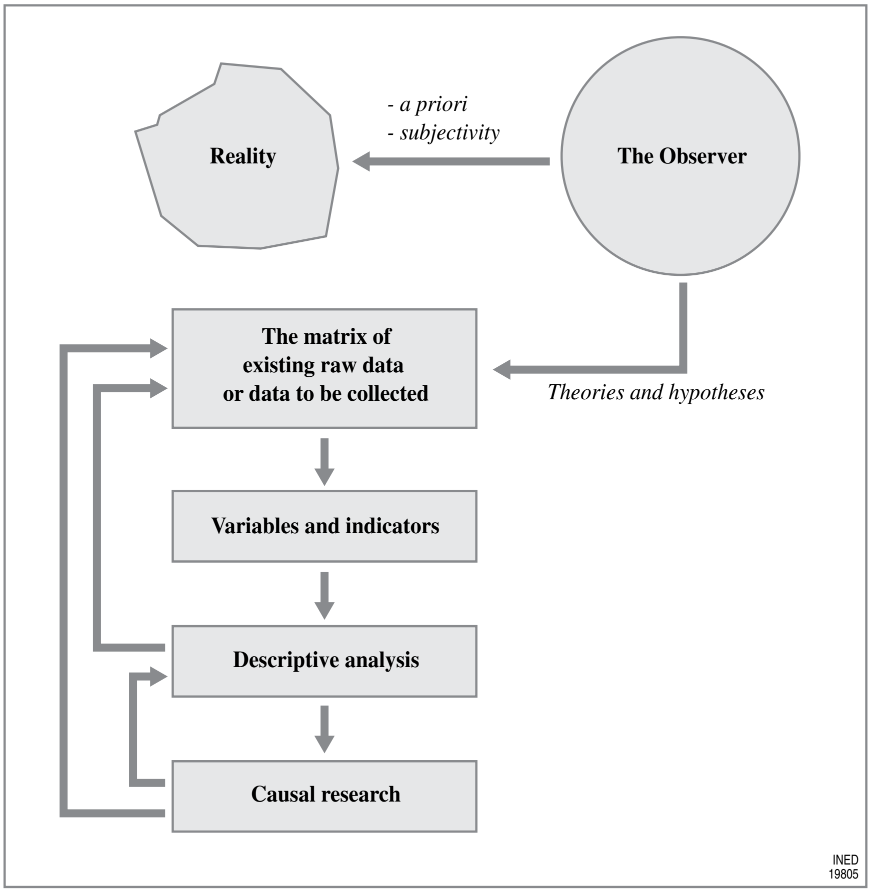


FIGURE 121–1 From reality to observation and analysis.

words, *strategies for the collection of data and analyses should be well thought out and integrated at the beginning of any study*.

II. HISTORY AND ROLE OF OBSERVATION IN DEMOGRAPHY

1. A Brief History

More so than statistics in the general sense, scientific observation in demography has only a brief history, even though before the appearance of the term demography (Guillard, 1855), and the constitution of a real technical corpus (20th century), a number of sources and collection operations useful to the subject had come into being. The majority of these were of an administrative nature, with frequently fiscal objectives and limited geographic coverage.

The first population *enumerations* go back to Early Antiquity (Mesopotamia, China, India, etc.), and the era of the first large centralizing empires (Tabutin, 1984b; Hecht, 1977; Vilquin, 1983; Dupaquier *et al*., 1985). In Europe, the Middle Ages was a period of eclipse, but it was with the mercantile system and the formation of states and centralized governments that the need to measure the population reappeared: surveys and censuses increased from the middle of the 17th century with mainly fiscal objectives. It was only at the end of the 18th and the beginning of the 19th century that we moved "*from administrative observation to scientific observation*" according to Louis Henry (1963). In Europe, the first real national censuses were mostly carried out during the 19th century[3] and their history is even more recent in the rest of the world. Apart from some countries under English rule, where censuses were organized from the end of the 19th century (1872 in India, 1881 in Pakistan), in most Latin American countries the first censuses were carried out

[3] With some countries like Sweden being in advance (1749), and others like Italy (1861) and Canada (1871) being delayed.

in 1900, 1910, or 1920. In the remaining Latin American countries and in Asia this only occurred after World War II, and in a considerable number of sub-Saharan African countries it did not happen until the 1970s.

In Christian Europe, parish registers were the origin of modern *civil registration data*. From the middle of the 16th century,[4] parish priests were to some extent obliged to keep records of baptisms, marriages, and burials, for accounting purposes and as legal proof (of religion and affiliations). Seldom exhaustive before the 18th century, these records are nevertheless the basis for many reconstructions of local population history. However, as was the case for censuses, modern civil registration systems did not cover the whole of Europe or leave the hands of the Church for those of the state[5] until the end of the 18th century. In parallel, the Netherlands (in 1829) and Belgium (in 1856) set up the first *population registers*. In developing countries, the history of civil registration data is entirely different: even though the obligation to register sometimes dates back to the 19th century (1884 in Brazil, 1882 in Algeria), it was rarely aimed, in the beginning, at the whole population, and in many countries the system did not really begin to be developed until the 1950s or 1960s, while to date the coverage rate for civil registration data records remains very poor in a considerable number of cases (notably in sub-Saharan Africa and South Asia).

Although the theoretical foundations of the *demographic survey* of representative samples of the population was established during the 18th century, resulting from the development of the theory of probability, there were hardly any concrete applications to social science before 1910.[6] In demography, this method was only really adopted after the World War II, initially in developed countries, to supplement data taken from civil registration, followed quite rapidly by developing countries to make up for the shortcomings of the civil registration data or instead of censuses, but also as an independent source of information—in particular on practices and attitudes regarding fertility. Today, it has become a major source of information in population research, and in many developing countries is even the main source.

2. Status of Observation in Population Science

Demography has a historical specificity: for a long time, observation was mainly carried out by administrative bodies (statistical offices, ministries) without any particular relationship to universities or research. These bodies had a near monopoly on observation, while in other scientific areas, as Louis Henry (1963) wrote: "*theoretical considerations, observation, analysis, and education are closely linked, if only because the same person considers, observes, analyses, and educates.*" In demography, teaching and research are, as a result of this, developed in universities,[7] cut off from observation, or, in any case, with little influence on it. The situation has obviously changed and the separation is less distinct than it was at one time: because of their main objectives, censuses, civil registration systems, and population registers still depend on the administration (they are even sometimes called *administrative sources*), but the relationship between those who produce the data and those who analyze them has become closer and has improved: statistical monopoly (or the idea of a monopoly) has lessened and cooperation between administration, research, and education has developed considerably, both at the level of dialogue on concepts and methods, and with regard to access to data and circulation of results. Not only does this trend fulfill the increasing needs of the administration and the government for information to be as objective as possible, which only a rigorous scientific approach can provide, but it has been significantly strengthened by the growing importance of survey methods which, while being more complex, are also lighter and less expensive than the traditional sources and the nearly endless growth in the analytical capabilities provided by the computer revolution. These developments certainly do not occur at the same time or at the same rhythm, but the trend is a general one.

Despite all this, researchers in demography retain less control over their observation sources, and the problems associated with data collection methods is not their main concern. To achieve the scientific recognition of their peers in demography it is better to be an analyst or a theorist than a specialist in data collection. This is verified by the low number of articles on observation published in international journals and the near

[4] 1539 in France, 1538 in England, 1563 in Belgium. See, among others, for the history of parish registers: Guillaume and Poussou, 1970.

[5] Here again, the Scandinavian countries (Finland, Norway, Sweden and Denmark) were pioneers, as their national systems were constituted around 1730. In France, parish registers were replaced by civil registers in 1792. In the United States, the system was implemented in the whole of the Federation in 1841.

[6] Arthur L. Bowley in 1913 was the first to use the survey method for a sociological study on the condition of the working class in five English towns.

[7] Except, for example, in France where it is mainly carried out at the Institut national d'études démographiques (INED).

absence of this issue in international conferences.[8] This is particularly true in developed countries where researcher-demographers frequently remain far from the field.[9] The problem is represented differently in developing countries, where administrative sources are rare and sometimes poor and where surveys are all the more numerous as researchers must frequently create their own collection tool: designing questionnaires, checking the quality of information, and adjusting incomplete data are all part of their daily work. Here the risk for the researcher is to be monopolized by data collection and not concerned enough with detailed analyses.

3. The Dependence of Developing Countries with Regard to Information Systems

Each developed country controls its population information system, determines its budget, priorities, the periodicity of activities, etc. There is a certain heterogeneity within a region such as Europe. In developing countries, even though the situations vary, few countries, particularly among the poorest ones, have the means to carry out large-scale data collection with their own funds. The majority depend largely (and sometimes completely) on international donors or bilateral partners, international data collection programmes (such as the World Fertility Survey of the 1970s, or the Demographic and Health Surveys of the 1980s and 1990s). This financial dependence, which in itself results in considerable scientific dependence, increases with the cost of operations and economic inequalities. In these conditions, how is it possible to establish a national medium- and long-term collection strategy?

III. A TYPOLOGY OF DEMOGRAPHIC OBSERVATION SYSTEMS

Table 121–1 provides a general overview of the different data collection sources and methods currently used all over the world, and indicates the importance of each one for each region (developed or developing), the observation unit that it uses, and the main objective assigned to it (movement or structure). It highlights six very different large systems, and the largest ones (continuous recording, censuses, and surveys) are accompanied by a certain number of subsystems or specific methods.

1. **Continuous registration systems**, of which the best known are *civil registration*, the *population register*, and the *demographic surveillance system* or *population laboratory*, i.e., a kind of permanent observatory of population dynamics covering a small geographic area, instantaneously (or almost instantaneously) measure basic demographic events.[10]
2. The main objective of **censuses** is to enumerate the population and obtain a cross-sectional knowledge on its demographic, social, and economic structures, etc., at a given time, even though they nearly always include questions regarding migration, and quite frequently in developing countries (direct or indirect) questions regarding fertility and mortality. The basic principle of the census is to exhaustively cover the target population, even though a census survey carried out on a relatively high percentage of the population in question is sometimes incorrectly called a *microcensus*. Censuses remain common practice, even though recently some European countries are attempting to replace it with population estimates supplied by information from registers.
3. **Demographic sample surveys** are extremely diverse both in their objectives and in the size of the samples used, the observation methods implemented, and naturally the cost of the activity. Six large categories can be distinguished according to the method used to observe the events: the *retrospective survey* over a short period, collecting events which occur over a short period, generally 12 months preceding the interview; the *biographical survey*, collecting all target events which have occurred over the life course of each respondent; the *multiround survey*, collecting events at a given round which have modified the observations made during the preceding round;

[8] Of a total of 269 sessions organized during seven conferences of the International Union for the Scientific Study of Population (IUSSP) from Mexico in 1977 to Salvador de Bahia in 2001, only 12 dealt specifically with data collection or information systems. Of the 88 sessions of the Salvador de Bahia conference, only 3 dealt with this subject. The themes were "The collection of demographic statistics," "Qualitative methods in demography," and "Census data in the 21st century," respectively.

[9] Even though today, in the increasing number of studies on international comparisons, they are more concerned with the problems of information systems, which could hamper the comparability of definitions, concepts, indicators, and therefore the results.

[10] It must however be noted that, generally, civil registration data do not cover migration, which in a few countries is subject to systematic obligatory declaration, as is the case in Italy, for example. On the other hand, contrary to the two other sub-systems, civil registration data do not provide population estimates that can be used as denominators to the rates.

TABLE 121–1 Typology of Information Systems According to Their Frequency, Observation Unit and Main Objective in Both Developed and Developing Countries

System	Developed Countries	Developing Countries	Unit of Observation	Main Objective
Continuous registration systems				
Civil registration data	xxx	xx	Individuals	Movement
Population registers	x	–	Individuals/Households	Movement /Structure
Demographic surveillance/ Population laboratory	–	x	Households/Individuals	Movement
Medical registers of births	x	–	Individuals	Movement
Censuses				
Traditional censuses	xxx	xxx	Households	Structures
Censuses based on registers	x	–	Households/Individuals	Structures
Microcensuses	x	x	Households	Structures
Surveys				
Short reference period survey	–	xx	Households	Structures/Movement
Retrospective or biographical surveys (births, marriage, child mortality, migration histories, etc.)	x	xxx	Individuals/Households	Movement
Multiround household surveys	–	xx	Households	Structures/Movement
Longitudinal prospective surveys monitoring births or individuals	x	xx	Individuals	Movement
Repeated or trend surveys	–	x	Households/Villages	Movement
Dual record system	–	x	Individuals	Movement
Medical and administrative registers	xx	–	Individuals	Movement
Aerial photography/Remote sensing	x	x	Districts/Villages	Movement
Qualitative surveys	x	x	Communities, households, and individuals	Understanding/Explanation

xxx = very frequent; xx = quite frequent; x = infrequent; – = never

cohort studies, a form of *longitudinal prospective studies*, where events in the lives of individuals born in the same year or who fulfill certain criteria are monitored over a number of years; *repeated* or *trend surveys*, which are much less widespread and which consist of repeating, after a relatively long period, a survey which is more or less identical to the preceding one; and finally the *dual record system*, the coupling of observations made during two independent surveys (or other information systems) carried out on the same population, in the same time period.

4. **Aerial photography and remote sensing** are geographers' methods that are useful for studying urban or rural populations and drawing up samples.
5. **Administrative registers** of schools, medical establishments, social security, and companies, etc., are also an additional data source, up-to-date but rarely exhaustive.
6. **Qualitative surveys** are more anthropological in nature and their objective is not measurement, but understanding, and are generally carried out on small samples.

The following sections will be particularly devoted to civil registration, population registers, censuses, and surveys. The other approaches will be the subject of a more succinct discussion, while the final section will deal with the problems of protecting privacy and some general comments.

IV. CONTINUOUS REGISTRATION SYSTEMS

Civil registration data, population registers and demographic surveillance systems aim above all to record events and, to ensure that the recording is as comprehensive as possible, they are designed in a way that minimizes the time period between the event and the recording of the event.

1. Civil Registration Data

Civil registration can be defined as a *continuous, permanent, and obligatory recording system of the events which establish or change an individual's civil status* (birth, death, marriage, divorce, separation, annulment, adoption, etc.) (Tabutin, 1984b, p. 63–64; United Nations, 1973, p. 155–172). It is a national and legal tool everywhere, under the responsibility of the public authorities. Civil registration is a precious resource for population studies but it also has its own limitations, while its spread to developing countries is far from generalized.[11]

a. Advantages and General Limitations

In addition to its primary function of providing legal proof (of an event, an identity, kinship ties, etc.), civil registration has a statistical function and this, of course, is what interests the population scientist. It is the only source that regularly provides[12] the numbers and main characteristics of births, deaths, stillbirths (i.e., late fetal deaths), marriages, and divorces at every geographical level (from the ward to the entire country). It also makes it possible to follow population trends (size and structure) and, notably, to undertake period and cohort analyses of fertility and mortality in all the more depth due to its secular nature. Most frequently however, the system ignores migration, and therefore limits the analysis of population dynamics to natural increase. Furthermore, it only records official marriages and divorces and excludes other unions (consensual and traditional), which are common in many developing countries and increasingly so in Europe and North America.

Almost everywhere in the world, civil registration is nevertheless the essential knowledge base for population change, causes of mortality (here it is irreplaceable), general characteristics of childbirth and children at birth (sex, legitimacy, weight, gestation period, type of childbirth, birth order, etc.,), and those of the parents (age, occupation, level of education, etc.). The types and number of variables selected in the statistical reports obviously vary considerably from one country to another but they make it possible to carry out numerous analyses, differential studies, and measure biological or social risk factors.

This source of abundant information does not however fulfill the growing needs of population research, which looks increasingly towards the explanation of phenomena. This requires numerous economic and social variables regarding the individual, his/her spouse, family, and environment, both with regard to the current situation and the past, which cannot appear on a simple statistical civil registration form.

Moreover, the requirement of exhaustiveness is not always compatible with the collection of detailed, good-quality data. For example, the recorded cause of death is frequently imprecise,[13] with confusion between primary and immediate causes, while changes in nomenclature cause comparability problems over time (Vallin and Meslé, 1988, 1998). Another example, weight at birth, which is available in many European certificates, is often only approximate and, again, the declared occupation (that of the parents for a birth, or that of the deceased for a death) is frequently too vague. The quality of data collected of course varies over time and space. It depends on the practices and diagnoses of the medical profession, the competence of the staff in civil registration offices, coding rules, and data management, but also frequently on the quality of the administrative or medical documents on the basis of which the statistical forms are filled.

Another problem that is crucial for comparison over time or space is the variability in the legal definitions of events, particularly regarding stillbirths and live births. In Europe alone, despite World Health Organization's (WHO's) recommendations and efforts at statistical homogenization in the region, registration criteria still vary considerably (Gourbin and Masuy-Stroobant, 1995). For example, although WHO has for a long time recommended that each birth where the child has shown any sign of life should be recorded as a live birth, in 1998, 5 out of 22 European countries still restricted the criteria by limiting the type of sign of life required or by requiring a minimum life span (Czech Republic). This has resulted in five different definitions. In the same year, only four countries followed WHO's 1975 recommendations on stillbirths[14] and, in

[11] For further information beyond these two points developed here, notably on the diversity of the operating methods and content of civil registration data, the reader can refer to specialized publications (United Nations, 1973, 1985; Tabutin 1984b, Chapt. 3; Linder and Moryama, 1984; Lohlé-Tart and Clairin, 1988).

[12] The results are generally published by year, and less frequently by quarter. Publication most often occurs within a relatively long time scale: from 1 to 2 years, and sometimes even more in some developed countries, such as Belgium.

[13] In Europe, for example, badly defined causes of death varied according to country—from 1% to 12% in 1990 (Meslé, 1995).

[14] According to these recommendations, a miscarriage should be considered a stillbirth (or late fetal death) if the gestation period lasts for at least 22 weeks or if the child weighs more than 500 g. Only Finland, Portugal, Belgium, and Poland adhere to these recommendations.

2003, there were still seven different definitions of stillbirth in Europe. This results in problems and requires the greatest prudence in international comparisons, notably when relating to rare phenomena that are sensitive to changes in definition, such as the rate of stillbirths, perinatal mortality, or infant mortality in developed countries. Depending on which definitions are used it is even possible to obtain different classifications of the countries (Masuy-Stroobant, 1994; Höhn, 1981).

b. The Situation in Developing Countries

These few problems that are inherent to every civil registration system are however negligible compared to those faced by the majority of developing countries. Without going into too much detail here,[15] let us say that there are problems of variable intensity according to region in the level of coverage, the type of data collected, and their use.

The coverage of events by the system is frequently incomplete and sometimes even unknown, and is rarely above 90%.[16] The situation is particularly bad in South Asia and sub-Saharan Africa where, in some cases, civil registration only really functions in urban zones. Generally, the coverage rate is always lower in rural areas than in towns. It is frequently lower for deaths (15% to 60% in sub-Saharan Africa) than for births (30% to 80%) and it is lower again for marriages. In a certain number of countries, the coverage has improved over time (in North Africa and Latin America, for example), while in others it is stagnant or has even regressed.

The data collected varies from one country to another. However, it is often limited to the minimum of information, sex, date of event (or of registration), place of the event, age of the mother (for a birth) or the deceased (for a death). Sometimes marital status is also included (of the mother or the deceased), or the father's nationality or profession. It is much rarer to find variables such as parity or education of the mother (or the father), the length of the gestation period, and the weight of the child at birth (United Nations, 1985, p. 41).

The processing, publication, and analysis of civil registration data is generally only very partial and incomplete, in the best of cases only resulting in the publication of some basic data such as the annual number of events or distribution according to sex or age group (of the mother or the deceased). Much of the information is not processed, and when it is used it is frequently with considerable delay (2 to 5 years).

The weaknesses and shortcomings of civil registration systems in developing countries are due to a large number of combined factors, such as the system itself (lack of infrastructure, resources, competent staff in the area of registration and analysis), the population (ignorance or lack of motivation), or the authorities (who give low priority to the problem). Numerous projects, whether stimulated from outside or not, have tried to improve the system, notably in sub-Saharan Africa, but the majority has had little effect. In reality, the civil registration system cannot function correctly without a strong and well-organized territorial administration, without any real legal obligation and sanction, without the population becoming aware. What can be done in the meantime? We can use and analyze existing data (data from towns, for example), motivate statistical institutes, train population scientists to specialize in the area, rehabilitate as much as possible the image of this information system. But in many countries it will take a long time before the civil registration system becomes a reliable source of national population measures and before it will be able to provide information for proper population registers.

2. Population Registers

From the 19th century, population registers were set up in some European countries with wider objectives than those of the civil registration system. The population register can be defined as *a permanent and continuous system of demographic and social information on each individual and household in a given territory*. Initially local, now frequently national (it is now called a National Register), it is progressively becoming a paramount and promising source of administrative and scientific information that obviously has its advantages and its limitations.[17]

[15] For further information, please refer to specialized publications that either provide an overview of the world situation (United Nations, 1985) or perspectives by major regions, like Africa (François, 1988; Gendreau, 1993; Garenne and Zanou, 1995).

[16] According to an already old survey carried out on 112 countries, the United Nations (1978) estimated that of 93 developing countries, only 28 for births, 25 for deaths, and 19 for marriages had a coverage rate of over 90%. This rate is unknown for 10 others (United Nations, 1985, p. 37). There is no precise recent data available, but generally, it would seem that the situation has changed little since then.

[17] For some publications regarding population registers, see United Nations (1969), Verhoef and van de Kaa (1987) or Poulain *et al.* (1991), for example.

a. Local Population Registers: Philosophy and Content

The local register system is not new. It dates back to 1749 in Sweden, 1847 in Belgium,[18] 1850 in the Netherlands, and 1924 in Denmark. In 2003, 9 of the then 15 member-states of the European Union had registers,[19] along with 6 Central and Eastern European countries.

The principle of registers is to gather certain information on each individual living in a community. The content varies from one country to another. Generally, names and given names, sex, date and place of birth, address, marital status, nationality, or profession, date and place of death are included in the register; the marriage date, the previous address, and the relation to the household head are also sometimes included. All changes in situations are recorded there through a mandatory declaration or through linkage with other forms: births, deaths, changes of address, marriages, divorces, or changes in occupation. For each migration, for example, the Belgian register records the date of arrival in the area and the previous place of residence as well as the date of departure and the destination.

Their principle of operation is relatively simple. The register is created from an initial census of the *de jure* population: selected information on each individual is recorded. It is organized by household and housing unit. Thereafter, the register (which was previously manual, but is now computerized) is regularly and rapidly updated by the recording of observed changes, either continuously with information collected in the civil registration system and by the authorities responsible for recording changes in address, or periodically (every 10 years in Belgium from 1866 to 1981) during each successive census.

In addition to the various administrative uses (identification of individuals, updating of electoral rolls, taxation, files linked to social schemes), the register is also of obvious interest for statistics and demography.

1. It can provide at any time an estimate of the population of each community and its composition according to various characteristics, as well as population dynamics (natality, mortality, and migration, internal and international).
2. It is a monitored observation, a kind of biographical surveillance system, as it collects and preserves information on each individual (and household) on the various demographic events that take place over the course of the life span. It makes it possible to combine period and cohort approaches.
3. It facilitates the preparation and evaluation of population censuses, and it is a good sample frame for surveys. Some countries that continuously record migration have even stopped carrying out censuses, as they consider that their system for recording events is sufficient to correctly update the population register.

But there is often a wide difference between the principle and reality.

Approximately 30 countries have population registers at present. The basic principles are the same everywhere, but the content, the organization, and the quality vary. It is possible to simplify by comparing the very reliable and well-organized systems of the Scandinavian countries, Belgium and the Netherlands, to the less-rigorous systems of countries such as Germany and Italy, where the updating (sometimes still manual) of data is carried out with delay and suffers from omissions and where definitively the register is not yet the central axis of the local system of information and management.

In this respect, it is advisable to distinguish the *centralized systems* at the level of the whole of the country (the Benelux and Scandinavian countries) from *local systems*, whether computerized or not, where information only exists at a local level (Poulain, 1995). Centralization, which leads to a kind of superregister covering the whole of the country, invariably favors the quality of the information (i.e., removes double counts, promotes more effective verification). It also facilitates linking the central population register to other administrative registers (Housing, Social Security). Under certain conditions, it even makes it possible to replace the traditional census. In Belgium, extending the communal register to a national one took 17 years between the adoption of the idea (1966) and its legal and practical implementation (1983). All information declared in the *Local Population Register* (*Registre Communal de Population*) is now transferred electronically to the *National Register* (*Registre National*), which instantaneously modifies its corresponding individual file and sends an acknowledgment of delivery with the new individual file. This system is reliable and quick.

b. Limitations and Dangers

Although it is a very valuable resource for population scientists, the population register also has its

[18] Shortly after independence (1830), Belgium followed an old regulation from the French Revolution (of 1791) of keeping *nominal lists of inhabitants* by municipality, a regulation which was practically never adhered to in France.

[19] Only Austria, Greece, France, Ireland, Portugal, and the United Kingdom do not have such a system.

limitations (notably with regard to the nature and quality of the data) and can even constitute a danger (risks of infringing on respect of privacy).

The nature of the data. A register naturally contains, above all, information which is of interest to local (and central) administration. It includes demographic variables but very little economic, social, medical, or cultural data (for example, in Belgium stillbirths are not included). The register alone therefore is not comprehensive enough to enable real differential demographic analysis to be made.

The quality of data. The quality of the information collected depends on resources, the competence of local staff, and on the motivation of staff and that of informants. Emigration (particularly beyond national borders) is generally badly recorded. It is indeed quite common for departures not to be declared, while the reasons for declaring an arrival are more binding. If the register is centralized, some of the gaps regarding departures can be filled by the recording of the arrival in another area, but this procedure is nearly impossible to implement for departures abroad. Changes in occupation also frequently suffer from underregistration.

Access to information. Access to the register data should obviously be strictly regulated and limited, in order to respect privacy.[20] On the one hand, in places where a register exists, legislation is strict and access to the data is generally difficult. On the other hand, when there is no register, such as in France for example, this is partly due to concerns regarding the capacity of the legislation to effectively preserve the confidentiality of the data.

Ultimately, a good population register is an incomparable source of information on local or national population dynamics and on internal migration. It can even sometimes call into question the usefulness of censuses. On the other hand, it will doubtless never remove the need for surveys.

3. Demographic Surveillance Systems or Population Laboratories

A *population surveillance system* means different things to different people, as well as to the reality covered by them.[21] In any case, the *demographic surveillance system* or *population laboratory* can be defined as *a method of continuous and long-term observation of a small area or a specifically defined society, the objectives of which are both the measurement and understanding of sociodemographic changes and health changes*. This will be discussed in more detail by Gilles Pison in chapter 124, which is devoted to this very particular and original method of collecting demographic data. Initially, in the absence of a reliable civil registration system, the idea is simply to place the entire population of a given geographic zone under constant, or at least very regular, observation. These geographic zones are generally small rural zones, as is the case of the demographic surveillance sites in Gambia and Senegal,[22] more rarely larger zones such as Matlab in Bangladesh.[23]

Even the methods of observation can vary enormously, from frequent multiround surveys (every 3 or 6 months) to permanent recording systems combined with periodic censuses. Moreover, multiple surveys on extremely varied themes are generally added to this basic observation. The interest and the value of these demographic surveillance sites depend largely on their durability, which makes it possible to trace changes over a long period: 10, 20, 30 years, or even more. The numbers concerned are, on the other hand, low (3,000 to 8,000 households, with Matlab being an exceptional case), and nonrepresentative.

Let us look at some examples. In Burkina Faso, in the context of a research project on levels, trends, and factors of child mortality and health, three quite different rural zones from economic, ethnic, and health points of view, made up of a total of 37 villages, approximately 3,800 households, and 34,000 individuals were monitored from 1986 to 1994. After the initial census, visits took place regularly every 6 months (multiround surveys), while in parallel, more specific operations were carried out at household (on migration in 1994 to 1995) and village level (on social, agricultural, and land characteristics).

Another example is when Institut national d'études démographiques (INED) set up a new population laboratory in Mlomp (Casamanca) in 1985, which supplements the older sites of Niakhar and Bandafasi (Pison *et al*., 2001; see also Chapter 124). The objective, which was initially mainly demographic, was to measure levels and trends in fertility, mortality (including causes of death), nuptiality, and migration (permanent or seasonal) of this small rural population

[20] Even more so as in the most advanced countries, such as Belgium, each individual has only one national identity number which enables the various administrative systems to communicate among themselves. In 1995, the European Union officially drew attention to the risks to individual freedom that are presented by these possibilities of linking individual data files.

[21] The term *Prospective Community Studies* is also found in English language publications (for a good summary, see Das Gupta *et al*., 1997).

[22] The first was established in 1963 by ORSTOM (which has now become Institut de Recherche pour le Developpement; IRD) in Niakhar in a zone of 30 villages with a total of 25,000 inhabitants.

[23] Some 200 villages with a total of 200,000 inhabitants that have been monitored for over thirty years.

of 11 villages (6,220 people in 1985, 7,600 in 2000). After an initial census carried out at the end of 1984, the data collected on the population was regularly updated by an annual survey, while in parallel, more specifically targeted surveys were carried out on very diverse subjects: conditions of childbirth, the feeding and growth of children, contraception, acquired immune deficiency syndrome (AIDS), etc.

The usefulness of these demographic surveillance systems no longer need to be demonstrated: they provide very valuable demographic data that enable longitudinal analyses to be made from individual and collective data (households, villages)—to which numerous economic, social, and health observations can be added, thus making explanatory analyses possible. They are particularly useful for studying changes over the long-term. The information collected is generally of good quality.

The limitations are nevertheless just as obvious. The main one is the very local nature of the information which, although perfectly representative as ideal for the area studied, does not go beyond the borders of the examined zone. In no case can the results obtained be extrapolated to the national level, or even the regional level. On the other hand, if there are several population laboratories, as is the case in Senegal, the information obtained can give an idea of the scale of the geographic variations. But the main advantage of the demographic surveillance system is found in its contribution to the understanding of phenomena and its evolution, rather than in the measurement of intensity levels. Furthermore, the duration, the cost, the delicate procedures of matching surveys, the complicated nature of data management, the problems associated with small numbers, and the complexity of the analyses (not always carried through to their conclusion) confine this technique to a rather experimental approach (hence the name *population laboratory*), preliminary or complementary to more basic procedures but can claim to be representative.

In spite of their limitations, demographic surveillance systems are an essential tool for research and studies in the developing countries. Of course they require competence, stability, motivation of scientific staff, financial resources, and time (an aspect which is not sufficiently taken into consideration by funders). But, in many contexts, it is only at this cost that we can (or could) improve our understanding of reality and changes.

V. CENSUSES

The census is doubtless the oldest tool for collecting demographic data (Hecht, 1977; Tabutin, 1984b), the best known by the general public today, and also the most debated from several points of view (i.e., cost, social usefulness, respect of privacy). Following a brief reminder of some general points, it is useful to examine in more detail the reasons for its existence, issues of quality, and the need to adapt the collection of information to the realities of the modern world, which naturally leads to the search for possible alternatives to the traditional census.

1. A Definition and Some General Points

The United Nations (1992b, p. 3) define population censuses[24] as "*a group of operations which consist of collecting, grouping, evaluating, analyzing, publishing, and disseminating demographic, economic, and social data relating to all the inhabitants of a country, or of a specific part of a country, at a given time*. This definition is essentially a reiteration of previous ones (United Nations, 1958, 1970), apart from the fact that it now includes the concept of disseminating information. This may appear banal, but it is revealing of the scale of the proportion of information never used and often definitively lost even though collected at great cost and with the support of the United Nations or other donors. Above all, a census aims at obtaining a kind of snapshot of the population: it makes it possible to know, at a given time, the size, geographic distribution, as well as distribution according to a certain number of socioeconomic and demographic characteristics. Furthermore, it is quite frequently used to obtain certain information regarding population movement. It nearly always includes questions on migration as, in the absence of a continuous registration system, it is the most effective tool available for collecting data on internal migration at the lowest geographical level. In developing countries, censuses are also frequently used to collect information on fertility and mortality. In the cases of migration and fertility, they can simply be status data (place of birth or residence at the previous census, number of children already born, etc.) of the same nature as other traditional census questions, although direct questions on events that have occurred during a recent period (generally the last 12 months) are also sometimes included.

The four main principles of the census are: *exhaustiveness* (it should cover all the inhabitants of a given territory), *individual enumeration* (it gathers informa-

[24] There are also other terms, such as census, general population census, population and housing census (all censuses include questions on housing, the basic statistical unit).

TABLE 121–2 Dates and Types of the Three Most Recent Enumeration Operations in Europe

Country	Mode of Operation (Most Recent)	Dates (Month/Year)		
Traditional censuses				
Austria	Delivery—collection of questionnaire	05/81	05/91	05/01
Belgium	Sending—return of the questionnaire by post	03/81	03/91	10/01
France	Delivery—collection of questionnaire	03/82	03/90	03/99
Germany	Delivery—collection of questionnaire	05/87		
Great Britain	Delivery of questionnaire, return by post	03/81	04/91	04/01
Greece	Interview by census taker	04/81	03/91	04/01
Ireland	Delivery—collection of questionnaire	04/81	04/91	04/02
Italy	Delivery—collection of questionnaire	10/81	10/91	10/01
Luxembourg	Delivery—collection of questionnaire	03/81	03/91	02/01
Portugal	Delivery—collection of questionnaire	03/81	03/91	03/01
Spain	Delivery—collection of questionnaire	03/81	03/91	11/01
Enumeration from registers or surveys				
Denmark	From registers	01/81	01/91	01/01
Finland	From registers	11/80	12/90	12/00
Germany	From surveys		04/91	
Netherlands	From registers and surveys	04/81	04/91	01/01
Sweden	From registers	09/80	11/90	

tion on each individual),[25] *simultaneity* (the information collected refers to the situation at the same single date and in order to carry out this exercise should be completed in a few days) and, lastly, *periodicity* (the census should be repeated at regular intervals). Let us add to this that, in contrast to surveys, a census can only be carried out by the government, not only because of the enormous human, technical, and financial resources that it requires, but also because it is closely associated with major administrative and political stakes.

Today, *censuses are conducted universally* with two exceptions. On the one hand, some developed countries such as Denmark, Finland, the Netherlands, and Sweden estimate their population using their population registers and no longer carry out censuses. On the other hand, certain developing countries have not been able, for various reasons, to carry out censuses. Between 1975 and 1984, 191 countries, of a total of 213, carried out a census; therefore 95% of the world population were subjected to a census during that decade.[26] Since 1990, with the exception of countries at war (i.e., Democratic Republic of the Congo, Angola, Afghanistan) or countries where the census is a political taboo (Lebanon), nearly every country has carried out a census. The African countries, which for a long time were very late in this field, have by now carried out at least two censuses.[27] North Africa has had four or five since the 1950s. The time interval between censuses is usually 10 years in America (North and South), in the Lusophone countries, and frequently in the former English colonies. Elsewhere, it is more variable, currently between 11 and 12 years in Africa, and 8 to 10 years in Europe during the 1980s and 1990s (Table 121–2). Only Japan and Ireland enumerate their population every 5 years.

The content of census questionnaires and therefore the wealth of information varies enormously from one region or country to another. Besides the essential variables of age, sex, and marital status, questions on nationality, level of education, and occupation can frequently be found. The quantity and nature of questions posed are very variable. It is also necessary to take into account this significant variability and the changes in definition, which sometimes make comparisons in space and time difficult. Weighty censuses (household questionnaires and individual questionnaires with 50 to 70 questions) can be distinguished from light censuses (a collective form with approximately 20 questions, as is the case in Africa and South Asia).

2. Advantages and Disadvantages of the Traditional Census

Although the traditional census, as it has just been described, remains the most common method of

[25] It is *inter alia* that distinguishes it from the *administrative census*, a practice inherited from colonization and still sometimes used in Africa, which consists of collecting some aggregated information from each household (number of persons by sex, number of children and adults, number of foreigners).

[26] For a world view, see United Nations, 1991; for Europe, see Eggerickx and Begeot, 1993; Redfern, 1987.

[27] Only Djibouti, Ethiopia, Chad and Congo R.D. have still carried out only one census.

observing a population and has become nearly universal due to the huge efforts made in developing countries over the last few decades, its fate is under debate in the most advanced countries due to the evolution of requirements and the appearance of alternative methods.

The *advantages* of the census are considerable. Its exhaustiveness enables it to provide total statistical coverage, from the broadest level (the country), to the most local which is the parish, district, or even zone of enumeration, and finally, the household[28] (or house unit). It makes it possible to study all kinds of ethnic, social, or occupational subgroups at the finest level. It is also the only source for analyzing internal migrations up to local level. It is an essential sample frame for any population survey, which only the few European countries that possess population registers can do without. Definitively, and perhaps especially, it is, far beyond the needs of demographic research, an essential tool for regional planning and management (both public and private) of national and local economic and social affairs.

But the census no longer enjoys unanimous consensus: it is increasingly being questioned both by the producers of data as well as by the users and even sometimes by the population itself. Indeed, despite its advantages it also has *weaknesses and disadvantages*, of a technical, methodological, and ethical nature. We will refer briefly to the most important points.[29]

1. **Periodicity.** The time interval between two censuses, which is generally approximately 10 years, is too long for taking into account ongoing changes rapidly enough.
2. **Publication intervals.** Too frequently, the time interval between analyses and particularly publication is very long; often 5 years and sometimes more.[30] When the results are published (notably at a local level), they are often completely out of date.
3. **Cost.** The traditional census is the most expensive collection operation of all, at least in internal budget.[31] Of course, the cost varies greatly from one country to another.[32] The current trend in Europe (and elsewhere no doubt) is to try and reduce it as much as possible by shortening the questionnaire, by reducing the frequency or, even more radically, by finding alternatives. This financial problem is particularly acute in the poorest countries, precisely where the census remains the main source of demographic information.
4. **Content.** Because of the scale of the operation, the questionnaire of a traditional census should be as simple as possible. There should be few questions. They should be easy to ask and the responses should be easy to provide. Complex questionnaires and delicate questions such as those concerning income or certain aspects of personal life, reproductive history, migratory history, and occupational history of individuals are excluded from it. It is essentially a diachronic tool.
5. **Quality.** The census is not perfect, no more than any other information system. Its objective of exhaustiveness is always confronted by coverage problems (over- or underestimates of some subpopulations), while the responses do not always strictly reflect reality.

To these technical and organizational problems is increasingly added an obstacle which is specific to modern democratic societies: the *respect of privacy*. The census is certainly useful and confidential,[33] but it is increasingly difficult to convince the population of this. The growing fear of intrusion into private life leads to the census being seen as an additional bureaucratic hold that, furthermore, is expensive to the taxpayer. Numerous pressure groups and anticensus organizations were established in Europe during the 1980s and 1990s (in Germany, the Netherlands, Switzerland, Belgium, etc.). This dispute, while never leading to the total abandon of the census (except perhaps in the Netherlands), has sometimes seriously affected its quality (failure to respond, falsification of responses).

Faced with all these problems, more or less radical alternatives are being sought.

[28] For example in this respect, see the work of Françoise Bartiaux (1991) on the households of elderly persons using the American, Australian, and Italian censuses.

[29] For more details, see notably the analysis of Philip Redfern (1987) on the problems and the future of censuses in Europe.

[30] In spite of the very widespread practice of percentage sample analyses (25% sample, to 5% sample, etc.) and of the undeniable progress in data-processing (acceleration of data capture and verification, optical character recognition, etc.).

[31] If it were necessary to take into account the total cost of the continuous registration of events for a population register, this statement would certainly need to be reviewed.

[32] From $2 to $5 per inhabitant in Europe, from $5 in Canada and the United States in the 1980s (Redfern, 1987). In Belgium, they increased from 12.4 million Euros in 1981 to 25 million Euros in 1991, in the United States they increased from $1 billion dollars in 1980 to $2.6 billion in 1990 ($10 per inhabitant). The cost of censuses increases considerably in developing countries.

[33] It was nevertheless used in Belgium to update the population registers.

3. The Alternatives to the Traditional Census

On a worldwide scale, the traditional census certainly continues to hold first place and it will remain for a long time essential in the majority of developing countries, although its content might need to be revised occasionally. However, some countries, often amongst the richest and most advanced in the area of statistics (registers, coupling of files, etc.), during the 1980s started to implement other data collection methods that could replace the census due to economic and effectiveness concerns (time, nature, and quality of data).

1. **The microcensus.** Some countries have attempted from time to time to replace a traditional census with a microcensus, which in fact is nothing more than a demographic sample survey on a varying scale. For example, in Germany in 1981 and 1991 a national survey on 1% of the population was carried out either side of the traditional general census of 1987. A microcensus was also carried out in Russia in 1994, as a general population census was not possible soon enough after the collapse of the USSR to be able to detect the demographic upheavals that followed it. The term microcensus leads to confusion as, due to its lack of exhaustiveness, it obviously cannot replace a census. It would be more appropriate to call it an intercensus survey as its objective is basically to reduce the disadvantage of the long time interval between two censuses whose cost prevents the periodicity from being increased. This pseudoalternative should therefore be classified among surveys.

2. **The light census accompanied by an extensive survey.** The population as a whole only receives a short questionnaire limited to a few basic questions on population and housing, but for a significant percentage of the population (from 10% to 25%), the questionnaire is much more comprehensive and deals with numerous variables concerning education, employment, migration, fertility, etc. This is the system that has been adopted in North America, in some Latin American countries (Brazil, Venezuela), and in Hungary and Russia. It accelerates the procedure for analysis and makes it possible to look deeper into some areas with a large sample.[34] This method has been proposed to reduce the cost and improve the performance of African censuses (Gubry *et al*., 1996), but it continues to be confronted by resistance from statistical offices that are accustomed to rely on the financial bonanza provided by the traditional census.

3. **The reduced census combined with administrative registers.** In the same spirit, some countries, like the Nordic countries (Sweden, Finland, and Norway) only carry out very light censuses that are immediately enriched by being matched with information from different administrative registers (education, housing, income) that are interconnected. One could even go as far as to say that these minicensuses are only *ad hoc* operations that are aimed at periodically verifying the exhaustiveness of the population estimates based on registers. In any case, this system is faster and the marginal cost is relatively low.

4. **Enumeration operations based exclusively on registers and surveys.** Two countries, Denmark since 1980 and the Netherlands since 1981, no longer carry out a general population census in the traditional meaning of the term (Eggerickx and Begeot, 1993; Redfern, 1987). In Denmark, the census has been replaced by the combined use of the data from seven different registers (population, housing, companies, salaries, employment, income, and education). In the Netherlands, the same is done using three sources: population register, a survey on employment, and a survey on housing. Following the 2001 census (its last), Belgium will change to a similar system based on the population register and three data bases (housing, education, socioeconomic characteristics) that will be created using information from the 2001 census.

The advantages of the two alternatives above are clear: variety and wealth of information, speed of execution, lower cost, possibly greater frequency of the operations, and no direct participation of the population. But this is only possible in the few countries that have a well-established tradition of high-quality centralized registers and can link files while guaranteeing the confidentiality of information.

5. **The French attempt at a "renovated census."** In France, the ongoing renovation of the census method is due to the paramount concern of adapting public statistics to the expectations of the users, with, notably, the need for a faster update of information (Institut National de la Statistique et des Etudes, INSEE; 2000). Instead of a traditional census every 8 or 9 years, since 2004, the renovated census is based on an annual collection of information that only affects a part of the population each time. The procedure used differs according to the size of the area. In areas with fewer than 10,000 inhabitants a traditional census must be carried out every 5 years, at a rate of annual operations dealing each time with a different fifth of the area. Areas with more than 10,000 inhabitants are each divided into five large groups of buildings. Each year,

[34] In Hungary in 1990, 20% of households received a long questionnaire (63 questions instead of the 29 in the short questionnaire); in the United States in 1990, this was the case for 17% of households (with 59 questions instead of 14 in the short questionnaire).

the housing units of a group are enumerated and a census conducted in 40% of the housing units. Therefore, in 5 years, all the housing units in the area will have been enumerated but only 40% of the population will have taken a census. The supporters of this new system believe it will make it possible to gain 3 to 4 years on the provision of results, with a better distribution of cost over time, and possibly improved quality (INSEE, 2000).

4. Quality and Comparability

The general public, users, and sometimes even demographers forget that every census includes errors, gaps, and shortcomings: regardless of the method used, the operation is of scale and should be completed in a short time. The scale of errors varies considerably from one country to another, from one census to another, according to the past experiences of the authorities in charge, the care taken in the preparation of the operation, the concern with verifying and correcting the data, and also according to the abilities of the census takers, and the population's education level. Didactically, these errors can be classified into two large categories: those which stem from the responses to the questions asked (observation errors), and those stemming from the census coverage (enumeration errors). Here we will only discuss the case of the traditional census.

1. **Observation errors.** Some examples are inaccuracy (whether voluntary or not), confusion, false declarations, nonresponses, etc.—of course anything is possible (even declaring fictional persons or children born after the census date[35] or persons who died before this date). There are of course variables that, for various reasons (technical, psychological, political) are more prone to error than others: economic activity (occupation, status, qualification), nationality, sometimes marital status, residency status (when asked in terms of present/absent/visitor), migratory status (date of arrival for example), and mortality (during the previous 12 months in the developing countries).

2. **Enumeration errors.** All users should be able to know their scale and nature. No census is completely exempt from omissions and double counts, with omissions generally being more numerous than double counts. What is the rate of net omissions? Who has been omitted or counted in error? How are the results rectified? Measuring these errors is the objective of *postcensus verification surveys* that consist of carrying out the census again on a sample of districts and comparing the results with those of the general census.[36] Unfortunately, not all countries carry out this type of verification. Table 121–3 presents some results for Europe. The omissions are between 0.5% and 6%, while the duplications (less well-known) are between 0.5% and 5%.

TABLE 121–3 Omissions and Double Counts in the Traditional Censuses of Some European Countries Circa 1991[1]

Country	Omissions	Duplications
Austria (1991)	0.5%–0.8%	0.5%–0.8%
Belgium (1991)	3.4%	ε
Spain (1991)	2.9%	3.0%
France (1990)	1.8%	0.7%
Ireland (1991)	ε	ε
Luxembourg (1991)	ε	ε
Portugal (1991)	6.3%	5.3%
FRG (1991)	<1.0%	?
United Kingdom (1991)	2.0%	?
Switzerland (1990)	1.1%	ε

Source: Eggerickx and Begeot (1993)
ε = very low; ? = data not available
[1] Data per official of corresponding country.

The coverage rate of censuses varies greatly according to the population category or the living environment. In Belgium in 1991, for example, the omissions varied from 2% among Belgians to 12% among foreigners, from 9% in Brussels to 4% in other cities, and to 2% in the rest of the country. In the United States in 1991, the omissions varied from 2% among whites to 5% among blacks and hispanics. In developing countries, there are generally no precise measures, but where they exist the omission rate is between 5% and 10% and is particularly high among very young children, men aged 15 to 30, and elderly women.

Certain population categories can also almost completely be omitted by censuses: people with no fixed address, illegal immigrants, the homeless, nomads, etc. These groups should be the subject, and sometimes are, of specific enumerations or estimates.[37]

3. **Comparability problems.** The diversity of these errors, as well as other factors of discordance, cause a

[35] These were estimated to be around 16,000 in the French census in 1990.

[36] For further detail on the procedure and the results obtained, see Hogan, 1993, and Pauti, 1992, for the United States; Coëffic, 1993, for France; Boudreau, 1989, for Canada; Redfern, 1987, for 12 European countries.

[37] In some countries, attempts are made to enumerate the homeless in nocturnal operations. Nearly 250,000 were enumerated in this way in the United States during the night of March 20–21, 1991. In some countries such as Mauritania, for example, the nomads are the subject of specific operations (a survey of a sample of wells for example, during a period of regrouping herds and households).

certain number of problems in the comparability of census data both in time (for the same country) and in space (between countries). Differential errors but also differences or changes in the definition of concepts used (occupation, living environment, etc.) or variations in the composition of the questionnaire and the formulation of questions (addition, removal, new formulation) may introduce a bias in comparisons. In their study of European censuses, Thierry Eggerickx and François Begeot (1993) conclude "*that apart from variables relating to structures* (*age*, *sex*, *marital status*, *and nationality*), *the comparison of results in time frequently seems impossible to achieve with the other demographic topics.*" International comparisons are even more delicate.

5. What is the Future for Censuses?

The future of the census is frequently evoked and discussed, both in the Northern and Southern regions, but in different contexts and for different reasons. The majority of demographers and many other users recognize *a priori* the importance of censuses: it is necessary from time to time for every country to have a detailed image (geographic or social) of its population, its structure, and sociodemographic characteristics, its housing stock, etc. There is however debate on the best means of achieving this objective.

In developed countries, the debate is centred, on the one hand, on the durability of the traditional census compared to other sources likely to satisfy the same needs at a lower cost, and on the other hand, on the adequacy of responses, which it can provide on major concerns, and its social usefulness. Abandoning the traditional census can, in effect, be envisaged in countries that have excellent population registers, and this step has already been taken in some Nordic countries and will also doubtless be taken shortly by some other countries (Belgium, Luxembourg). But the censuses based entirely or partly on registers themselves do not only have advantages: they are less flexible (changes or adaptation of questions), rather rudimentary (for employment or education, for example), and not always of excellent quality. In addition, they can clash as much, and maybe even more, with the sensitivity of a population that feels observed without really knowing why or how.

While the traditional census remains essential in other countries (the vast majority), some would like its content to be better adapted to the major concerns of our time (unemployment, youth, elderly people, part-time employment, situations of precariousness and poverty, etc.). But this is asking too much of a census. Every subject cannot be addressed in a census whose main objective is exhaustiveness. This is what enables a census to provide a complete and precise sample frame for every type of survey that is better able to approach these same questions in more depth than the census itself.

The problems raised by the census in *developing countries* are of a different nature: they are variable from one region to another (the statistical situation in Brazil or Mexico cannot be compared with that of Burkina Faso or Nepal). But the census remains essential everywhere and it would be out of the question to abolish it. In the statistically poorest regions, on the contrary, it is necessary to increase the frequency, decrease the time delays in analysis and publication, and improve the analyses and the dissemination of the results. The most serious impediment remains the increasingly high cost of the operation, which very few countries can meet. For example, in Africa it has certainly become time to use light censuses, which are more rapid from all points of view and doubtless of better quality, together with a solid program of detailed surveys in the context of a national strategy of coherent data collection (Gubry *et al*., 1996).

VI. SAMPLE SURVEYS

The survey is a more recent tradition (in demography, the method became generalized from the 1950s) and is, for various reasons, a preferred and ideal tool for collecting data in demography, and more generally in social science. It exists in developed countries and in developing countries, but with different objectives, which relate to the performance of other information systems. Basically, in Europe and North America the survey is a supplement to qualitative and quantitative information (i.e., behavior, opinion, individual life courses) compared to the source data provided by censuses and permanent registration systems; elsewhere it is still frequently a palliative to the lack of basic sources and its main aim is to measure levels and trends of components of demographic change, but it must also fulfill the needs of explanatory analysis.

The problems of population surveys[38] are vast and diverse.[39] After a brief reminder of their advantages and limitations, and an attempt at creating a typology, I will then focus particularly on two major types of

[38] Here I will not discuss specific surveys on health, budget, and household consumption, or the workforce and employment, even though they are sometimes indirect sources of demographic information.

[39] For more details, see, for example, Tabutin, 1984b, Chapter 5; Committee on Population and Demography, 1981; United Nations, 1992a; Lohlé-Tart *et al*., 1988.

survey: the retrospective survey and the multiround survey.

1. Definitions, Advantages, and Limitations

The sample survey estimates certain characteristics of a population through observing only a fraction of that population.[40] This is a completely general definition, which applies to population surveys as well as to other social science and health surveys. Basically, it is the fact of using a sample that distinguishes a sample survey from a census, which is an exhaustive survey. But because of this, its objectives can also be different from those of a census.

Its main *advantage* is the fact that it is far less costly, which leaves a large margin to invest in other directions, which, in turn, contribute to the advantages of the method. Firstly, the data quality can be improved (the very small number of interviewers allows for high-level recruitment and a more comprehensive *ad hoc* training), the room for maneuver provided to the researcher with regard to the definition of concepts is nearly limitless (in any case incomparable to the rigidity of statistical administrative systems), the quantity of possible variables (frequently between 100 and 200) is high, and the nature of the questions asked (sometimes specific or delicate) is very open. Far from being merely a process of replacement or a substitute to the gaps in administrative statistics, the sample survey is above all an essential tool for advancing knowledge: it makes it possible to reconstitute the past; establish family or individual biographies; and study the interference between events, phenomena, or the interrelations (individual or aggregated) between population and social, economic, or cultural factors, to provide quantitative (unions outside marriage, contraceptive prevalence, etc.), and qualitative (opinions, practices, motivation) information that would otherwise be impossible to obtain.

Of course, on the other hand, it has its *limitations* and poses some problems. The level of geographical analysis can only be fairly rough: the majority of national surveys, at least in developing countries, are only representative at the level of two or three living environments, or of four or five large regions. Let us not forget that every estimate through a sample survey is affected by a sampling error,[41] which is greater when the sample size is small. The latter must obviously be in connection to the statistical frequency of the events studied (rather low in demography) and the precision required. But furthermore, the greater the sample, the higher the cost of the operation and the greater the risk of observation errors. Finally, a problem sometimes difficult to solve, in particular in the developing countries, is the precise, complete and up-to-date sample frame required for the constitution of the sample. The census, which is the natural basis for this, is sometimes too remote to be effective enough. Sample surveys are not without their cost, or delays in completion, or problems of quality, and all these aspects vary significantly according to the type of survey and the strategies of observation.

[40] A similar definition to that of the 1981 Multilingual Demographic Dictionary (IUSSP, 1981).

[41] Sampling errors (or sample errors) add on to observation errors (Desabie, 1965; Tabutin, 1984b, Chapter 5; Clairin and Brion, 1996). Demographers have an annoying tendency to only seldom clarify it in their publication of results.

2. Objectives and Typology

But *what do we understand exactly by population or demographic survey?* Let us say that it is a survey whose main objective could be measurement of population composition and changes, or the understanding and explanation of levels and trends of the demographic phenomena (fertility, nuptiality, mortality, migration). Considered in this way, the line between demography, sociology, and health is sometimes narrow. We could for example include family surveys or health surveys, at least when they relate to individuals or households and seek (among other things) to quantify. On the other hand, we should doubtless exclude operations specific to employment, education, living conditions, nutrition, and consumption from the demographic field, even if they can indirectly contribute to sociodemographic knowledge.[42]

Thus defined, we can classify population surveys into a few large categories according to their main objective:

1. **Measurement surveys.** Many surveys carried out in developing countries aim above to compensate for the shortcomings of the civil registration system. They therefore measure one or all of the phenomena that contribute to population growth.
2. **Opinion or behavioral surveys.** Other surveys have a main objective of knowing and understanding the motivations, aspirations, and practices of individuals or couples, who condition the evolution of demographic phenomena.
3. **Mixed surveys.** However, the most common case today is that of surveys with two simultaneous objectives. For example, we wish to measure not only

[42] However, some authors prefer to classify surveys on employment or schooling within the field of demography (Desabie, 1965, p. 30–31).

fertility and its intermediate variables, but also to discover why or how some practices (breast-feeding, postpartum sexual abstinence, contraception, for example) exist and evolve, or to discover the fertility plans and aspirations of couples or women. They are particularly interested in the interaction between factors, in social inequality, and in geographic disparities. In short, the objective is focused greatly on explanatory research.

The term *population survey* consequently covers extremely diverse operations both at a technical level and at a thematic level. First of all, with regard to objectives, Table 121–4 proposes a general typology of the most common surveys, with the main characteristics of each category, i.e. the method of observation, geographic coverage, and sample size.

Having given the population survey a rather broad definition, I use the term *household surveys* here for what many demographers call *population surveys*. Common in developing countries, these are often surveys with multiple objectives: measuring the sociodemographic composition of the population and its dynamics (fertility, mortality, migration). In the past, their aim was to replace both the civil registration system and the census. Nowadays, they are mainly carried out in the middle of the intercensus period. They require large samples (often around 30,000 households).

4. **Fertility** was, by far, the most studied phenomenon (due to greater funding), notably in the context of large international programs: the Knowledge, Attitude and Practice (KAP) program of the 1960s, the World Fertility Survey (WFS) program of the 1970s, the Demographic and Health Surveys (DHS) program of the 1980s, 1990s, and 2000s, for developing countries, or the American program of contraceptive prevalence surveys of the 1980s. These surveys obviously vary from one program to another, but they are all national and are all based on samples of 3,000 to 8,000 women with a predominant objective: fertility and family

TABLE 121–4 Typology of the Most Common Population Surveys

Survey Category	Observation Method[1]	Geographic Coverage	Sample Size
Multiple objectives household surveys (numerous developing countries since the 1960s)	1 round or 3 rounds (over 1 year)	National	20,000 to 40,000 households
Fertility/child health surveys (WFS program of the 1970s, DHS program of the 1980s, 1990s, and 2000s; and the GCHS for the Gulf countries; ±210 in total)	1 round	National	4,000 to 8,000 women (1,500 to 3,000 men sometimes)
Fertility/family surveys (FFS program; 24 developed countries, from 1990 to 1999)	1 round	Most frequently national	3,000 to 6,000 women; 1,000 to 4,000 men
Contraceptive prevalence surveys (e.g., CDC program; 30 in total since 1985)	1 round	National	2,000 to 3,000 women (and men sometimes)
Migration/rural migration/urbanization survey (e.g., REMUAO Project in West Africa, 1990)	1 round	National or regional	Varies
Child mortality survey (e.g. IFORD Project in Africa, 1980s)	Multiround (over a period of 1 to 3 years)	Local (frequently cities)	10,000 to 15,000 births
Pregnancy and Childhood survey (WHO-Europe ELSPAC Project, 1990s)[2]	Multiround (from birth up to age 3 or 7 years)	Regional	±5,000 pregnancies
Survey on birth (e.g., INSERM-France)	1 round	National	5,000 to 6,000 births
Biographical surveys (migratory, occupational, and family itineraries; e.g., INED's 3-B survey, RIVASS Project in the Sahel)	1 round	National or cities	3,000 to 6,000 individual adults

DHS = Demographic and Health Surveys; ELSPAC = European Longitudinal Study of Pregnancy and Childhood; FFS = Fertility and Family Survey; GCHS = Gulf Child Health Survey; INSERM = Institut national de la santé et de la recherche médicale; REMUAO = Network of Migration and Urbanization in West Africa; RIVASS = Research on Integration into Cities in Africa South of the Sahara; WFS = World Fertility Survey.

[1] The characteristics presented are those most frequently encountered. They do not exclude other alternatives.

[2] The ELSPAC surveys from the beginning of the 1990s are at the limits of demography as their objectives are broader: the health and the development of the child.

planning—even through the most recent project (the DHS survey) is also interested in child health. The European Fertility and Family Surveys (FFS) project of the 1990s (on 24 countries) also deals with fertility but is mainly focused on the family.

Neither *migration* nor *mortality* was the object of such attention. They were the subject of only purely national operations, or at best, some regional programs, such as those on migration in West Africa (REMUAO Project of the beginning of the 1980s). The majority of these operations are only concerned with one region of a country, or often even one city (the capital); very few countries, such as Tunisia in 1989, have carried out a national survey on infant mortality.

Although it is a major factor of fertility, *nuptiality* is the poor relation in demography, in developed countries as well as in developing countries. It is the subject of very few specific surveys. We can certainly approach it in fertility surveys through the reconstitution of the marital life of each woman. The questionnaire must enable this, as was the case of the WFS, but not, unfortunately, that of the DHS.

5. **Biographical or life course surveys**, a more recent and innovative approach, have a specific category in Table 121–4 because they integrate, in the full meaning of the word, migratory, residential, family (fertility and nuptiality included), and occupational histories. Some good examples of these are the 3B (triple biography), a pioneering survey carried out by INED in 1981,[43] the urban insertion surveys in Africa (Dakar 1991, Bamako 1992, Yaoundé 1996, and Antananarivo 1998),[44] and finally, the survey carried out in 2000 in Burkina Faso (the first at a national level) on migration, urban insertion, and the environment. All these surveys aim to understand and measure the interactions between the various aspects of the family, migratory, and occupational lives of the respondents.[45]

[43] *Family, occupational, and migratory biography* survey on a national sample of 4,602 persons aged from 45 to 69 years (Courgeau and Leliévre, 1989, 1990). The length of each interview was between one and two hours (one hour and ten minutes on average).

[44] The Dakar survey was carried out by Institut Fondamental d'Afrique Noire (IFAN) and ORSTOM, that of Bamako by Centre d'Etudes et de Recherche sur la Population pour le Développement (CERPOD) and the Department of Demography of the University of Montreal. The procedures were similar and the results were relatively comparable. The Bamako survey consisted of three questionnaires: household, biography, and health. The biographical questionnaire was made up of no less than 6 modules: "family origin and childhood," "migratory and occupational life in Bamako," "residence of more than six months in Bamako," "active life in Bamako," "married life," and "live-born children." It covered 2,141 individuals aged 25 to 54 years.

[45] Concerning the advantage of these surveys with regard to analyzing the interaction between phenomena in Africa, see: Antoine, 2003; Antoine and Dial, 2003.

3. Observation Methods and Types of Survey

In any collection of information on demographic events (births, deaths, migration, marriages, illnesses, etc.), on practices (vaccinations, attendance of health centres, contraception, etc.), or on the evolution of certain individual characteristics (marital status, occupational situation, level of education), the observation method can be retrospective or prospective and the operation can be specific or permanent.

1. **Retrospective or prospective survey.** In a *retrospective* survey, individuals are questioned on what they have experienced over the last weeks, the last year (short reference period, Fig. 121–2a), or throughout their lives since a given age or a given event (Fig. 121–2b). They are generally single-round surveys. In a *prospective* survey, the same population sample (individuals, households, or geographic zone) is monitored over a period of varying duration by visiting them[46] at regular intervals to record changes that have occurred since the last visit (Figs. 121–2c and 2d). The latter are called *multiround surveys* (or sometimes *panel surveys* when individuals are the observation units).

2. **Ad hoc or permanent survey.** A survey is called *ad hoc*, even if it is repeated (in that case, it can be called a *trend survey*), when each observation is disassociated from the others and no individual monitoring is possible. It is only at the aggregate level that evolution from one point to another can be observed (Fig. 2e). For example, this is what occurs when the results of surveys of the same nature that were carried out successively are compared, e.g., the two or three DHSs of a given country. On the contrary, the survey is called *permanent* if a population (frequently a group of villages or districts) is regularly monitored (through a multiround survey or even continuous observation) for a long period (up to several dozen years). This is the basic principle of demographic surveillance systems or population laboratories discussed earlier.

Let us now look more closely at the retrospective single-round survey and the multiround survey to discuss their advantages and disadvantages.

4. The Single-round Retrospective Survey

When visiting the sample only once, we should necessarily ask questions concerning the past if we wish to observe changes in the population and its compo-

[46] Or, as is the case now in most developed countries, by post or by telephone.

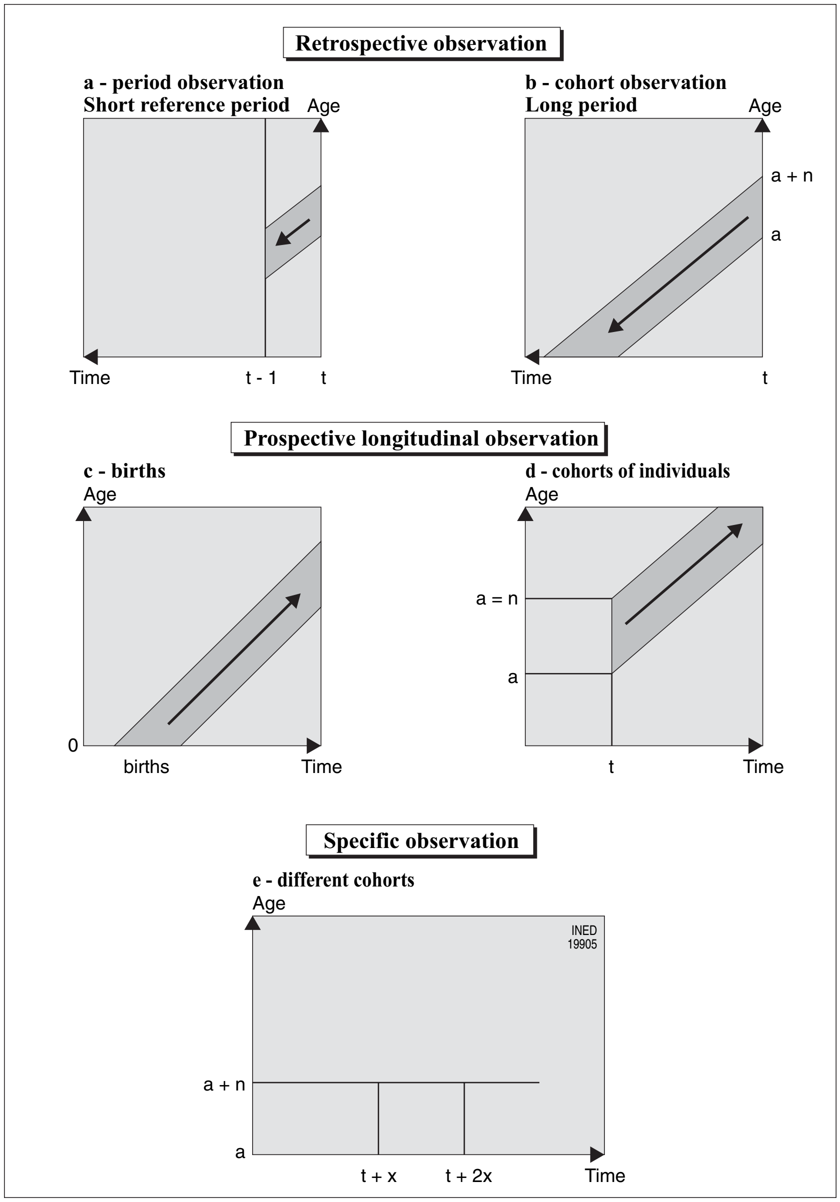


FIGURE 121–2 Schematization of the different types of observation.

nents. But we should distinguish two possible strategies with different results according to whether one short and recent reference period is sufficient, or whether we wish to trace all the demographic events experienced by the respondents.

a. Retrospective Surveys over Short Reference Periods

To grasp events such as births, deaths, marriages, and illnesses through a retrospective survey,[47] we should question individuals on what occurred during a reference period. Here, it is a short and recent period. Regarding births and deaths, for different reasons (annual rates, suitable numbers, eliminating seasonal effects), it is most common to ask about events in the 12 months preceding the interview; regarding illnesses, this can be, according to the case, between a few days to a few weeks or months. The method of the last 12 months, frequently used in household surveys (and censuses) in developing countries, is *simple* at the conceptual and analytical levels (it leads to calculations of the period indicators, rates, or incidence), and *relatively inexpensive* (it only requires a few questions),

[47] Let us recall that the last 12 months are also much used in censuses in developing countries for estimating fertility and mortality.

but, as has been demonstrated many times (Vallin, 1969; Tabutin, 1984b, 1984a; Arretx, 1984), poses many problems. First of all, the method requires large samples (from 20,000 to 40,000 households, Table 121–4) to counteract the scarcity of demographic events. Furthermore, it is then susceptible to considerable sources of bias, whether from a *telescoping* effect, due to errors in localizing events in time (the respondent wrongly includes or excludes an event in the period), or that of omissions (conscious or not). These errors, which are sometimes modest for natality, are frequently enormous for mortality, sometimes making the results unusable.[48] I have been asking myself for a long time why, in spite of such problems, such a method is still being used, notably for mortality and for censuses! Attempts have been made to reduce the errors of the last-12-months method, notably those which are due to the memory of the respondents, e.g., by using a fixed, easily identified date (a memorable and popular festival, for example) as a starting point for the reference period, but this has not totally solved the problem (Tabutin, 1984a).

b. Retrospective Surveys over Long Reference Periods

These errors are much easier to manage in long retrospective surveys and make it possible to better ensure the coherence of the information collected on the same individual. All the respondents are questioned on their experiences since a given change in status (marriage, motherhood, residence in a community, unemployment, for example), since reaching a given age, or since birth. In other words, from an event or a starting age, respondents entirely reconstitute their reproductive lives, matrimonial course, or migratory, residential, educational, occupational experiences, and sometimes even several of these components or all at the same time, as is the case in biographical surveys. For example, this is the strategy used for the majority of fertility surveys (women are interviewed on their reproductive and marital lives from the age 15 or from their first union formation) or in migration surveys (for example, men aged 40 to 65 are interviewed on their history since their first migration or since leaving school). It is a *demanding* process both for the respondent and the interviewer (many dates, durations, causes), but *rich in information*. For example, the complete reproductive histories of women aged 15 to 49 years (including questions on the survival of each child, or on the duration and result of every pregnancy) makes it possible to reconstitute the fertility trend over 15 years or infant and child mortality over the 20 or 25 years preceding the survey. The results can obviously be analyzed by period, but they lend themselves quite naturally to longitudinal analysis, even if the duration of individual observation varies from one respondent to another according to age at the time of the survey.[49]

This approach also has its *disadvantages and risks of error*. Retrospective observation by definition operates a selection in the cohorts: we can only interview individuals who are still alive and who have not emigrated (we must therefore make a hypothesis of independence between the phenomenon studied and other events excluding the individual from the population). The long retrospective survey also calls upon the *memory* and recollection of an event that is sometimes far in the past. Inaccuracies (in the dating of events, notably) and omissions are likely to increase with the amount of time past of the event. The method therefore may distort the observed trends, for example, underestimating the scale of a reduction in fertility or infant mortality. These recall lapses arise differently according to values and perception of time but they exist in every social context, in both developed and developing countries (Courgeau and Leliévre, 1989, p. 17–19). Finally, the *duration of the interview* is long, particularly in multiple biographical surveys (1 to 2 hours).

5. Multiround Surveys

These were developed during the 1960s, particularly in Francophone countries, in response to problems (omissions and inaccuracies) in retrospective observation, notably over the last 12 months. They very rapidly attracted supporters and opponents due to the fact that, while they certainly have many advantages, they also pose significant problems. A multiround survey[50] is *a survey where following an initial complete observation, the same population sample is interviewed again several times and at short time intervals*. As in the other demographic surveys, it can involve a sample of individuals, households, dwellings, or geographic zones, but also, in some cases, a sample of events (pregnancies, births, marriages).

[48] There are of course correction methods, but they have their limitations. They are notably based on considerable hypotheses (constancy of omissions according to age or social group, for example).

[49] On the *truncation effects*, see Tabutin, 1984b, p. 120–124; Courgeau and Lelièvre, 1989, p. 21–22. In fact, there is generally double truncation: on the *left* as the account only begins at a given moment in the life cycle and we have no information on the preceding period, and on the *right* with the interview date.

[50] Which is also sometimes called prospective survey or follow-up survey.

a. Monitoring Households

Multiround household surveys carried out since 1965, particularly in Africa or Asia (20 in total, most of them national),[51] aim above all to correctly measure population change and avoid the disadvantages of the last-12-months method mentioned earlier. The reference period is determined by the time passed between the first and last interview. In this case, the area sample (based on districts of around 500 inhabitants, for example) is preferable (to better manage the problem of sample ageing), but samples of dwellings (households) have also been used. Most frequently, the sample is visited 3 or sometimes 4 times a year. At the first visit, a census provides an initial inventory on the basis of which any changes that occurred since the last visit can be observed, notably arrivals (births, immigrations) and departures (deaths and emigrations). The area sample makes possible the observation of collective migration (entire households), not only at departure but also on arrival (Tabutin, 1984b; United Nations, 1992a). Indeed, this method considerably reduces the risk of omissions[52] and significantly increases the precision of the observation. However, unless the period of reference and therefore the number of visits (and the cost) are increased, the samples should be as large as they are for retrospective surveys over a short reference period (between 20,000 and 40,000 households). The cost is high due to the number of visits. The duration of the operation is relatively long (at least 18 months of collection, if 6 months on the ground for each of the three visits are counted). Preparation of data and analysis are complex. The risk of total failure (when fieldwork is stopped) or partial failure (when the analyses that enable full advantage to be taken of data is abandoned) is not negligible. In a period when resources are increasingly rare and in view of certain failures in developing countries, the number of skeptics has increased and includes the author of this chapter who was however formerly in favor of this technique, which is full of potential.

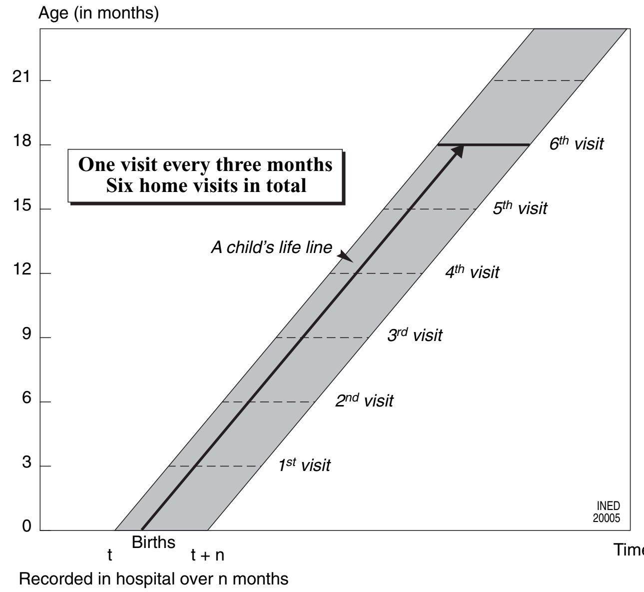


FIGURE 121–3 Monitored observation of a sample of births over 18 months of age.

b. Monitoring Individuals or Births

Multiround surveys can have more specific objectives and resort to monitoring individuals, either since their age of arrival to the population exposed to the risk (e.g., 18 years, if we wish to study marriage, or 1 to 4 years for the study of child morbidity and health) or since an original event (i.e., school entry, first visit to a family planning, Figs. 121–2c and 2d; or since their birth, Fig. 121–3).

After a first visit, when the interview pertains to a fundamental event, the past, or on the family or social environment, the individuals concerned are visited at regular intervals. Comparison of the situation at various visits and questions on what has occurred between two interviews[53] makes it possible to measure and understand the changes. The total duration of observation, the number of visits, and the interval between visits are all variable from one survey to another, according to the objective (from a few weeks for perinatal mortality, to 5 or 10 years for monitoring the health and nutrition of children, or conditions of access to marriage) and, of course, according to the human and financial resources available.

[51] To my knowledge, the most recent was that of Morocco in 1987–1989 (two years of observation, five rounds at six month intervals).

[52] There is certainly a risk of omission for two contrary events that are monitored over a very short interval, e.g. an entry followed rapidly by a departure by migration, or a birth followed by a death at a very young age. The individual could not be identified during the preceding visit and is no longer there during the current visit. For migration, the issue can be resolved by defining the permanent residence according to a duration of presence longer than the interval between visits. For death at a young age, the situation is greatly improved if the pregnancy was declared during the previous visit. More generally, it is also possible for the initial itinerary to be incorrect (omission or under-enumeration) but the later visits often make it possible to identify and correct them.

[53] In developing countries, the method always used is that of the interview, while in developed countries questionnaires sent by post can be used as well as interviews over the telephone.

The most common example in demography, at least in developing countries, is monitoring a *sample of births* in order to study childbirth, characteristics at birth, mortality, morbidity, and maternal behavior with regard to health and nutrition (vaccination, breast-feeding). For example, a sample of childbirths or newborn infants (recorded in maternity wards or in the civil registration system, where it works efficiently) are regularly visited at home. In the *Les enquêtes sur la mortalité infantile et juvénile* (EMIJ) program (surveys on infant and child mortality) carried out by Regional Institute for Demography Training and Research (IFORD) in the 1980s in five African capitals (Courbage, 1979), samples of around 12,000 births were selected in maternity wards and then observed during seven home visits over 2 years. A similar scheme was adopted for the EMIS programme (surveys on infant mortality in the Sahel). Numerous similar surveys, although generally less ambitious, were carried out in Latin America and Asia.

Theoretically perfect, the prospective observation method, which indeed generally leads to better results than the retrospective short-period method, is however not without its methodological and operational problems.

c. The Limitations and Problems of Prospective Surveys

Without going into detail here, there are two different types of problems (Tabutin 1984a, 1984b; Riandey, 1988; van de Walle, 1988): some are inherent to the procedure, others are linked to the field and to analysis.

The major problem of any multiround monitoring of individuals (or households) is *loss to observation* through migration[54] and refusal to respond. These sample losses can introduce bias and affect individuals (and households) whose behavior differs from that of the population of respondents. They also lead to the attrition of the observed population, which can affect the precision of the final measurements. This phenomenon of *sample slippage*, towards more sedentary individuals and towards more favored social classes, through the progressive exclusion of the most mobile populations, the poorest, the most marginal, increases with the duration of monitoring. Moreover, the multiplication of observations during successive visits causes *increased problems* at the operational level of the survey (duration of collection, fatigue of respondents and interviewers, cost of the operation), and also at the level of analyzing the results, which is much more complex.

6. Retrospective/Prospective: An Outdated Debate?

During the 1970s and 1980s, supporters and opponents of the prospective survey clashed, with the former stressing the advantages of multiround surveys[55] and the latter stressing its constraints compared to the flexibility and the rapidity of single-round surveys.[56] The major aim was to improve collection, which was then in full expansion in Third World countries.

The debate gradually died out, partly due to lack of combatants and the fact that the problems and methods of collection were no longer of great interest to many, but also because, for financial reasons, retrospective observation was imposed on the large international programs (United Nations and United States). The developing countries are following the movement because they do not have sufficient resources to explore alternatives. Indeed, very few of them have the financial and human resources to carry out a national multiround survey. A good retrospective survey is better than a bad multiround survey.[57] Nevertheless, the retrospective survey cannot be a replacement for a multiround survey: the correct measurement of mortality at different ages, collective migration, and spatial or social inequalities require prospective observation. But this procedure is more costly, longer, and riskier, at least when used at a national level. The retrospective survey is fashionable, financial resources are rare, international pressure is strong, and prudence and security are required (often rightly so). However, the debate should not be considered definitively closed.

VII. EXTRA TOOLS: AERIAL PHOTOGRAPHY AND REMOTE SENSING

For a long time limited to military strategy and disciplines such as geography, geology, and agronomy,

[54] For examples on France, see Riandey 1988; for an example on Africa, see van de Walle, 1988.

[55] Whether at the theoretical level or because of some clear successes in this period, such as the National Statistical Population Study (ENSP) in Algeria in 1970–1971, based on a three-round survey (Vallin, 1975; Tabutin, 1984b).

[56] Indeed, one of the great misunderstandings of this debate was that the supporters of the multi-round survey were fighting against the proven ineffectiveness of retrospective surveys of the last 12 months, while those of the retrospective survey were relying on the wealth of biographical surveys to safeguard the last 12 months.

[57] By this we mean the common under-use of this type of operation.

aerial photography and remote sensing recently entered into the panoply of tools for demographic data collection, notably in the context of urban observation systems and developing countries.[58]

Aerial photography, an old technique which is very common today, can only correctly record, notably due to the curvature of the earth, microsurfaces. Therefore, a large number of photos must be combined in order to obtain comprehensive aerial coverage of a small territory or a city, but it provides many details, and this is one of its advantages. When it is well executed, it makes possible the selection of districts, compounds, and even buildings, and it clearly provides useful reference points (wells, schools, etc.).

Remote sensing by satellite is much more recent[59] and makes it possible to observe much larger zones (a whole city, a region, even a country), to rapidly obtain the required satellite image (in 20 days, for example, with the SPOT system (Satellite Probatoire d'Observation de la Terre) or even to order it for a specific date, and with classification techniques,[60] to rapidly carry out an initial stratification of the city according to the type of vegetation, densification, the type of habitat, or soil. On the other hand, the satellite image does not yet have the precision of an aerial photograph. For example, it is still often difficult to distinguish the limits, function, and nature of buildings, to distinguish a building from a street, or a large market from a residential zone if the roofs resemble each other.

Remote sensing can be of great benefit to demography in the establishment or implementation of a sample frame, in other words, the constitution of a sample of households, which is a thorny and expensive problem in many countries. The satellite image today makes it possible to quickly and cheaply define[61] an aerial sample frame of a city,[62] by proceeding successively to a stratification according to the density of the frame, then to a selection of small enumeration areas (1st degree) and finally on the ground to a selection of households in these small enumeration areas (2nd degree). Research by ORSTOM[63] (now IRD) on Marseille, Quito, and other towns have demonstrated, since the beginning of the 1990s, the effectiveness of this approach (Dureau, 1995b). In addition to the rapidity of implementation of a sampling frame in cities that do not have a traditional frame, "*the selection of a sample by satellite imagery also favors the description of the internal differences of a city and the analysis of the dynamics of the city's subpopulations*" (Dureau, 1995b, p. 284).

Remote sensing is also very useful in studying urban planning and management, for which censuses and traditional surveys are, and will no doubt remain, unsuitable. They make it possible to continuously and precisely observe the morphology of a city: occupation of the ground, vegetation, densification, buildings, etc., even if all the technical problems have not as yet been solved and that there are non-negligible margins of error (Baudot, 1994). The approach needs to be developed for studying rural settlements, at least in nonarboreus zones.

Going even further, researchers in remote sensing are now attempting to indirectly estimate (on the sole basis of data from the images) the number of households and population, and even the size of families, population density, and sometimes even certain socioeconomic characteristics. Will it be possible one day to dispense of all the other collection methods and notably of censuses for obtaining this type of information? This would truly be revolutionary, but it is still far in the future. The few studies that have compared satellite estimates, with results from the ground, show that the margin of error is still very wide (Baudot, 1994).

Remote sensing is no doubt a tool for the future due to the availability of images, the rapidity of their treatment, and their relatively low cost.[64] They increasingly fulfill certain needs for geographic and spatial analysis. In demography, we can expect it to become an essential tool in the construction of samples and the study of settlements (urban and also rural), at least in countries where the landscape, the climate, and the vegetation allow it, but it will not supplant the essential collection of individual data, which is increasingly varied and numerous and which only the traditional sources of demographic information can provide. This goes without mentioning the technology and the competence necessary to interpret and analyze the images from which many developing countries will suffer.[65]

[58] For more information on this subject, see Dureau *et al*., 1989; Dureau and Weber, 1995; Baudot and Wilmet, 1992.

[59] Use of remote sensing by satellite in town planning began with the appearance of the first high-resolution satellites at the beginning of the 1980s.

[60] Complex techniques which relieve specialists of the interpretation of satellite images.

[61] At least with regard to traditional cartography of the terrain.

[62] So far, this has mainly been done in urban zones.

[63] *Office national de recherche scientifique et technique outre-mer*, recently renamed *Institut de recherche pour le développement* (IRD).

[64] A standard satellite image with spatial high resolution (20 m × 20 m) is far less costly ($1,900 for this type of SPOT image in 2003) than complete aerial coverage of a territory (a city, for example).

[65] Technology which, if it were essential, would once more reinforce the scientific dependency of the poorest countries.

VIII. NEGLECTED SOURCES: ADMINISTRATIVE INFORMATION

By administrative system, we refer to *every source of statistical information currently used by an administration (public or private) in a management objective*. The sources that are of interest to us here are obviously those which contain individual data relating to children in schools, mothers or children monitored by health centres, women giving birth in maternity wards, subscribers to pension funds, those paying social security contributions, patients in hospitals, etc. Of the multitude of administrative registers and files, here are some examples:

- school registers
- hospital registers, maternity clinic registers, and registers of maternal and infant health centers
- social security files
- social insurance fund files, pension fund files
- military, police, and judiciary files

Each of these registers, whether managed at a central or local level, computerized or not, contains information (of variable number) on each individual concerned and sometimes on his/her family environment. They can also be an additional source for studying sociodemographic issues, as long as they are up-to-date and of good quality. They are rarely representative of an entire population and it is sometimes difficult to access them (due to statistical confidentiality, for example). These administrative sources are still rarely sought out by demographers.

IX. A FEW WORDS ON QUALITATIVE APPROACHES[66]

Demography, as a social science based on measurement from the beginning, moving more or less towards comprehension, explanation, interaction between factors, cannot avoid the need for more qualitative approaches. In other words, they are observation methods whose main objective is not measurement, but understanding a phenomenon, a process, or a mechanism by placing it in its social context.[67] As Hubert Gérard (1988) writes, "*the qualitative approach appears as a process of questioning and seeking a response, in the same way as the quantitative approach; it can at the same time therefore be autonomous and sufficient, and consequently it must meet the same demands in rigour and quality as the latter, even if, for this, it must necessarily approach it differently.*"

Generally in demography, qualitative observation is considered as a complementary approach to quantitative observation, and takes one of the following three forms.

1. **Structured or semistructured interviews** are practiced on a few dozen target individuals (generally thirty or so), selected logically, with no concern for a strict statistical representation. During these interviews (there can be two or three), freely or semi-freely (if there is an interview guide), the people speak about their vision of the problem under consideration, of themselves, and of their environment. The content of these interviews, which often last several hours, is then analysed.
2. **Interviews of resource persons** which can be individual or collective (focus-group); the individuals are chosen according to their role, experiences, and their competence with regard to the subject under consideration. For example, with regard to fertility, these would be doctors, midwives, traditional midwives, and welfare officers.
3. **Participant or direct observation** is rarely used in demography. By residing somewhere for varying periods of time (from a few months to several years), the observer integrates as much as possible into the district, the village, or community which is the object of the study and notes all the daily events (both individual and collective) likely to clarify the issue under consideration. This type of work and approach relates more to anthropology than demography.

Different qualitative methods can obviously be combined, notably the first two: life histories or semi-structured interviews can indeed be carried out in parallel with focus groups.

Demographers still seldom use the qualitative approach. Some (generally sociologists) take the risk but the majority do not consider it to be of interest or are unaware of it, if they do not challenge it.[68] It does however have a role in our discipline, either as a preliminary to a quantitative survey (identification of issues, problems, variables, or key factors), or as a complement to quantitative analyses for which they provide an explanation for certain facts or certain statistically established relationships. But it is true that

[66] This aspect will be discussed in greater depth in Chapter 136 by Michel Bozon.

[67] For a broader view on qualitative problems and methods in social science and in demography, see *Chaire Quetelet 1985* (Gérard and Loriaux, 1988). See also Chapter 136 hereafter.

[68] Moreover, this is not surprising when we consider the content of numerous training programs in demography, that are essentially if not entirely focused on quantification, number worshipping.

one cannot just suddenly become a qualitative expert. The collection as well as the (complex) analysis of results require a specific and well established methodological arsenal.

X. PRIVACY AND DATA COLLECTION

The deontological issue of protecting privacy is currently in the headlines in North America and Europe, particularly in countries where computerization and the coupling of information systems is very advanced, where individual freedom is of primary importance, and where private (and frequently commercial) statistics meet public statistics. The problem is not a new one (in a number of countries, the first legislation for regulating surveys dates back to the 1950s and 1960s), but it is accentuated by the multiplication of data collection organizations, the increasing number of operations, the diversification of subjects of study, the opening of borders, the power of new information technologies, and the increasing sensitivity of public opinion (Eurostat, 1986; Riandey, 1996; Moreau, 2001).

In addition to the chapter already dedicated to ethics in demography in Volume 7 (Chapter 99, by François Héran), let us return here to the three main elements composing the problem of relations between statistical collection[69] and privacy: the obligation to respond, confidentiality (or statistical confidentiality), and the content of information requested. It will become particularly clear that the seriousness of the problem varies according to the type of survey, but also that every survey is inevitably an intrusion of privacy.

1. The Obligation to Respond

During surveys and even currently during censuses, some people refuse to respond in the name of respecting their privacy, or that of their household, sometimes through fear of future use of the information provided, other times by the simple refusal of an intrusion into their personal lives. The more sensitive the subject (income, sexuality, marginal behavior . . .), the greater the likelihood of refusal. Should a survey be obligatory? Can it even be so, to the extent that a forced consent necessary leads to bias in the information collected? "*The freedom to refuse and the information provided by the respondent are high priorities in the legal texts, the French Law of 1978 (Informatique et Libertés), and the 1981 Convention (of the Council of Europe in relation to protecting privacy)*" (Riandey, 1996, p. 451).

[69] These problems obviously go well beyond demography.

2. Confidentiality or Statistical Confidentiality

Confidentiality and secrecy are practically always guaranteed in a survey, to the individual participating in a census, or the informant in an administrative register. No personal information is provided to third parties (whether commercial or not), to organizations other than the one to which the declaration was made. But the concept of confidentiality today goes well beyond this simple preliminary guarantee. It includes (or should include) the power for every individual to know, control, and, in the extreme, to refuse any future use of the information provided. The scientific surveys, and those on public (or parapublic) funds mentioned here, and traditional censuses do not cause any major problems from this point of view. In Europe and in North America, research organizations assure anonymity and the protection of computerized files. The case of registers is *a priori* a little more complex: how can the use of population registers be controlled at a local level? How can the fusion of administrative files be controlled in countries where there is only one single national identity number? Who can have access, and under which conditions, to the statistical wealth that is a computerized national register? Northern Europe and Canada completely protect their files and there is complete confidence in their statistical organizations (Eurostat, 1986), while "*French jurisprudence tends to divide the identifiers into sectors in a way that prevents the communication of data between administrations*" (Riandey, 1996, p. 444). In 1995 the European Commission adopted a directive relative to "*the protection of persons with regard to the processing of personal data and the free circulation of these data*", which came into force in October 1998. It states the rights of individuals and the duty of those responsible for the files, but it also includes many exemptions. The risk of personal data being transferred to the commercial private sector, "*the business of population data*" (Moreau, 2001), is more than ever an issue in a world of free exchange, of excess marketing, and fierce competition between large commercial companies.

3. The Type of Information

The protection of privacy includes the types of questions included or excluded in questionnaires. The most typical example in demography concerns variables of cultural identity (ethnic group, language, religion), and variables of behavior (sexual life, relation between spouses, etc.). The sensitivity of these questions, which varies greatly according to situation and country, can lead to the self-censorship of researchers in their topics of study and in their questions.

"*All citizens have the right to be protected against any excesses, including the carelessness or incompetence of researchers*" (Leridon, 1993, p. 399). Ultimately, the problem is finding a balance, a working arrangement between the collective and essential protection of individual rights and the need for reliable and varied statistical information. Exceeding the national context, Europe (now including 25 countries), is committing itself little by little,[70] and is trying to harmonize national legislations (that are still very diverse), while at the same time ensuring the free movement of data, two objectives that are difficult to reconcile. Statistical collection and scientific research can only gain from working within a legislative context which is both clear and adapted to the knowledge needs of contemporary societies.

XI. SOME FINAL THOUGHTS

Everyone will agree that observation is a crucial phase of every study, in demography as in all the other social sciences. The results of analysis, descriptive and causal, depend on the nature and quality of data available (administrative registers, civil registration system), or data collected (censuses, surveys). But societies evolve, mentalities change, new issues emerge, technology progresses, and financial constraints are considerable. Are the current information systems still adapted, appropriate, and relevant?

1. New Demands and Needs of Society

For a long time, too long a time undoubtedly, demography was confined to observing and measuring the traditional demographic events of births, deaths, marriages, and internal migration, measuring growth, and making projections on population dynamics. Today it is obliged to widen its field of observation and research, to assure its social credibility and its role in the system of scientific production. It faces the challenge of "*impossible issues*" (Leridon, 1993, p. 390), that are often burning issues, both politically and socially: illegal induced abortion, illegal immigrants, international migration, omissions in censuses, population ageing, etc. It must also deal with the increasing diversification and complexity of individual life courses, notably of family structures and residences. It must become more committed to the interactions between professional activity, migration, and family life. In short, *it must increasingly deal with the new problems of society* and here it is obvious that the old, traditional demographic information systems are inadequate or totally outmoded, without forgetting that social demands are becoming more diverse, from stakeholders (users) that are more numerous now than in the past (public and private sectors, associations, NGOs, local managers, mass media, other sectors of social science research).

2. Technology and Cost of Information: Progress, But . . . ?

In a few years (the 1970s and particularly the 1980s), the spread and progress of computerization has transformed demographic data collection, regardless of method: acceleration of data entry, data management and processing, increase in record linkage, direct linking of one level (e.g. a town) with another (the region, the country). We can even foresee little by little tools for portable and direct data entry on the ground, which will even change the nature of the interviewer's work. But paradoxically, apart from some rich countries, this has not resulted in the hoped for revolution in the waiting period for scientific production, whether it be in the context of civil registration, censuses, or surveys. Progress is being made but we still depend on cumbersome and slow machinery.

The technological progress has not led to a drastic reduction in cost either. In the best cases, the unit cost falls, but not the total cost, which sometimes even increases. The collection of population data remains expensive, as it requires increasingly detailed questionnaires and the samples must be relatively large in order to be representative. With regard to the census, it may soon be considered as a luxury that very few developing countries can afford, without mentioning recent technologies, such as remote sensing which is promising but financially and technically difficult to access for many. A large number of developing countries, notably the poorest ones, are already in total or near total financial dependence.

3. International Recommendations and Programs—Warning: Danger!

One cannot *a priori* deny the usefulness of *international recommendations* in the field of demographic information. Generally proposed by the various specialized organizations of the United Nations, or by regional organizations such as the European Union,

[70] See notably the directive on privacy and the movement of data voted by the European Parliament in July 1995.

they vary from the definitions of concepts to the issues and variables to include within observation systems. The praiseworthy objective is to succeed in generating demographic statistics that are comparable between countries. But by standardizing too much, and by imposing too much (willingly or not), the diversity of local or regional initiatives is reduced, the freedom of choice in each country is restricted, scientific knowledge is definitively impoverished. Without mentioning the fact that reference models and expertise come mainly from developed countries. Regional recommendations do not fundamentally change anything.

One cannot either deny the usefulness for countries, or for the international scientific community, of *large international survey programs*, such as DHS and FFS. Due to the DHS programme notably, in sub-Saharan Africa alone, 69 national surveys in 33 countries (many of them amongst the poorest in the world) were carried out between 1986 and 2003. Without them, knowledge of fertility and child mortality/health in Africa would be lacunary. But every coin has its flip side. Concepts, procedures, measurement, reports: everything is standardized (or almost) at an international level. As Thérèse Locoh (1995) writes about Africa: "*the standardized drafting of reports only leaves a little room for understanding the specificities of a society and makes it possible to standardize the interpretations that are proposed with a view to orienting the actions with regard to population. The necessity for rapid production and of providing a clear message (the need for family planning) results in simplifications that can frequently be criticized.*" Under no circumstances should these programs supplant national operations and research, but the risk is very real in the increasingly numerous countries that are stripped of resources.

4. And the Quality?

Much is discussed in terms of the quantity of information: increasing and diversifying the number of surveys, carrying out censuses more frequently, developing civil registration systems. There is no lack of recommendations, plans, and ideas. And on this aspect some progress has been made all over the world during the last two decades. On the other hand, much less is spoken of the quality of data, notably for example, problems on the ground, the pertinence of questionnaires, the degree of reliability of the results, the margins of error, the procedures for verifying and correcting data, the nonresponses. This is due to the lack of methodological research on data collection, as much in developed countries as in developing countries. Studies of quality and methodology should once again become priority.

5. The Future of Training and Imagination?

In conclusion, it seems to me that the area of information in demography suffers from a lack of imagination and new ideas. It is not a coincidence but a consequence of the fact, already mentioned at the beginning of this chapter, that data collection is the poor relative in demography, due as much to the procedures in force for evaluating research and researchers, as to the directing of programs by funders. Strengthening training in this field seems to me essential if we wish to introduce methodology, a critical attitude, and imagination within the new generation of demographers. In other words, in order to make progress and to be innovative *it is necessary ultimately to restore data collection to the position it held at the beginning of our discipline*.

Acknowledgment

This chapter was translated from French by Accenta Ltd.

References

Affichard Joëlle (ed.), 1977. *Pour une histoire de la statistique.* Paris, INSEE and Economica, vol. 2, 593 + 969 p.

Antoine Philippe, 2003. Nuptialité et conditions de vie dans les villes africaines, *in:* Thierry Eggerickx, Catherine Gourbin, Bruno Schoumaker, Christophe Vandeschrick and Eric Vilquin (eds.), *Populations et défis urbains*, p. 581–604. Louvain-la-Neuve and Paris, Academia-Bruylant and L'Harmattan, 778 p.

Antoine Philippe and Bocquier Philippe, 1995. Le temps et l'analyse des biographies, *in:* Jacques Vallin (ed.), *Clins d'oeil de démographes à l'Afrique et à Michel François*, p. 157–166. Paris, CEPED, 244 p. (Documents et Manuels du CEPED, no. 2).

Antoine Philippe and Dial Fatou Binetou, 2003. *Mariage, divorce et remariage à Dakar et à Lomé.* Dakar, IRD, 22 p. (Paper presented at the *5es Journées scientifiques du réseau Démographie de l'AUF*, Marseille, June 23–26, 2003).

Arretx Carmen, 1984. Comparison between prospective and retrospective demographic surveys to estimate levels and differentials of mortality, the experience of CELADE, *in:* Jacques Vallin, John Pollard and Larry Heligman (eds.), *Methodologies for the collection and analysis of mortality data*, p. 27–48. Liège, IUSSP and Ordina éditions, 510 p.

Auriat Nadia, 1991. Who forgets? An analysis of memory effects in a retrospective survey on migration history, *European Journal of Population/Revue européenne de démographie*, vol.7(4), p. 311–342.

Auriat Nadia, 1996. *Les défaillances de la mémoire humaine. Aspects cognitifs des enquêtes rétrospectives.* Paris, INED and PUF, 204 p. (Travaux et Documents, Cahier no. 136).

Bartiaux Françoise, 1991. *Formation et transformations des ménages des personnes âgées Une analyse par ménage de différents recensements.* Louvain-la-Neuve, Académia, 256p.

Battagliola Françoise, Bertaux-Wiame Isabelle, Ferrand Michèle and Imbert Françoise, 1993. A propos des biographies: regards croisés sur questionnaires et entretiens, *Population*, vol. 48(2), p. 325–346.

Baudot Yves, 1994. *Télédétection aérospatiale et analyse géographique de la population des villes dans les pays en développement.* Louvain-la-Neuve, Université catholique de Louvain, Faculté des Sciences, 356 p. (Doctoral thesis).

Baudot Yves and Wilmet Jules, 1992. Quantification de la croissance urbaine au moyen des satellites à haute résolution: application à Lubumbashi, Marrakech et Ouagadougou, *in:* CRTS (ed.), Marisy '92 Conference Proceedings, *"Evaluation of natural resources with remote sensing techniques / Évaluation des ressources naturelles par télédétection", Rabat 7–9 October 1992*, p. 279–291. Rabat, Centre royal de télédétection spatiale, xii + 370 p.

Boudreau Jean-René, 1989. Mesure des erreurs de couverture du recensement canadien, *Cahiers Québécois de Démographie*, vol. 18(2), p. 265–284.

Clairin Rémy and Brion Philippe, 1996. *Manuel de sondages, application aux pays en développement*. Paris, CEPED, 103 p. (Documents et Manuels du CEPED, no. 3).

Cleland John, 1996. Demographic data collections in less developed countries, 1946–1996, *Population Studies*, vol. 50(3), p. 433–450.

Clignet Rémi (ed.), 1998. *Observatoire du développement, observatoires pour le développement*, Paris, ORSTOM Éditions, 302 p.

Coëffic Nicole, 1993. L'enquête post-censitaire de 1990, une mesure de l'exhaustivité du recensement, *Population*, vol. 48(6), p. 1655–1681.

Committee on Population and Demography, 1981. *Collecting data for the estimation of fertility and mortality*, Washington, National Academy Press, 291 p. (Report, no. 6).

Courbage Youssef, 1979. Enquête sur la mortalité infantile et juvénile à Yaoundé. Présentation méthodologique, *in:* Département de démographie (ed.), *La mortalité des enfants dans le tiers-monde*, p. 107–130. Louvain-la-Neuve and Liège, Département de Démographie and Ordina Editions, 226 p. (Chaire Quetelet 1979).

Courgeau Daniel, 1985. Les enquêtes de migration dans les pays développés, *in:* Département de Démographie (ed.), *Migrations internes, collecte des données et méthodes d'analyse*, p. 53–81. Louvain-la-Neuve, Cabay, 453 p. (Chaire Quetelet 1983).

Courgeau Daniel and Lelièvre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p.

Courgeau Daniel and Lelièvre Éva, 1990. L'approche biographique en démographie, *Revue Française de Sociologie*, vol. 31(1), p. 55–74.

Courgeau Daniel and Lelièvre Éva, 1992. *Event history analysis in demography.* Oxford, Clarendon Press, 226 p.

Das Gupta Monica, Aaby Peter, Garenne Michel and Pison Gilles (eds.), 1997. *Prospective community studies in developing countries.* Oxford, Clarendon Press, 350 p. (International studies in demography).

Département de démographie, 1985. *Migrations internes. Collecte des données et méthodes d'analyse.* Louvain-la-Neuve, Cabay, 458 p. (Chaire Qutelet 1983).

Desabie Jacques, 1965. *Théorie et pratique des sondages.* Paris, Dunod, 461 p.

Duchêne Josianne and Wunsch Guillaume (eds.), 1995. *Collecte et comparabilité des données démographiques et sociales en Europe.* Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Dupâquier Jacques and Dupâquier Michel, 1985. *Histoire de la démographie.* Paris, Librairie Académique Perrin, 462 p.

Dureau Françoise, 1995. Images satellite et démographie: présentation d'une nouvelle méthode de production d'information sur les populations citadines, *in:* Françoise Dureau and Christian Weber (eds.), *Télédétection et systèmes d'information urbains*, p. 221–284. Paris, Anthropos, 379 p. (Villes).

Dureau Françoise, Barbary Olivier, Michel Alain and Lortie Bernard, 1989. *Sondages aérolaires sur image satellite pour des enquêtes socio-démographiques en milieu urbain.* Paris, Editions de l'ORSTOM, 8 p. + 15 sheets. (Collection didactique, 15 instalments).

Dureau Françoise and Weber Christian (eds.), 1995. *Télédétection et systèmes d'information urbains.* Paris, Anthropos, 379 p. (Villes).

Eggerickx Thierry, 1995. Le recensement de la population et la comparabilité internationale des données de migrations internes et internationales en Europe, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 527–552. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Eggerickx Thierry et Bégeot François, 1993. Les recensements en Europe dans les années 1990: de la diversité des pratiques nationales à la comparabilité internationale des résultats, *Population*, vol. 48(6), p. 1705–1732.

Eurostat, 1986. *Protection de la vie privée, informatique et progrès de la documentation statistique.* Luxembourg, Statistical Office of the European Communities, 347 p. (Theme 9, Series C, special issue).

Festy Patrick and Prioux France, 2002. *An evaluation of the Fertility and Family Surveys Project.* New York, Geneva, UNECE-UNPF et INED, 55 p.

François Michel, 1988. L'état civil, *in:* Louis Lohlé-Tart and Rémy Clairin (eds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 91–111. Paris, CEPED, UIESP, IFORD, 330 p. (Les études du CEPED, no. 1).

Garenne Michel and Zanou Benjamin, 1995. L'état civil en Afrique: que peut-on en tirer?, *in:* Jacques Vallin (ed.), *Clins d'oeil de démographes à l'Afrique et à Michel François*, p. 29–42. Paris, CEPED, 244 p. (Documents et Manuels du CEPED, no. 2).

Gendreau Francis, 1993. *La population de l'Afrique, manuel de démographie.* Paris, Karthala and CEPED, 463 p.

Gérard Hubert, 1988. Discours d'ouverture, *in:* Hubert Gérard and Michel Loriaux (eds.), *Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie*, p. 15–52. Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quetelet 1985).

Gérard Hubert and Loriaux Michel (eds.), 1988. *Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie.* Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quetelet 1985).

Gourbin Catherine, 1995. Critères d'enregistrement des naissances vivantes et des mort-nés en Europe, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 219–242. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Gourbin Catherine, 1996. Inégalités sociales en santé et mortalité foeto-infantiles. Les sources nationales disponibles en Europe, *in:* Godelieve Masuy-Stroobant, Catherine Gourbin and Pierre Buekens (eds.), *Santé et mortalité des enfants en Europe. Inégalités sociales d'hier et d'aujourd'hui*, p. 67–96. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 560 p. (Chaire Quetelet 1994).

Gourbin Catherine and Masuy-Stroobant Godelieve, 1995. Registration of vital data: are live births and stillbirths comparable all over Europe?, *Bulletin of the World Health Organization*, vol. 73(4), p. 449–460.

Groupe de Réflexion sur L'Approche Biographique (GRAB), 1999. *Biographies d'enquêtes. Bilan de 14 collectes biographiques.* Paris, INED, IRD, Réseau Socio-Economie de l'habitat, PUF, 336 p. (Méthodes et Savoirs).

Gubry Patrick, Lefranc Christophe and Vallin Jacques, 1996. Sauver les recensements africains, *La Chronique du CEPED*, no. 22, p. 1–4.

Guillaume Pierre and Poussou Jean-Pierre, 1970. *Démographie historique.* Paris, Armand Colin, 409 p. (Collection U).

Guillard Achille, 1855. *Éléments de statistique humaine, ou démographie comparée.* Paris, Guillaumin, 376 p.

Hecht Jacqueline, 1977. L'idée de dénombrement jusqu'à la Révolution, *in:* Joëlle Affichard (ed.), *Pour une histoire de la Statistique, Vol. 1: Contributions*, p. 21–81. Paris, Economica and INSEE, 593 p.

Henry Louis, 1963. Réflexions sur l'observation, *Population*, vol. 18(2), p. 233–262.

Hertrich Véronique, 1997. Les réponses des hommes valent-elles celles des femmes? Une double collecte sur les questions génésiques et matrimoniales dans une population du Mali, *Population*, vol. 52(1), p. 45–62.

Hogan Howard, 1993. The 1990 post-enumeration survey: operation and results, *Journal of the American Statistical Association*, vol. 88(423), p. 1047–1060.

Höhn Charlotte, 1981. Les différences internationales de mortalité infantile: illusion ou réalité?, *Population*, vol. 36(4–5), p. 791–816.

Institut National d'e'tudes d'emographiques (INED), 1999. Données et enquêtes sensibles, *Population*, vol. 54(2), p. 225–330 (5 articles).

Institut National d'e'tudes d'emographiques (INED), 2000. Dossier: un projet de réforme des recensements français, *Population*, vol. 55(4–5), p. 803–829 (3 articles).

Institut National d'e'tudes d'emographiques (INED), 2001. Dossier: Méthode de collecte et d'analyse pour les enquêtes, *Population*, vol. 58(6), p. 959–1072 (5 articles).

Institut National de la Statistique et des Etudes (INSEE), Département de démographie, 2000. Présentation du recensement rénové, *Population*, vol. 55(4–5), p. 803–812.

Langevin Bernard and Begeot François, 1995. Comparabilité et synthèse des données européennes: l'expérience d'Eurostat, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 43–64. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Laroche Benoît, 1993. The future of population census, *in:* IUSSP (ed.), *International Population Conference, Montreal 1993*, vol. 3, p. 151–155. Liège, IUSSP, 495 p.

Leridon Henri, 1993. Informer, s'informer: le démographe et la population, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European Population. II Demographic Dynamics*, p. 387–402. Paris, INED, John Libbey Eurotext, 525 p.

Linder Forrest and Moriyama Iwao (eds.), 1984. *Improving civil registration.* Maryland, International Institute for Vital Registration and Statistics, 242 p.

Locoh Thérèse, 1995. Parlons des EDS: l'analyse statistique est-elle neutre?, *in:* Jacques Vallin (ed.), *Clins d'oeil de démographes à l'Afrique et à Michel François*, p. 143–155. Paris, CEPED, 244 p. (Documents et Manuels du CEPED, no. 2).

Locoh Thérèse and Omoluabi Elizabeth, 1995. Où sont donc passés les 30 millions de nigérians manquants?, *in:* Jacques Vallin (ed.), *Clins d'oeil de démographes à l'Afrique et à Michel François*, p. 57–75. Paris, CEPED, 244 p. (Documents et Manuels du CEPED, no. 2).

Lohlé-Tart Louis and Clairin Rémy (eds.), 1988. *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique.* Paris, CEPED, UIESP, IFORD, 330 p. (Les Études du CEPED, no. 1).

Lohlé-Tart Louis and Francois Michel, 1999. *État civil et recensements en Afrique francophone.* Paris, CEPED, 555 p. (Documents et Manuels du CEPED, no. 10).

Martelli Cristina, 1999. Information systems for a complex approach to demographic analysis, *in:* Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie*, p. 117–135. Louvain-la-Neuve, Paris, Academia-Bruylant, L'Harmattan, 670 p. (Chaire Quetelet 1997).

Masuy-Stroobant Godelieve, 1994. La mortalité infantile en Europe et au Canada. Un problème résolu?, *Cahiers Québecois de Démographie*, vol. 23(2), p. 297–340.

Meslé France, 1995. L'enregistrement des causes de décès en Europe, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 401–431. Louvain-la-Neuve, Paris, Academia/L'Harmattan, 671 p. (Chaire Quetelet 1991).

Moreau Françoise, 2001. *Commerce des données sur la population et libertés individuelles.* Paris, INED, p. 22. (Dossiers et recherches, no. 91).

Nations unies/United Nations, 1958. *Manuel des méthodes de recensement de la population. Volume I. Aspects généraux d'un recensement de la population.* New York, United Nations, Statistical Office, vi + 180 p. (Études méthodologiques, Série F, no. 5, rev. 1).

Nations unies/United Nations, 1970. *Manuel des méthodes de recensement de la population et de l'habitation.* New York, United Nations, xi + 166 p. (Études méthodologiques, Série F, no. 6).

Nations unies/United Nations, 1980. *Principes et recommandations concernant les recensements de la population et de l'habitation.* New York, United Nations, 130 p. (Document ST/ESA/STAT/SER.M/67).

Nations unies/United Nations, 1985. *Manuel de statistiques de l'état civil, vol. II: Étude des pratiques nationales.* New York, United Nations, 115 p. (Études Méthodologiques, Série F, no. 35).

Nations unies/United Nations, 1990. *Principes et recommandations complémentaires concernant les recensements de la population et de l'habitation.* New York, United Nations, 195 p. (Document ST/ESA/STAT/SER.M/67/Add/1).

Nations unies/United Nations, 1991. *Tendances et thèmes nouveaux des recensements de la population et de l'habitation.* New York, United Nations, 41 p. (Études méthodologiques, Série F, no. 52, Document ST/ESA/STAT/SER.F/52).

Nations unies/United Nations, 1992a. *Les enquêtes de suivi pour la mesure de la fécondité, de la mortalité et de la migration.* New York, United Nations, 167 p. (Études Méthodologiques, Série F, no. 41, Document ST/ESA/STAT/SER.F/41).

Nations unies/United Nations, 1992b. *Manuel des méthodes de recensement de la population et de l'habitat.* New York, United Nations, 2 volumes, 164 p. and 182 p. (Document ST/ESA/STAT/SER.F/54).

Ouedraogo Dieudonné and Piché Victor (eds.), 1995. *L'insertion urbaine à Bamako.* Paris, Karthala, 207 p.

Pauti Anne, 1992. Les enjeux du recensement américain de 1990, *Population*, vol. 47(2), p. 468–477.

Pison Gilles, Desgrées du Loû Annabel and Langaney André, 1997. Bandafassi: a 25 year prospective community study in rural Senegal (1970–1995), *in:* Monica Das Gupta, Peter Aaby, Michel Garenne and Gilles Pison (eds.), *Prospective community studies in developing countries*, p. 253–275. Oxford, Clarendon Press, 350 p. (International Studies in Demography).

Pison Gilles, Gabadinho Alexis and Enel Catherine, 2001. *Mlomp (Sénégal): niveaux et tendances démographiques 1985–2000.* Paris, INED, 181 p. (Dossiers et recherches, no. 103).

Poulain Michel, 1995. Mesurer les migrations à l'aide d'un registre de population: quelle comparabilité européenne?, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 553–571. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

Poulain Michel, Riandey Benoît and Firdion Jean-Marie, 1991. Enquête biographique et registre belge de population: une confrontation des données, *Population*, vol. 46(1), p. 65–88.

Poulain Michel, Riandey Benoît and Firdion Jean-Marie, 1992. Data from a life history survey and the Belgian population register: a comparison, *Population. An English Selection*, vol. 4, p. 77–96.

Redfern Philip, 1987. *A study on the future of the census of population: alternative approaches.* Luxembourg, Statistical Office of the European Communities, 265 p. (Theme 3: Population and social conditions, Series C: Accounts, Surveys and Statistics).

Riandey Benoît, 1988. La qualité du suivi des échantillons dans les enquêtes démographiques, un bilan, *Population*, vol. 43(4–5), p. 829–854.

Riandey Benoît, 1989. Répertoire des enquêtes démographiques: bilan pour la France métropolitaine, *Revue Française de sociologie*, vol. 30(2), p. 279–298.

Riandey Benoît, 1996. Démographie, statistique et vie privée, *Population*, vol. 51(2), p. 439–456.

Saboulin (de) Michel and Seibel Claude, 1993. L'avenir de la collecte de l'information démographique, *in:* Alain Blum and Jean-Louis Rallu (eds.), *European Population. II Demographic Dynamics*, p. 403–420. Paris, INED, John Libbey Eurotext, 525 p.

Tabutin Dominique, 1984a. Comparison of single and multi-round surveys for measuring mortality in developing countries, *in:* Jacques Vallin, John Pollard and Larry Heligman (eds.), *Methodologies for the collection and analysis of mortality data*, p. 11–25. Liège, IUSSP, and Ordina Éditions, 510 p.

Tabutin Dominique, 1984b. *La collecte des données en démographie. Méthodes, organisation et exploitation.* Liège, Ordina Editions, 257 p.

Théodore Gérard, 1985. Similitudes et différences dans la méthodologie des recensements de population dans les pays industrialisés et les pays en développement, *in:* IUSSP (ed.), *International Population Conference, Florence 1985*, vol. 4, p. 103–117. Liège, IUSSP, 555 p.

Union Internationale pour l'Etude Scientifique de la Population (UIESP/IUSSP), 1981. *Dictionnaire démographique multilingue (Français: 2nd édition par Louis Henry).* Liège, Ordina Éditions, 180 p.

United Nations, 1969. *Methodology and evaluation of population registers and similar systems.* New York, United Nations, 80 p. (Studies in Methods, Series F, no. 15).

United Nations, 1973. *Principles and recommendations for a vital statistics system.* New York, United Nations, 135 p. (Document E.73.XVII.9).

United Nations, 1979. *The development of integrated data bases for social, economic and demographic statistics.* New York, United Nations, 80 p. (Studies in Methods, Series F, no. 27).

United Nations, 1991. *Emerging trends and issues in population and housing censures.* New York, United Nations, 110 p. (Studies in Methods, Series F, no. 52).

Vallin Jacques, 1969. Le temps facteur d'omission dans les enquêtes rétrospectives, *Population*, vol. 24(3), p. 548–549.

Vallin Jacques, 1975. L'enquête à passages répétés, un outil efficace: l'exemple algérien, *in: Actes du IIe colloque de démographie maghrébine, Oran Les Andalouses, 21–25 avril 1975*, p. 79–96. Oran, Secrétariat d'État au Plan, 639 p.

Vallin Jacques, D'Souza Stan and Palloni Alberto (eds.), 1988. *Mesure et analyse de la mortalité: nouvelles approches.* Paris, INED, PUF, 460 p. (Travaux et Documents, Cahier no. 119).

Vallin Jacques, D'Souza Stan and Palloni Alberto (eds.), 1990. *Measurement and analysis of mortality: new approaches.* Oxford, Clarendon, 415 p. (International studies in demography).

Vallin Jacques and Meslé France, 1988. *Les causes de décès en France de 1925 à 1978.* Paris, INED, PUF, 608 p. (Travaux et Documents, Cahier no. 115. Publication accompanied by 7 appended volumes).

Vallin Jacques and Meslé France, 1998. Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925 à 1993, *in:* Gérard Pavilion (ed.), *Enjeux des classifications internationales en santé*, p. 113–156. Paris, Éditions INSERM, 220 p. (Questions en santé publique).

van de Walle Étienne, 1988. Avantages et limites des enquêtes IFORD: le cas de Bobo-Dioulasso, *in:* Jacques Vallin, Stan D'souza and Alberto Palloni (eds.), *Mesure et analyse de la mortalité, nouvelles approches*, p. 33–45. Paris, INED, PUF, 460 p. (Travaux et documents, Cahier no. 119).

Verhoef Rolf and van de Kaa Dirk, 1987. Population registers and population statistics, *Population Index*, vol. 53(4), p. 633–642.

Vilquin Éric, 1983. Théorie et pratique des dénombrements français au XVIIe siècle, *Revue d'Histoire Economique et Sociale*, vol. 55(3–4), p. 325–358.

Wagner Michael, 1995. Problems of measurement and comparability in migration survey research, *in:* Josiane Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 578–592. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).

C H A P T E R

122

Internal Migration: Observation and Concepts

BRIGITTE BACCAÏNI

Institut national d'études démographiques (INED), Paris, France

Migration is undoubtedly the demographic event that poses the most challenges for observation and conceptualization. It assumes a double reference, both spatial and temporal: a residential migration is a move between two places situated a certain distance apart and in which the duration of residence has been long enough for the move not to be considered temporary. These criteria are not entirely clear, nor are they unanimously accepted (see Chapter 22 of Volume I). Short-distance residential mobility, as between two dwellings in the same municipal district or same city, for example, is not always considered migration. Use of the term *residential migration* assumes also that a place of residence or successive places of residence are clearly identifiable, which can be problematic for certain population categories or in the case of multiple residence.

The sources of observation are disparate and rarely free of defects, and are themselves at the origin of assorted concepts and measures, often of poor comparability. Measures of mobility can vary even within the same country, depending on the source of observation used. At the international level, the harmonization of sources and of measurement methods for internal mobility remains an as-yet unrealized goal.

For several decades, however, migratory movement, rather than natural change, has been responsible for the differentials in population growth between subnational areas, especially at finer spatial scales and for this reason deserves particular attention. The contrasts in growth between rural and urban areas, between city centers and suburbs, between economically dynamic and declining regions, result primarily from migration exchanges. More importantly, the internal contrasts of the natural balance are themselves often merely a consequence of migration. With levels of fertility and life expectancy tending to uniformity at the national level, the disparities observed in birth and death rates arise primarily from differences in age structure. Now the contrasts in age structure between regions or between different neighborhoods of a conurbation originate in the age composition of the migration flows, which vary widely according to the zones of origin and destination. A region such as the Ile-de-France, attractive for young people who come to study or to look for a first job, but strongly propulsive for other age groups, enjoys a strongly positive natural balance thanks to migration. Somewhat paradoxically, therefore, the relatively fast growth of the Ile-de-France population is driven by the migration flows affecting the region despite its own balance being negative (Baccaïni, 2001).

Migration not only contributes to changing the spatial distribution of the population inside a national territory, it modifies the sociodemographic structures of the spatial subdivisions that make up this territory.

An initial definition of *internal migration* is a permanent change of residence inside the boundaries of a country. It is therefore a fundamentally spatial event, one that is observed at a variety of scales, stretching from a simple change of address within the same municipal district or *commune* (the smallest administrative unit in France), to a long-distance move to another region. These movements are very different in nature, and so necessarily are the concepts associated with them.

Concepts evolve also in response to social change and changing lifestyles. New concepts develop when new kinds of migration flows emerge, like those involved in the processes of suburbanization (periurbanization) or of counter-urbanization, for example.

Section 1 of this Chapter shows how the conceptualization of internal migration is related to different modes of approach and research questions, still largely bound to disciplinary divides. The question of the shortcomings in sources for internal migration, and hence in its observation, is addressed in Section 2.

I. DIVERSITY IN APPROACHES AND CONCEPTS

From the start of this discussion, we exclude the various journeys undertaken daily or less frequently but that are not associated with changes of residence—journeys by commuters and holiday makers, for example—and confine attention to residential migration. Two main approaches are then possible. According to whether the focus of study is the territory (spatial analysis of migration) or the individual (demographic or sociological analysis of migration), the concepts employed are different. In the economic analysis of migration, the main emphasis can be on the territory (spatial economics) or on the individual (microeconomics), depending on the case.

1. Internal Migration as a Revealing Factor in Spatial Interaction

Spatial analysis of migration approaches the migratory phenomenon in terms of flows that create interactions between distinct places. Attention focuses not on the individual migrant but on the territory and on the spatial interaction that migration can help to create or reveal. The term "flow" here denotes the "*measurement of spatial interaction obtained by summing the individual moves or the material or immaterial quantities exchanged between two geographical zones for a given length of time*" (Pumain and Saint-Julien, 2001). Geographers and planners study the distribution of flows of migrants between the geographical subdivisions of a country, the migration balance of the units under analysis and their net migration rate (obtained by dividing this balance by the population of the zone). The phenomena brought to light include specific flows, barrier effects, poles of attraction and places of repulsion.

This line of research led to the introduction of spatial interaction models, usually based on the gravity model and in which distance plays a key role. This model was first proposed in the late 19th century by many authors who observed a pattern in the organization of migrant flows: the closer together the zones and the larger their populations, the greater the flows between them. In this model, F_{ij} the number of interactions between geographical zones i and j is proportional to the product of the masses M_i and M_j of each region and inversely proportional to the distance between them:

$$F_{ij} = kM_iM_jd_{ij}^{-a}$$

Parameter a in this equation represents the frictional effect of distance. Other things being equal, the higher the value of this parameter, the fewer exchanges there are at a given distance. Parameter k, which controls the relationship between the volume of flows and that of the masses, depends on the rate of mobility, i.e., the proportion of migrants in the population.

The gravity model often gives a good account of migration flow patterns, but "*the theoretical framework underlying this model* [. . .] *is still weak*" (Pumain and Saint-Julien, 2001).

In fact, this model is valuable—less for showing the effects of distance and of the populations at risk, often regarded as trivial, than for permitting through examination of the residual flows an analysis of the migration flows that eliminate these effects. The divergences between the empirical observations and the model may thus reveal less familiar explanatory factors for interregional flows and bring to light the existence of "barrier effects" or "preferred flows" (Pumain, 1986; Baccaïni, 1993).

Over the longer-term, the flows exchanged between the regions of a country result in a spatial redistribution of the population. However, it is possible for two regions to exchange large numbers of migrants without these movements having any major impact on their populations if both flows are of similar size. Conversely, migration at a low intensity can have a relatively large impact on the population of the regions concerned if the two flows are highly unbalanced. The concept of migration "efficiency" expresses the proportion of the flows between two areas that have a real effect on the spatial redistribution of the population. It is measured as the ratio between net migration (the difference between in- and out-flows for an area) and gross migration (the sum of in- and out-flows). The closer this indicator is in absolute values to 1, the more *efficient* the migration movements have been in terms of redistributing the population.

It is frequently observed that population movements are in large part simple substitution movements. For example, in France between 1990 and 1999, only one quarter of the migration between the 137 urban areas with over 50,000 inhabitants qualified as efficient. Put another way, three quarters of the migrants who left one town or city for another had no effect on the population profiles of these towns and cities, in strictly numerical terms, even though they may have altered the sociodemographic composition of their populations, as when a 10-year-old child replaces an economically active 45-year-old, for example (Baccaïni, 2003).

2. Varied Processes Contribute to Population Redistribution

Processes of various kinds contribute to the redistribution of the population within a country, processes that often succeed each other over time and in this way lead to the introduction of new concepts. We shall draw on the case of France to illustrate these concepts.

In the industrialized countries, the evolution of migration flows is closely bound up with changes in town-country relations broadly defined. The concept of "urban transition" developed to parallel that of the demographic transition, summarizes the general pattern observed in most western countries since the industrial revolution (Baccaïni *et al.*, 2000). In this pattern, a population distribution dispersed between numerous small rural localities has given way to one concentrated in urban centers that are larger and more differentiated in size. At the level of cities, after a period of concentration, the population distribution began to decentralize, through a reduction in inner-city densities and an extension of urbanization in the suburbs and semirural periphery from the 1970s. In Europe, these stages have appeared at different times in different countries: earlier in the United Kingdom and the countries of northwestern Europe, later in those of southern, central and eastern Europe.

In western countries, the period comprising the 19th century and the first half of the 20th century was characterized by rural depopulation or "rural exodus." In many countries of the developing world, this process is still under way. Rural depopulation occurs when the population of zones in which agriculture is the main economic activity migrates to towns and cities. Because of its selective nature, this out-migration produces distortions in the age and sex structure of the population. Those leaving are primarily young, economically active people, and frequently include more females than males. The rural zones thus have to contend with both an ageing population and with a rise in the proportion of men never marrying. For France, this phenomenon has been analyzed in detail by Pierre Merlin (1971). He found that 30% of the rural population, or around 7.5 million persons, left the countryside between 1872 and 1931.

The rural communes of France had a negative migration balance, i.e., departures outnumbered arrivals, until the early 1970s (the annual rate of change due to net migration still stood at –0.8% between 1962 and 1968, and at –0.2% between 1968 and 1975). It was only from this date that the direction of migration reversed and rural communes taken as a whole became attractive, acquiring a positive migration balance. The rural communes are defined by Institut National de la Statistique et des Etudes (INSEE) as communes that do not belong to any urban unit. An urban unit is a commune or group of communes in which is located a built-up area of at least 2,000 inhabitants where no dwelling is more than 200 m away from its nearest neighbor. Other definitions of "rural," based on more functional criteria, exist—in particular the one which stems from the classification of zones as urban areas (Bessy-Pietri and Sicamois, 2001).

Urbanization is a corollary of rural exodus and corresponds to the concentration of population in urban settlements of varying sizes. Usually it is accompanied by growth of the urban population in absolute terms. Depending on the stage of the process, different levels in the urban hierarchy are affected. In France during the 1960s, a decade of rapid urban growth, the larger the size of the cities, the greater their "pulling" power. Between 1962 and 1968, the annual rate of change due to net migration rose steadily from the smallest urban units (+0.4% for those with under 5,000 inhabitants) to the largest (1.3% for those with over 200,000 inhabitants, except the Paris urban area, which grew more slowly). At the start of the 1970s, urban units except Paris still had a positive migration balance, but the smallest among them had become the most attractive. Since 1975, the medium- and large-sized urban units of over 20,000 inhabitants have been losing population in aggregate terms due to migration (though they continue to grow because of a net natural increase), and only the smallest retain a positive migration balance, though lower than in the rural communes.

Does this justify use of the term *counterurbanization* for the case of France? The concept of counterurbanization was introduced in the United States in the 1960s. It denotes a systematic loss of population from urban centers, especially the largest ones, that contrasts with gains for the population of smaller towns and cities and rural areas. The only countries in Europe for which this process is confirmed are the United Kingdom and, to a lesser degree, the Netherlands (Fielding, 1986). For France, such a hypothesis is poorly established, since the least accessible rural areas and some small towns continue to lose population. Furthermore, analysis of migration rates by city size shows that the highest net losses are not from the major cities (over 200,000 inhabitants, excluding the Paris urban area), which have seen a substantial reduction in their migratory deficit over the last decade, but from the intermediate cities with between 50,000 and 200,000 inhabitants. The medium-sized cities, with populations between 20,000 and 50,000, also record heavier losses from migration than the very large cities.

At the scale of urban areas, the phenomenon that has prevailed in most of the industrialized countries since the end of the 1960s (slightly earlier in the United States and United Kingdom) is a process of urban deconcentration that sometimes goes under the name of "periurbanization." This process is also observed, albeit in different forms, in the countries of the developing world (Dupuy and Sajous, 2000).

Periurbanization (or suburbanization) can be defined as the relocation of population within an urban area away from zones in or near the center, to zones on or beyond the periphery. In many cases, the members of this "periurban" population continue to work in the city center zones, thus causing an increase in both the volume and length of home-workplace commuting journeys (Baccaïni, 1997). In France, the peak of periurbanization occurred between 1968 and 1975 for Paris, and between 1975 and 1982 for most of the other urban centers (Le Jeannic, 1997). During this period, the annual rate of net migration gain in the periurban rings was 2.4%, whereas the core zones or "urban poles" experienced negative net balances (–0.43% per year). This process has since slowed, with the periurban rings recording a relative annual gain from migration of 0.6% between 1990 and 1999; however, urban sprawl is affecting zones further and further from the city centers. Most urban poles and their periurban rings have thus registered a migration deficit since the period of 1975 to 1982.

3. The Migrating Population Perspective

When attention focuses more on the population itself than on the territory, the question of internal migration leads to the different concepts of, first, the level of mobility, and second, the selectivity of migration.

By *level of mobility* one means the proportion of the population in a zone undertaking a move during a specified period. The mobility in question can therefore concern every residential move or, on the contrary, be limited to moves from one administrative unit to

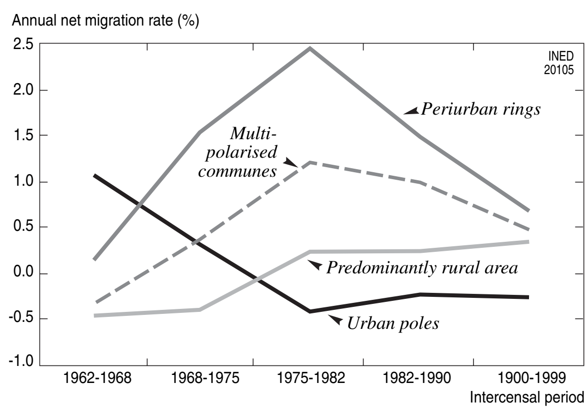


FIGURE 122–1 Trend in migration balances by type of spatial unit, France. (Source: INSEE, Population censuses).

TABLE 122–1 Net Migration Balance by Size of Urban and Rural Units, France

Rate of Annual Variation due to Migration Balance (%)	Intercensal Periods 1962–1968	1968–1975	1975–1982	1982–1990	1990–1999
Rural communes	–0.79	–0.22	0.95	0.73	0.48
Urban units <5,000 inhabitants	0.44	0.70	0.84	0.49	0.45
Urban units 5,000–10,000	0.81	0.75	0.54	0.36	0.30
Urban units 10,000–20,000	0.87	0.76	0.20	0.14	0.06
Urban units 20,000–50,000	1.12	0.45	–0.22	–0.19	–0.13
Urban units 50,000–100,000	1.27	0.64	–0.35	–0.27	–0.36
Urban units 100,000–200,000	1.20	0.40	–0.58	–0.47	–0.17
Urban units 200,000–2,000,000	1.29	0.43	–0.37	–0.16	–0.08
Paris urban area	0.63	–0.11	–0.60	–0.27	–0.64

Source: INSEE, Population censuses.

another. The degree of mobility of the population will thus vary considerably. In France, for example, the annual rates in the period from 1990 to 1999 were 8.1% for changes of residence, 5.3% for between-commune moves, 2.5% for between-department moves, and 1.6% for changes of region (Baccaïni, 2001).

The sensitivity of mobility measurement to the form and/or scale of the geographical or administrative division creates problems for international comparisons in particular. Having shown that the rate of mobility varies as a logarithmic function of the number of units in the division used, Daniel Courgeau in 1973 proposed an index for quantitative evaluation of internal migration that is independent of the geographical unit of measurement (Chapter 22, Volume I).This index has recently been used to compare mobility levels in ten European countries, showing that mobility is higher in northwestern Europe and lower in eastern Europe (Rees and Kupiszewski, 1999b).

Migration motivations affect individuals to varying degrees. As it does for other demographic phenomena, selection operates on the basis of age, sex, marital status, educational level and occupational status. The concept of differential migration expresses this unequal propensity of populations to migrate according to their sociodemographic characteristics. Most research in this domain shows relatively high mobility between ages 20 and 30, a greater mobility among men than among women, and among highly qualified individuals compared with those who are not. But while selectivity by age follows the same pattern in most settings, it is harder to generalize about the role of selection by sex (in the developing countries, for example, women are often more mobile than men) or by educational level (the higher mobility of qualified individuals is mainly for long distance migration, far less for short-distance moves).

To capture the qualitative effects of migration into and out of a particular zone independently of the overall attractiveness of the zone, a coefficient of differential migration can be used. For a subpopulation x, this coefficient is defined as follows: If I_x is the number of arrivals in the zone belonging to subpopulation x, and E_x the number of departures from the zone belonging to subpopulation x, then I and E the total number of arrivals and departures, respectively, for the zone:

$$c_x = \frac{I_x/E_x}{I/E}$$

We say that the zone is selective for subpopulation x when this coefficient is greater than unity, and conversely that it has a repulsive effect for the subpopulation when the coefficient is less than unity.

4. Analysis of Individual-level Migration Behavior

If we adopt the standpoint of the individual, for the study of individual behavior, other concepts, both temporal and spatial, have to be employed. Here we discuss those of life cycle, residential strategy, residential trajectory, and change of life space.

Demographers and sociologists are devoting a growing proportion of their research to studying spatial mobility, for the most part explaining it by reference to the motivations and attributes of individuals. The objective is to understand and explain why the propensity of individuals to migrate varies with their sociodemographic characteristics, and to identify the reasons for their migration. This approach has been enriched by the concept of the life cycle, introduced into demographic and sociological research 20 years ago (Courgeau, 1984). This substitutes criteria of social age (the position of the individual in relation to family and labor market careers) for the traditional considerations of variations in mobility by age, thus improving the interpretation of migration at the individual level. Residential migration by individuals is examined in relation to their complete life histories, in a longitudinal perspective. Attention focuses on the interactions between the different events of the life cycle (Baccaïni, 1991). A better understanding of residential mobility has been gained through the development of new methods of analysis with the capacity to handle individual (micro) longitudinal data and to allow for the interactions between events occurring over the life course of individuals (Courgeau and Lelièvre, 1989; Baccaïni and Courgeau, 1996). The *residential strategies* research theme first emerged in the mid-1980s. It was the subject of a seminar held in October 1988, organized by Institut National d'E'tudes De'mographiques (INED) and the Plan Construction et Architecture (PCA). As Jacques Brun (1993) has pointed out, "use of this concept is based on the assumption that the bulk of individual actors have the possibility of expressing a minimum of independence in managing their residential life." Thus, some authors reject use of the term for the practices of individuals, arguing that institutional and macroeconomic actors alone have the possibility of implementing genuine strategies. Residential strategies can be considered from three angles: (1) strategy related to housing occupancy status and in particular home ownership (Cribier *et al.,* 1990); (2) strategies related to certain periods in the life cycle (nest leaving, divorce, retirement, etc.); (3) and strategies related to locational choices (Brun, 1990). Using the notion of strategy is not to deny that migrants face constraints, notably economic ones, or to credit them with unlimited freedom in their residential decisions; the objective is to capture the social and cultural determinants of these decisions in their full diversity; more attention is given to the aspirations of individuals and to residential behavior.

The concept of *residential trajectory* represents a direct continuation of the two previous concepts of life cycle and residential strategy. Use of the term trajectory implies envisaging residential migration in a life history perspective, as a succession of residential choices, of places occupied, and with linkages to family and labor market trajectories. This notion thus takes into account the spatial and temporal dimensions of the phenomenon. *Residence* is defined as the place where the individual usually lives. The *life space* is defined, then, as that part of space in which the individual carries out all his or her activities (Courgeau, 1980; Courgeau, 1988). A simplified life space may correspond to the individual's home and place of work or education, but it can also be made to include his or her relational network, thereby delineating a more complex life space. Conceptualizing migration as a change of life space considerably enriches the notion of residential migration, since the life space of an individual can evolve in several ways: extension to new points of spatial attachment, spatial relocation or transplantation, or contraction.

5. The Economist's Perspective and Concepts

In the macroeconomic perspective, migration is an indicator of disequilibrium. Migration balances are interpreted as measures of spatial economic disequilibrium, hence migration acts as an adjustment mechanism, particularly for income and employment levels between regions, helping to bring about an optimal economic equilibrium.

This theory is the basis upon which are constructed most aggregate-level econometric models of migration. These models proceed by estimating a migration function in which the dependent variable is the flow of migrants (or migration rate) from origin zone i to destination zone j, with the explanatory variables most often introduced being the income differentials between the zones (Greenwood, 1975), levels of unemployment, occupational structure, level of job skills, etc. If the wages- (or income level–)migration relationship often yields significant results, that between unemployment and migration is less certain. As Michael Greenwood (1975) notes, many studies into the effect of the unemployment rate on migration have produced results that do not conform to expectations. The relationship between unemployment and migration is complex and cannot be reduced to a direct correlation, as is shown by the example of the French regions at the present time. In the northern regions where the negative migration balance goes with high levels of unemployment, one might be tempted to attribute the strong out-migration and weak in-migration to unemployment and lack of job opportunities. But in the regions of southern France (Languedoc-Roussillon in particular), a high level of unemployment coexists with a strongly positive migration balance. In these regions, it is as if migrants move in independently of labor market opportunities, bringing new demand, but also, with their savings, a capacity for expenditure or investment favorable to growth on an appreciable scale, albeit insufficient to provide jobs for everyone. Unemployment thus appears to be at once a cause (in the declining regions) and a consequence (in the growing regions) of migration (Aydalot, 1985, Chapter 5).

The microeconomic approach to migration is based on the idea that for individuals the decision to migrate is the consequence of a comparison between their present situation (employment, income, etc.) and the situation they hope to achieve by changing place of residence. The notion of migration for the pursuit of well-being has been given theoretical form by Larry Sjaastad (1960, 1962) with the so-called "human capital" models. Individuals migrate to improve their human capital, i.e., their lifetime capacity to create the conditions for achieving higher income. In other words, individuals will migrate if they believe that the benefits will exceed the costs. Human capital models take into account economic costs and benefits, but also introduce nonmonetary measures (psychological costs like that of leaving one's friends, the benefits to be derived from a better climate, moving closer to friends, etc.). The costs and benefits of the migration become apparent not immediately but over a period of time, so the migration can be seen as an investment made by the individual. By formalizing this approach, one can define a value for the investment associated with migration from i to j and, therefore, suppose that an individual will migrate from i to j only if this value is positive. Moreover, the individual will tend to select a destination j for which this value is highest. These models thus take the form of a subtraction: the hoped-for gain from migrating from i to j is a function of the difference between the expected utility at j and that existing at i, minus the costs of migrating.

William Milne (1991) has produced a good synthesis of these human capital models. He formalizes the individual decision to move to live in a region

j as an indirect utility function (v), in the following way:

$$v(P_j, W_jOY_j, A_{ij}) > v(P_i, W_iOY_i, A_{ij}),\ i \neq j$$

where:

P_i = a vector of prices of goods in region i.
W_i = individual wages in region i.
OY_i = individual benefits (other than wages) in region i.
A_{ij} = a vector of other variables (climate, job opportunities, etc.) characterizing region j from the perspective of an individual residing in region i.

The general formulation of the human capital model, used in empirical studies, considering that the flow of migrants from i to j can be written:

$$M_{ij} = f_{ij}\ (X_i, Y_j, c_{ij})$$

where:

X_i = a vector of variables in origin region i that influence the decision to migrate.
Y_j = a vector of variables in the destination region j that influence the decision to migrate.
c_{ij} = the cost of migrating from i to j
f_{ij} = a functional form (for example, linear or log linear).

These models have been extended for interpreting family migration. In this case, the move is viewed as a family investment in human capital (Mincer, 1978).

II. MULTIPLE BUT DEFECTIVE SOURCES

In contrast to the other demographic phenomena (births, marriages, divorces, deaths), migration is not the subject of recording by civil registration systems.

Four sources of information can be used to measure and analyze the residential migrations of populations. They fall into two main groups (Poulain, 1994). First, censuses and surveys, whose purpose is essentially statistical. Second, population registers and other administrative documents for which statistical analysis, particularly the study of mobility, is not the principal objective. Data from censuses and registers may have the advantage of giving complete coverage of the population of a country, but they often prove inadequate for a satisfactory analysis of residential mobility, particularly when one wishes to examine individual behavior. Surveys alone can capture the stages of the process that leads an individual to move.

1. Censuses

In the past, the bulk of information on the residential mobility of individuals came from censuses and was obtained by cross-tabulating the question on place of birth with that on place of residence. Over the last 50 years, new questions have been included in censuses: the date of the last move, the place of residence prior to this last move, and the place of residence at a particular date in the past.

A simple method for estimating the migration balance of a geographical area between two censuses is often used, since it is based on a minimum of information.

If the natural balance of a zone (B_{t1}–$D_{t1,t2}$) is subtracted from the total population change (P_{t2}–P_{t1}), we obtain an estimate of net migration:

$$I–E = P_{t2}–P_{t1}–B_{t1,t2}–D_{t1,t2}.$$

This estimate has the disadvantage of lumping together domestic and international migration, as well as making measurement dependent on the uneven quality of the censuses conducted at t_1 and t_2, and on the possibly defective recording of births and deaths by the civil registration system (Baccaïni, 1999).

However, this method does also permit estimation of migration balances by groups of cohorts (for example, the migration balance for persons born between 1960 and 1970), births not then being included in the equation.

The question on place of birth presents the advantage of supplying a basic statistic of high spatial and temporal availability, having long been included in most census questionnaires. Over time, however, a falling off has been observed in the use of this question for measuring residential migration, probably due to the introduction of questions that permit a more refined approach.

This question can be used in a given territory to identify the persons whose current place of residence is different from their place of birth, referred to as *lifetime migrants*, but it raises several types of problem. The results are hard to interpret because they amalgamate the moves that have occurred at any age in a long and ill-defined period extending from the birth of the persons enumerated up to the date of the census.

Nothing is known about multiple intervening moves between birth and the census, or about return moves away from and back to the place of birth. Thus, individuals counted as residing at their birthplace could in fact have made many moves in their lifetime but have returned to the birthplace by the date of the census.

Identifying the place of birth of elderly people can be difficult because of changes to territorial divisions that may have occurred since their birth. Moreover, it seems that a nonnegligible proportion of the population, even in the industrialized countries, is unaware or has only hazy knowledge of its place of birth.

Nowadays, births in the industrialized countries usually occur in clinics and hospitals, most of which are situated in towns and cities. Consequently, the place of birth question is of limited value for studying migration, in particular between rural and urban zones. A question on the place of residence of the mother at the birth of the enumerated individual would be more useful.

Despite these limitations, however, the place of birth question can supply valuable information on residential mobility. Thus, when working on age groups not yet at retirement (for which returns to place of birth are assumed not to be on a significant scale), comparison of the place of residence with the place of birth gives a good idea of the trend in the mobility of the population. In France, the percentage of people aged 45 living outside the *département* in which they were born has risen from 20% in cohorts born at the start of the 19th century, to 25% in those born mid-century, to 35% in those born at the start of the 20th century, and to nearly 50% in cohorts aged 45 at the end of the 20th century (INED, 1989).

The question on birthplace can also be valuable when used in conjunction with the question on place of residence at a fixed date in the past. In this way, the spatial location of the individuals is known at three dates, and although any multiple moves that occurred during the two periods are not observed, examination of these data is richer than when using a single question (Courgeau, 1982).

In some countries (though rarely those of Europe), a question is asked on previous place of residence, often in association with a question on duration of residence at the address occupied at the time of the census. These questions thus provide geographical and temporal information on the last move of the individuals. Analysis of internal migration using these questions faces numerous pitfalls. The last moves of individuals may have occurred at very different times: for some individuals the last move may have been very recent, for others many years ago. This aggregate of last moves can thus provide no information about the volume of migratory movements in a country at a particular point in time. Furthermore, individuals who have been in the same place of residence for long periods may give incorrect answers, due to recall failure. Frequently, therefore, attention is limited to last moves that occurred in the year prior to the census, by considering only the individuals who reported a length of stay of less than 1 year at their present address. A spatial analysis of internal mobility in the course of the year is then possible.

An increasingly common practice is for national censuses to include a question on the place of residence at some fixed date in the past. The date selected in the past is usually 1 or 5 years before the census, which is the basis recommended by the United Nations. In some countries, a reference date is chosen that has connection with a major event in national history, so as to limit memory recall errors. In France, since the 1962 Census, individuals are asked about their place of residence on January 1st of the year of the previous census. In 1999, therefore, the question concerned the place of residence on January 1, 1990.

However this question is formulated, it allows identification of migrants (individuals who have changed residence in the course of a fixed period), not of moves (events occurring at specific points in time).

The number of migrants is underestimated due to the omission of any individuals who, in the course of the period, have gone abroad, or died, or made a return move between two zones. For children born in the period, the United Nations (1988) recommends that they be assigned "*migration characteristics derived, preferably, from data for the migration characteristics of the mother*." These children will thus be considered migrants if their mother is, even if they were born after their mother's move. It may be preferable to exclude these children from published migration statistics, as is done by some countries (including United States, Canada, and Australia).

Because of multiple moves, the number of migrants is less than the number of moves. The underestimation of mobility using this question is especially large for short-distance moves, which are more frequent than long-distance moves. In addition, the real direction of the migration flows is poorly identified. What we obtain by comparing the place of residence of individuals at two dates separated by a relatively long interval is the result of all the moves made by individuals during that period. In an extreme case one could count the migrants between i and j without any real move ever having occurred between these two places, if all the migrants had in fact merely made one or several intermediate stages. When the reference period is 1 year, the number of migrants is close to the number of moves, since multiple moves are fewer the shorter the period. Note also that migration balances *are* well measured, if we overlook however the impact of mortality.

It can also be noted that for understanding the role of migration in the process of spatial population redistribution, information on migrants between two dates is probably more useful than information on moves. If the objective is to compare the spatial distribution of the population at two dates, by revealing the relative contributions of migration movement and natural increase to the observed changes, information on the resultant moves and on migration balances is quite adequate. On the other hand, using this information on the migrants is more problematic when the objective is to analyze the mobility (frequency of moves) of individuals. Recourse to migration behavior models then becomes necessary (Chapter 22, Volume I).

2. Surveys

Among the many surveys used for analysis of residential mobility, we must distinguish between those in which migration is the central focus and those whose objective is quite different but include questions on mobility.

Among the specific surveys that seek to give a full account of residential migration, two main types can be identified: retrospective surveys and prospective surveys.

a. Retrospective Surveys

Retrospective surveys are the most widely used and the easiest to conduct. The French *Triple Biography* (or "*3-B*") survey, a retrospective survey conducted by INED in 1981 and directed by Daniel Courgeau, probably represents the best approach in this domain. This retrospective survey had objectives going well beyond the study of mobility, though this came within its scope. It recorded the main lifetime events of respondents in the areas of family, employment and migration. The many analyses based on this survey have increased the understanding of the frequently complex interactions between moves by individuals and family and labor market events (Courgeau, 1984; Courgeau, 1985a; Baccaïni, 1991). The "*3-B*" survey marks a direct extension of surveys conducted by INED from the early 1960s. Mention can be made of Guy Pourcher's 1961 survey, entitled *Le peuplement de Paris. Origine régionale. Composition sociale. Attitudes et motivations* ["The Population of Paris. Regional origins. Social composition. Attitudes and motivations"], or the survey conducted in 1972 by Alain Girard and Henri Bastide called *Mobilité de la population et motivations des personnes* ["Population mobility and personal motivations"]. The *Peuplement et dépeuplement de Paris* ["Population and depopulation of Paris"] survey conducted by Catherine Bonvalet in 1986 was modelled on the "*3-B*" survey, though it put the emphasis more clearly on the links between family and housing. The "*3-B*" survey has provided the model for numerous surveys conducted in other countries.

These retrospective surveys raise a number of problems that are often hard to overcome. The most intractable is certainly that of misreporting by individuals, who have to rely on their memory, resulting in omitted events, errors on dates of moves or places of residence. Also, to the extent that a survey can only reach the individuals who have survived and are present in the country, the sample is never representative of the cohorts studied.

b. Prospective Surveys

Prospective surveys are widely used in the developing countries, little so in the industrialized countries. This type of survey overcomes the problem of recall failure. A distinction needs to be made between multiple-round surveys, which do not really track individuals spatially but at best question people who have not moved to obtain information on those who have, and surveys which do follow the initial population. These surveys, though they yield extremely rich information on mobility, are costly to execute, and the risk of being unable to trace the majority of migrants is high.

In France, surveys on residential mobility have been conducted using the records of the French electricity distribution company (EDF) on its account holders. The methodology for these surveys was developed by Observatoire des loyers de l'Agglomération parisienne (Observatory of Rents in the Paris region; OLAP) in 1995, in response to a request from the Direction générale de l'urbanisme, de l'habitat et de la construction (DGUHC) of the *Ministère de l'équipement, des transports et du logement*. A national survey and six local surveys have been conducted using this method. The main disadvantages with these surveys, besides their high cost, are that they cannot be used to calculate mobility rates, compare the population studied with the total population, or carry out detailed cross-tabulations.

c. Other Surveys

Other surveys not primarily intended for analysis of residential mobility are nonetheless suitable for this purpose. Here we will mention two such surveys.

The labor force surveys conducted in most European countries provide valuable information on spatial mobility. The French version of the labor force survey, the *Enquête emploi*, is conducted by INSEE. It is an annual survey, conducted on a continuous basis since 2003 (interviewing is spread over the whole year), on a sample of some 60,000 households in which all individuals aged 15 or more are interviewed. One third of the sample is replaced each year. The survey includes a question on the place of residence 1 year ago. Few individuals change address more than once a year, so this measure of movers in 1 year corresponds closely to the number of moves. In addition, a parallel can be established between residential change and occupational change, since a series of questions describes the main activity of the respondent 1 year before the survey date.

These data cannot be used for studying migration flows at a detailed geographical scale (though the survey is representative at the regional level). However, they do permit monitoring of the phenomenon on a yearly basis and are helpful for detecting the most recent trends in mobility.

The *Enquête logement* (Housing Survey) is one of the main surveys conducted by INSEE both by the length of time it has been running (since1955) and its periodicity (conducted, on average, every 4 years), and by the size of its sample (roughly 40,000 households). It is the main statistical source in France for describing the housing stock and the occupancy conditions of households in their main residence. Several items from this survey can be used for analysis of residential mobility, in particular the question on place of residence 4 years ago. For households that have moved, the survey records the main characteristics of the housing occupied 4 years ago and the occupancy status at that date. Households that have moved more than once in the last 4 years are also asked about their occupancy status prior to the last change of address and about the reasons for this last move. The survey also asks respondents about their mobility intentions (Baccaïni, 2000). This survey can thus be used to analyze mobility and mobility intentions, and to bring out the connections between past mobility and intended mobility. Residential histories, in particular, can be studied in detail (Bessy and Tabard, 1996). On the other hand, unless regional or local extensions of the survey are set up, it is not possible to study mobility across a territory.

3. Population Registers

Population registers are administrative records whose purpose is to provide an up-to-date record of the population in the various administrative units that make up a country. Births, deaths, arrivals and departures of population are recorded. Population registers do not exist in all countries; 20 countries in Europe operate them. Furthermore, where they do exist, these registers are not always used as a source of data on internal migration.

The comprehensive review conducted by Philip Rees and Marek Kupiszewski for the Council of Europe (1996) shows that of the 28 member-countries of the Council of Europe, 14 use their population register as the main source on migration.

Population registers supply information on changes of residence and thus on movement, and in this way provide a dynamic picture of the population. The data on internal migration contained in the registers allow two kinds of approach to the phenomenon: (1) an overview of migration trends within the country from year to year and (2) a tracking of individuals from place to place over their entire lives.

TABLE 122–2 Main Sources of Migration Data in Council of Europe Countries

Population Registers	Censuses	Registers and Censuses	Migration Surveys	Censuses and Migration Surveys
Germany	Austria	Hungary	Slovak Republic	Ireland
Belgium	France	Romania	Czech Republic	Portugal
Bulgaria	Greece	Slovenia		
Denmark	Turkey	Spain		
Estonia		Switzerland		
Finland		United Kingdom		
Iceland				
Italy				
Latvia				
Lithuania				
Norway				
Netherlands				
Poland				
Sweden				
14 countries	4 countries	6 countries	2 countries	2 countries

Source: Rees and Kupiszewski, 1996.

By combining the information present in these registers it is possible to produce matrices of annual interregional migration, disaggregated by subpopulations (by age, sex, etc.). In some countries, the information is available every 3 months or even monthly (in Lithuania). So for these countries where population registers are in use, we have long series on interregional exchanges at a detailed geographical scale.

The possibility of tracking individuals over their entire lives is one of the main advantages of population registers. Their chief drawback comes from the limited amount of information they contain on the individual characteristics of migrants. The Norwegian register is one that lends itself most to this type of longitudinal analysis. This central computerized register has been combined with individual-level data collected in the 1960, 1970 and 1980 censuses (an operation no longer possible today, as the census has been abolished in Norway). The result is a rich statistical database containing individuals' successive changes of address and births, plus a large variety of sociodemographic information collected at the date of the census (educational level, occupation, marital status, income, etc.). Methods of longitudinal analysis are particularly well suited here, notably for revealing the effects of individual characteristics on the probabilities of making specific kinds of residential relocation, for example, from a rural area to an urban center (Baccaïni and Courgeau, 1996).

In Europe, centralized and computerized registers on these lines exist in the Nordic countries, Belgium, Poland, the Czech and Slovak Republics, Hungary, Slovenia and in Liechtenstein (Poulain, 1994). These population registers are maintained by means of two procedures. One involves updating civil registration records at the place of legal residence; the other is the compulsory declaration and registration of changes of address in the localities of origin and of destination.

Data from the registers are not free from measurement problems. As Michel Poulain (1994) stresses, some changes of address are not recorded (for example those occurring within a single smallest administrative unit). Also, what movers report may be at variance with the facts, influenced perhaps by the existence of advantages or disadvantages attached to declaration. Finally, the collection of information may introduce errors, particularly when the system is not centralized.

In many countries, moreover, legislation regulating access to the registers has made them hard to use. In some countries where there is no population register, census and civil registration data have been used to set up panels. In France, the *Echantillon démographique permanent* (Demographic Longitudinal Sample; EDP) can be used to construct the family, labor market and migration histories of individuals. The EDP is a longitudinal base of individual microdata maintained by INSEE for the last 30 years, and contains data on nearly 900,000 individuals who are or have been residents of metropolitan France. Information recorded in successive census forms (since 1968) is stocked and matched to civil registration records, to compile a vast panel. The study base currently contains 886,458 individual records, of which 617,195 concern persons in the 1999 population census. In contrast to population registers, the EDP does not supply all the successive addresses of where individuals have lived: "*The dates observed precede and follow the migration event one wishes to study, which in effect means that one has a fragmentary life history*" (Courgeau *et al.*, 1998). It does, however, constitute an extremely rich resource for the study of mobility and residential histories. The same kind of data set exists in the United Kingdom, the Longitudinal Study (LS), which was started the same time as the EDP in France.

4. Administrative Records

In every country, the records kept by various official departments and organizations can be a source of valuable, though rarely flawless, information on residential migration.

In the United Kingdom, the central register of the National Health Service (NHS) is an important source of information on the redistribution of the population between administrative units (the unit used is specific to the NHS), which has been used since the 1950s (Stillwell *et al.*, 1994). When patients who have moved home register with a doctor in their new area of residence, their previous area of residence is recorded. The main problem with this source is an underrecording of short-distance moves.

In France, a number of administrative records, access to which is often restricted, provide information on residential migration. Similar records are available in most industrialized countries. The records of the French electricity company (EDF) on new accounts and those of the telephone company, France Télécom, are extremely rich sources since they contain the household's previous address. These sources provide information on current locational trends, but the information is poor in qualitative terms. EDF also has a source on household mobility that supplies longitudinal information.

The fiscal records relative to income tax and local taxes can also be used. With these it should be possible to produce crosstabulations at the level of individual communes between the housing and sociodemographic characteristics of households, and the date of moving to the address.

The annual data returns *Déclarations annuelles de données sociales* (DADS) are the official documents by which employers notify the tax authorities and social security organizations, individually for each of their employees, of the amount of wages paid yearly. Except for people employing domestic staff, any employer, whether a physical person or a business, based in France and who pays salaries or wages is required to fill in an annual DADS-type return. INSEE, which administers this register, currently excludes farm employees and public sector employees, whose returns relate to special social security schemes and are specific, from processing for statistical purposes. Since 1998, the French data protection agency National Council for Data Processing and Civil Liberties (*Commission national informatique et liberté*; CNIL) has authorized INSEE to reconstitute a register of salaried workers using the national identification number, so as to group together the different positions an individual has held. Hence, it is possible to monitor all the changes of address and employment by salaried workers. Because the geographical reference unit is the commune, only intercommunal mobility is counted.

Research conducted at INSEE by Cédric Blum in 2000 pointed up the value of using the DADS for studying residential migration. This source offers a number of new elements compared with the population census, especially regarding the link between residential mobility and occupational mobility. The periodicity (annual) of the DADS register is also a serious advantage.

Electoral registers can be used for residential migration study, but with caution. They only cover the population old enough and eligible to vote, thus there is no coverage of foreigners and individuals who have lost their right to vote. Furthermore, even if registration is compulsory, a nonnegligible proportion of the population is not in fact inscribed on the register. Lastly, in France individuals are not required to register to vote in their commune of destination and can remain registered in another commune, where, for example, they own a weekend or holiday home.

CONCLUSION

At the end of this nonexhaustive review of the principal concepts and sources of observation relative to internal migration, two questions arise.

Could it be possible to bring together or reconcile the different approaches to internal migration, with the aim of constructing a migration theory (or theories) of a universal nature? The purely spatial approach that deals with geographical entities would clearly gain from including the preoccupations of demographers and sociologists, focused on populations and individual behavior. Vice versa, analysis centered on the individual, by ignoring that individual behavior, is always spatially defined and thus largely dependent on contextual effects, if only because an individual lives in one place rather than another, would be enriched by inclusion of the geographical environment.

From a methodological point of view, the approach based on multilevel models contributes to bridging the gap between these different levels of analysis, that of the individual and that of geographical entities, or indeed intermediate levels such as the family or the individual's relational network (Chapter 24, Volume I). From a more theoretical point of view, however, much remains to be done to disentangle the extremely complex linkages between the different levels of aggregation, in order to produce a coherent theory for these models and for human behavior in general.

A second question that demands attention concerns the reliability of the measurements of internal migration that can be achieved through the various sources currently available. On this point it must be recognized that, although the trend is toward improvement, the data suitable for characterizing spatial mobility within a country are often of poor quality. As a consequence, the available data are hard to compare across countries. Very often, comparison can apply only to processes and trends but not to levels of mobility and measured flows. On the other hand, increasingly complex surveys and improved analytical methods offer the prospect of a better understanding of individual mobility behavior.

Acknowledgment

This chapter was translated from French by Godfrey I. Rogers.

References

Aydalot Philippe, 1985. *Économie régionale et urbaine.* Paris, Economica, VIII + 487 p.

Baccaini Brigite, 1991. *Mobilité, distances de migration et cycle de vie.* Paris, Université de Paris I, UFR de Géographie, 520 p. (Doctoral thesis).

Baccaïni Brigitte, 1993. Régions attractives et régions répulsives entre 1982 et 1990. Comparaison avec la période 1975–1982 et spécificités des différentes classes d'âges, *Population*, vol. 48(6), p. 1791–1812.

Baccaïni Brigitte, 1997. Les navettes des périurbains d'Île-de-France, *Population*, vol. 52(2), p. 327–364.

Baccaïni Brigitte, 1999. Analyse des migrations internes et estimation du solde migratoire externe au niveau local à l'aide des données censitaires, *Population*, vol. 54(4–5), p. 801–816.

Baccaïni Brigitte, 2000. Projet de mobilité et cadre de vie, *in:* Marie-Flore Mattei and Denise Pumain (eds.), *Données urbaines, 3,* p. 91–112. Paris, Anthropos, 441 p.

Baccaïni Brigitte, 2001. Les migrations internes en France de 1990 à 1999: l'appel de l'Ouest, *Economie et Statistique*, vol. 344(4), p. 39–79.

Baccaïni Brigitte, 2003. Les migrations interurbaines entre 1990 et 1999, *in:* Marie-Flore Mattei and Denise Pumain (eds.), *Données urbaines, 4,* p. 323–340. Paris, Anthropos, 433 p.

Baccaïni Brigitte and Courgeau Daniel, 1996. The spatial mobility of two generations of young adults in Norway, *International Journal of Population Geography*, vol. 2(4), p. 333–359.

Baccaïni Brigitte, Pumain Denise, Rees Philip and Kupiszewski Marek, 2000. Migrations et urbanisation dans les pays européens, *in:* Marie-Flore Mattei and Denise Pumain (eds.), *Données urbaines, 3,* p. 351–361. Paris, Anthropos, 441 p.

Bessy Pascale and Tabard Nicole, 1996. *Trajectoires résidentielles et organisation socio-économique de l'espace. Une typologie des ménages ayant changé de commune entre 1988 et 1992.* Paris, INSEE, Direction des Statistiques Démographiques et Sociales, 47 p. (Working Paper, no. F9603).

Bessy-Pietri Pascale and Sicamois Yann, 2001. Le zonage en aires urbaines en 1999: 4 millions d'habitants en plus dans les aires urbaines, *Insee Première*, no. 765, p. 1–4.

Brun Cédric, 2000. *Etude des migrations résidentielles dans la zone d'emploi de Saint-Etienne à partir du fichier DADS 1997.* Paris, INSEE, Direction Régionale de Rhône-Alpes, 27 p. (Working Paper).

Brun Jacques, 1990. Mobilité résidentielle et stratégies de localisation, *in:* Catherine Bonvalet and Anne-Marie Fribourg (eds.), *Stratégies résidentielles*, p. 299–311. Paris, INED, 460 p. (Congrès et Colloques, no. 2).

Brun Jacques, 1993. La mobilité résidentielle et les sciences sociales, *Les Annales de la recherche urbaine*, no. 59–60, p. 2–13.

Collomb Philippe, 1979. Les émigrants de l'ouest audois, dix-neuf ans après, *Population*, vol. 34(1), p. 65–90

Courgeau Daniel, 1973. Migrations et découpage du territoire, *Population*, vol. 28(3), p. 511–537.

Courgeau Daniel, 1980. *Analyse quantitative des migrations humaines.* Paris, Masson, 225 p. (Anthropologie physique, no. 12).

Courgeau Daniel, 1982. Premiers migrants, migrants secondaires, et retours (France 1968–1975), *Population*, vol. 37(6), p. 1189–1193.

Courgeau Daniel, 1984. Relations entre cycle de vie et migrations, *Population*, vol. 39(3), p. 483–514.

Courgeau Daniel, 1985. Les enquêtes de migration dans les pays développés, *in:* Département de Démographie (ed.), *Migrations internes, collecte des données et méthodes d'analyse,* p. 53–81. Louvain-la-Neuve, Cabay, 453 p. (Chaire Quetelet 1983).

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life-cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel, 1988. *Méthodes de mesure de la mobilité spatiale: migrations internes, mobilité temporaire, navettes.* Paris, INED, 301 p.

Courgeau Daniel and Lelièvre Éva, 1989. *Event history analysis in demography,* Oxford, Clarendon Press, 1992, 226 p.

Courgeau Daniel, Lelièvre Éva and Wolber Odile, 1998. Reconstruire des trajectoires de mobilité résidentielle. Éléments d'une analyse biographique de l'EDP, *Économie et Statistique*, no. 316–317, p. 163–176.

Cribier Françoise, Duffau Marie-Luce and Kych Alexandre, 1990. Stratégie résidentielle et statut d'occupation, *in:* Catherine Bonvalet and Anne-Marie Fribourg (eds.), *Stratégies résidentielles*, p. 157–176. Paris, INED, 460 p. (Congrès et Colloques, no. 2).

Desplanques Guy, 1994. ConnaÎtre les migrations, *Espace, Populations, Société*, no. 1, p. 31–39.

Dupuy Gabriel and Sajous Patricia, 2000. L'étalement périurbain: perspectives internationales, *in:* Marie-Flore Mattei and Denise Pumain (eds.), *Données urbaines, 3,* p. 141–151. Paris, Anthropos, 441 p.

Fielding Anthony J., 1986. Counterurbanization in Western Europe, *in:* Allan Findlay and Paul White (eds.), *Western European population change,* p. 35–49. Londres, Croom Helm, xv + 253 p.

Greenwood Michael J., 1975. Research on internal migration in the United States: a survey, *Journal of Economic Literature*, vol. 13, p. 397–433.

Haynes Kingsley E. and Fotheringham A. Stewart, 1984. *Gravity and spatial interaction models.* Beverly Hills, London and New Delhi, Sage publications, 87 p.

Institut National d'E'tudes De'mographiques (INED), 1989. *Dix-huitième rapport sur la situation démographique de la France.* Paris, INED, 66 p. (II. Migrations internes en France durant la période 1954–1982 et tendance récentes).

Le Jeannic Thomas, 1997. Trente ans de périurbanisation: extension et dilution des villes, *Économie et statistique*, no. 307, p. 21–41.

Merlin Pierre, 1971. *L'exode rural.* Paris, INED et PUF, 350 p. (Travaux et Documents, Cahier no. 59).

Milne William, 1991. The human capital model and its econometric estimation, *in:* John Stillwell and Peter Congdon (eds.), *Migration Models. Macro and Micro Approaches,* p. 137–151. London and New York, Belhaven Press, 329 p.

Mincer Jacob, 1978. Family migration decisions, *Journal of Political Economy*, vol. 86(5) (October), p. 749–773. (Chicago, Illinois).

Nations unies (United Nations), 1988. *Principes et recommandations pour les recensements de la population et des habitations de 1990 dans la région de la CEE.* New-York, Nations unies/United Nations, 116 p. (Normes et études statistiques, no. 4).

Poulain Michel, 1981. *Contribution à l'analyse spatiale d'une matrice de migration interne.* Louvain la Neuve, Cabay, 226 p. (Recherches démographiques, Cahier no. 3).

Poulain Michel, 1994. La mobilité interne en Europe. Quelle données statistiques? *Espace, Populations, Sociétés*, no. 1, p. 13–30.

Puig Jean-Pierre, 1981. Le rôle des villes dans les migrations d'actifs, *Économie et statistique*, no. 133, p. 57–75.

Pumain Denise, 1986. Les migrations interrégionales de 1954 à 1982: directions préférentielles et effets de barrière, *Population*, vol. 41(2), p. 378–388.

Pumain Denise and Saint-Julien Thérèse, 2001. *Les interactions spatiales.* Paris, Armand Colin, 191 p. (Cursus géographie).

Rees Philip and Kupiszewski Marek, 1996. *Internal migration and regional population dynamics: what data are available in the Council or Europe Member States.* Leeds (United Kingdom), University of Leeds, School of Geography, 43 p. (Working Paper, no. 96/1).

Rees Philip H. and Kupiszewski Marek, 1999a. Internal migration: what data are available in Europe? *Journal of Official Statistics*, vol. 15(4), p. 551–586. (Statistics, Sweden).

Rees Philip and Kupizievski Marek, 1999b. *Internal migration and regional population dynamism in Europe: a synthesis.* Strasbourg, Council of Europe, 113 p.

Rees Philip, Stillwell John, Convey Andrew and Kupiszewski Marek (eds.), 1996. *Population migration in the European Union.* New York, John Wiley & Sons, 390 p.

Sjaastad Larry A., 1960. The relationship between migration and income in the United States, *Papers and Proceedings of the Regional Science Association*, no. 6, p. 37–64.

Sjaastad Larry A., 1962. The costs and returns of human migration, *Journal of Political Economy (Chicago, Illinois)*, vol. 70(5), p. 80–93.

Stillwell John, Boden Peter and Rees Philip, 1994. Internal migration trends in the United Kingdom from National Health Service Registration data, *Espace, Populations, Sociétés*, no. 1, p. 95–108.

C H A P T E R

123

Conceptual Framework and Data Collection in International Migration

CORRADO BONIFAZI AND SALVATORE STROZZA

Istituto di Ricerche sulla Popolazione e le Politiche Sociali-Consiglio Nazionale delle Ricerche, Rome, Italy

Università degli Studi di Napoli Federico II, Napoli, Italy

"Migration is a fuzzy concept with many definitional problems" (Willekens, 1984, p. 2–3). *"Trying to piece together a coherent statistical picture of any of the five key groups*[1] *(of migrant workers) is similar to trying to put together a large puzzle based on photographs of a rapidly changing reality, with important pieces missing and many of the available pieces out of focus"* (Hoffmann and Lawrence, 1996, p. 2). Expressions like *fuzzy concept* and *puzzle* are not often encountered in scientific literature, and their use therefore strongly underlines the difficulties facing those attempting to conceptualize, define, and measure migratory processes.

The issues are well known. Competent international organizations and the scientific community have been working on them for more than a century and searching for adequate answers (Krali and Gnanasekaran, 1987). In recent years, efforts in this direction have increased, under pressure from the political resonance of the phenomenon and the increasing concern of governments and public opinion in the face of the dynamics and consequences of international migration. The results have been really important.

The United Nations has made new recommendations for the production of statistics on international migration (United Nations, 1998b). The International Labour Office (ILO) has also stimulated research aimed at improving the data collection processes for migration in general and, more particularly, on employment-based movements (Bilsborrow *et al.*, 1997; Hoffmann and Lawrence, 1996). To the classic report of the Systeme d'Observation Permanente sur les Migrations (SOPEMI) (continuous observation system for migration) set up by the Organisation for Economic Co-operation and Development (OECD) have been added the *International Migration Report* of the United Nations Population Division, the first edition of which was published in 2002, and the *World Migration Report* of the International Organization for Migration (IOM), published for the first time in 2000 and the second time in 2003. Publications and databases are currently produced by Eurostat, the statistical office of the European Community, while the Council of Europe has considerably increased the space reserved for international migration in its yearbook. In many cases, the geographical coverage of information gathered by these international organizations has increased. The information is no longer limited to traditional immigration countries within or outside Europe, but now offers a more complete view on essential aspects of the international migration panorama. Beyond this reminder, it is of course impossible to review the entire catalogue of more specific work undertaken at national, bilateral, and multilateral levels, but many of them focused on or are still based on the comparability of data and statistical cooperation to improve the quality of data produced in the countries of immigration and emigration (Poulain, 1991, 1993 and 2001; Salt *et al.*, 1994).

[1] Immigration of foreign workers, emigration of national workers, return migration of migrant workers, stock of foreign workers, and stock of national workers abroad.

Ultimately, never has Hania Zlotnik's statement (1987, p. v–vi) that *"Although the charge of inadequacy may be appropriate, scarcity is not necessarily a trait of international migration data. Indeed, statistical information potentially relevant to the measurement of international flows of migrants is more commonly available than one would expect"* been more applicable to the current situation. Indeed, despite the positive results of the growth in available data, we are still far from having satisfactory and definitive answers to all the problems related to the statistical measurement of international migration. However, current scientific research seems to have modified its own strategy and opted for a more pragmatic approach aimed at using and developing what already exists today by placing it in a conceptual framework that enables the understanding of the articulation of different types of migratory flows rather than looking for definitive answers to issues of definition or theoretical framework of the measurement of the phenomenon (Bilsborrow *et al.*, 1997).

This approach nevertheless takes some fundamental characteristics of the migratory process as a given. First of all, we must recognize its complexity and hence the corresponding difficulty at arriving at a definition and precise and univocal measure, because of the spatial and temporal variety and variability of its forms, its characteristics, its causes, its consequences, and the contexts in which it develops. Secondly, the existence of a strong tie between statistical records and the political dimension of the migratory phenomenon should be noted. On the other hand, if the complexity of the phenomenon and political interest work together to make the measurement of mobility and the comparison of data difficult, there are elements which promote the coexistence of a plurality of statistics and which make it possible to highlight the various aspects of migration (Costanzo *et al.*, 2002). This situation may produce a statistical picture of the phenomenon as fuzzy and ill-defined, but this may prove itself useful for constructing another, more articulate, and, finally, more realistic and interesting one. Especially in the current period (marked by great changes in the intensity, characteristics, and even the direction of the flows), the availability of many measurements of the same phenomenon, taken from different perspectives and different angles, is a favorable factor for the study and analysis of international mobility (Willekens, 1994). It is therefore judicious, at a practical level, to assemble the different statistics available, as elements of a mosaic or pieces of a puzzle to be built, on the condition that one starts by recognizing the real significance and the explanatory power of each piece.

The user, and before that the producer, of international migration statistics must first of all clarify the meaning of the data by taking into account the concepts and definitions used, as well as the procedures put in place for collecting information. In this regard, we wish to offer to the reader a series of elements to note for the correct usage of the major migration statistics. Both the United Nations and the ILO (Bilsborrow *et al.*, 1997; United Nations, 1998b), at the heart of the two major in-depth analysis projects in the last decade, have themselves given up on the exhaustive reconstruction of the characteristics of the sources used for measuring the phenomenon. They prefer to consolidate our knowledge on regional entities or explore in depth the pathways already traced by publications of the *Demographic Yearbook* of 1977 (United Nations, 1978) and of 1989 (United Nations, 1991), whose *special topics* focused on international migration.

The first section of this chapter therefore focuses on an overview of the problem. We will recall the principal definitions used at the international level while distinguishing between data of flow and data of stock. In the second section, we will study, separately for flow and for stock, the issue of identification of the major categories of migrants and of immigrant populations. In the third section, dealing with data collection modalities, we have chosen, for practical reasons, to work according to the sources of statistics rather than to the type of data. We will therefore successively study censuses, population and foreigners registers, statistics of border movements, other administrative sources and finally, sample surveys. Specific flows, like that of the movement of refugees or that of migrant workers, have not been isolated as such, for most of the general remarks apply also to these particular cases.

I. THE OVERALL FRAMEWORK

The starting point of the problems of measuring mobility is in the concept of migration itself, whose definition is far from obvious. The definition of the phenomenon requires taking into account the relationship linking man with space and time and thus supposes that one has conceptual instruments capable of distinguishing among all space-time movements—those that are migrations and those that are not (Courgeau, 1980). We are confronted here with a type of problem that the analysis of a population's natural dynamics does not generally encounter, because it is based on events (births and deaths) that are much simpler to define and identify and necessarily occur in the life of each individual once and only once. First, an individual, on the contrary, may never migrate, or may migrate many times. There is therefore no univocal relationship between these events (migrations) and the individuals to whom they occur (migrants). Secondly, migration implies a relationship between one geographic space and another. Consequently, its observation always has a precise territorial dimension, which must be defined at the moment of data collection and made known when published. And this complicates demographic research when territorial reality is often disregarded within the framework of the conceptualization and theoretical formalizations.

Because of the complexity of the phenomenon, quantitative analysis is forced to simplify. The first distinction to make is obviously the separation of internal migration from international migration, according to whether an international border has been crossed or not. In a way, this distinction may appear superfluous as the context, the causes, effects, and mechanisms of the two phenomena show similar characteristics. However, because of its very high dependence on political factors, international migration dynamics are really a different phenomenon from internal migration. Better still, it is precisely the control exerted by each state on its borders that defines international migration and makes it a social process independent of other movements (Zolberg, 1989). And consequently, as we shall see, the different international migration data collection operations are articulated around this state control.

Efforts to conceptualize migratory processes on which to base data collection procedures seem to have followed two directions. First, especially among international organizations, attempts were made to reach a consensus on a definition that would serve as a reference point for various producers of statistics. Secondly, efforts were made to construct typologies capable of describing the main aspects of the phenomenon. These two perspectives have crossed on more than one occasion in recent years in an attempt to find effective operational solutions to issues raised.

One could say that the first international recommendations for the improvement of migration statistics were made in 1932 with the International Conference of Migration Statisticians, organized by the ILO (Bilsborrow *et al*., 1997; Krali and Gnanasekaran, 1987). According to these recommendations, *"In principle, every act of removal from one country to another for a certain length of time should be included in the statistics of migration, with the exception of tourist traffic"* (United Nations, 1949, p. 3–4, cited by Bilsborrow *et al*., 1997, p. 1). From this statement, the United Nations then later developed a first set of recommendations, adopted in 1953, centered on the definition of *permanent immigrants* as *"non-resident (both nationals and aliens) arriving with the intention to remain for a period exceeding a year"* and of *permanent emigrants* as *"resident (nationals and aliens) intending to remain abroad for a period exceeding one year"* (United Nations, 1998b, p. 5). The inaccuracy of the concept of residence and especially of its temporal dimension proved to be the weak point of this definition. Consequently, in the 1976 review of these recommendations, a *long-term immigrant* was characterized as *a person who has entered a country with the intention of remaining for more than one year and who either must never have been in that country continuously for more than one year or, having been in the country at least once continuously for more than one year, must have been away continuously for more than one year since the last stay of more than one year"* (United Nations, 1998b, p. 5).

The coherence of this last definition was not enough to assure its effective usage in the production of migration statistics, so much that the United Nations had to again revise their recommendations in 1998. The major objective of this revision was to simplify the definitions and admit that *"An international migrant is* [. . .] *any person who changes his or her country of usual residence"* (United Nations, 1998b, p. 9). The country of usual residence is, according to the definition used in the census, one *"In which a person lives, that is to say, the country in which he or she has a place to live where he or she normally spends the daily period of rest."* Trips abroad, *"for purposes of recreation, holiday, visits to friends and relatives, business, medical treatment or religious pilgrimage"* (United Nations, 1998b, p. 10) are not regarded as a change in the country of usual residence. The duration of stay abroad, based on a 1-year threshold, determines the distinction between *long-term* and *short-term migrants*.

Beyond the changes in definitions, which, moreover, are far from covering all the real problems of migration statistics, the last recommendations of the United Nations are especially remarkable for their application modalities. Indeed, based on the premise that it would be difficult to obtain from states a change in their definitions and an adoption of those of the United Nations, it has proposed a two-stage approach. In the first, the available statistics are supplied with instructions explaining the differences and similarities between the different sources of the same country. In the second, each country is expected to define specific strategies for improving the quality and comparability of its data. Fundamentally, this two-stage process is founded on establishing a well thought out taxonomy of international mobility (not only migration) on the basis of the legal status established by the country of arrival. The typology is established with regard to four occurrences (two relating to nationals and two to foreigners): departure from a given country, arrival in a destination country, departure from a destination country, and return to the country of origin. To simplify, the United Nations assumes that the flows always start from the country of nationality of the migrant and are directed towards the country of destination; migrants are classified into different categories that do not change during their stay in the host country. These four occurrences are combined with 18 categories of persons, which thus has lead to a definition of 72 types of mobility. However, among these categories of travelers, some are not relevant for international migration (inhabitants of border zones or transit passengers), others are important for international tourism, and some (diplomatic staff, the military, and nomads) are generally excluded from migration statistics. Lastly, there are only eight relevant categories for migration statistics (students, trainees, workers, staff of international organizations, free establishment migrants, settlers, family reunion migrants, refugees), whereas for two other categories (asylum seekers and illegal migrants), the duration of stay is uncertain.

In the same perspective, to classify international migration with regard to real situations, the ILO has also prepared *Guidelines for Improving Data Collection Systems*) (Bilsborrow *et al*., 1997). By tabulating nationality (nationals and foreigners) across the legal basis of entry, the ILO has defined 10 types of international migrants and 16 subcategories:

1. return migrants
2. returning ethnics
3. free circulation migrants
4. foreigners admitted for special reasons (students, trainees, retirees)
5. settlers
6. workers (seasonal, project, contract, temporary, permanent, highly qualified)
7. economic migrants (business travelers, investors)
8. refugees and asylum seekers (refugees, persons admitted for humanitarian reasons, persons granted a temporary protected status or a stay of deportation)
9. irregular migrants
10. family reunion migrants (ILO has defined specific categories for this group).

These two examples show the complexity of the migratory process and the difficulties in establishing a typology for adequate collection of data on its various components. This was already made evident in previous attempts at classifying migratory flows, from that of William Petersen in 1958 to the more recent work of Antonio Golini (1987). The latter, working from causes (grouped into three different types of mobility: migratory, semi- or pseudomigratory, and nonmigratory mobility)—taking into account the spatial dimension (distance or the crossing of administrative or political borders), temporal dimension (periodicity and duration), and the legitimacy of movement—identified 55 types of population movement.

Definitions and ordering criteria can therefore take into consideration different elements: nationality, residence, duration of stay, reasons for stay, place of birth, the spatial dimension, the cause of the move, legal status, etc. The first five factors are generally considered the most important (Zlotnik, 1987b). In different combinations, they contribute to defining migration statistics. In the next paragraph, we shall discuss in detail the limitations and characteristics of some of the elements, but before then, it is useful to make a comment on the reasons why, despite all previous efforts, there is a major gap between international recommendations and national practice.

First of all, it must be stressed that apart from the difficulties mentioned earlier, there is also the question of scale. The ease, speed, and low cost of movement from one point of the globe to another today, have led to an unprecedented rise in the mass of international movement, especially of what is not migration in the sense indicated earlier. It only makes it more difficult to put in place, at an acceptable cost, statistical data collection instruments, which will enable migrants to be identified among the set of individuals who cross borders daily.

But the major reason for this gap is because the modalities for the collection of international migration data have a specific political dimension, resulting from complex processes determining the role given to immigration within each nation state (Bonifazi, 1998). From then on, one cannot but be surprised at the preponderance of administrative sources, not only for the more than obvious economic reasons, but for the greater possibilities offered for the study of the different categories defined in migratory policies (Birindelli and Gesano, 1991) and the measurement of the effectiveness of policy. This also explains the medium and long-term instability of the systems of data collection, which undergo the changes in the characterization of the phenomena and the influence of the political and cultural debate. In addition, it should not be forgotten that statistical data on migration are a central element of the process (including migratory history, perception of the phenomenon, migratory policy) with which each state builds its own relationship with immigration (Kritz, 1987). In this sense, *"nomenclatures and categories continue to interact with the entities which they* *represent. As 'convention', they produce reality by the frames of reference which they place at the disposal of the public debate. Reciprocally, they are in tune with the social world and record its evolution"* (Simon, 1998, p. 541).

From a strictly demographic point of view, one must be interested in all migratory flows, whether it is of foreign citizens or citizens of the country in question, precisely because each one of these two movements contributes to the variations in the size and structure of the various aggregates forming the population. As we have shown, the United Nations's definitions have always towed this line. In practice however, national definitions are, on the contrary, often very different (Kujisten, 1995). The principal concern of statistical offices is to measure the presence and the movements, in particular that of the arrival of foreigners, for the simple reason that—especially in these times of strict control of immigration—the demands of the policies and governments go in this direction without considering that it is even sometimes easier statistically to measure the movement of foreigners than that of nationals because of the many administrative traces that foreigners, at least those who move legally, are obliged to leave in their wake (Bonifazi and Strozza, 2002).

It is obviously important to distinguish between the data on flows and the data on stock. The former enable the study of migration in a continuous manner in time, the latter are useful for measuring the result of this in an arrival zone. Indeed, as with the other demographic phenomena, and as with the majority of economic and social phenomena, the first thing to look for in the area of migration is the number of events that occurred during a given time interval. In fact, the principal objective should thus be to know the extent and the nature of entry and exit flows; an adequate accounting of this would also provide an information base on the size of the immigrant population at a given time. However, from a theoretical and practical point of view, the observation of migratory flows requires a well-organized and structured statistical system which is always very difficult to set up. This situation has led to the use of migratory statistics of stock, not only jointly (as one would have wished) but sometimes in an alternative manner, because of the impossibility of direct access to statistics of flow.

In reality, it is not only a question of two different measures of the same phenomenon but of two radically different phenomena. The second, the immigrant population, depends not only on the past dynamics of the other, but also on the effects of mortality. In addition, stock data rarely cover the concept of migration alone because of the very different status accorded to different categories of migrants. Thus, in the United Kingdom and in countries of permanent immigration (North America and Oceania) primary emphasis is on the population born in a foreign country and, possibly, their descendants. In continental Europe, on the other hand, the stress is laid on the foreign population, with a vast range of criteria arising from the various legislative frameworks governing entry and the conditions of residence of foreigners as well as access to naturalization. In France for example, because of the greater ease with which immigrants were able to obtain nationality, the data on the foreign population was not very representative of the extent of the migratory phenomenon. This obliged statistical research to take stock of the concepts used and to consider the direct and indirect effects of immigration (Tribalat *et al*., 1991; Tribalat, 1994). In Germany, on the contrary, the almost insurmountable obstacles erected to prevent naturalization, even for the children of immigrants born in Germany, created the opposite situation, by inflating the foreign population with people born in Germany and thus nonimmigrant. Thus, the more the stock depends on old immigration flows the greater the difference between what is administratively regarded as the immigrant population and the real stock of immigrants in the demographic sense of the term.

II. IDENTIFICATION CRITERIA FOR THE DIFFERENT COMPONENTS OF MIGRATION

Before examining the different systems of observation of international migrant flow and stock, it is useful to introduce some further clarification. First of all, much more than for internal migration, the study of international migration is confronted with the problem of hidden flows, qualified as irregular, illegal migration. But, generally, the various types of migration are perceived differently from one country to another, or even in the same country from one period to another, and the systems and data collection criteria adopted depend on these (Zlotnik, 1987b; Krekels and Poulain, 1996).

At first, in a very schematic way, a country's international migration can be distinguished according to the direction of the move (entry or exit) and the *relationship* of the migrant to the country, defining four components. Table 123–1 illustrates this framework in cases where nationality is taken as the criterion of membership (Natale and Strozza, 1997). For entries, one can for example compare nationals to foreigners or persons from the host country (repatriates, for example) to individuals from other countries. The dis-

TABLE 123–1 Direction of Migration and Relationship of Migrants with Host Country Determined by Nationality

Relationship	Direction of Movement	
	Entry	Exit
Nationals	Immigration of nationals	Emigration of nationals
Foreigners	Immigration of foreigners	Emigration of foreigners

tinction can thus be made on different criteria such as country of birth, nationality or place of residence, each one preferring a different notion of membership of the host country. The nationality criterion is perhaps that which, at least in theory, makes the most useful distinction possible, not only because of its legal and objective nature (at least when it is recorded on the basis of passports or other official documents) but because of its clear limits. On the contrary, the country of birth can be fortuitous and the home country, even when residence is the subject of a legal definition, depends on very different definitions from one country to another, resulting in annoying confusion (Zlotnik, 1987b). Especially for countries with a recent history of emigration abroad, the criterion of nationality appears to be more coherent and more interesting in distinguishing migration flows according to the migrants' relationship to the host country.

The following pages will demonstrate several times that emigration and immigration flows pose different measurement problems. Let us add that each country tends to focus its attention on the most significant migratory component at the time of data collection. Lastly, this results in a better quality of the entire immigration statistics for several reasons. First of all, countries have historically been more concerned about measuring the arrivals and the presence of foreigners on their territory than their departure or the stay of their nationals abroad. Secondly, the principal immigration zones are the more developed countries where statistical systems are more effective. Thus, the criteria for measuring migration flows of arrival have been the object of the greatest care, on the one hand because they attract the attention of the authorities more, and on the other hand because the arrivals occur more in the countries best equipped to observe them. The statistical information available in this area is thus more abundant and of better quality.

The identification of the stock of immigrants during a general data collection exercise (for example, a population census) is not a simple problem to solve. Indeed, it depends as much on the definition of the target population as on the availability of necessary identification information (Strozza *et al*., 2002). The United Nations has many times recommended the definition of the *immigrant population* as people born abroad and residing for more than 1 year in the considered country. More recently, based on the definition of *long-term international migrant*, the United Nations indicated it as the *"stock of international migrants present in a country . . . the set of persons who have ever changed their country of usual residence, that is to say, persons who have spent at least one year of their lives in a country other than the one in which they live at the time the data are gathered"* (United Nations, 1998b, p. 83). As foreigners are almost never accorded the same social, economic and political benefits as nationals, it is important to count the number of persons who do not have the nationality of their country of residence, and even the number of those who, although they have acquired the nationality of the country of residence, did not have it from birth. On the basis of these considerations, the United Nations defined two subpopulations: (1) *"foreigners living in a country . . . [that] includes all persons who have that country as country of usual residence and who are citizens of another country"*; and (2) *"persons born in a country other than the one in which they live . . . [that] includes all persons who have that country as country of usual residence and whose place of birth is located in another country"* (United Nations, 1998b, p. 83).

The definition based on the *criterion of country of birth* makes it possible to take into account all immigrants, irrespective of whether or not they have obtained the nationality of the host country. It excludes, of course, the children of immigrants born in the host country (the "second-generation immigrants," who, contrary to the impression given by this term, are not at all immigrants), whereas it naturally includes the nationals of the host country born abroad, a category which, in a country that has undergone a high level of emigration in the past, could constitute a large group (Strozza *et al*., 2002). This definition makes it possible to identify the population born abroad from one year to the next with the following equation:

$$^{n}P_{t} = {}^{n}P_{t-1} - {}^{n}D_{\Delta t} + {}^{n}I_{\Delta t} - {}^{n}E_{\Delta t},$$

where ($^{n}P_{t}$), the number of persons born in a foreign country present at the end of year t, is equal to the same number ($^{n}P_{t-1}$) present at the end of year ($t - 1$) reduced by the number of deaths ($^{n}D_{\Delta t}$) or emigrations ($^{n}E_{\Delta t}$) and increased by the immigrations ($^{n}I_{\Delta t}$) during year t of persons born in a foreign country.

In the same way, the definition based on the *nationality criterion* makes it possible to identify a foreign population made up of persons who do not have the nationality of the country where they live, irrespective of their place of birth. This population is thus made up of foreigners born abroad and those born in the national territory. It excludes, on the contrary, persons born in a foreign country who have since acquired the nationality of the host country, as well as nationals born abroad. The nationality criterion thus identifies a group made of two components: the first, generally the larger one, is made up of first generation immigrants (immigrant foreigners not yet naturalized), while the second is made up of immigrants of the second or subsequent generations (foreigners born in the host country and not yet naturalized). The portion of descendants of immigrants born and who remain foreigners depends obviously on the national legislation regarding nationality and naturalization: it is minimal when *ius soli*, which enables children of foreigners born in the host country to automatically acquire the nationality of the country, is applied and it is maximal where *ius sanguinis*, which attributes nationality only by filial relation, prevails. Besides acquisition of nationality (${}^{s}A_{\Delta t}$), the size of the population depends also, of course, on births (${}^{s}B_{\Delta t}$) and, in the absence of cases of loss of the country's nationality, one can write:

$$P_t = {}^{s}P_{t-1} + {}^{s}B_{\Delta t} - {}^{s}D_{\Delta t} + {}^{s}I_{\Delta t} - {}^{s}E_{\Delta t} - {}^{s}A_{\Delta t}.$$

Thus, the nationality and the country of birth identifies very different subpopulations. Indeed, *immigrant foreigners*, who at the beginning of the migratory process made up practically the entire population (except in cases of major national repatriations), are reduced with time to a part of the total flanked by a second, that of the *naturalized immigrants*, and a third, that of the *descendants of immigrants* whose number grows with time. This is how in reality the mixing of simple concepts like *immigrant* or *foreigner*, which never coincide, moves them farther and farther away from one another over time and especially only covers one section of the population that one is interested in (Strozza *et al*., 2002).

Furthermore, in countries where a high level of emigration had occurred in the past, as was the case in Europe but especially in the recent past in southern Europe (Italy, Spain, Greece, and Portugal), the differences between the criteria of country of birth and those of current nationality are still very significant. Take the case of Italy. Recent censuses have recorded a considerable difference between the resident population born in a foreign country (almost 1,000,000 in 1981 and around 1,500,000 in 1991) and the foreign population (210,000 in 1981 and 360,000 in 1991). Among those born in a foreign country, there were in reality in those years a great majority of Italians and a minority of foreigners as naturalized persons were few.

Some European countries have decided to simultaneously use the nationality and the country of birth variables, thereby reducing some of the inconvenience related to the separate use of the two criteria.[2] The combination of the criterion of current nationality and that of place of birth is still not enough to identify the foreign population. Among nationals born abroad, it does not enable the distinction between those of foreign origin and those of the national origin.

This difficulty can be illustrated in a diagram in which current nationality is matched with previous nationality (or nationality at birth) with country of birth (Strozza *et al*., 2002). In this way, it is possible to identify eight different groups, four of which are related to the population of foreign origin (S) and four to that of the national origin (N). This diagram (Table 123–2) shows how the foreign population observed at a given time consists not only of foreigners of the first generation (SI) and the second generation (S2), but of emigrants of national origin (N3) or children of emigrants (N4) who have returned to their country of origin but who no longer have the nationality of that country. Furthermore, the national population is composed not only of people who have the nationality from birth, whether they were born in the country (N1) or not (N2), but of naturalized persons of the first generation (S3) and second generation (S4). Consequently, the two central columns of Table 123–2 form what can be called a "grey area" made up of persons who, according to the adopted criteria, can be considered as belonging to the country or not (Strozza *et al*., 2002).

At the European level, the concept of *population of foreign origin* has been proposed. The concrete determination of this should be *based on the combination of many criteria*, which should enable a real comparison between the data from various countries and at the same time make it possible to monitor the whole social process associated with immigration (Krekels and Poulain, 1996 and 1998). The observation of this target population is complex for technical and political reasons. In fact, it requires asking each person his/her nationality at birth or their previous nationality (that which they had in their country of origin), and children of immigrants born in the host country, the nationality or country of birth of their parents. Politically, it is increasingly difficult to ask this type of question, which citizens refuse to answer, as they consider it a statistical intrusion in their private life and an

[2] In particular it is possible to better identify the origin of migrants and thus the membership group especially in host countries which have an important colonial history that is reflected in the nationality of immigrants.

TABLE 123–2 Categories of the Resident Population of a Country: Data Obtained by Matching Current Nationality With Previous Nationality (or Nationality at Birth) with Country of Birth

Country of birth[1]	Previous Nationality or Nationality at Birth			
	Foreign Current Nationality		National Current Nationality	
	Foreign	National	Foreign	National
Foreign	S1 Immigrant foreigners	S3 Immigrant foreigners (first generation) who are naturalized	N4 Born of emigrated nationals (second generation) and have returned to the country with a foreign nationality	N2 Nationals born abroad or born of emigrated nationals (second generation) who have returned to the country
Country studied	S2 Foreigners born of immigrants (second generation)	S4 Born of immigrants (second generation) who are naturalized	N3 Emigrated nationals who have returned to the country after having lost their nationality of origin	N1 Resident nationals who did not migrate

Source: Strozza *et al.*, 2002.
[1] The variable country of birth can be likened to the *de facto* situation (country of actual birth) or to the *de jure* situation (the mother's country of residence at the person's birth).

exposure to the risk of being included in a national database, even if these concerns, widely shared, are in contradiction with the desire or demand by each minority for recognition of its specificity (Golini *et al*., 2001).

In the following section, we will focus on the possibilities of access, through the available sources, to the immigrant population understood as the population born abroad or the foreign population, without forgetting either the risks of incoherence in these criteria for international or temporal comparison or their limitations for the analysis of the process of integration of immigrant communities.

III. A CRITICAL ANALYSIS OF THE MAJOR DATA COLLECTION SYSTEMS

Before examining the possibilities offered by each of the available sources, we must emphasize the fact that most of them cover only one section of the flows and stocks of international migrants: that of legal migrants. The legal conditions of entry and residence thus form another criterion of classification of migratory flows and the immigrant population (foreign or born abroad) on the basis of the legitimacy of the immigration or residence according to the prevailing laws of the host country. In a diagrammatic way, two components can be distinguished at the level of flows and stocks: one legal and the other illegal. For the flows, the distinction is based on the *status* of migrants with regard to the rules of entry into the host country (with or without an entry visa or, more generally, with or without the necessary documents), and for the stock it is based on the norms regarding residence in the territory (with or without a valid resident permit).[3] Naturally, entry without the necessary documents (illegal entry) implies presence in the host country at least temporarily outside the limits of the prevailing laws (illegal presence), and a legal entry should normally, at least in theory, go hand in hand with residence within the limits of the existing norms (legal presence).[4] Statuses do not however remain fixed in their initial state (Briggs, 1984; Strozza, 1996). They can, on the contrary, change during residence in the host country. The longer the duration of stay, the weaker the link

[3] Naturally, the legal situation of immigrants is related to the legislation and even to the policy adopted by each host country with regard to international migration. These provisions can differ significantly from one country to another and can provide for exceptions or variations with regard to certain categories of immigrants or particular immigration zones.

[4] In the case of a given host country, effective ad hoc schematisations have been suggested (see for example, for France that of Yann Moulier-Boutang, Jean-Pierre Garson and Roxana Silberman (1986) in 23 categories and for Greece, that of Angeliki Grammatikopoulou (1998)) which also takes into account certain specific criteria of the prevailing legislation. In addition, Daniel Delaunay and Georges Tapinos (1998) have proposed a dynamic classification of the situation of foreigners based on entry, residence and work.

between entry conditions (legal or illegal) and the residence status: among illegal residents we also find immigrants who entered illegally and are residing outside the prevailing legislation as well as legal immigrants who, beyond the period of validity of their residence permits (initial or subsequent), have remained in the host country without renewing their authorization. On the other hand, a section of the immigrants who entered illegally, may obtain a residence permit, for example after a regularization law has been passed. There again, the situation is different according to whether we are looking at the flow of illegal migration during a given time interval or at the population illegally present in the territory at a given date. Many specific studies, particularly where the phenomenon of illegal residents is very common, have tried to estimate the stock using various methods depending on the available statistical material. In the following analysis, we will examine various sources and determine whether they permit the observation of the illegal component of migrant population flows and stocks.

1. The Census

Despite the improvements recorded in recent years in international migration data collection and dissemination, the census still remains the only source for the large-scale measurement of the intensity of the phenomenon at a global level and of its long-term trend.[5] The census especially makes it possible to highlight and to quantify the stock of immigrant population through questions on country of birth or place of residence at a given date prior to the observation.[6] The latter question, whose main objective is to measure internal migration, also makes it possible to quantify (at least generally) the flow of immigration in the considered time interval (generally 1, 5, or 10 years). Specific questions on nationality make it possible to identify the foreign population, and questions on the acquisition of the nationality of the host country often enable the identification of persons of foreign origin. In some countries, questions on the ethnic group of respondents are also asked. The census can be used to analyze particular flows such as that of return migrants, or to improve knowledge on languages, reasons for stay, etc. In certain cases, data on emigration have been collected (Bilsborrow *et al*., 1997). The latest recommendations of the United Nations (1988a) on censuses list among the obligatory questions useful for the collection of data on internal migration or demographic and social characteristics of the population the duration of presence at the current residence, the place of previous residence, the place of residence at a given date prior to nationality. The optional questions that could be useful in the analysis of migration include: language, national or ethnic group, and year or period of arrival.

These recommendations are not followed by all countries. In fact, it is mainly the developed countries who deviate from them the most by using questionnaires designed to meet the specific information needs of their countries and therefore coincide less with the United Nations recommendations. Besides, for obvious economic reasons, the number of questions must be limited. The census, compared to other sources, has the advantage of being a complete observation capable of collecting from the entire population demographic, economic, and social information that is not available in administrative statistics. These characteristics are particularly useful in the case of relatively rare phenomena like international migration. Admittedly, the census cannot provide all the information, and in-depth analysis of migration demands more specific observation tools. But the collection of certain information from sample surveys and the increasing use of representative samples of census data should, in the case of international migration analysis, be evaluated according to the level of reliability of the data and the extent of information lost in the passage from the total population to a sample. We should also remember that very often the distribution of the immigrant population, according to elementary variables like sex, age, marital status, family composition, level of education, occupation, etc., are only available through the census or, at least, with greater detail than other sources can provide. Furthermore, the census is probably the only official source that enables the collection of information on illegal immigration. Lastly, we must remember that the detailed territorial level of analysis provided by the census makes it a rare instrument for the study of the geographic distribution of immigrant populations, a particularly important point in the study of large metropolitan areas (Champion, 2002).

This instrument, however has well-known limits. They include underenumeration, which for immigrants may be particularly high, especially in the case of illegal immigrants, and have many potential errors, especially relating to the more difficult questions because of the enormity of the task of administering a questionnaire to all the inhabitants of a country. Think

[5] We must not forget that with data from two successive censuses and an estimation of the intercensal natural increase we can obtain an evaluation of the balance of migration, which in its great simplicity often remains the only or the best estimate possible.

[6] In certain cases, questions on previous place of residence and date of arrival at current residence have been used.

of recall lapses, of the difficulties in countries that have had border changes for the determination of place of birth (for older persons especially, for example). Furthermore, repeated migration is not taken into consideration. Whereas, the persons who return to their country before the observation are not counted, as is the case for migrants who die before the census. In the case of the question on residence at a date before the census, we know nothing of the possible migration of the children born in the considered interval. The analysis should be adapted to the duration of the reference period: with the question on the place of birth, it depends for each individual on his/her age, but at the level of the entire population, it yields a measure of immigration over a period of around 100 years.

In conclusion, according to the United Nations, the census is *"the most comprehensive source of data on the types of population stock that are most relevant for both the measurement and the study of international migration. In countries that are major receivers of international migrants, census information provides the basis for many types of in-depth analysis of the economic and social aspects of international migration."* [. . .] *Census data have provided the basis for the most comprehensive set of estimates of the "total migrant stock" in countries of the world"* (United Nations, 1998b, p. 87).

2. Population Registers

Population registers,[7] in the broadest sense of the word, may be defined as systems of continuous observation of demographic and social information on every person resident in a given country or territory (United Nations, 1969). Even if the aim of the observations is administrative, these sources make statistical information available on the dimension and characteristics of the recorded population, as well as on a certain number of events (essentially births, deaths, migration). It is important to stress the fact that the registered population is the resident (*de jure*) population made up of nationals and foreigners. Thus, population registers make it possible, in theory, to measure internal migration. They should also make it possible to measure international migration. Records deleted because of departures abroad, as well as those entered as arrivals from abroad, could indeed be assimilated to international emigrations and immigrations, and in both cases one can separate the national and foreign components. These flows can also be distinguished according to source country in the case of immigration, or destination country in the case of emigration. In addition, it is also possible in many cases to obtain a measure of the stock of foreign population (or immigrants) resident in the territory by using place of birth and nationality as the identification criteria.

However, it must be noted that the countries that have population registers are few[8] and that among these various elements of differentiation make comparability of data difficult and raises questions about their depth and reliability. The few countries with population registers capable of providing information on international migration are also all European. The comparability of recorded information depends on the degree of homogeneity of the provisions adopted in the various countries to determine who is resident and who must be entered into or deleted from the register. The minimum duration of residence required for registration or erasure[9] varies. It could be the effective duration or the programmed duration, depending on whether reference is made to actual permanent residence in the territory or merely declared intentions. At the end of the careful analysis of information gathered by Michel Poulain, Marc Debuisson, and Thierry Eggerickx (1990) on European countries with data on international migration from population registers, *"A problem that is immediately evident is that there are a number of cases where persons moving between country A and country B may be considered emigrants according to A but not immigrants according to B and vice versa"* (Bilsborrow *et al.*, 1997, p. 87). This situation has already been highlighted by other authors (Poulain and Wattelar, 1983; Zlotnik, 1987b). In addition, the elements required for registration are often different not only from one country to another but within the same country, from one population category to another (Poulain *et al.*, 1990). In many cases, apart from the difference between nationals and foreigners, there is also diversity in the treatment of the latter, especially in the case of countries that are members of the same free circulation zone, like the European Union (United Nations, 1998b). Consequently, the categories of migrants (for example, nationals and foreigners) are often problematic because of the differences in registration modalities and the definition of migrants adopted in the population registers (OECD, 2002). The United Nations has proposed that *long-term immigrants*

[7] We are grateful to Alessio Cangiano for collaboration in writing this paragraph.

[8] According to the United Nations, during the 1960s, 65 countries had population registers, 35 of which used them to produce statistics of which only 13 contained information on international migration (Bilsborrow *et al.*, 1997).

[9] This duration varies in general from 3 months to one year. Certain countries distinguish *between short-term* and *long-term migrants*, precisely on the basis of this period, as suggested by the new recommendations of the United Nations on international migration statistics (United Nations, 1998b).

be considered as persons registered in a given year who are still in the register the following year, and *long-term emigrants* as those who have been deleted from the register in a given year and not reregistered the following year (United Nations, 1998b).

It is essential to measure the degree of reliability of available information so as to avoid making very serious mistakes. Indeed, the nature of these sources is inevitably reflected in the typology and quality of the data provided, as well as the major problems associated with the utilization of this type of statistical material need to be recalled. As has been rightly noted (Salt *et al*., 1994), emigration is more difficult to study than immigration, because, in general, recording an arrival in a given territory confers on the immigrant a series of rights and benefits, whereas there is no comparable incentive to inform the authorities of a departure. In addition, it is possible that people remain in the records long after they have left the country, either because they initially had the intention of returning after a short while and then postponed their return, or in the case of foreigners, it is in their interest to remain registered to ensure their readmission after a period of absence (United Nations, 1998b). Thus, because of lack of effectiveness of control procedures, the underestimation of exit flows results in an obvious overestimation of the stock of resident immigrants.

It is also essential to underline the consequences of the very nature of the existing population registers for the quality of information: in certain cases, the system is centralized (Nordic countries, Iceland, Belgium and Luxembourg); in others, it consists *of local systems that are not completely computerized* (Poulain, 1995). While in the latter the archives are managed by local administrators, in the centralized systems information is collected in a national archival center, thereby improving the quality of data available through greater internal controls (for example, in identifying and removing double counts) and matching with other administrative archives (such as tax, medical, social security, etc.). These record-linking procedures were made possible by the attribution of a personal identification number, or PIN, to each individual to be included on every administrative document concerning the person, sometimes along with other codes of more restricted use. The effectiveness of the population registers of northern European countries in estimating the effective migratory flows of nationals and foreigners is also dependent on the considerable efforts made by the national statistical institutions to counter the problem of noncancellation through periodic archival review operations carried out in collaboration with various public organizations and through specific verification surveys on "suspected" cases (Statistics Norway, 2000). In addition, the adoption of a single certificate of a change of residence for movement between the Nordic countries makes further verification of this particular component of migratory movement possible (Poulain, 1995; Kruijsten, 1995). In various countries, the population registers, whether centralized or not, are updated at each population census by linking archived data with those of the census. The ensuing automatic cancellations are obviously not linked to a date in the past, but they make it possible to obtain an estimate of the number of immigrant residents corresponding to the real situation. Furthermore, data on foreign immigration often undergo a time lapse between the moment of arrival in the country and the moment of registration in the population register, sometimes postponed for a few years until the effective stabilization of their presence in the territory or the regularization of their legal situation after a special regularization program has provided illegal immigrants with first temporary then permanent residence permits (Natale and Strozza, 1997).

Lastly, a word must be said regarding the registers of foreigners in certain countries. These registers operate in the same manner as population registers (through a continuous update of the stock of residents on the basis of recorded demographic events), but they perhaps have the advantage of focusing more on the aim of observation. In addition, registers of foreigners also record information on the legal situation of each person (for example, information on the type and duration of the residence permit), which at least should potentially enable an analytical distinction between different categories of international migrants. They also make it possible to more easily observe international migration of foreigners for short periods, which is rarely the case for population registers (except when the period of residence required for registration is less than 3 months). In Germany in particular, the register of foreigners, the *Ausländerzentralregister* (AZR), and in Switzerland the *Zentrales Ausländerregister* (ZAR), are usually used as sources of statistics on the foreign population (Salt *et al*., 1994).

Population registers and registers of foreigners are therefore great sources for the continuous measurement of (long-term) international migration, and it is recommended that countries like those of central and eastern Europe, which inherited this instrument from the past, should try to take these out of their historic role as instruments of population control and into that of providing quality data on external migration. An improvement of the information on the categories of immigrants, which varies from country to country (for example, that of asylum seekers), and the adoption of standardized tabulations among the given countries are simple objectives that can guarantee better comparability of data from the registers, even if doubts persist as to the heterogeneity associated with the different prevailing norms.

3. Visas, Residence Permits,[10] and Other Administrative Sources

Other international migration statistics result from administrative operations intended to control migratory flow (entries and exits) and the presence of immigrants in the national territory. *"Clearly, there is a wide variety of possible administrative sources which differ considerably in terms of modes of operation and types of statistics produced. However, all of them share a common trait: statistics derived therefrom usually reflect administrative procedures rather than people"* (United Nations, 1998, p. 20). Contrary to population registers or registers of foreigners, the statistical unit of observation for the other administrative sources is not the individual but the administrative procedure. In other words, the statistics produced refer to authorizations or other control documents but not directly to individuals concerned. This is one of the key problems of meaning and of reliability associated with data from such archives.

The major administrative records relate to the delivery of passports or exit authorizations from the country, which provide information on the emigration of nationals. The reliability of this depends on the respect of the rules imposed by the country and the category of moves considered. In host countries, the focus is on entry visas and on temporary or permanent residence permits.

Entry visas delivered to foreigners by diplomatic representations of the host country abroad are permission (*nihil obstat*) to cross the border, which must be obtained before migration. Consequently, it is possible to find a difference between the moment the visa is delivered and that of its effective use, just as it is possible for the visa to remain unused. It must also be noted that nationals of countries belonging to the same free circulation zone do not need a visa for their movement in the countries within the zone, and this is also the case between countries with bilateral agreements with a waiver on the visa obligation. Inversely, a visa, the most important limit, is required for a whole series of moves that are not considered migration (in particular, tourist or business moves, etc.). It is often very difficult to use these sources of information because of the complexity of separating migrants from nonmigrants, and in addition, international comparison is impossible because of the variability in the provisions made by the different countries.

Chronologically, after the entry visa comes the temporary or permanent residence permit for foreigners intending to remain in the territory of a host country for a period greater than a given number of months (this duration differs from one country to another; in certain cases, it is limited to 1 month). It must be stressed that the temporary residence permit, contrary to the entry visa, is most often required even for nationals of countries belonging to the same free circulation zone as the host country. Nationals of member countries of the European Union for example, do not need an entry visa for movement within the community, but they need to request a residence permit for a stay of over 3 months. The statistics of residence permits therefore constitutes a useful information base for measuring the flow of immigrant foreigners and, under certain conditions, give information on the size of legal foreign immigration in the country. Information on the type of the permits granted and on their length gives the opportunity to single out specific categories of immigrants and, sometimes, of estimating long-term and short-term flows.

As has been stressed earlier, the main limitation of administrative statistics and, in fact, that of the residence permit, is in terms of the procedures adopted for the management of the archives, and these are different from those used for population registers. There is no continuous update of the status of each foreigner with a residence permit. It is possible for the same person to have several residence permits delivered successively during a period between which no link is made (double counts are possible). It is rare to find procedures for checking that foreigners with valid residence permits are still present in the territory. Finally, even when foreigners leaving the country are required to return their residence permit, this obligation is often not fulfilled. Consequently, the total number of valid residence permits at a given moment[11] generally overestimates the real number of foreigners legally present in the country, at least, those whose legal presence require a residence permit. Indeed, people above a certain age, of variable threshold from

[10] The type and names of permits granted to foreigners vary considerably between countries; we use the general expression *residence permits* to indicate all these kinds of permits.

[11] Naturally, it is not possible to deduce the stock of foreigners legally present in cases where information on the date of expiry of the residence permit is not recorded. Even when this information exists, in certain cases, published national and international data include residence permits showing considerable overestimation of the phenomenon. This is the case for example, of Italy which, before the ISTAT revision of the archives of the Ministry of internal affairs, used to publish data with 20% to 40% expired residence permits (ISTAT, 1998).

one country to the other but generally between 14 and 18 years, are rarely given their own residence permit but depend on that of their parents. This population is thus almost always underestimated, as was indicated earlier, because only persons with documents are taken into account. For these reasons, only a few countries use data on stock of residence permits: in particular, in the SOPEMI report (OECD, various years) data on foreign population were treated from residence permits only from the new host countries of Mediterranean Europe (Italy, Spain, Portugal, and, at least in the past, Greece).

In addition, there are many interpretation problems associated with the use of residence permits to measure the flow of legal foreign immigration. In various cases, the available data on flows covers not only new permits delivered during the year to foreigners who have just arrived in the country, but includes permits delivered to foreigners already present in the territory as renewal or as change of status (change of type of residence permit). We must not forget that the number of permits delivered in one year is not equivalent to the number of persons admitted in that year either, as one person can receive more than one permit in the course of the same year, just as inversely, the permit delivered to the head of family can be used by other members of the family.

When information on the outflow of foreigners is inadequate or nonexistent it has consequences for the reliability of measures of stock. In fact, it is the entry flows that we try to match with the number of new permits and the renewals of permits, or a combination of both. In general, the countries that publish statistics on the number of residence permits delivered in a year do not provide enough information on the exact significance of the published data. The available statistics are often difficult to interpret and in any case are less able to represent the reality of the flow of persons than those of change of status of holders of residence permits (United Nations, 1998b). In some countries, however, information derived from residence permits or other documents regulating entry and residence of foreigners are the only available source of continuous observation of international migration and the estimation of the stock of legal foreigners.

"There is considerable heterogeneity in the statistics derived from residence permits, visas or other equivalent documents. [. . .] Because residence permit statistics are administrative in nature, they are closely related to the application of laws and regulations relative to the admission and stay of foreigners and reflect the idiosyncrasies of such regulations. A full understanding of the statistics available usually demands some knowledge of how those laws and regulations operate. [. . .] Statistics derived from the issuance of residence permits can be used to provide two type of measures: those relative to stocks, which usually reflect the number of valid residence permits at a given point in time, and those relative to the different flows that bring about changes in the stock issuance" (Bilsborrow *et al*., 1997, p. 128).

4. Statistics of Border Flows

The control of border movements is the most direct source of information on a country's international entries and exits. For a long time, it was also the focus of United Nations recommendations concerning international migration statistics. The major limitation of this type of observation resides in the difficulty in identifying migrants among all the international movements of persons including movements for other reasons. This problem has become more difficult to resolve as movements of persons have become more intense and the crossing of borders have increased in size. Data collection operations are even more complex along land borders, where the intensity of movements is often higher and the customs formalities are reduced to a minimum or are virtually nonexistent and can therefore be easily circumvented. The solutions used to adapt the source to the growth in volume of border crossings are a combination of administrative procedures, selective checks, or the use of samples as is the case with the *International Passengers Survey* of the United Kingdom. The first case consists of a mixed system of observation, which at the same time as border controls, uses data produced by different administrative procedures. In the case of Canada and the United States, this second element seems to take precedence over the first, as neither of the two countries produces statistics on the emigration and immigration of its own nationals or of foreigners. On the contrary, Australia and New Zealand *"gather border statistics that reflect the full range of possible movements, including those of citizens and former settlers, and those of foreigners admitted on a temporary basis"* (Bilsborrow *et al*., 1987, p. 121).

The main problem is identifying migrants among all travelers who cross borders. The usual criterion is based on length of stay, actual or assumed, in order to distinguish short- or long-term migrants from other travelers. In practice however, the purpose of the stay is often taken into account, especially for foreigners on the basis of their type of visa or entry permit. An efficient system for collecting data is based on a form to be filled in, in two steps: the first time at entry and the second time at departure. A problem here is the possibility of counting several times migrants who cross the border multiple times during the same year.

The countries without land borders have obvious advantages in the use of this system of observation, since the concentration of flows at ports and airports can be checked more easily. The collection of information can be based on administrative or statistical criteria. In the first case, the characteristics of migrants are recorded from documents (passports, visas, residence permits, etc.). In the second case, the information is derived from forms filled by the person at entry. The data collected at border checks reflect precisely the moment when the migration event took place. It is improbable that they can guarantee the same precision in the measurement of the intensity of flows, especially exit flows. *"In practice, statistics derived from border collection rarely provide the best measures of international migration flows because of the difficulties involved in gathering reliable information from a large volume of people subject to different degrees of control depending on their citizenship, mode of transport and port of entry"* (United Nations, 1998b, p. 23). Let us add that these statistics tend in any case to reflect only the legal situation and those migrants who cross the border at authorized points. It is highly improbable that at a border crossing one declares a status different from that established by the entry or exit documents.

The strict dependence on legal conditions is at the same time the price and the limitation of this type of statistical observation. On the one hand, it does indeed enable the collection of precise and detailed information on legal migration, on the other hand it is hardly possible to draw a credible conclusion on the exact extent of the real migration flows. In particular, entry flows are generally underestimated in statistics of border crossings, thereby giving low reliability to the estimation of the migratory balance, whereas the typology used for the data collection tends to reflect almost exclusively the targeted categories of migration policies. Besides, these categories are not always easy to define on the basis of the vocabulary used in official statistical publications, where the same word can relate to different categories. The result is a false impression of comparability between countries and of continuity within the same country where this type of observation is in use, as has been shown, for example, by a study of passenger cards in Australia and New Zealand (Carmichael, 1993).

5. Sample Surveys

Very detailed information on migrants' characteristics and, in certain cases, the causes and consequences of international migration, can be collected through sample surveys. Actually, in countries with a high immigrant population, it is possible to isolate within large family surveys the subpopulation of persons born abroad or who are foreigners. The most significant example is certainly that of the *Current Population Survey* (CPS), executed monthly in the United States. This survey enables the estimation, among other things, of the size of the population born abroad, including its illegal component, and even allows identification of the various countries of birth. Labor force sample surveys standardized in the European Union should make it possible to obtain representative data on foreigners, at least in countries where this population constitutes a large part of the total population.[12] In general, although these routine observations facilitate comparison on the dimension and characteristics of migrant and nonmigrant populations, they do not however allow covering very specific aspects of the migratory process. *"Thus, although it is common to use census or survey data to compare the status of international migrants with that of non migrants in terms of occupation, income, unemployment level, fertility, etc., and to interpret similarities between the two groups as indicative of the migrants' success, in reality such similarities reflect, at best, only the process of integration"* (Bilsborrow *et al*., 1997, p. 240). For a more detailed analysis of the related processes of international migration, it is necessary to have access to information on the situation prior to the migratory event in order to compare with the situation at the moment of observation. This information can hardly be derived from a general data collection operation aimed above all at representing the entire population. In addition, only in a few countries is the population of immigrants sufficiently large enough for a credible representation to be drawn from a general sample, and rarer still are the cases where it is possible to collect detail on the country of origin and the nationality of immigrants.

For further analyses, *ad hoc* surveys are necessary and some examples of operations of this type conducted either in the migrant's country of origin or the host country will be given here. Let us examine three specific surveys, which, in a certain manner, illustrate three different approaches. The first is the *Encuesta sobre migración en la frontiera norte de México,* initiated in 1993 within the context of a collaboration between Mexico and the United States (STPS, 2001). The objective of this survey is to study the dynamics of population movements between Mexico and the United States with a particular focus on temporary migration and return migration. Within the framework of a complex

[12] The situation is different when the proportion of foreigners is small. In Italy, for example, ISTAT collects information on resident foreigners from labour force surveys, but these data are not published as their level of significance is too low.

sampling method, specific weights were applied to observations made in each of the principal border crossing points between the two countries in proportion to the passage flows. A control group, which would have captured the specificity of the characteristics of the migrant populations, is however missing in this case. This reference population would have by comparison made it possible to draw some conclusions on the determinants of migration and the conditions of evolution of migrant populations.

The second example is provided by a system of sample surveys funded by Eurostat and coordinated by the *Netherlands Interdisciplinary Demographic Institute* (NIDI). It was carried out among migrants in some countries of origin (Egypt, Ghana, Morocco, Senegal, and Turkey) and some host countries (Italy and Spain) with the aim of analyzing the causes of international migration (Schoorl *et al*., 2000; Birindelli *et al*., 2000). Each survey employed a specific technique. In particular, the survey on Ghanaian and Egyptian immigrants carried out in Italy used the technique of centers and points of aggregation that will be discussed later.

The third example is that of the *Geographic Mobility and Integration* sample survey, carried out in France in 1992 by Institut national d'études démographiques (INED) in collaboration with Institut National de la Statistique et des Etudes Économiques (INSEE) on the living conditions of immigrants and their children. Three categories were considered: foreign immigrants (from seven zones: Algeria, Spain, Portugal, Morocco, Tunisia, Black Africa, and Southeast Asia), the children of immigrants born in France (persons below 30 years of Algerian, Spanish, or Portuguese origin), and a control sample of French people. The detailed questionnaires were built around a series of retrospective questions aimed at reconstructing the migratory, residential, marital, reproductive, and professional biographies of each respondent (Tribalat, 1995).

In recent years, especially in traditional European host countries, there has been a growing interest in designing and carrying out longitudinal (multiround) surveys on immigrants (Bilsborrow *et al*., 1997). In this sense, the *New Immigrant Survey* (NIS) set up in 2003 in the United States is particularly interesting (Jasso *et al*., 2003). It is a nationally representative longitudinal survey of immigrants who have obtained a permanent residence permit during the years of observation. The sample is extracted from the administrative archives of the *Immigration and Naturalization Service* (INS), and in 2003 consisted of around 11,000 persons interviewed immediately after obtaining their permanent residence permit and then re-interviewed periodically. This survey was aimed at collecting information on the multiple aspects of migration and measuring the process of integration of new arrivals. The use of the longitudinal approach is meant to control for the problems of selection posed by emigration outside the United States by some of the immigrants (Jasso *et al*., 2000).

In the new European host countries of international migration, many field surveys limited to specific contexts or particular foreign communities have been conducted in the past 15 years. In particular, in many countries like Italy, the extent of the presence of illegal immigrants has led to the testing of innovative sampling techniques aimed at observing rare populations or populations that are normally difficult to access. From sampling frames already adopted in other circumstances,[13] a new technique called *"sampling by centers and points of aggregation"* has been developed which, based on identification of the major points of passage of immigrants, makes it possible to allocate weights on respondents depending on their presence at these passage points (Blangiardo, 1996, and 2000). This technique has enabled the quantitative estimation of illegal migration and improved knowledge on its characteristics, and specific dynamism by comparison with legal migration. Nevertheless, the question regarding whether the data will make it possible to compare these characteristics and behaviors with those of the general population still remains.

CONCLUSION

A fully satisfactory solution is yet to be found to the problems of conceptualization and observation of international migration statistics, even in statistically advanced countries. Here, we have focused our observations mainly on the latter, although some conclusions are more general. The reason is that for the study of international migration, these statistically more advanced countries are also generally host countries, where it is not only possible to observe the greatest flows and have access to the most reliable sources, but it is also possible to do so on at least one of the two angles under which migration can be measured.

We have tried to provide a panorama of the problems encountered, but we are far from covering all the details of the various sources of statistical information available on different aspects of the phenomenon. Official sources provide measures of the extent and the characteristics of the migratory flows and of the immigrant population, but the level of detail, the quality,

[13] Like the *snowball sampling* for illegal Mexicans present in the United States (Cornelius, 1982).

and the degree of reliability, although generally quite low, varies reliability. Negative factors are many and varied. A major aspect comes from the very nature of the sources of information (statistics or administrative) and the prevailing norms regarding immigration and the presence of new arrivals in the national territory. It is rare, for example, that administrative sources enable access to illegal immigrants, who can be accessed, albeit at an unsatisfactory coverage, by censuses or certain types of sample surveys.

We have highlighted the existence of many grey areas around statistical information on international migration. The definitions and concepts are far from being stable in time and space. Obviously, this fluidity does not facilitate the task of the user and makes the use of statistics built on slipping concepts (migrant or migration, stock or flow, native or foreigner, etc.) of variable definitions unavoidable. This state of affairs inevitably reduces the degree of comparability of data and, in the absence of a careful evaluation of these difficulties, makes them unsuitable or even incorrect for use.

It is essential to undertake a meticulous evaluation of the real information from each source according to the categories of migrants (or of migration) and its adequacy to the studied phenomenon. Administrative records hardly allow anything beyond a descriptive analysis of the phenomenon, making it possible, at best, to follow certain dimensions of the migratory process more or less regularly. Only *ad hoc* surveys make it possible to contemplate looking further into the analysis of the migration and going from the purely descriptive dimension to research on causal relations and functional linkages. We must however not forget the usefulness of traditional sources for the construction of indicators capable of providing a general description of the phenomenon and an evaluation of basic trends. This information does not definitively allow for a full understanding of the causes and effects of the phenomenon, but that notwithstanding, it retains its importance especially if we succeed in using it within a conceptual framework capable of developing its specificities and synergies.

In the future, it will be possible to check if the new recommendations of the United Nations have had greater success than previous ones. In any case, we can expect a growing interest in problems of measurement of international migration. Indeed, at the national level, migratory movements play an increasing role next to natural increase within the total dynamics of population. Without it, many countries that would have been undergoing population decline continue to witness population growth due to the arrival of foreigners, while others are witnessing a reduction in their population growth due to emigration. More generally, the contribution of migration is that of a greater selection of the population compared to the population of origin or the host population. Migrants are different in their demographic, social, and economic structure, factors that encourage us to pay more attention to migratory phenomena.

Acknowledgment

This chapter was translated from French by Accenta Ltd.

References

Bilsborrow Richard E., Hugo Graeme, Oberai Amarjit S. and Zlotnik Hania, 1997. *International migration statistics. Guidelines for improving data collection systems.* Geneva, International Labor Office, 441 p.

Birindelli Anna-Maria, Blangiardo Gian Carlo, Bonifazi Corrado, Caruso Maria G., Cesarini Sforza Letizia, Farina Patrizia, Gesano Giuseppe, Grillo Francesca, Kouider Samia, Savioli Maria and Terzera Laura, 2000. *Push and pull factors of international migration: country report for Italy.* Luxembourg, Eurostat, xx + 122 p. (Working Papers on Population and Social Conditions, no. 5).

Birindelli Anna-Maria and Gesano Giuseppe, 1991. La mobilità interna ed internazionale della popolazione, *in:* Giorgio Alleva and Maria G. Ottaviani (eds.), *La statistica italiana per l'Europa del 1993. Atti del Convegno di studi, Roma 21–23 maggio 1990*, p. 527–556. Rome, Litaled, 696 p.

Blangiardo Gian Carlo, 1996. Il campionamento per centri o ambienti di aggregazione nelle indagini sulla presenza straniera, *in: Studi in onore di Giampiero Landenna*, p. 13–20. Milan, Giuffré Editore, xviii + 541 p.

Blangiardo Gian Carlo, 2000. Appendix: methodological notes on sampling technique, *in:* Anna-Maria Birindelli, Gian Carlo Blangiardo, Corrado Bonifazi, Maria G. Caruso, Letizia Cesarini Sforza, Patrizia Farina, Giuseppe Gesano, Francesca Grillo, Samia Kouider, Miria Savioli and Laura Terzera, *Push and pull factors of international migration: country report for Italy*, p. 107–117. Luxembourg, Eurostat, xx + 122 p. (Working Papers on Population and Social Conditions, no. 5).

Bonifazi Corrado, 1998. *L'immigrazione straniera in Italia.* Bologne, Il Mulino, 293 p.

Bonifazi Corrado and Strozza Salvatore, 2002. International migration in Europe in the last fifty years, *in:* Corrado Bonifazi and Giuseppe Gesano (eds.), *Contributions to international migration studies*, p. 33–105. Rome, IRP-CNR, 159 p.

Briggs Vernon M. Jr, 1984. Methods of analysis of illegal immigration into the United States, *International Migration Review*, vol. 18(3), p. 623–641.

Carmichael Gordon A., 1993. Beware the passenger card! Australian and New Zealand data on population movement between the two countries, *International Migration Review*, vol. 27(4), p. 829–849.

Champion Anthony G., 2002. *Use of census data for measuring and monitoring international migrations in the United Kingdom.* Newcastle upon Tyne, University of Newcastle, 20 p. (Presented to the Euro-Mediterranean workshop on International migration Statistics, Rome, September 30–October 2, 2002).

Cornelius Wayne A., 1982. Interviewing undocumented immigrants: methodological reflections based on fieldwork in

Mexico and US, *International Migration Review*, vol. 16(2), p. 378–411.
Costanzo Joseph M., Davis Cynthia J. and Malone Nolan, 2002. *Guide to international migration statistics: the sources, collection, and processing data at the U.S. Census Bureau.* Washington, US Census Bureau, Population Division, 31 p. (Population Division Working Paper, no. 68; http://www.census.gov/population/www/documentation/twps0068/twps0068.html)
Courgeau Daniel, 1980. *Analyse quantitative des migrations humaines.* Paris, Masson, 225 p. (Anthropologie physique, no. 12).
Delaunay Daniel and Tapinos Georges (eds.), 1998. *La mesure de la migration clandestine en Europe, Vol. I e II.* Luxembourg, Eurostat, vol. 1, 104 p.; vol. 2, 172 p. (Eurostat Working Papers, no. 3).
Golini Antonio, 1987. Population movements. Typology and data collection, trends, policies, *in:* IUSSP, EAPS and FINNCO (eds.), *Plenaries of the European population conference 1987: issues and prospects*, p. 113–200. Helsinki, Central Statistical Office of Finland, IX + 434 p.
Golini Antonio, Strozza Salvatore and Amato Flavia, 2001. Un sistema di indicatori di integrazione: primo tentativo di costruzione, *in:* Giovanna Zincone (ed.), *Secondo rapporto sull'integrazione degli immigrati in Italia*, p. 85–153. Bologne, Il Mulino, 838 p.
Gramatikopoulou Angeliki, 1998. La Grèce, *in:* Daniel Delaunay and Georges Tapinos (eds.), *La mesure de la migration clandestine en Europe, Vol. II*, p. 1–57. Luxembourg, Eurostat, 172 p. (Eurostat Working Papers, no. 3).
Herm Anne and Poulain Michel, 2003. *Le registre de population, principale source de données démographiques au sein de l'Europe des 25.* Louvain-la-Neuve, Université catholique de Louvain, 25 diapos. (PowerPoint presentation at the Quetelet seminar 2003: "Collecte, qualité et harmonisation des données socio-démographiques").
Hoffmann Eivind and Lawrence Sophia, 1996. *Statistics on international migration: a review of sources and methodological issues.* Geneva, ILO, V + 110 p. (http://www.ilo.org).
International Organization for Migrations (IOM), 2000. *World Migration Report 2000, vol. I.* Geneva, IOM, and United Nations.
Istituto Nazionale di Statistica (ISTAT), 1998. *La presenza straniera in Italia negli anni '90.* Rome, ISTAT, 550 p. (Informazioni, no. 61).
Jasso Guillermina, Massey Douglas S., Rosenzweig Mark R. and Smith James P., 2000. The new immigrant survey pilot (NIS-P): overview and new findings about U.S. legal immigrants at admission, *Demography*, vol. 37(1), p. 127–138.
Jasso Guillermina, Massey Douglas S., Rosenzweig Mark R. and Smith James P., 2003. The new immigrant survey in the US: the experience over time, *Migration Information Source*, January. (http://www.migrationinformation.org/feature/display.cfm?ID=81).
Kraly Ellen P. and Gnanasekaran Kottai S., 1987. Efforts to improve international migration statistics: a historical perspective, *International Migration Review*, vol. 21(4), p. 967–995.
Krekels Barbara and Poulain Michel, 1996. Population d'origine étrangère. La compatibilité internationale des divers concepts utilisés, *Espace Populations sociétés*, no. 2–3, p. 257–270.
Krekels Barbara and Poulain Michel, 1998. *Stock de migrants et population d'origine étrangère. Comparaison des concepts dans le pays de l'Union européenne.* Luxembourg, Eurostat, 89 p. (Eurostat Working Papers, no. 4).
Kritz Mary M., 1987. International migration policies: conceptual problems, *International Migration Review*, vol. 21(4), p. 947–964.
Kuijsten Anton C., 1995. International migrations: harmonization of data collection, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 593–624. Louvain-la-Neuve, Academia/L'Harmattan, 671 p. (Actes de la Chaire Quetelet 1991).
Moulier-Boutang Yann, Garson Jean-Pierre and Silberman Roxane, 1986. *Economie politique des migrations clandestines de main-d'oeuvre.* Paris, Publisud, 276 p.
Natale Marcello and Strozza Salvatore, 1997. *Gli immigrati stranieri in Italia. Quanti sono, chi sono, come vivono?* Bari, Cacucci, 507 p.
Nations unies/united nations, 1949. *Problèmes de la statistique des migrations.* Lake Success (New York), United Nations, VIII + 66 p. (Études démographiques, no. 5, ST/SOA/Série A/5).
Nations unies/united nations, 1998. *Recommandations en matière de statistiques des migrations internationales. Première révision.* New York, United Nations/Nations unies, 90 p. (Études statistiques, Série M, no. 58, rev.1).
Petersen Jørn K., 2000. *The Danish demographic database: Longitudinal data for advanced demographic methods.* Odense, Danish Center for Demographic Research, 28 p. (Research Report, no. 15).
Petersen William, 1958. A general typology of migration, *American Sociological Review*, vol. 23(3), p. 256–266.
Poulain Michel, 1991. Un projet d'harmonisation des statistiques de migration international au sein de la communauté européenne, *Revue Européenne des Migrations Internationales*, vol. 7(2), p. 115–138.
Poulain Michel, 1993. Confrontation des statistiques de migrations intra-européennes: vers plus d'harmonisation? *European Journal of Population*, vol. 9(4), p. 353–381.
Poulain Michel, 1995. Mesurer les migrations à l'aide d'un registre de population: quelle comparabilité européenne?, *in:* Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démo-sociales en Europe*, p. 553–571. Louvain-la-Neuve, Paris, Academia, L'Harmattan, 671 p. (Chaire Quetelet 1991).
Poulain Michel, 2001. *Comparing data sources for measuring international migration in Europe.* Louvain-La-Neuve, Département SED, 28 p. (Presented to "MED-Migr workshop on administrative and statistical sources for migration statistics", Paris, April 25–27, 2001).
Poulain Michel, Debuisson Marc and Eggerickx Thierry, 1990. *Projet d'harmonisation de statistiques des migration internationales au sein de la Communauté Européenne.* Louvain-la-Neuve, Université Catholique de Louvain, 5 volumes.
Poulain Michel and Wattelar Christine, 1983. Les Migrations intra-Européennes: a la recherche d'un fil d'Ariane au travers des 21 Pays du Conseil de l'Europe, *Espace Populations Sociétés*, vol. 1(2), p. 11–26.
Salt John, Singleton Ann and Hogarth Jenifer, 1994. *Europe's international migrants. Data sources, patterns and trends.* London, HMSO, 223 p.
Schoorl Jeannette J., Heering Liesbeth, Esveldt Ingrid, Groenewold George, van der Erf Rob F., Bosch Alinda M., de Valk Helga and de Bruijn Bart J., 2000. *Push and pull factors of international migration: a comparative report.* The Hague and Luxembourg, Eurostat, XXI + 161 p. (Theme 1: General Statistics).
Simon Patrick, 1998. Nationalité et origine dans la statistique française, *Population*, vol. 53(3), p. 541–568.
Système d'Observation Permanente des Migrations (SOPEMI), 2002. *Trends in international migration: continuous reporting system on migration. Annual report: 2002 edition.* Paris, Organisation for Economic Co-operation and Development (OECD), 372 p.
Statistics Netherlands, 1998. *Statistics of the population with a foreign background, based on population register data.* The Hague, Statistics Netherlands, 8 p. (Presented to the Joint ECE-Eurostat-ESCWA Work Session on Migration Statistics, Geneva, March 25–27, 1998).

Statistics Norway, 2000. *New developments in defining and counting international migrations in population registers.* The Hague, Statistics Netherlands, 4 p. (Presented to the Joint ECE-Eurostat-ESCWA Work Session on Migration Statistics, Geneva, May 8–10, 2000).

Strozza Salvatore, 1996. Possibilità di quantificazione della presenza straniera in Italia: il punto della situazione, *in:* Luigi Di Comite and Antonio Cardamone (eds.), *Crescita demografica e migrazioni nel bacino Mediterraneo. Quaderni no. 11*, p. 227–261. Bari, Università di Bari, Dipartimento per lo studio delle società mediterranee and Cacucci Editore, 264 p.

Strozza Salvatore, Natale Marcello, Todisco Enrico and Ballacci Francesca, 2002. *La rilevazione delle migrazioni internazionali e la predisposizione di un sistema informativo sugli stranieri. Rapporto finale di ricerca per la Commissione per la Garanzia dell'Informazione Statistica.* Rome, Commissione per la Garanzia dell'Informazione Statistica, 229 p.

Secretaria del Trabajo y Prevision Social (STPS), 2001. *Encuesta sobre migración en la frontera norte de México 1998–1999.* Tijuana, Consejo Nacional de Poblacion, El Colegio de la Frontera Norte, 428 p.

Thierry Xavier, 2004. *L'utilisation des micro-données relatives à la délivrance des titres de séjour en France, un changement de cap?* Paris, INED, 7 p. (Paper presented at the AIDELF meeting on "Les Migrations internationales. Observation, analyse et perspectives" Budapest, September 20–24, 2004).

Tribalat Michèle, 1994. Mise au point, *Population & sociétés*, no. 291, p. 1–4.

Tribalat Michèle, 1995. *Faire France. Une enquête sur les immigrés et leurs enfants.* Paris, La Découverte, 232 p.

Tribalat Michèle, Garson Jean-Pierre, Moulier-Boutang Yann and Silberman Roxane, 1991. *Cent ans d'immigration, étrangers d'hier français d'aujourd'hui.* Paris, INED and PUF, 302 p. (Travaux et document, Cahier no. 131).

United Nations, 1949. *Problems of migration statistics.* Lake Success (New York), United Nations, Department of Social Affairs, 71 p. (Population Studies, no. 5, doc. ST/SOA/SER.A/5).

United Nations, 1969. *Methodology and evaluation of population registers and similar systems.* New York, United Nations, 80 p. (Studies in Methods, Series F, no. 15).

United Nations, 1978. *Demographic Yearbook 1977.* New York, United Nations, x + 1024 p. (Twenty-ninth issue, Special topic: International Migration Statistics).

United Nations, 1991. *Demographic Yearbook 1989.* New York, United Nations, x + 926p. (Forty-first issue, Special topic: International Migration Statistics).

United Nations, 1998a. *Principles and recommendations for population and housing censuses. Revision 1.* New York, United Nations, xvi + 274 p. (Statistical papers, series M, 67, rev. 1).

United Nations, 1998b. *Recommendations on statistics of international migration. Revision 1.* New York, United Nations/Nations unies, 96 p. (Statistical Papers, Series M, 58, Rev. 1).

United Nations, 2002. *International migration report 2002.* New York, United Nations, Department of Economic and Social Affairs, xiv + 324 p.

Verhoef Rolf and Van de Kaa Dirk J., 1987. Population registers and population statistics, *Population Index*, vol. 53(4), p. 633–642.

Willekens Frans, 1984. *Comparability of migration. Utopia or reality?* Vooburg, NIDI, v + 36 p. (Nidi Working Paper, no. 47).

Willekens Frans, 1994. Monitoring international migration flows in Europe, *European Journal of Population*, vol. 10, no. 1, p. 1–42.

Zlotnik Hania, 1987a. Introduction, *International Migration Review*, vol. 21(4), p. v–xii. (Special issue edited in collaboration with IUSSP, "Measuring international migration: theory and practice").

Zlotnik Hania, 1987b. The concept of international migration as reflected in data collection system, *International Migration Review*, vol. 21(4), p. 925–946. (Special issue edited in collaboration with IUSSP, "Measuring international migration: theory and practice").

Zolberg Aristide R., 1989. The next waves: migration theory for a changing world, *International Migration Review*, vol. 23(3), p. 403–430.

C H A P T E R

124

Demographic Surveillance Sites: A Tool for Studying Demographic and Health Trends in Developing Countries

GILLES PISON

Institut national d'études démographiques (INED), Paris, France

As explained in Chapter 121, a demographic surveillance site is a system for gathering population data on a town or set of villages over a period of several years or decades by regularly gathering information on the life events—births, deaths, marriages and migration—that take place there. A demographic surveillance site differs from a cohort study in that the entire population of the area is monitored rather than a fraction of it.

A growing number of demographic surveillance sites have been established in developing countries since the late 1980s, in response to two requirements.

In many countries, precise data on population dynamics are lacking because public record systems are inadequate. Surveys, particularly demographic and health surveys (DHSs), partly fill the gap by indicating trends, especially where several successive observations are available. But these surveys only provide observations from two or three time points at best, and provide little information about factors underlying the trends. An in-depth study of a small area over a long period, although it gives only local information, can provide better measurement and understanding of the causes underlying major demographic trends.

Secondly, many developing countries have no reliable health statistics, and research based on hospital data gives a biased picture of the health situation since part of the population never, or only exceptionally, go to a hospital. Demographic surveillance improves knowledge of a population's health, as it allows studies of a whole population that can provide more accurate data on health issues. Demographic surveillance maintained for several years or decades can serve as a watchdog mechanism for the public health system. It is also an excellent context for testing and assessing new vaccines, therapies or healthcare strategies.

Following the brief presentation in Chapter 121, this chapter examines this innovative data gathering method in greater detail, with the kinds of results it can produce and its usefulness for health development and for studying demographic change. However, every demographic surveillance site has its own system and these systems vary widely. The first part of this chapter therefore reviews the demographic surveillance sites in developing countries, examining some of their particular methodological features in greater detail. The second part presents two specific examples, at Bandafassi and Mlomp, in Senegal.

I. A GROWING NUMBER OF DEMOGRAPHIC SURVEILLANCE SITES IN DEVELOPING COUNTRIES

There were several tens of demographic surveillance sites in operation in developing countries in

2003. Table 124–1 lists 22 of them, with their main characteristics (size of monitored population, start date and duration of monitoring, frequency of visits, etc.) and two mortality indicators: mortality in children under 5 years of age (per thousand) and life expectancy at birth for males and females. All sites in the list are members of the International Network of field sites with continuous Demographic Evaluation of Populations and Their Health (INDEPTH) network, which was formed in 1998 and had 31 member sites in 2003 (INDEPTH network, 2002). Table 124–1 is therefore by no means exhaustive, especially as not all monitoring studies are members of INDEPTH. However, the dividing line between a demographic surveillance site proper and a simple epidemiological surveillance system is not always clear, and the list in Table 124–1 gives a good idea of the characteristics of systems that can definitely be classed as demographic surveillance sites.

1. Renewed Interest in Demographic Surveillance Sites in the 1990s

Few existing demographic surveillance sites were established before 1990 (see Table 124–1). They were long regarded as relatively costly studies for the results they produced. They were particularly criticized for covering only subpopulations that were not represen-

TABLE 124–1 Current demographic surveillance sites in developing countries, 2003 (selection of INDEPTH network member sites)

Country	Name of study	Population in 2000	Setting: rural (R), urban (U) or periurban (P)	Start date	Duration of study to 2003 (years)	Interval between visits (months)	Mortality (1995–1999)		
							$_5q_0$ (per 1,000)[1]	e_0 (years)[2]	
								Female	Male
East and Southern Africa									
Ethiopia	Butajira	40,000	R, U	1987	16	3	119	57	56
Mozambique	Manhica	36,000	P, R	1996	7	4	148	58	47
South Africa	Agincourt	67,000	R	1992	11	12	51	69	63
South Africa	Digkale	8,000	P	1995	8	12	45	68	62
South Africa	Hlabisa	85,000	R	2000	3	4	—	—	—
Tanzania	Dar es Salaam	70,000	U	1992	11	6	114	50	50
Tanzania	Hai	154,000	R, P	1992	11	12	92	63	56
Tanzania	Ifakara	60,000	R	1996	7	4	124	58	56
Tanzania	Morogoro	120,000	R	1992	11	12	183	46	44
Tanzania	Rufiji	85,000	R	1998	5	4	179	52	53
Uganda	Rakai	42,000	R	1988	15	10	—	—	—
West Africa									
Burkina-Faso	Nouna	55,000	U, P	1992	11	3	137	53	54
Burkina-Faso	Oubritenga	100,000	R	1993	10	12	188	55	52
Gambia	Farafenni	16,000	U	1981	22	3	170	55	51
Ghana	Navrongo	140,000	R	1993	10	3	183	51	47
Guinea-Bissau	Bandim	100,000	U, R	1978	25	3	226	39	36
Senegal	Bandafassi	10,500	R	1970	33	12	253	48	45
Senegal	Mlomp	7,600	R	1985	18	12	89	65	60
Senegal	Niakhar	29,000	R	1962	41	3	209	54	49
Asia									
Bangladesh	Matlab	215,000	R	1966	37	1	70	66	65
Bangladesh	ORP	127,000	U, R	1982	21	3	86	66	62
Vietnam	Filabavi	52,000	R	1999	4	3	—	—	—

[1] Probability of newborn child dying before the age of five.
[2] Life expectancy at birth.
Source: INDEPTH network, 2002. (INDEPTH is an international network of field sites with continuous demographic evaluation of populations and their health in developing countries.)

tative at national or even regional level, at a time when basic demographic indicators for the population as a whole were lacking and the priority was to find reliable methods for estimating them. With very few exceptions, surveillance sites were established for only a few years, usually to study a particular disease or health problem, or to assess a control program. Surveillance stopped once these goals were attained. These numerous short-lived surveillance studies are not shown in Table 124–1, which includes only studies that were still in operation in 2003.

Of the 22 studies in the Table, only 6 began more than 20 years ago. These have been maintained beyond the period initially planned because it was found that there were advantages to be gained from a longer-term study. For example, the Matlab study in Bangladesh was originally set up to run vaccine trials, testing the effectiveness of different cholera vaccines in a drive to eliminate the disease from regions where it was endemic, particularly South Asia. A series of trials in the Matlab area began in 1963 and continued until the late 1980s (Aziz and Mosley, 1997).

In the early years, observation was limited to identifying all episodes of severe diarrhea by means of daily monitoring in a set of villages with a total population of several tens of thousands. Three years later, in 1966, it was clear that a demographic survey, periodically recording of births, deaths, marriages and migration, was needed to provide the study with up-to-date population data at all times. At the same time as the demographic surveillance was being organized, the study area was extended; by 1974, it covered a population of 277,000. As well as precisely identifying episodes of diarrhea in connection with the vaccine trials, the study also sought to improve knowledge of nutritional and diarrheal diseases in general. But it was hard to justify maintaining surveillance long-term for the cholera and diarrhea research alone, and in 1975 a major change was made to the goals of the work. A new research program was launched to assess the effectiveness of various birth control programs, including one project distributing oral contraceptives and another offering family planning services. In 1977, to reduce costs, surveillance coverage was reduced by nearly 40% (84 villages with a total population of 105,000 were excluded).

The seven surveillance sites in Table 124–1 that date from before 1985 are in Bangladesh (Matlab and Operations Research Project; ORP) and in the three countries at the westernmost tip of Africa—Senegal, The Gambia, and Guinea-Bissau. None of these early surveillance sites are in East Africa or southern Africa. The oldest was founded by Pierre Cantrelle in 1962, in Niakhar, Senegal (Cantrelle, 1969).

The mid-1980s marked a turning point, with many new surveillance sites established, particularly in East and southern Africa—a sign that the method was now recognized as a useful tool for monitoring demographic and health trends. Some surveillance sites were established specifically to measure mortality and causes of death and to supply healthcare policy makers with information, enabling them to monitor the situation. In some countries where several surveillance sites were established, such as Senegal and Tanzania, the locations were chosen so as to cover different regions or socioeconomic situations. Having a network of surveillance sites partly compensates for the fact that no individual site is representative. Another reason for the increasing number of surveillance sites during the 1980s and 1990s, especially in East and southern Africa, was the need to gather information on the general population regarding several emerging or re-emerging diseases, especially acquired immune deficiency syndrome (AIDS) and malaria, for which only partial and poorly reliable sources were available.

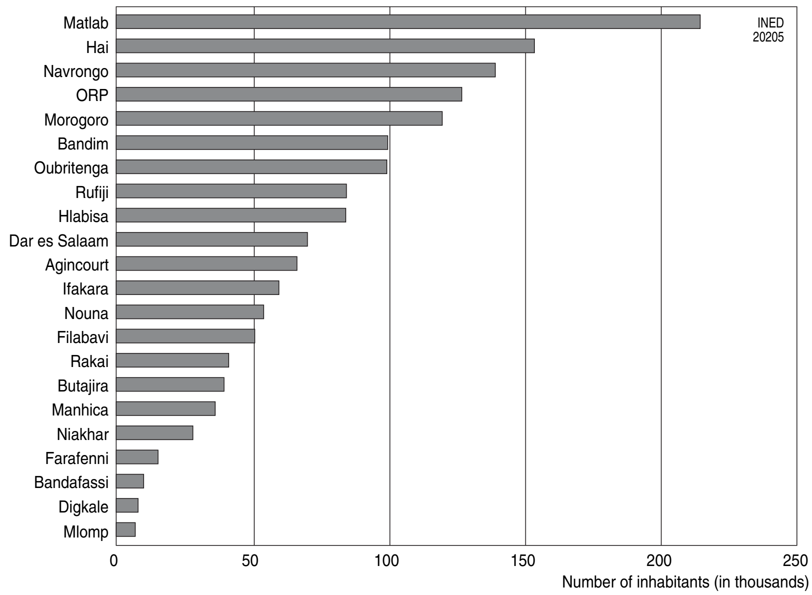
2. Characteristics of the Surveillance Studies

a. Population Size

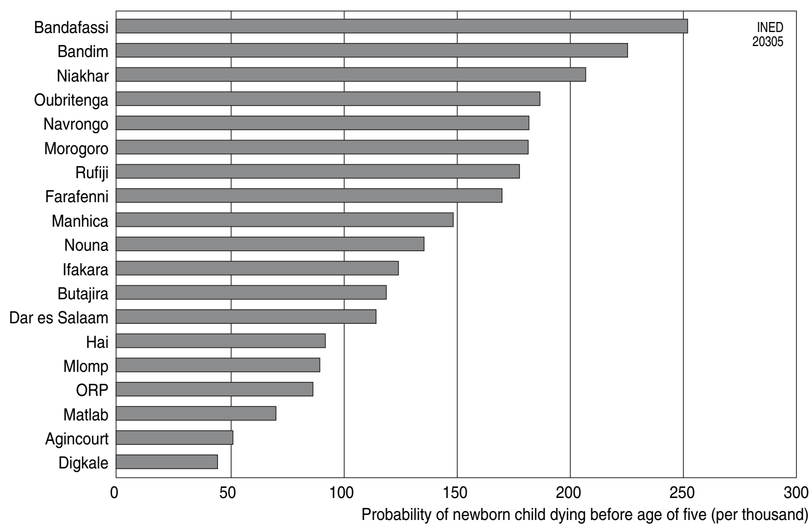
The studies in the INDEPTH network vary widely in the size of the population they monitor (Fig. 124–1a). The smallest, at Mlomp in Senegal, covers 7,500 people and the largest, Matlab in Bangladesh, 215,000. There are studies throughout the range between these two extremes, the median population size being just over 60,000. Surveillance sites set up in recent years tend to be large. For many of them, one purpose is to study mortality and causes of death among adults, including time trends, particularly in relation to the AIDS epidemic. And a population of at least 200,000 is needed to measure these correctly (Hayes *et al.*, 1989). Matlab, with a population slightly larger than this benchmark, is often regarded as a model.

The optimum size obviously depends on the purpose of the study and the resources available. But there is no guarantee that measurement precision will increase with population size; maintaining high-quality data gathering becomes more difficult as numbers of investigators and levels in the supervision hierarchy increase.

The examples of Mlomp and Bandafassi, examined in the second part of this chapter, show that studies covering only 10,000 residents each can reveal the main differences between the two sites for mortality and causes of death (children and adults).



A

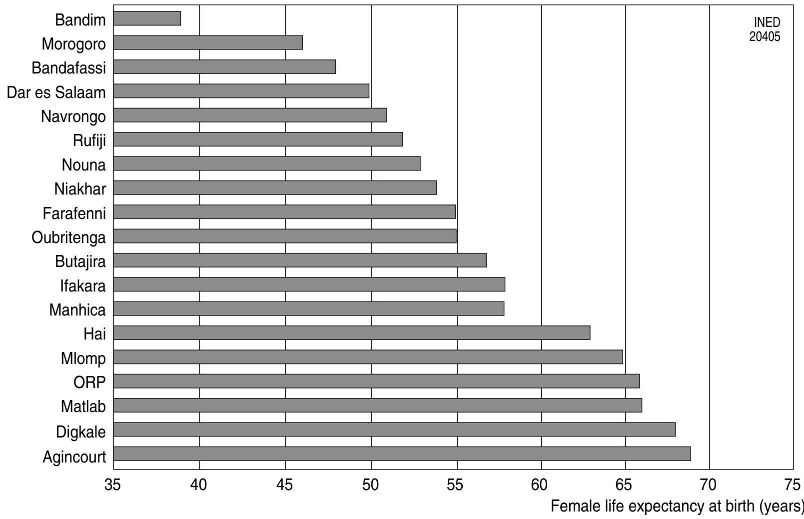


B

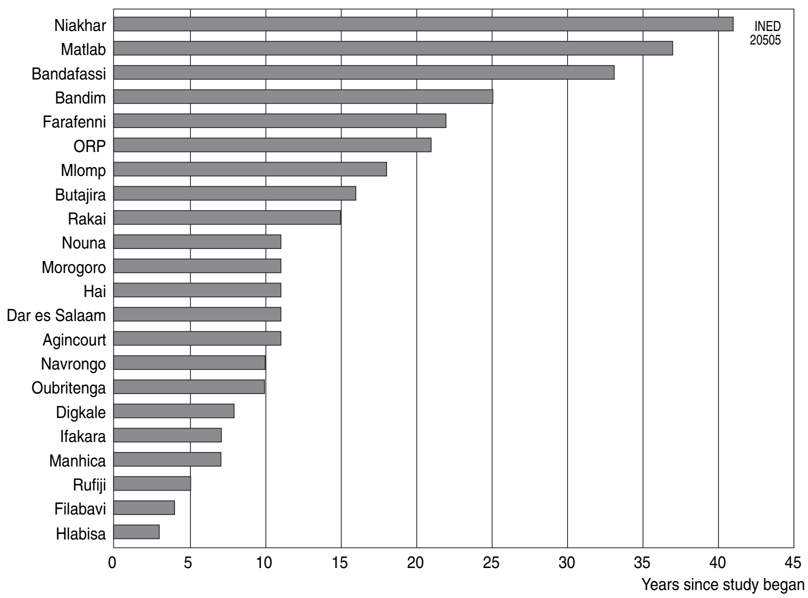
FIGURE 124–1 **(A)** Size of population monitored in 2000 (number of inhabitants). **(B)** Under-five mortality per thousand (5q0, probability of a newborn child dying before the age of 5 years of age), from 1995 to 1999.

b. The Health Situation

Table 124–1 gives two health indicators for each surveillance site: under-5 mortality rate (probability of a newborn child dying before the age of 5 years of age) (Fig. 124–1b) and female life expectancy at birth (Fig. 124–1c), both estimated for the period 1995 to 1999. The health situation seems very variable from site to site. In some populations, under-5 mortality rates are around 250 per thousand or more (Bandafassi in Senegal and Bandim in Guinea-Bissau); yet there are sites in South Africa where it is only one fifth that figure, at around 50 per thousand (Agincourt and Digkale). Figure 124–1b shows the contrasts between continents and regions, with child mortality markedly higher in Africa than in Asia and higher in West Africa than in East and southern Africa. However, exceptions can also be seen: the child mortality rate at Morogoro in Tanzania is close to those in West Africa; Morogoro is a poor, thinly populated rural area and was chosen precisely to analyze this type of exception. Conversely, Mlomp in Senegal has a particularly low child mortality rate for West Africa. As we shall see below, the population there has the benefit of



C



D

FIGURE 124–1 (*Continued*) **(C)** Female life expectancy at birth (years) (1995–1999). **(D)** Duration of surveillance up to 2003 (years since study began).

health programs and facilities that are rare in rural Africa.

The ranking order of the surveillance sites by female life expectancy at birth is much the same (Fig. 124–1c), since child mortality weighs heavily on life expectancy. The very short life expectancy in Bandim, Guinea-Bissau—39 years—is due to a combination of three factors: a low level of socioeconomic development, the AIDS epidemic, and the civil war that has ravaged the country since 1998. Several sites in Tanzania also have relatively low life expectancy, owing to low levels of socioeconomic development and the AIDS epidemic.

c. Age of the Surveillance Studies

Figure 124–1d ranks the surveillance sites in Table 124–1 by the age of the study. It illustrates what we already know: the majority are less than 10 years old and many of the recent ones are in East and southern Africa.

d. Frequency of Investigators' Rounds

At a rate of one surveillance site in three, the update rounds are annual; in just over half, they take place once every 3 or 4 months (Fig. 124–1e). It is generally recommended that the interval between visits be kept

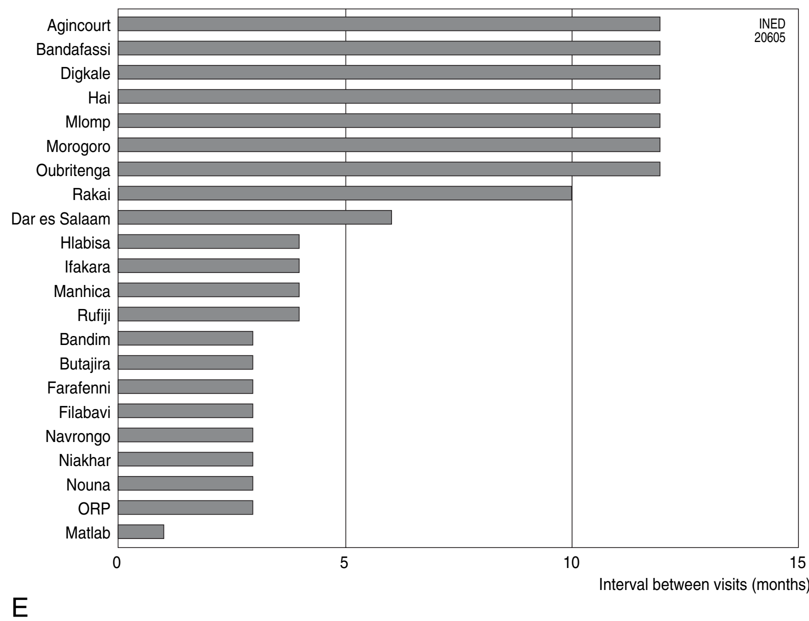


FIGURE 124–1 (*Continued*) **(E)** Interval between visits (months).

to about 3 months to make sure that such events as a birth quickly followed by death, or a brief outmigration and return, do not slip through the investigators' net. But here, again, actual practice is the result of a compromise between goals, resources and local limitations, and increasing the frequency of visits does not necessarily improve data gathering.

The data gathering method used at demographic surveillance sites, the types of finding they produce, and their usefulness for studying demographic and health trends in developing countries are illustrated later in the next section by detailed presentations of the studies at Bandafassi and Mlomp in Senegal.

II. THE DEMOGRAPHIC AND HEALTH SURVEILLANCE SITES AT BANDAFASSI AND MLOMP, SENEGAL

Bandafassi and Mlomp are rural districts of Senegal, with populations of about 10,000 and 8,000, respectively, in 2000. These populations have been monitored since 1970 and 1985,[1] respectively. Since the first year's baseline survey, both populations were visited annually without a break until 2003. At each visit, births, deaths, marriages, and migrations occurring since the previous annual visit were systematically recorded, cause of death being determined by a postmortem interview or *verbal autopsy.*[2]

1. The Bandafassi Demographic and Health Surveillance Site

Bandafassi is in the department of Kedougou, Tambacounda region, Senegal, near the border of Guinea (Fig. 124–2). On April 1, 2002, the population numbered 11,067. It is an entirely rural population, the main livelihoods being crop farming—sorghum, maize and rice, groundnuts and cotton—and cattle breeding.

The region is relatively disadvantaged compared to the average for rural parts of Senegal, owing to its remoteness from the main urban centers: it is 700 km from the national capital, Dakar, and 250 km from the regional capital, Tambacounda. The Tambacounda region itself is less well endowed with infrastructure and health programs than the rest of Senegal (Pison *et al.*, 1997b).

Bandafassi's population is comprised of three ethnic groups living in separate villages:

- Bedik (28% of the population)
- Malinke (16%)
- Fula (57%)

[1] Research program run by the National Institute of Demographic Studies (INED), Paris, in partnership with the Senegalese Ministry of Health.

[2] Before 1984, a brief information on cause of death was gathered. Thereafter, a structured questionnaire was introduced, as explained later in the chapter.

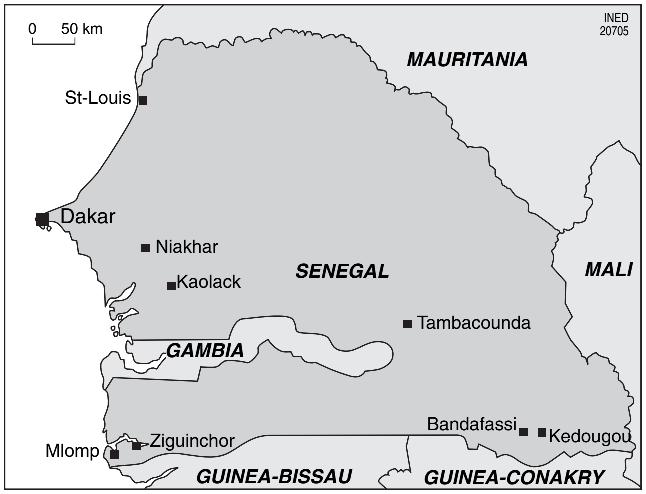


FIGURE 124–2 Map of Senegal showing locations of the Bandafassi, Mlomp and Niakhar demographic surveillance sites.

The baseline population survey and the monitoring were initially undertaken as preliminaries to a blood sampling campaign designed to study genetic diversity in the region's populations and measure differences in mortality rates in relation to individuals' genotypes. The study was intended to provide information about the mechanisms of genetic differentiation between population groups and the impact of differential mortality rates. Monitoring of the population, and particularly the recording of deaths, was designed to measure the variations of mortality rates according to individuals' genetic subgroups. However, that objective was soon abandoned, and the study's main objective now is to collect the necessary data for measuring demographic levels, trends and factors (Pison *et al.*, 1997; Guyavarch, 2003).

a. Demographic Monitoring

The gathering of demographic data began at Bandafassi in 1970. After the baseline survey, each village was visited yearly, usually between January and March. At each visit, the list of people present at the previous visit is checked and information is gathered on births, marriages, migration and deaths that have occurred since then (including causes of death). The information is usually given by the head of each compound or a few knowledgeable informants in the village.

At the start of the study in 1970, only the Malinke villages in the study area were covered. The study was extended to the Fula villages in 1975, and to the Bedik villages in 1980. The observation periods are thus different for the three groups of villages: 33 years (1970 to 2003) for the Malinke villages, 28 years (1975 to 2003) for the Fula villages, and 23 years (1980 to 2003) for the Bedik villages.

The documents used for the annual update survey include

- a list of households, updated at each annual visit
- a personalized questionnaire or nominative list of members of each household (see Box 1).

This list is also updated at each annual visit; thus, the list used for year N includes all the residents present in the household at the time of the visit in year N-1. All information required to identify each individual is included—first names, last names, parents' names, name of spouse, etc. (see Box 1).

The questions to be asked are also included. Some are the same for all residents (e.g., Is the person present or not?), while others are specific to each category, depending on sex, age, marital status and life history. For example, only women of childbearing age are asked whether any pregnancy or delivery has occurred since the last visit. In fact, the list mentions their last two deliveries, in order to make information gathering easier and avoid mistakes.

b. Determining Cause of Death by the Verbal Autopsy Method

The Bandafassi study aims to identify not only deaths but also their causes. But there are few cases where a doctor has certified the death and diagnosed its cause. During the early years of the study, cause of death was established by asking the deceased's close relatives what they died of. Since 1984, a verbal autopsy questionnaire has been used, asking the family questions about the illness or circumstances that preceded the death (Desgrées du Loû *et al.*, 1996). The verbal autopsy questionnaire covers a pre-established list of easily recognized symptoms such as fever, diarrhea, dehydration, vomiting, coughing, and convulsions. Each symptom is taken in turn, asking the family whether the person presented that symptom or not, and if so in what form; for this step, a second list of pre-established questions is used. By way of example, the questions asked in cases of diarrhea or dysentery are listed in Box 2.

Once the verbal autopsy questionnaire has been filled in it is read by two doctors who each give an independent diagnosis. In the event of disagreement between them, a third doctor reads the questionnaire and decides between them.

Diagnoses obtained in this way are more precise with some causes of death than with others. Neonatal tetanus, for example, is fairly readily identified; the signs are quite characteristic and mothers remember

Box 124–1 Example of Personalized Questionnaire, or Nominative List, Used in Bandafassi Annual Survey

This list gives the composition of the household—in this case at 1 March 2001—and was used as the questionnaire for the next annual survey in February–March 2002.

Note: the names and data given below are fictitious and are not those of any real household.

Village: ETYES 4
Compound: 2

Head of compound: Mama Yale KEITA Batialka 9782

MALE
9782 1 31 yrs Tama Mama Yale KEITA Batialka Born at Mangama

Father: DCD 9340 Dondo Byadine
Mother: 9341 Kuma Sira Numu SAMURA Bandiel
Married to: 9570 Meta, Sophie
Present? yes
no –> Deceased date:
Away place: since when?

FEMALE
9570 2 27 yrs Niafo Meta, Sophie KANTE Bandiel Born at Iwol

Father: DCD 9302 Charo Bakari
Mother: 9306 Niano Manata KEITA Bamana
Married to: 9782 Mama Yale

Last two births:
96878 1 born 15/12/98 Dondo Tungumun father: 9782 Mama Yale
95708 2 born 15/09/96 Kuma Niitabi father: 9782 Mama Yale
Present? yes
no –> Deceased date:
Away place: since when?
Since last year- has she given birth to a living baby? yes no
Has she given birth to a stillborn baby? yes no
Has she aborted? yes no
Is she pregnant? yes no

90782 1 11 yrs Charo Yata KANTE Bandiel Born at Iwol
Present? yes
no –> Deceased date:
Away place: since when?

94133 2 7 yrs Niano Machend KAMARA Batialka Born at Etyes
Present? yes
no –> Deceased date:
Away place: since when?

95708 2 4 yrs Kuma Niitabi KAMARA Batialka Born at Etyes
Present? yes
no –> Deceased date:
Away place: since when?

96878 1 2 yrs Dondo Tungumun KAMARA Batialka Born at Etyes
Present? yes
no –> Deceased date:
Away place: since when?

Box 124–2 Questions Asked at a Verbal Autopsy When Signs of Diarrhea or Dysentery Have Been Reported, to Pinpoint Symptoms More Precisely

DIARRHEA OR DYSENTERY

How long did it last?
When did it begin?
When did it end?
How many stools per day?

Were the stools like water (colorless)?		yes no
	like sputum?	yes no
	with blood?	yes no

Source: Bandafassi demographic study, Senegal.

them clearly. So a diagnosis of neonatal tetanus can be made without much risk of error if the child died between the 5th and 28th days after birth and at least three of the following symptoms are identified at the interview:

- rigid body, legs stretched out and arms bent
- convulsive movements
- mouth clenched tight, child refused to suckle (sign of "trismus," contraction of mastication muscles causing mouth to close and clench)
- hypersalivation

Measles is also easily identified. Although the symptoms are less specific and less easily identified than with neonatal tetanus, the disease has a name in all the local languages and is well known to the people, who can identify it when an epidemic occurs. If a mother is asked whether her child died of measles (giving the name used in her own language), she is rarely mistaken. However, with the spread of immunization since the 1980s, measles epidemics have become less frequent and reduced familiarity has made self-diagnosis less reliable.

For some diseases, the verbal autopsy method does not give a very reliable diagnosis. Malaria is one example, as it is easily confused with other diseases involving fever.

c. *Other Data Collected*

Along with the routine annual collecting of basic demographic data, specific surveys are run on an *ad hoc* basis on subjects of particular interest concerning population, family or health. The population files act as a baseline survey for these studies. In Bandafassi, surveys have been conducted on

- Immunization (two surveys on vaccination coverage were conducted, in 1992 and 2001, to identify the proportion of children vaccinated and study variation factors)
- Breast-feeding and weaning (in 2000, a survey was conducted among a sample of women with small children, to provide new information about infant feeding practices)
- Sexual behavior and sexually transmitted diseases (in 1998, a survey was conducted among a sample of adult men and women to provide new information about sexual behavior, measure the prevalence of HIV, syphilis, gonorrhoea and other sexually transmitted infections, and analyze risk factors of getting infected. This survey combined a medical consultation, blood and urine tests and an interview on marital and reproductive history and sexual behavior. This was followed, between 2000 and 2003, by a survey among a network of informants on the effectiveness of prevention)
- Fertility (in 2000, a survey was conducted among a sample of adult men and women, gathering information on fertility and contraception. Combining this information with the local dispensary's family planning register and the demographic monitoring data, the year-by-year progress of contraception was tracked from the time it began, and the factors involved examined)
- Malaria (entomological, parasitological and pharmacological measurements were taken on several occasions, to improve knowledge of the epidemiology of malaria and monitor changes, particularly regarding the spread of drug resistance).

These quantitative surveys are often preceded (and sometimes followed) by more qualitative surveys on the same subjects, designed to prepare for data gathering or to examine particular issues in more detail.

2. The Mlomp Demographic Surveillance Site

In 1985, a new demographic and health surveillance site was established at Mlomp, in the Ziguinchor region of southern Senegal (Fig. 124–2). The aim was to complement the two existing rural demographic surveillance sites in Senegal (Bandafassi in the southeast and Niakhar in the center-west) with a third, in the southwest of the country (in the Casamance region) which has a different history, ethnic makeup and economic situation than the regions with existing surveillance sites. Data from the three surveillance sites should provide better coverage of the country's demographic and epidemiological diversity.

On January 1, 2000 the population of the Mlomp area numbered 7,591, divided among 11 villages. Population density was 108 persons/km^2. In ethnic terms, the population are Diola; the main religion is animist, with a strong Christian minority and some Moslems. Although educational levels are low (in 2000, 55% of women aged 15 to 49 had attended school for at least a year), they are significantly higher than at Bandafassi. The population also has much better health programs and facilities. Since 1961, there has been a private dispensary in the area run by French Catholic nurses and since 1968, there has been a village maternity clinic where most women go to give birth. Most children are fully immunized and are involved in a growth-monitoring program (Pison *et al.*, 1993, 2001).

Apart from a few points of detail, the Mlomp demographic surveillance study uses the same data gathering method as Bandafassi.

3. An Example of Findings: Child Mortality Trends

Findings on child mortality will serve to illustrate the potential of demographic surveillance sites for studying demographic and health trends.

a. Bandafassi: After Significant Improvements, Has Progress Ground to a Halt?

From the data gathered at Bandafassi, mortality rates and trends can be described, the main causes of death identified, and changes analyzed. For example, Figure 124–3 shows the trend in under-5 mortality rates ($_5q_0$). As not all villages had been monitored for

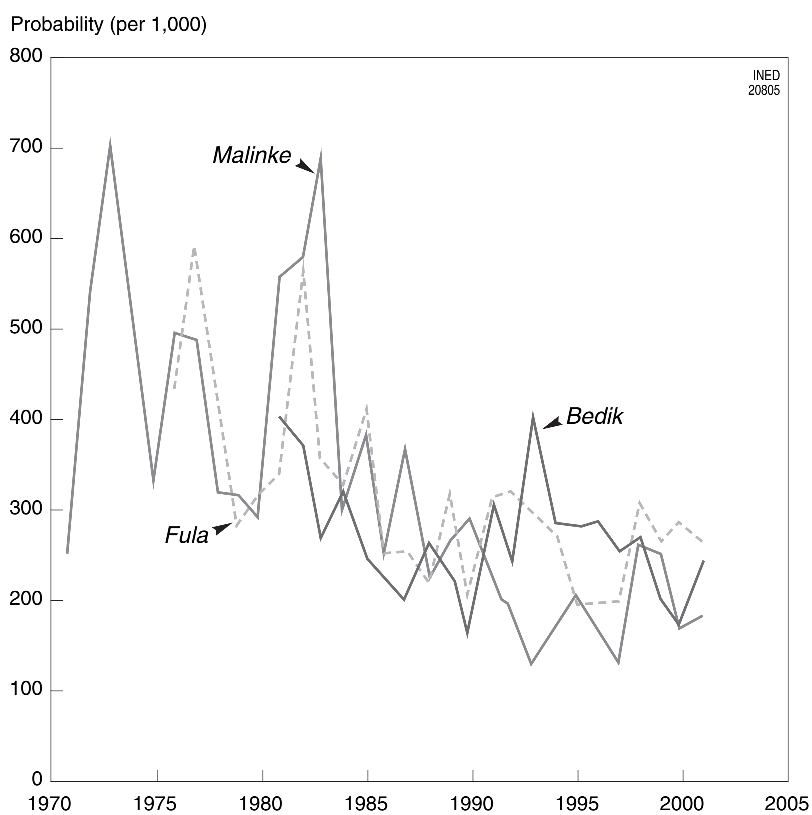


FIGURE 124–3 Trend in under-5 mortality rates ($_5q_0$, probability of a newborn child dying before the age of 5 years) in Bandafassi, by group of villages.

the same length of time, trends can be traced from 1971 for the Malinke, 1976 for the Fula, and 1981 for the Bedik. During the 1970s and the first half of the 1980s, 4 or 5 in every 10 children born died before the age of 5, though there were marked fluctuations from year to year. In the mid-1980s, mortality began to decline, until in the early 1990s the under-5 mortality rate was only 2 in 10, and year-to-year fluctuations became much sharper. Improvement ceased during the 1990s, and at the end of the decade there was even a slight increase in under-5 mortality.

The trend is similar, with slight differences for all three village groups. In the 1980s, under-5 mortality was slightly higher in Fula villages than in the other village groups, but during the 1990s it fell slightly below the others.

This pattern raises three questions. Why did under-5 mortality fluctuate so sharply before the mid-1980s? Why did it drop in the second half of the 1980s? What stopped that decline in the 1990s?

b. The Impact of Vaccination

Before 1987, the population of Bandafassi was not vaccinated on a regular basis. It was involved in a few national vaccination drives, such as smallpox vaccination as part of the global eradication of that disease, and the measles vaccination drives of 1967 and 1969. Apart from that, only some children of public servants—teachers and nurses—were vaccinated on a regular basis, and overall vaccination coverage was very low.

Vaccinations on a regular basis began in Bandafassi in 1987, with a national campaign to achieve country-wide coverage for the Extended Program of Immunization (EPI) that had been running in the cities since 1981. In the space of 3 months in the first quarter of 1987, vaccination coverage in Bandafassi rose to 48% (i.e., the percentage of children aged 12 to 35 months fully vaccinated, having received Bacillus Calmette-Guerin vaccine (BCG) and vaccination against measles, yellow fever, diphtheria, whooping cough, tetanus, and poliomyelitis) (Desgrées du Loû and Pison, 1994). The vaccination program was continued, with a view to gradually making it an integral part of the standard healthcare provided by the health services. The 1992 survey showed that although the vaccination drive continued after 1987, it varied from year to year with a downward trend overall: in 1992, 5 years after the start of the program, vaccination coverage was only 39%.

The population of Bandafassi entered the vaccination era abruptly. As the demographic surveillance system had already been in operation for several years when vaccination began in 1987, advantage was taken of this almost experimental situation to assess the possible impact of vaccination on under-5 mortality. Between the 6 years before the vaccination program began (1981 to 1986) and the 6 years after (1987 to 1992), under-5 mortality fell by 40%. The extent of the drop varied according to age group: 31% in the neonatal period (0 to 28 days), 20% between 1 month and 8 months, and 48% between 9 months and 59 months. As no other health change occurred during this period and the socioeconomic situation scarcely varied, vaccination clearly seemed to be the main explanation for under-5 mortality drop in the latter half of the 1980s.

A drop in under-5 deaths is hardly surprising, since the vaccinations gave protection against several common causes of death among children. But a detailed study of causes of death produced a surprising result: the drop was sharper than could be explained by even total elimination of deaths due to diseases covered by the vaccination program (Desgrées du Loû and Pison, 1996), and yet these diseases were still causing some deaths.

c. Measles: A Rapidly Declining Cause of Death

The decline in deaths due to measles was particularly dramatic. Because measles is a disease that the people of Bandafassi can identify easily, we have data on measles deaths from the start of the demographic monitoring in 1970, even before the verbal autopsy questionnaire was introduced in 1984. Table 124–2 compares deaths from measles between four periods, two prior to the start of vaccination in 1987 (1970 to 1979 and 1980 to 1986) and two after (1987 to 1989 and 1990 to 1993).

The change in 1987 is radical. Before that date, measles was responsible for about one death in seven (14% in 1970 to 1979 and 15% in 1980 to 1986) in children aged 1 to 20 months, and one in three (35% and 30%) in children aged 21 to 59 months. It was the foremost cause of death in these age groups. From 1987 on, measles accounted for only 3% and 5% of deaths in these age groups respectively.

Figure 124–4, which shows bimonthly variations in the 0- to 5-year age group's measles mortality rate gives a clear illustration of the epidemic nature of the disease. There were five epidemics between 1970 and 1995: in 1973, 1976 to 1977, 1981 to 1982, 1985, and 1992. Of the few children who died of measles between these epidemics, most were infected (and died) outside the study zone, when traveling with their mothers. Further, each epidemic affected only some villages. Before 1987, there were often intervals of over 10 years between epidemics in the same village. Because of the infrequency, when an epidemic did hit a village the

TABLE 124–2 Variations in Measles Mortality, by Age and Time Period

Period	Age group			
	1–20 months		21 months–59 months	
	Annual measles mortality rate (per 1,000)	Measles deaths as % of total deaths	Annual measles mortality rate (per 1,000)	Measles deaths as % of total deaths
1970–1979	21.0	15	25.3	35
1980–1986	14.6	14	14.9	30
1987–1989	0.0	0	0.0	0
1990–1993	2.9	4	2.6	8

Source: Bandafassi, 1970–1993.

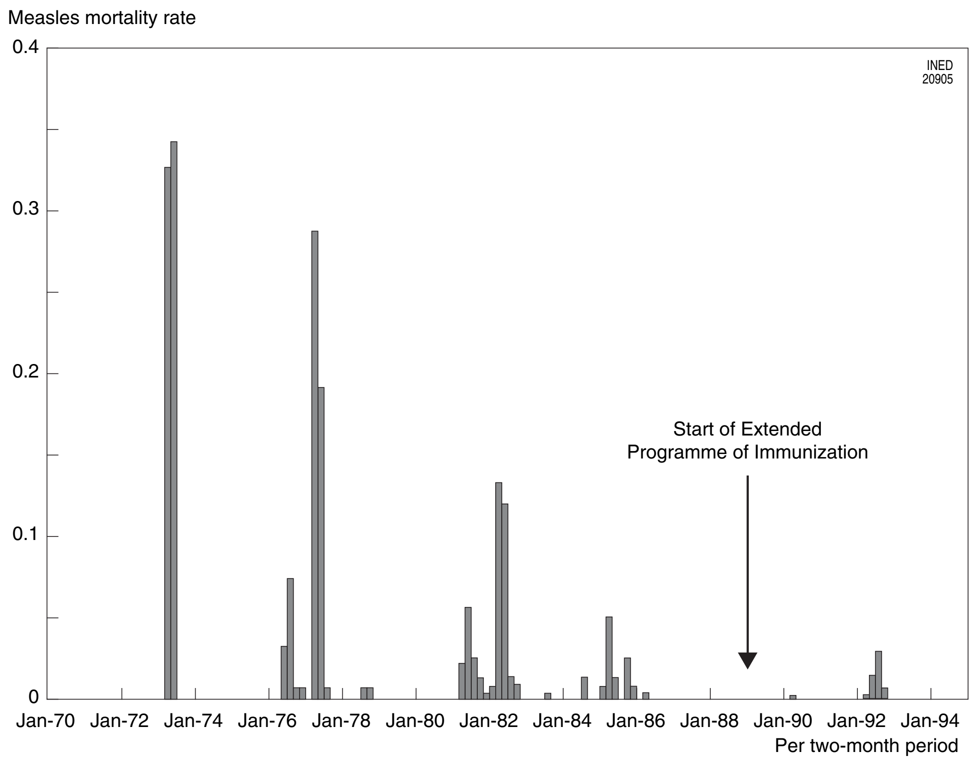


FIGURE 124–4 Bimonthly fluctuations in mortality due to measles in children under 5 years of age, Bandafassi, 1970 to 1994.

impact was severe: within a few weeks, nearly all children born since the last epidemic fell ill and the lethality rate was high—over 15% in the 1976 to 1977 and 1981 to 1982 epidemics.

From 1986, mortality due to measles plummeted. The virtual disappearance of one of the main causes of death in children is the most striking result of the introduction of vaccination in the study zone. But, as mentioned earlier, the decline in child mortality was greater than expected from the elimination of diseases covered by the vaccination program. This may be a sign that vaccines have a nonspecific beneficial effect, the measles vaccine in particular (Aaby *et al.*, 1995); like measles itself, the measles vaccine seems to stimulate the immune system in a nonspecific manner.

d. Mortality and Causes of Death at Mlomp Compared With Bandafassi

From an analysis of reproductive histories gathered during the baseline survey in 1985, child mortality trends in Mlomp can be traced back to the 1930s. Until the mid-1960s, the probability of death before age 5 was roughly stable, at 350 to 400 per thousand

(Fig. 124–5). It began to fall in the mid-1960s, and in the second half of the 1980s fell below 100 per thousand—far lower than that recorded in Bandafassi at that time. This very rapid decline is linked to the development of health programs and facilities in the 1960s and 1970s (Pison *et al.*, 1993).

From an analysis of the monitoring data for the two study zones, infant and child mortality per cause of death in Mlomp and in Bandafassi can be compared for the second half of the 1980s, distinguishing between neonatal mortality (Fig. 124–6a) and mortality between 1 month and 5 years (Fig. 124–6b). There are no data on cause of child deaths in Mlomp before 1970, when mortality rates were still high, but the causes recorded in Bandafassi in the recent period are likely to have been the predominant ones in Mlomp before 1970.

In the second half of the 1980s, neonatal mortality in Mlomp was less than half that of Bandafassi (36 per thousand compared to 87). This difference reflects differences in mortality rates for all major causes of death. For example, in Bandafassi neonatal tetanus was responsible for one neonatal death in 4 in that period, killing nearly 20 newborns per thousand compared to a mere 1 per thousand in Mlomp. Deaths due to premature birth or low birth weight show the same pattern: 2 per thousand in Mlomp compared to 21 per thousand in Bandafassi. For obstetric trauma, the figures are 3 per thousand and 12 per thousand.

After the neonatal period (Fig. 124–6b), the difference in overall mortality rates is even greater (45 per thousand compared to 277) and here too it is linked to differences in mortality for each of the main causes of death: diarrhea and malnutrition, pneumopathy, malaria, etc. In this period, measles was a declining though still significant cause of death in Bandafassi, but does not figure as a cause of death in Mlomp.

Assuming that mortality patterns in Bandafassi reflect those that were current in Mlomp until the 1960s, it appears that mortality in Mlomp too has waned because of a decline in each of the major causes of death.

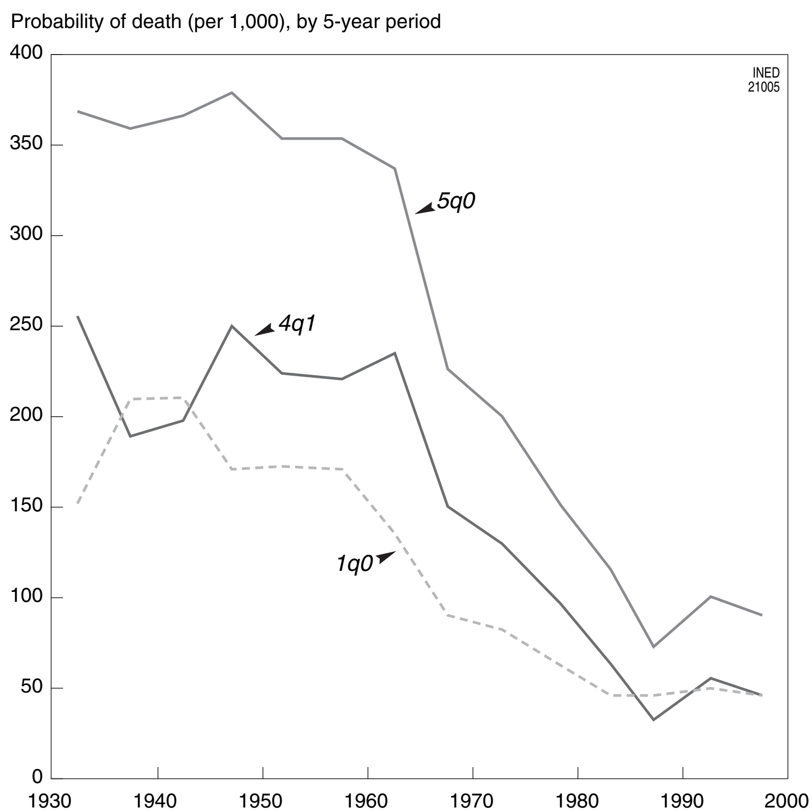


FIGURE 124–5 Child mortality rates in children age 0 to 5 years in Mlomp, since 1930. ${}_1q_0$: probability that a newborn child die before reaching the age of 1 year of age; ${}_4q_1$: probability that a 1-year-old child will die before the age of 5; ${}_5q_0$: probability that a newborn child die before the age of 5; *NB*: before 1985, measured by birth cohort (5-year age group) from reproductive histories gathered in 1985. From 1985 on, measured by 5-year period from demographic surveillance data.

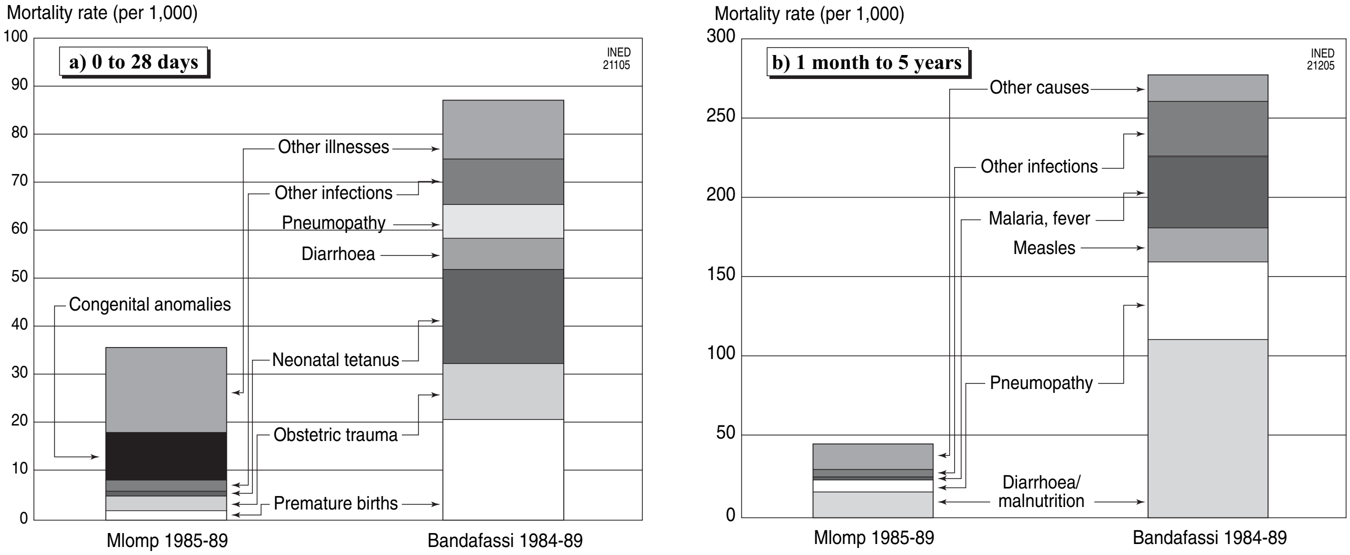


FIGURE 124–6 Child mortality: comparison between Mlomp and Bandafassi.

e. Malaria: A Cause of Death That is Regaining Ground

As shown earlier, child mortality ceased to decline in the early 1990s and even started to climb again at the end of the decade, in both Mlomp and Bandafassi. One reason is the considerable increase in deaths due to malaria (Fig. 124–7).

The verbal autopsy method is not efficient for identifying malaria as a cause of death because it is difficult to distinguish malaria from other illnesses involving fever. However, malaria mortality in Mlomp can be traced with some accuracy, as many children who died following a bout of fever had attended the local dispensary, where the malaria diagnosis was confirmed or invalidated by laboratory tests on thick blood films.

A sign of the success of the dispensary's antimalaria campaign in the 1970s and 1980s, malaria mortality was very low at Mlomp in 1985 and remained low until 1990. But it rose sharply between 1990 and 1992, and has remained quite high ever since. This is due to the arrival and spread of malaria strains resistant to chloroquine, the drug responsible for the earlier sharp reduction in malaria, mortality which is widely used by local people against fever attacks, both curatively and preventively (Trape *et al.*, 1998). Malaria mortality also increased again in Bandafassi and in Niakhar, Senegal's third demographic surveillance site (Fig. 124–7). This is one factor in the increase in overall child mortality rates in the 1990s.

Although many African countries have been hard hit by AIDS, Senegal has not yet been seriously affected and AIDS is not a major cause of death there. The recent increase in child mortality in the demographic surveillance sites is not AIDS-related.

CONCLUSION

As we have seen, demographic surveillance sites are a very effective way of monitoring changes in demography and health in countries with no reliable, routine data collection. In particular, they provide detailed information about mortality rates and causes of death in children under 5 years of age. At Mlomp, child mortality fell from the mid-1960s, in connection with the development of health programs and facilities. The demographic surveillance system has made it possible to measure the scale of the improvement and show that the rapid, dramatic reduction in mortality was due to significant reductions in mortality from every cause of death—a sign of good coordination and complementarity among the various health measures. The fact that the decline in mortality halted in the 1990s does not mean that change has ceased. Underlying the plateau in overall mortality there are contradictory trends, with some causes of death declining or disappearing and others gaining ground again.

While demographic surveillance sites are one of very few ways of monitoring mortality and causes of death in countries without health statistics, they are useful in other fields as well. For example, they can be used to monitor demographic and social change and study the processes involved. In this regard, they are a useful complement to demographic and health surveys. While DHSs provide estimated levels and

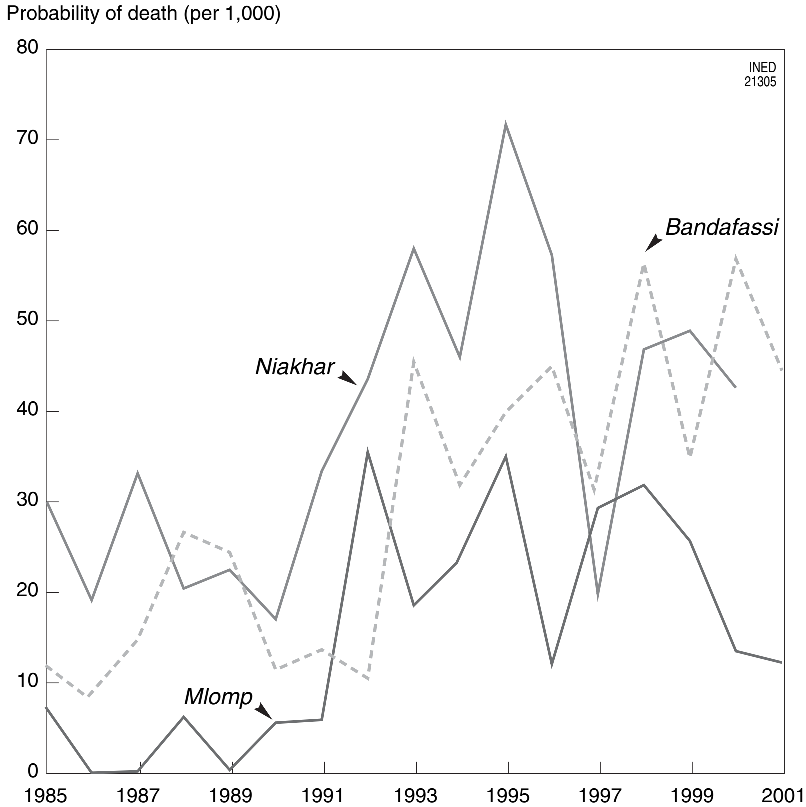


FIGURE 124–7 Malaria mortality in children under the age of 5 (${}_5q_0$) in Bandafassi, Mlomp and Niakhar, from 1985 to 2001.

trends that are representative at the national level, demographic surveillance sites supply detailed information about the mechanisms involved in those changes.

Acknowledgment

This chapter was translated from French by Harriet Coleman.

References

AABY Peter, SAMB Badara, SIMONDON François, COLL SECK Awa Marie M., KNUDSEN Kim M. and WHITTLE Hilton, 1995. Non-specific beneficial effect of measles immunisation: analysis of mortality studies from developing countries, *British Medical Journal*, vol. 311(7003), p. 481–485.

AZIZ K.M.A. and MOSLEY W. Henry, 1997. The history, methodology and main findings of the Matlab project in Bangladesh, *in:* Monica Das Gupta, Peter Aaby, Michel Garenne and Gilles Pison (eds.), *Prospective community studies in developing countries*, p. 28–53. Oxford, Clarendon Press, 350 p. (International studies in demography).

CANTRELLE Pierre, 1969. Étude démographique dans la région du Siné Saloum (Senegal). État civil et observations démographiques, 1963–65. Paris, ORSTOM, 121 p. (Travaux et documents de l'ORSTOM, no. 1).

DAS GUPTA Monica, AABY Peter, GARENNE Michel and PISON Gilles (eds.), 1997. *Prospective community studies in developing countries*. Oxford, Clarendon Press, 350 p. (International studies in demography).

DESGRÉES du Loû Annabel and PISON Gilles, 1994. Barriers to universal child immunization in rural Senegal, five years after the accelerated Expanded Program on Immunisation, *Bulletin of the World Health Organization*, vol. 72(5), p. 751–759.

DESGRÉES du Loû Annabel and PISON Gilles, 1996. The role of vaccination in the reduction of childhood mortality in Senegal. *Population An English Selection*, vol. 8, p. 95–121.

DESGRÉES du Loû Annabel, PISON Gilles, SAMB Badara and TRAPPE Jean-François, 1996. L'évolution des causes de décès d'enfants en Afrique: une étude de cas au Senegal avec la méthode d'autopsie verbale, *Population*, vol. 51(4–5), p. 845–881.

GUYAVARCH Emmanuelle, 2003. *Démographie et santé de la reproduction en Afrique sub-saharienne. Analyse des évolutions en cours. Une étude de cas: l'observatoire de population de Bandafassi (Sénégal)*. Paris, Museum national d'histoire naturelle, 349 p. + annexes. (Doctoral thesis).

HAYES Richard, MERTENS Thierry, LOCKETT Geraldine and RODRIGUES Laura, 1989. *Causes of adult deaths in developing countries: a review of data and methods*. Washington, D.C., World Bank, 41 p. (Working Paper no. 246).

INDEPTH, 2002. *Population and health in developing countries*. Ottawa, International Development Research Centre, 339 p.

PISON Gilles, DESGRÉES du Loû Annabel and LANGANEY André, 1997. Bandafassi: a 25 years prospective community study in rural Senegal (1970–1995), *in:* Monica Das Gupta, Peter Aaby, Michel Garenne and Gilles Pison (eds.), *Prospective community studies in developing countries*, p. 253–275. Oxford, Clarendon Press, 350 p. (International studies in demography).

Pison Gilles, Gabadinho Alexis and Enel Catherine, 2001. *Mlomp (Senegal): niveaux et tendances démographiques 1985–2000*. Paris, INED, 181 p. (Dossiers et recherches no. 103).

Pison Gilles, Hill Kenneth H., Cohen Barney and Foote Karen A. (eds.), 1995. *Population dynamics of Senegal*. National Academy Press, Washington, 254 p. (http://www.nap.edu/catalog/4900.html).

Pison Gilles, Trape Jean-François, Lefebvre Monique and Enel Catherine, 1993. Rapid decline in child mortality in a rural area of Senegal, *International Journal of Epidemiology*, vol. 22(1), p. 72–80.

Trape Jean-François, Pison Gilles, Preziosi Marie Pierre, Enel Catherine, Desgrées du Loû Annabel, Delaunay Valérie, Samb Badara, Lagarde Emmanuel, Molez Jean-François and Simondon François, 1988. Impact of chloroquine resistance on malaria mortality, *Comptes rendus de l'Académie des sciences de Paris. Sciences de la vie*, vol. 321, p. 689–697.

C H A P T E R

125

Comparing to Understand

VIVIANA EGIDI AND PATRICK FESTY

Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," Rome, Italy

Institut national d'études démographiques (INED), Paris, France

I. AGE AND DIVERSITY OF THE PRACTICE

Comparative analysis has always been an especially fertile branch of research—in demography as well as in the other social sciences. Demographers have always been concerned with comparing a population's demographic behavior with what is occurring in neighboring countries or countries that belong to the same cultural or economic zone. The main objective has varied according to era and demographic way of thinking: to expose the specificity in the characteristics and behavior within a country, or to find the universal factors or *laws* that may be hidden behind an observed heterogeneity. Comparison makes it possible to avoid the risk of giving specific and national interpretations to international phenomena, and to expose precursory signs indicating the direction that the population will take in the future.

In today's globalized world, regulated by increasingly uniform economic laws from one end of the planet to another, where countries associate to create vast common economic and social zones and where at the same time the concept of nationality is being defined at an increasingly detailed level, it is more imperative than ever to create a common language that will enable people to understand one another and communicate. The language adopted to this end is one that can, at a statistical level, make comparable information available and enable its analyses with methods capable of shedding light on similarities and differences. The aim is obviously not to construct comparable information by confining it to the limited field of demography or to the single level of societies and populations (macro level), as has been done for a long time. Following the development of studies aimed at the interpretation of and research into the determinants of demographic behavior, today's focus is on establishing the comparability of individual data (micro level) on an increasingly large number and range of descriptive demographic, social, economic, medical, and health variables, variables of individual behavior, and of the social and economic structure of different countries.

The definition of a common statistical language capable of ensuring comparability is faced with numerous obstacles, some of them are clearly legitimate, while others are linked to the varying quality of the information to be compared. The increasing demand for the development of comparable statistics necessitates approaching the issue with new tools, while attempting to identify the causes of differences and trying to reconcile the demands for quality at various levels, national (and increasingly infranational) and supranational, with the awareness that these demands can be conflicting as the objective of comparability can often contradict with the continuity of temporal series, which is another powerful knowledge tool.

Fortunately, the era of simplistic solutions (when attempts were made to solve the comparability problem through the rigid standardization of concepts, survey techniques and methods) underestimating the negative effect that such a strategy can sometimes have on the overall quality of information seems to have come to an end. We realize today that the solution is more complicated and that it requires a detailed knowledge of the concepts and structures of the various realities to be compared in order to represent of all of them. We have also realized that the construction of comparability, in the context of quality requirements, is a continuous process where each stage marks provisional progress towards better results, but where the objective of perfect comparability is only a theoretical aim.

II. THE CONDITIONS FOR COMPARISON: DATA COMPARABILITY

Comparative analysis should first of all be based on comparable data, i.e., data that can be compared to another in a common frame of reference: in other words, data that speak the same language. The harmonization process can be interpreted in light of the following objective: the development of a common language that makes it possible to obtain, for a determined variable, estimates that are not affected by the international diversity of concepts and social structure.

Data comparability thus requires, by its very nature, a compromise between the demand for the national specificity of information (which enables comparisons over time), and that of international comparability. It always has a cost in terms of informative capacity, insofar as losing specificity can imply that the information is less relevant for understanding local reality. This must be taken into consideration not only when the data are being produced but during the analysis and evaluation of the results.

Concern regarding data comparability and attempts at harmonization occurred very early in demography: the United Nations' recommendations for carrying out censuses and models for collecting and classifying information on causes of death are classic examples. Important progress has been made on the path to demographic data comparability but more remains to be done, in particular when comparative studies are based on data that are not strictly demographic. The process of harmonizing social variables, as well as socioeconomic variables, at the individual and family level, actually began much later: the first steps towards a common strategy were not taken until the second half of the 1990s. The *Siena Group* (one of the *City Groups* dedicated to social statistics launched by the United Nations), and in Europe, Eurostat (through various initiatives associating the managers of demographic and social statistics of various countries) concentrated their activities during this period on harmonizing demographic and social statistics, putting forward many propositions and tracing the course of a common path (Mejer, 2003).

1. The Notion of Harmonization

Harmonization is the search (never completed or definitive) for reasonable equivalences between concepts that translate the same idea into different social realities, the search for a common language that makes it possible to go beyond national specificities. All the elements of statistical language (concepts, definitions, classifications, collection methods, and methods for developing data) should be taken into consideration in the strategy for data harmonization, insofar as each element can facilitate or hinder comparability and thus influence the quality of the results.

a. Concepts and Definitions

When the objective is to compare demographic or socioeconomic variables measured in different cultural contexts, we should expect differences in the concepts used in diverse national realities, and consequently differences in the social and institutional structures that reflect these concepts (consider, for example, the concept of education and the educational structures set up in the different countries to fulfill the educational requirements of the population).

The divergences are particularly significant for certain complex concepts (such as well-being, poverty and social exclusion, health) that are based on theoretical structures deeply entrenched in the value system of each country and each era concerned.[1] In these cases, the construction of comparable measurements leads first of all to a search for the definition of the variable(s), which maximize(s) the similarities between the concepts (and/or between the national structures)—a sort of *greater common divisor*, which can become significantly removed from the most appropriate definition for each country. At the same time, it is necessary to protect the measurement as much as possible from all activities that can cause distortions in the comparison (collection, codification, correction, calculation).

[1] For example, after the adoption by the World Health Organization (WHO) of an international definition of health, the indicators used in studies on health have over time progressively replaced the traditional macro objective indicators (general mortality and by causes, child mortality), with clinical indicators (presence of diseases and clinically verified risk factors), and subjective indicators based on respondents' self-evaluation (perceived health).

The starting point of the harmonization process is therefore updating and going beyond the natural tendency towards ethnocentrism, through detailed knowledge of national concepts and the corresponding structures in order to develop a common definition of a compromise concept (an international or harmonized concept), which represents an agreement for comparisons between different countries (Ehling, 2003). This stage is simpler to carry out if it is possible to refer to a universal concept, or at least objective elements. Some of the variables most frequently used in demography are of this type: the concepts of *sex* and *age*, for example, are generally shared or they can at least be easily attached to international standards. Some recent developments should be carefully monitored to make sure that they do not distance us from this favorable situation. For example, the increasingly frequent use in demography of the *gender* concept, based on the social roles played by individuals, instead of the biological concept of *sex*, could cause progressive divergence in the operational definitions of different countries and loss of comparability.[2]

Concerning age, comparability problems are often linked to the recording method (age of the individual rather than the year or date of birth), but the distortion that can result from this is generally not likely to compromise possibilities for comparison, at least when the classification is detailed enough. Other concepts, also fundamental in demography, such as the vitality of a birth, parity, migration, nationality, and cause of death, pose more serious problems.

The problems are even greater for variables such as ethnic group, which, on the one hand, are more difficult to define, and which, on the other hand, are (in some countries more than in others) sensitive issues that cannot be broached with simple and direct questions (Hoffmeyer-Zlotnik, 2003). The same is true for variables that are more resolutely social (such as education, employment, or social condition) and that participate in demographic studies as possible determinants of the behavior of the individual or the family.

In these cases, the greater the difference between contexts, the greater the risk of the acceptable definition being too complex and too difficult to estimate with available tools (too many questions in the questionnaire, respondents' lack of understanding, lack of correspondence with administrative sources, etc.).

b. Definition of the Unit of Collection and Analysis: The Household

Some concepts and some definitions play a particularly important role, as they are attached to the units of analysis of the information collected, and, at the same time, to the units from which the samples and the information are taken. They should therefore be able to ensure complete coverage of the basic units. The household is an example of this type of unit: it is the unit of collection of numerous statistical surveys, as well as an active subject of demographic and social behavior, either autonomously, or as a framework within which individuals live their lives and make decisions in interaction with other members.

The household is a complex and dynamic unit, whose definition varies from one country to another (it can sometimes even refer to different concepts from one survey to another in the same country). Cohabitation (the concept of *household housing unit*) is the minimum condition, adopted by every country, for a group of individuals to constitute a household. Some countries, such as the United States, Canada, and France, adopt a definition of household based on this condition only, which guarantees complete coverage of the population, even if there are no population registers, and which enables samples to be selected on the basis of a list of addresses. But most frequently, the definition of household includes other conditions relative to the manner in which the members *live together*, and the way in which they use the resources available to them (*household budget* concept): if the household members pool all or part of their income regardless of their family ties, they have a common budget. These conditions can take various forms according to the habits and sensitivity of the population. In Australia, the household is a group of individuals who live and usually eat together, in the Philippines it is a group of people who sleep in the same residence and who prepare and eat their meals together, in New Zealand it is a group of people who usually live together and share the services of the residence (kitchen, bathroom, lounge), in Germany it is a group of people who live together and constitute an economic unit. In Great Britain, a group of individuals is considered to be a household if there is a combination of three elements: living together, eating together, and constituting an economic unit.

For censuses around the year 2000, the United Nations and its regional agencies recommended that

[2] "Although the above mentioned problems of distinguishing between men and women are real and may increase in the future, sex belongs to the variables most easily assessed. The reverse is true for gender and gender roles. Here researchers have still some way to go to yield fruitful clear-cut definitions and robust measures of these concepts that then can be adopted to cross-cultural studies." (Wolf and Hoffmeyer-Zlotnik, 2003). See also (Krieger, 2003).

countries use common housing and budget criteria[3] in information collecting operations. Working towards its objective of making European countries converge towards comparable demographic and social statistics, Eurostat suggests that the countries adopt the same conditions, not only for censuses, but also for social surveys from which statistical information useful to the European Community is extracted. Despite this, some countries continue, for internal consistency between the definitions used in their various national surveys, to use the single condition of co-residence (Mejer, 2003).

Of course, there is generally a nearly perfect correspondence between the households identified by the two definitions. The problem arises from particular situations such as that of domestic employees, subtenants, or lodgers, who live at the same address but who, generally, do not share their resources with other household members, or that of people living temporarily outside of the household for health reasons or study reasons. In the latter case, the recommendations suggest that a person's membership of the household should be judged according to length of stay (usually, at least 6 months out of the 12) and other information that makes it possible to understand the nature of his/her link with the household: for example, students who live far from their family and who do not receive any economic support should be excluded if their absence is longer than 6 months, but should be included, regardless of the length of absence, if the family is financially responsible for them.

The method of dealing with students living apart from their family is important, as numerous studies have been developed recently on this phase of the life cycle, on the methods and the rhythm of forming new family units, and on the consequences of the choice of autonomy on quality of life and well-being. Indeed, when units made up of young adults still in education are treated as autonomous households, there is a strong probability that they will be classified as poor. We will spend a little time on this example, as it clearly illustrates the complexity of choices that must be made in order to create comparability that goes beyond cultural specificities. The situation of students living away from their families and who have limited resources can correspond to very diverse situations in different countries, according to the way in which the parent-child link is experienced. In some contexts, leaving the parental home is experienced, by both the parents and the child, as the beginning of an autonomous life with all its advantages and disadvantages: independence but also economic fragility and the need for young people to find most of the resources that they need. In other contexts, in particular that of southern Europe, the departure of students is considered transitory, without the bonds of interdependence being distended. The children remain, in every way, members of their family and they share (from afar) the same level of well-being: the parents take direct responsibility for the majority of the cost while frequently providing the students with an income that enables them to pay their day-to-day expenses. In the first case, it is normal for households of students to be classified as being poor (according to their income and their expenses), but in the second case applying the same definition would lead to unduly overestimating the number of poor households.

Furthermore, the duration of absence from the parental home takes on a different meaning according to the tools available for choosing the sample. In many countries, the selection is made from lists of addresses and the criteria of duration of absence is necessary to avoid a person being selected twice. In other countries, there are population registers, particularly registers of the resident local population, from which the samples are selected. As a person cannot be recorded in several registers at the same time, individuals have the choice of defining their place of residence as the place where their interests lie, including emotional ones: children can maintain their place of residence at their parents' address even if they are away for longer than the 6 months recommended by the definition without changing their probability of being selected.

In general, therefore, in the European Union and more widely in Western countries, the adoption of a definition that only respects the condition of cohabitation does not create substantial differences in the structure of households (Bien and Quellenberg, 2003). Moreover, as the condition of a common budget is always combined with that of co-residence, the countries that adopt the most restrictive definition are able to aggregate their results according to the simpler criterion of cohabitation, while the reverse is impossible. This is why, for example, the comparative indicators and tables that the countries must provide the European Community based on their censuses only refer to the criterion of co-residence, although the *household budget* concept has been recommended (Eurostat, 1999a, p. 23). However, logical inconsistencies can

[3] "A household is classified as either: 1) A one-person household, defined as an arrangement in which one person makes provision for his or her own food or other essentials for living without combining with any other person to form part of a multi-person household, or 2) A multi-person household, defined as a group of two or more persons living together who make common provision for food or other essentials for living". (www.unstats.un.org/unsd/demographic/)

TABLE 125–1 Concept of Household in the European Household Budget Survey

Country[a]	**B**	**DK**	**D**	**EL**	**E**	**F**	**IRL**	**I**	**L**	**NL**	**A**	**P**	**FIN**	**S**	**UK**
Residence	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Expenditure	■	■	■	■	■		■	■	■	■	■	■	■	■	■
Income		■	■					■					■	■	
Emotionalties								■							

[a] B: Belgium; DK: Denmark; D: Germany; EL: Greece; E: Spain; F: France; IRL: Ireland; I: Italy; L: Luxembourg; NL: The Netherlands; A: Austria; P: Portugal; FIN: Finland; S: Sweden; UK: United Kingdom.

Source: Eurostat, 2003a.

occur between the definition adopted for family and the type of phenomenon analyzed. For all the phenomena relative to the economic condition of households, their revenue, consumption, or poverty, for example (Bauman, 1997), it is important that the household be defined by the dual condition of common housing and budget, as the economic situation of a group is only meaningful if all members share it. In the *European Household Budget Survey*, which plays a key role in the comparative study of the living conditions of families, the definitions adopted by the countries are more or less restricted (Table 125–1) to the combination of the four criteria (Eurostat, 2003a):

1. co-residence
2. sharing expenditure
3. pooling income and resources
4. the existence of emotional ties

At the two extremes are France, which has adopted the least restrictive definition, and Italy, which combines all four criteria for defining a household.

In the *European Union Survey on Income and Living Conditions (EU-SILC),*[4] the new harmonized survey intended to gather statistical information on the economic situation of families and social exclusion in Europe, the definition of a household has been specified in a European regulation which, by being formally imposed on the countries, aims at creating conditions for improved comparability. The regulation makes the provision that a *private household* means "*a person living alone or a group of people who live together in the same private dwelling and share expenditures, including the joint provision of the essentials of living*" (Article 2)[5] and it specifies the conditions that make it possible to establish whether a person living at the address of the household (or is temporarily absent) is a member of the household.[6]

[4] European Union Survey on Income and Living Conditions.

[5] Regulation (EC) No. 1177/2003 of the European Parliament of the Council of Europe, June 16, 2003.

[6] Commission Regulation (EC) No. 1980/2003, October 21, 2003.

The household is not only a unit of data collection, it is also a unit of analysis, either as an exclusive referent of behavior (in surveys on family budgets it is frequently impossible to attribute choices and practices to specific household members), or as an influential and often even determining context of the choices and behaviors of individuals. It is also fundamental, from this point of view, that common definitions delimit the borders of a household and determine the internal structure. Within a household, other complex units can also be used as analytical units: the *family* (group of co-residing individuals who are linked by affiliation, biological or adoptive, or by marriage (*de jure* or *de facto*), the *family nucleus* (limited to the relationships of couples and parents with unmarried children). The definitions of the family nucleus are generally shared by the countries, and the lack of comparability is mainly due to the tools for collecting information on the ties between individuals within a household, which enable the complete reconstitution of families or family units more or less effectively. There are two types of tools: those that collect information on family relationship, adoption, marriage, and cohabitation ties, relative to a single subject (reference person); and those that collect information on the ties of each person with all the others (relationship matrix). The second tool obviously collects more detailed information, which makes it possible, in theory, to more specifically define the relationship between the different individuals. The first tool is compelled to trace the entire network of relationships on the basis of those which attach each member to a single person. The method is an effective one when the reference person is at the center of the network (in some demographic surveys, it is advisable to choose a woman who belongs to the median generation).

The combination between the concept of household and that of family provides a hierarchical classification shared by many countries, who generally, in the presentation of results of their various surveys, adopt classifications at various levels of specificity according to the link between the phenomenon studied and the family structure.

c. Other Sources of Noncomparability

The differences between concepts and between definitions (of variables and units) only represent part of the problem to be solved in order to ensure comparability. Several other elements can prevent the comparison of estimates of the same phenomenon. Indeed, the construction of comparable statistical data (demographic, economic, and social) involves the whole process—from the appearance of the need for information, to making the data available. Harmonization should affect all the stages and all the tools for producing statistical information: instruments, nomenclatures, procedures, analytical methods, etc. In a demographic survey, for example, once the relevant variable has been defined, a set of questions must be decided upon, as well as the order in which they should be inserted and the whole of the questionnaire itself in order for this definition to be respected.

The questionnaire, its vocabulary, and its structure are key elements in constructing the comparability of variables collected in household surveys. Several efforts have been made to determine the best *translation* strategy in order to ensure better understanding on the part of respondents and to guarantee the highest possible equivalence of the measurements. Several methods are possible: the most common consists of adopting a questionnaire (or a module of a questionnaire) initially designed for a determined country, and then translating it into several languages. The implicit condition of this operation is that the national concepts that underlie the variables are similar to one another: in this case, the key element of the procedure is the quality of the translation, which should reproduce the precise meaning of the terms used in the original language and which suggest the same reactions in the respondent. Several methods were developed to improve translation quality and to overcome the tendency to place too much value on the formal equivalence of the questions in the different languages (precise equivalence of content) to the detriment of functional equivalence (the capacity that the question has of evoking the same concept to the respondent) (Brislin *et al.*, 1973). The most common method is that of back translation, in which the questions are first of all translated into the different languages (preferably several times by several translators) then translated once again into the original language (once more by different translators) in order to check for consistency.

In order to obtain good results, it is frequently necessary to forego strict adherence to the content of each question (Harkness, 2003; Harkness *et al.*, in press), but rather to take the meaning of a given word or expression in its everyday use and adapt it to different cultural and social contexts (Guyatt, 1993). The scale of responses is also sensitive to the context, insofar as the modes of response (whether they are numerical or verbal) can provoke reactions, and therefore produce different results. It would be useful to carry out experimental research on these themes, possibly resorting to psychometric tests and to the evaluation strategies that accompany them (WHOQOL Group, 1994, 1998), in order to reveal the communication mechanisms that regulate respondents' perceptions and reactions during the survey. These mechanisms are made up not only of questions, responses, structure, and sequence within a questionnaire, but also relations between interviewer and respondent and type of interview (face to face, *paper and pencil interview*, or *PAPI* type; *CAPI*, for *computer-assisted personal interviewing*; or *CATI*, for *computer-assisted telephone interviewing*). This methodological investment, common at the national level, has hitherto been missing from the development of comparative studies; its absence has in particular led to underestimating the importance of the period and the tools of the survey in the quality of statistical information, and therefore in comparability.

The circumstances that have just been described (adoption of a questionnaire for a single country, followed by translation) are undoubtedly not the most effective for obtaining comparable results. When the survey is designed from the start as an international project, it is undoubtedly preferable to carry out the design of the questionnaire (or of various thematic modules) in the different languages, with the participation of experts in statistical data collection to verify the similarities and record the differences between the national concepts and the social and institutional structures that correspond to these national concepts. The development of recommendations and standards for use during data collection can be a good strategy for reducing sources of incomparability, particularly when common tools result from a process of thorough dialogue. The *European Health Interview Survey* (EUROHIS) project provides a good example of a recommended procedure (Nosikov and Gudex, 2003a and 2003b). It was aimed at designing the necessary tools for the development of eight indicators essential for the *Health for All* program (chronic physical conditions, mental health, alcohol consumption, physical activity, recourse to curative medical services, consumption of medicine, prevention, quality of life), coordinated by the World Health Organization (WHO) European Office, with the cooperation of those in charge of statistical information on health in European countries. During the second half of the 1990s, a previous project had already led to the development of the tools necessary for producing six health indicators (perceived health, temporary incapacity, tobacco consumption, weight at birth, breast-feeding, body mass index) (de Bruin *et al.*, 1996). Thereafter, a *survey of surveys* on health carried out in Europe with the participation of 33 countries made it possible to select intervention zones for networking, with each network being responsible for developing a tool. The method adopted gave the project participants (each with specific responsibilities in their specialty) considerable co-responsibility, which enabled the emergence of a full consensus at every phase of the process, from the selection of indicators on which to intervene, the analysis of concepts used, and the development of operational definitions to field experimentation with the proposed tools (Nosikov and Gudex, 2003b).

Beyond the problems related to the questionnaire, a significant part of the lack of comparability between countries is due to the methods and modes of organization adopted for collecting and producing data. Let us list some of the most important ones (Egidi and Buratta, 1995; de Bruin *et al.*, 1996), by grouping them into four categories.

1. **Sampling factors**. The sample design and size are frequently very different from one country to another. The possibility of reliable comparison is often reduced when it is based on estimates taken from confidence intervals of unequal sizes. Even the reference population of surveys sometimes differs according to country, thus affecting the quality of comparisons: some surveys are representative of all the people living at a given time in a given territory in an ordinary household, with the exception of those living in institutions (hospitals, homes for the elderly, prisons, etc.), while others cover both groups. The sampling units can also differ according to country (individuals, families, households, housing units, etc.), the sampling frame can be constituted using nominative registers (population registers, electoral rolls, telephone directories, etc.) or it could follow a probabilistic strategy. We often underestimate the consequences of strategies for taking into account sample losses as much at the primary level (generally territorial units) as at that of secondary units (households, individuals, etc.). Some countries replace the missing units with substituted units, while others overestimate the initial sample so that the nonreplaced losses do not cause the sample to fall below the threshold necessary for statistical significance of estimates. Furthermore, in some countries, in the absence of the person who should have responded directly to the questionnaire, a *proxy* is interviewed, while in other countries this procedure is forbidden.

2. **Factors linked to the various data collection methods**. As discussed earlier, depending on the type of survey, the interview is carried out *face-to-face*, in direct interaction between the interviewer and the respondent, the questionnaire either being on paper (*Papi* method) or electronic format (*Capi* method); in other cases, the interview is carried out over the telephone (*Cati* method); lastly, the questionnaire can be self-administered (without any interaction with an interviewer). For some phenomena, the use of one method or another can cause variable reactions from respondents, either from one country to another, or from one subpopulation to another within the same country according to cultural specificities. When interviewers are used, the characteristics of their network can be determining factors: the best solution is obviously to use professional interviewers who have been specially trained and who are directly supervised by the person in charge of the survey; in many cases, at least one of these two elements is missing (interviewers who are not specialized or professionals, or monitoring fieldwork through a subcontracted company results in a screen being placed between the survey manager and the real conditions of data collection).

3. **Factors linked to time, to the distribution of interviews throughout the year, or as the reference period for the information collected**. For the same phenomenon, in some countries survey protocol requires interviews over different periods of the year (usually quarterly), and on one specific date in others. The reference period of some questions can also significantly influence estimates due to the effects of memory, which often acts differently in different cultural contexts.

4. **Factors associated with the procedures for collecting and analyzing data**. All the stages of the progress from raw collected data to statistical data are included in this group. The methods of identifying and correcting errors can differ considerably: countries use different consistency tests and different systems for correction based on deterministic or probabilistic models. Furthermore, the strategies develop over time with the introduction of innovations at both the conceptual and operational levels.

Tools, methodologies, and techniques can all encourage or impede comparability. In summary, comparative research should be able to rely on quality statistical information: precision, reliability, clarity and accessibility, consistency, and exhaustiveness are the conditions that, when they are respected in every country, result in good comparability (Eurostat, 1991).

Only one element of data quality can impede their comparability and that is the relevance of statistical information: the further the compromise concept adopted for comparative purposes moves away from the national concept and prevents satisfactory estimation, the more the information loses its relevance for knowledge of the phenomenon at national level conceptualization. It is necessary to take this into consideration in the design stage of the project as well as in the analysis of the results. This distancing has multiple negative effects: (1) in the current situation of limited resources available for statistical production, there is competition between statistical information that fulfills different requirements (national and international), (2) and the need for information at a higher level has suddenly been given priority (which could well occur under the effects of strong political pressure, such as that which is exerted today around the themes that are the subjects of community policies within the European Union). Information would lose consistency within its national context, at the risk of creating confusion among data users and reducing data validity, due to a reduced capacity to identify the errors. Where there is a valuable information heritage accumulated over time the conflict between comparability over time within a country and comparability in space with other countries is significant, and the reconstruction of chronological series making it possible to rehabilitate the previous heritage of information may be too onerous to be undertaken.

Thus, improvement in comparability is a costly process, both financially and in its effect on the overall quality of statistical information. The choice of a strategy of harmonization must take this into account and must resist the temptation of simplistic solutions that will inevitably lead to unsatisfactory results. Harmonization should always be considered a process that affects all of the stages of data production—from their definition, collection, and development to the construction of indicators—and a dynamic process that develops through successive stages respectful of the previous heritage.

2. Harmonization Strategies

The choice of a harmonization strategy depends on a number of elements, the first of which are the social, cultural, normative, and institutional distance between the national units to be compared.

The other elements to be taken into account are the characteristics of the object for comparison and the targeted objectives: comparison between indicators (macrocomparability or comparability of indicators), essentially with a descriptive goal or constructing integrated information bases from individual data in every country (microcomparability or comparability of individual data), leading to a microanalysis of the phenomena. Different levels of harmonization correspond to the two methods, imposing increasingly demanding conditions as we move from the aggregated level towards the individual level.

Traditionally two harmonization strategies are distinguished: that of *input* (or the harmonization of methods or sources), and that of *output* (or harmonization of results). Complex strategies have recently been developed, which borrow elements from both.

a. Input Harmonization

The input strategy requires that the different countries adopt the same definitions, and common methods and tools for collecting information (structure of the questionnaire and wording of questions, periodicity, sampling plan, sample size and rotation, form of interviews, methods of extrapolation, principles of checking and correcting data, coding methods, rules for imputing nonresponses, etc.; Grais, 1998, p. 12). This strategy makes it possible to eliminate or reduce as much as possible comparability problems related to the different methods of observation and the different procedures for the correction and analysis of data. It leads to good results if the social and cultural distance (and therefore the distance between the concepts and the structures) between countries is small and if it is possible to provide sufficiently clear and detailed definitions of the variables used. If this is not possible, the use of identical vocabulary (by way of an accurate translation) is often inadequate because of the diversity of institutional bases,[7] and the strategy leads to unsatisfactory results. Moreover, the operation can prove to be harmful insofar as it dissimulates the lack of comparability of the information that lies behind the apparent unity of the definitions, instruments, and methods.

Among the disadvantages of this approach is its considerable rigidity, which frequently imposes a choice between quality information for each country and information that is internationally comparable between the countries. Moreover, this is an expensive approach, as it generally results in the launching of new surveys, which can waste resources and frequently contradicts existing sources without necessarily improving the quality of the estimates.

But the method also has significant advantages, for when the quality is maintained, it guarantees comparability at the micro level and consequently makes it possible to create integrated information systems on individual data.

[7] This is the case, for example, of questions on income when taxation systems are different, or questions on the level of training and education when education systems are different.

The European Community Household Panel (ECHP) is an example of this type of approach: a socioeconomic longitudinal survey,[8] designed and coordinated by Eurostat and carried out between 1994 and 2001 in the majority of European countries (Eurostat, 1996a, 1996b). The experiment carried out on this occasion put into practice the strictest hypotheses of input harmonization, involving most of the factors according to a radical approach that was popular in the first years of the operation. Indeed, the strategy envisaged that countries adopt the same questionnaire or the same modules (translated into different languages), the same survey procedures, the same rules for monitoring households and individuals (then revised in some countries due to budgetary constraints), the same sample design, the same procedures for correction and imputation (further centralized by the services of Eurostat). Only certain aspects of the sampling could differ from one country to another (selection from registers or master samples compiled during censuses; the number of levels of sampling and stratification of the sample) (Peracchi, 2002). The experiment proved to be a positive one in many respects, but it also shed light on problems, which led to substantial modification of the strategy used thereafter in similar operations. First, the extreme complexity of the system implemented was quickly confronted by the impossibility, for the services of Eurostat, of ensuring effective and timely management and, consequently, in significant delays in the availability of the data. Second, aggregated indicators with a wide policy range were produced in significant contradiction to those taken from sources already available at national or European levels, which were frequently of far higher quality. More generally, the experiment revealed the excessive cost (and not only in financial terms) compared to the results that were frequently of inferior quality compared to those produced at the national level for the same variables and which, when they were the object of specific studies, revealed variations that were difficult to interpret without reference to persistent comparability problems (Robine *et al.*, 1998). An alternative policy for disseminating the information (obviously ensuring respect of statistical confidentiality) would have enabled useful investigation of these points, but the contractual rules and the high cost imposed by Eurostat on the researchers working on the ECHP data prevented the best use being made of the information collected, including of the comparative research.

[8] The survey covers a wide range of demographic variables and characteristics, employment behavior, income, health, level of education, housing conditions, migration, etc.

b. Output Harmonization

In the output strategy, only the final product (its definition and its characteristics) is precisely specified, but each country is free to choose the source(s) and methods of collection, correction, and analysis of the data they consider most appropriate.

There are in fact two strategies by which comparability can be constructed: *a priori*, when the output harmonization is designed at the same time the process data itself is produced; or at the end of the process, directly on the result. In the first case, called *ex-ante output harmonization*, the adaptation of the national concept to the international concept generally goes through definitions that make it possible to reconstruct the different aggregates in order to fulfill both national and international requirements. There are no constraints on the tools for collecting information and the countries are free to choose the source that provides the best quality information (statistical source, administrative source, combination of several sources). We thus reduce the risk of loss of relevance in the information at national level and we avoid breaking up temporal series, even if the disadvantage still persists of losing overall consistency of the information due to the possible dissemination of different estimates (corresponding to different definitions) of the same phenomenon. When it can be adopted, this strategy ensures the best comparability between the other requirements of quality, up to the micro level of data. The *European Household Budget Survey* (HBS) provides an example of this type of strategy: Eurostat puts forth strict and detailed recommendations on all aspects of the definition of analytical units, as well as the variables and classifications to be used (Eurostat, 2003a), and centralizes the individual data in order to create comparable aggregates and indicators.

Among the new operations, the harmonized EU-SILC adheres to this same strategy: strict and detailed rules are imposed on all the countries for the adoption of common definitions and classifications, while leaving them the possibility of choosing the best source (administrative source, statistical survey, combinations) from which to collect the required information. Moreover, the availability of microdata at the European level provides the possibility of carrying out *ex-post* harmonization on all that cannot be resolved during the collection phase (for example, the variables relative to gross and net income in countries with different taxation systems).

In the second case, *ex-post output harmonization*, intervention only affects the final result (micro or macro) of the production process. The quality of the result obtained depends largely on the degree of specificity of the national database and on its capacity to allow reconstruction of the information in the international definition. In the case of survey data, this strategy generally represents a transitory stage before an *ex-ante* intervention is possible. On the contrary, it is the most frequent method among those who do not have control over the production of information, and it is in general the only possibility for statistical data taken from administrative sources.

Classic examples of this strategy are provided by numerous compilations of demographic and social indicators published and disseminated by various international organizations. For example, the WHO's *Health for All* information systems or the Organization for the Economic Co-operation and Development's (OECD's) information system are valuable references for macrocomparative analysis, due to the range of the areas covered and the timeliness of their updates.

Output harmonization, whether it be *ex-ante* or *ex-post*, is far less costly than the input harmonization. The results, particularly those of the *ex-ante* variant are as good in terms of comparability as those obtained through input harmonization (for example, comparability is also possible at the micro level), while at the same time a good level of general quality of information is guaranteed.

c. Mixed Strategies and Targeted Strategies

Between the two extremes of input and output harmonization, numerous mixed solutions have been, and continue to be, tested, either on a purely transitory basis in a process that aims for complete harmonization of the input or possibly on a definitive basis.

An example of a mixed strategy is provided by the targeted harmonization of the new European survey on the workforce (*European Union Labour Force Survey*; *EU-LFS*), intended to provide information for Community policies on work and employment. Given the sensitivity of the theme, both at the national and community levels, a progressive advance towards better comparability has been planned, respecting as much as possible demands for the internal consistency of national information systems and the temporal continuity of the estimates (Mejer, 2003). To this end, strict rules have been established on the collection methods and characteristics of the survey, and a set of core variables of great importance at the community level have been defined, which must comply with the greatest comparability even for the short-term. The main objective of this survey is to subdivide the active-age population into three mutually exclusive groups (the employed, the unemployed, and the inactive) and to estimate the size and the main characteristics of these three groups. For these variables, the collection is totally standardized, to the extent that community law decides which questions to be asked and the order of the questions within the questionnaire. For the other variables, the definitions and characteristics of the quality required are fixed, but the countries are given the possibility of choosing the most appropriate source and method of collection. Little by little, taking advantage of the periodic reorganization of their surveys, the countries must ensure a progressive convergence towards the standards associated with every variable in terms of definitions and collection methods (Eurostat, 2003b).[9] For example, in 2002 a Commission regulation[10] considerably modified the concepts and measurements in the modules on school attendance and the level of educational attainment. In 2003, the survey in the different countries integrated the new modules, thus progressing further towards comparability. In the same year, the final phase in the progress—from an annual survey (quarterly in some countries) towards a continuous survey that provides quarterly estimates—was launched.

The level of comparability already reached by the harmonized survey EU-LFS is high and should continue to progress in the future. The most important aspect is the fact that, for the first time in the history of European statistics, the complexity of the comparability objective has been recognized, as well as the impossibility of attaining it through immediate and simplistic solutions, and the absolute necessity of simultaneously taking into account national and international requirements to ensure the statistical quality of the results. Harmonization through sharing and consensus is undoubtedly one of the main results of the experiment developed in Europe for creating an integrated and comparable information base on employment.

3. A Particular Problem: Statistics of Administrative Origin

In recent years, increasing economic constraints have moved the focus away from statistical sources towards administrative sources, in order to limit as much as possible the direct collection of data in fields in which sources of this type do not exist. Administrative sources have numerous advantages: apart from

[9] Council Regulation (EC) no.577/98; Regulation (EC) no.1991/2002; Regulation (EC) no.2257/2003; Commission Regulation (EC) no.1575/2000; Commission Regulation (EC) no.1897/2000; Commission Regulation (EC) no.1897/2000; Commission Regulation (EC) no.2104/2002.

[10] Commission Regulation (EC) no.2104/2002.

their low cost (at least when it is limited to the marginal cost of production), they generally provide universal data, which ensures detailed representation at various territorial levels, as well as population groups—even small ones; they do not require the good will of respondents, individuals, or families; they are continuously available and can, at least in theory, provide rapid results; when they are well used (for example if an individual identifier is available) they can make a longitudinal reconstitution of information possible.

However, there are also considerable disadvantages. The concepts used in the collection of the information respect administrative requirements and do not necessarily correspond to those that would be preferable for statistical aims or for definitions relevant to the international level. This poses problems at the national level, insofar as the relevance of statistical data is frequently limited, and more so when the aim is international comparability. Furthermore, data quality is largely dependent on the effectiveness of public administration and in the benefit of the data for the administration, the vigilance of the compilers being directly associated to the relevance of the data for their own activities. Moreover, even the goodwill of the individuals and families can affect the reliability of the information which they provide. This reflects the image which they wish to give to the public administration, which is sometimes far from reality (for example, statistics on income based on declarations made to tax services). Another disadvantage is the rigidity of administrative sources with regard to modifications and innovations, particularly in cases where the statistical authority does not have control over the forms used in administrative procedures.

a. Statistics on Causes of Death

Statistics on causes of death are often the only information available for comparing the health conditions of countries. They are exhaustive as they cover all the events recorded in a year, and their comparability is not affected by the problems related to sampling procedures. Moreover, as they are among the oldest statistics collected regularly by the countries, the collection procedures have become relatively homogenous over time: there is a standard model recommended by WHO, adopted nearly systematically by the different countries, there is a widely adopted nomenclature, the *International Classification of Diseases* (ICD) (WHO, 1995) which is in its tenth revision. Despite this favorable situation, problems persist with regard to the quality and comparability of data (EC and INSERM, 2001; Duchêne and Wunsch, 1995; Meslé, 1995; see also Chapter 42 of Volume III) and in many countries it would be necessary to update and modernize the collection procedures and tools.

Taking into account the considerable homogeneity of the concepts and definitions in this field, the comparability problems are mainly due to the differential quality of the data collected and analyzed in the countries. Two phases can prove to be critical: certification, i.e. the declaration made by the doctor of the series of causes that resulted in the death, and codification, i.e. the determination of the underlying cause and its translation into an ICD code (currently ICD-9 or ICD-10).

In 2001, the European project *Comparability and Quality Improvement of European Causes of Death Statistics* (CépicDc and Inserm, 2001) was concluded. It launched a detailed survey on certification practices and the quality and comparability of statistics on causes of death in Europe. The survey revealed the main problems for comparability (ISTAT, 2004; Pace *et al.*, 2004a). The first problem is due to the use of different death certificates: there are countries which still do not use the four lines advised by international recommendations (one for the initial cause, two for any intermediate causes, and one for the immediate cause). The second problem is the diversity of strategies for collecting information on infant mortality: many countries have adopted a special certificate (as recommended by WHO), while others, frequently those who have registers of births and deaths or hospital data that they can pair through the intermediary of an identifier, do not use different certificates for infant deaths but provide the information necessary to differentiate them in the registers. Furthermore, the definition of the infancy period is not the same everywhere: in some countries it is considered as the perinatal period (with some variants: between 154 and 196 days of gestation, up to 6 or 7 days after the birth), in others it is the neonatal period (27 to 28 days), and in others it is the first year of life (364–365 days). Moreover, the distinction between stillbirth and abortion has various definitions: in some countries stillbirths are determined by the gestation period, which can vary from 0 (Turkey) to 154 days (22 weeks) or 196 days (28 weeks); in others by the weight of the fetus, which can vary from 500 to 1000 grams.

The survey also revealed differences in procedure which can affect comparability: for example, differences in methods of diagnosis and an unequal proportion of autopsied deaths in different countries (in some countries, the autopsy of all violent deaths is obligatory, but not in others), from 7 to 38% (WHO, 1998; Pace *et al.*, 2004a and b), with the possibility of diverse consequences on death certificates. In cases of violent or suspicious deaths for which a specific procedure is adopted that includes the involvement of specialists (for example, the coroner in the United States and in Great Britain), the process is longer and more complex and can result in a deterioration in data quality that, as it cannot take into account the successive investigation of the information, results in a high number of deaths being attributed to an unknown cause, particularly for external causes and for young people.

Coverage problems can also occur. As a general rule, the resident population is the reference population for statistics on causes of death. But this is not always the case and most of the time it concerns the deaths of residents who die in their country of residence. The deaths of residents who die abroad does not generally include reference to the cause, thereby contributing to the number of deaths attributed to an unknown cause. Furthermore, it is not always clear if deaths of nonresident foreigners are counted. To overcome this type of problem, the creation of a centralized clearing house which would redistribute information on causes of death of residents dying abroad between the countries of residence is envisaged.

The coding phase can also have an effect on the quality of information and its comparability. The first problem appears when a new version of the ICD is published, which normally causes interruption in series (at least at the level of the detailed sections which have three or four numbers: Vallin and Nizard, 1978; Nizard and Vallin, 1980; Vallin and Meslé, 1988, 1998) and problems in comparability between the countries that use different versions of the nomenclature: these problems should become one of the main aspects that the WHO should take into account in its development of a new revision. The interruption of series considerably reduces the value of information on causes of death and this element should be explicitly included in the evaluation of costs and benefits when the best compromise between innovation and continuity of information is being sought.

An important contribution to the quality of data and to their harmonization can result from the introduction of new technology: automatic coding, particularly if it is designed in a coordinated manner during the initial introduction or in the continuous updating process of optical recognition and coding forms, limits errors and improves comparability. There is much to be done in the years to come in order to accelerate the conversion to automatic coding in the countries which have not yet adopted it and in order to facilitate the convergence towards homogeneous tools.

The European project devoted to comparability and the quality of statistics on causes of death formulated 39 recommendations which should serve as model principle for the countries which, in the next few years, are planning to modernize the collection of information through the introduction of new tools and procedures. The recommendations come in eight chapters devoted respectively to:

1. coverage
2. confidentiality
3. organization of services in charge of causes of death
4. certification of infant mortality
5. certification of mortality at other ages
6. complementary surveys
7. training
8. collecting information on mortality in Europe

Many recommendations relate to certification, and particularly to the training of certifiers, in faculties of medicine or in the context of the continuous training of medical personnel (Recommendations 30–36). Sessions in basic training (e-learning tools) adaptable to national specificities (Recommendation 33) have been developed and are currently being adapted (Pace *et al.*, 2004b) to facilitate training and to provide the certifiers with a common knowledge base. Furthermore, the importance is recognized, as much in terms of quality as comparability, of applying new technology to certification: the development of an electronic and harmonized death certificate is encouraged (Recommendation 22), as well as the creation of a common site for the networks of Eurostat and WHO on the certification of causes of death, destined for certifiers from every country (Recommendation 36).

The strategy of the European Union with regard to causes of death is therefore a long-term strategy and it takes into consideration the complexity of a task which largely eludes those in charge and the rules of statistics and often falls to personnel who are not frequently aware of the relevance of the task of certification. The full harmonization of input, through the progressive convergence of all the countries towards common procedures and tools will therefore take time. Even though the process could be accelerated through imposing a European Regulation on all countries, raising the level of awareness and training certifying doctors which is essential to solving some problems of data quality requires time. In the meantime, output harmonizing, on a minimum of information (abridged European list of 65 causes of death, age, concept of resident population), as well as the progressive introduction of various innovations in certification practices and tools as well as in the coding of information, should result in an improvement in the situation.

b. Administrative Registers and Statistical Sources: Problematic Harmonization

The creation of administrative registers is generally a long, complex, and costly process, but once the system is functioning correctly the production of statistics is relatively simple and economical. The countries of northern Europe in particular are showing increasing resistance to carrying out surveys on issues whose responses are already included in their administrative registers. These countries have long since harmonized their systems for recording the main demographic and social events, on the basis of a central population register and a single personal identification number (PIN) used in all contact with the administration. This number makes it possible to couple all kinds of information recorded on different public administration databases (civil registration, education, tax, health, social security, etc.).

Resistance to carrying out statistical surveys in countries with effective systems of administrative registers is often reinforced by laws prohibiting statistical authorities from asking questions which might not be strictly essential, in order to spare citizens innumerable demands from the public administration.

The position of these countries was clearly asserted within the group responsible for the *Generations and Gender Survey* (*GGS*) and a subgroup, Administrative Records Group (ARG), was created to study how some countries could extract some of the responses to the survey's questionnaire from existing information bases (Statistisk sentralbyrå, 2004). Indeed, it seems that in the northern European countries, a substantial part of the information (30% to 50% of the questions) can be obtained from existing registers, but using them requires a preliminary detailed analysis of the quality, validity, and coverage of the registers, as well as effective correspondence between the events included in the registers and the definitions used by GGS. A detailed analysis of the cost of extracting these data is also needed as, frequently, the marginal cost of administrative information is only low in theory and extraction can require complex and costly data processing procedures, as well as long administrative procedures to obtain the necessary authorizations.

4. Harmonization and Quality of the Information

The classification of statistical sources and the strategies for harmonization illustrated to date have above all a theoretical value. It is very rare for the distinctions to be so great in practice. The realities are multiple and complex between a step towards *ex-ante* harmonization of tools and methods (input harmonization) and the *ex-post* harmonization of output. They can go through all the intermediate combinations in the search for the difficult balance between cost, quality, and comparability.

Harmonization always has a cost, not only in economic terms but also in terms of the informative capacity and quality of the data. The adoption of concepts which decrease the relevance of national information, or which create inconsistencies in the data bases, can have negative effects on data quality while creating a dangerous illusion of comparability which can lead the most careful researcher to errors. Therefore, the problem lies in the level of comparability which should be aimed at in order to reconcile (at a given time, and for the targeted objective) comparative necessity, relevance of the information, and quality.

To progress towards more comparability, it is necessary to improve scientific knowledge of the obstacles to comparability between countries: nothing is obvious and everything should be verified and evaluated. Indeed, it is not always the case that the greatest comparability results from the adoption of strictly identical definitions and standardized procedures. As long as the institutional and normative characteristics of countries differ, without mentioning their social and cultural structures, the best comparability for all the phenomena influenced by these characteristics and structures will be obtained by the most flexible strategies for collecting and analyzing data. Such a strategy requires efforts in the area of theory and definitions which will make it possible to clarify the characteristics of the required result. Once the objective has been defined, strategies for improving comparability will generally coincide with those of improving the quality of statistical information. There is no shortcut to this aim, as numerous attempts at creating comparability through standardized procedures and tools have demonstrated. Total harmonization is only possible between countries whose economic, social, normative, cultural, and institutional contexts are similar.

More generally, the strategy that the United States has adhered to for a long time is the one which can lead to the most satisfactory results: a considerable effort at harmonizing the definitions and classifications is required, as well as improvements in and specification of the performance of the procedures in order to improve results and to reduce the risk of researchers mistaking the real possibilities of making good comparisons between countries.

III. METHODS OF COMPARISON

Quantitative demographic analysis, as we know it today, developed according to two streams, and at two distinct periods.

The first, and the oldest, is based on recording the volume and the characteristics of populations, as well as their transformations, by measuring the contribution of various phenomena (mortality, fertility, mobility) to these situations and their trends. It is based on aggregated data, taking the form of statistical tables. The indicators designed from these data reflect the frequency of the events or the prevalence of the situations and makes it possible to monitor their evolution and to model their dynamics. This way of thinking based on macrosocial aggregates was more or less exclusive from any other during the first half of the 20th century in demography but also in economics and sociology at the time of Alfred Lotka, John Maynard Keynes, and Emile Durkheim.

From the 1950's, focus shifted to a better understanding of the individual behavior at the source of the social situations and trends. The development of household surveys was an essential expression of this. This has resulted in a modification of the type of statistical analysis applied to the data collected. Focus was on the diversity of human behavior with regard to health, family, and a search for the reflection of the determinants affecting demographic phenomena. The base was thereafter constituted of individual data files rather than statistical tables: with the stress being placed on variability rather than on central values, on the individual rather than on social facts.

The difference between the two approaches is not in the nature of the information used. Statistical tables from censuses or civil registration data were based, at least from the middle of the 19th century, on collecting information on individuals, which does not fundamentally differ from sample survey procedures. But focusing on society in one case, and on the individual in the other, gives a very different position to the nation in global analysis. International comparisons are therefore approached very differently according to whether they use indicators based on aggregated data or the direct analysis of individual data. Therefore, at the time of analysis, there is a dichotomy already introduced when the various stages and strategies of the harmonization process are described.

1. The Comparative Analysis of Indicators

Demographic analysis aims at measuring behavior which contributes to the renewal of populations and understanding its influence on population growth and structure. It achieves this first by defining increasingly small *risk groups*, in order to find pure indicators, and by neutralizing the size of these groups by calculating rates or proportions. It then proceeds to synthesize elementary indicators by using models which combine, within various hypotheses, a series of rates or proportions (tables, comparative rates, etc.).

Because they eliminate size from the risk groups, these operations create the minimal conditions necessary for international comparison. If the absolute numbers of death are not comparable from one country to another due to different population sizes, the crude mortality rates are more comparable as they apply to the same number (generally 1,000 people). In the next phase, rates by age, as well as their synthesis in mortality tables are even more comparable as they overcome differences in age structure (see Chapters 5, 11, 14, and 15 of Volume I).

The indicators which lead to demographic analysis thus lend themselves directly to an international comparative analysis. But their use is not without some problems. For example, the complex indicators call upon hypotheses which need to be verified. In the calculation of comparative rates, the choice of standard population can influence the result when a country's age-specific rates are not systematically higher or lower than those of another country: for example, if two populations are distinguished from each other because one has a lower mortality rate before the age of 50 than the other, and that the opposite is true for later ages, the comparative rates will list the two countries in a different order according to whether the model population gives more weight to the young ages or the older ages. Moreover, syntheses which use a fictitious cohort can be affected, unequally according to country, by the effects of cohorts likely to distort the interpretation of the differences, from one country to another: later on we will see an example where the total fertility rate is affected both by the variations of mother's age at birth and the completed fertility of the cohorts, and the classification of countries can thus differ according to whether fertility is summarized by the total fertility rate or completed fertility.

a. Comparative Demographic Analysis is Descriptive

Demographic analysis reveals the effect that a specific group of factors has on demographic behavior and on the size and structure of the population. In addition to sex, they are mainly categories of temporality, whether they are the age of the individuals, their membership of a cohort, or the location of phenomena in history (duration, cohort, period). Although the web of these different scales of time can be extremely complex and its disentangling very informative with regard to the mechanisms of demographic behavior, this exercise and its results are often considered as descriptive, compared to that which the introduction of socioeconomic, cultural, or biological aspects contributes to the interpretation of behavior.

Comparative demographic analysis, by observing the diversity of behavior across international space is a step towards this interpretation. But it still remains descriptive and restricts itself to noting differences between countries before attempting to explain why demographic behavior varies in a given national context through socioeconomic, cultural, or other variables.

At the beginning of the 21st century, the population pyramids of the different countries of the European Union were thus compared (by bringing them to a standard of 100,000 inhabitants), to show their unequal levels of ageing, due to different fertility trends, to unequal progress in combating mortality at old ages, and to the differentiated consequences of international migration. "*The population pyramid of the twenty five Member States of the European Union provides an image of the secular history of these countries. The marks of the First World War are becoming blurred, but the deficit of births due to the 1939–1945 war is still visible in many countries where natality declined during the conflict. For the second half of the XXth century changes in natality have sculpted the pyramids in peaks and troughs, but these variations reflect national economic situations that no longer have the European dimension of the two World Wars. In addition, these variations cancel out and no longer appear on the pyramid for the whole of the European Union which is essentially distinguished by a major and general phenomenon: the fertility decline observed since the 1960's, at variable dates according to the countries*" (Monnier, 2004).

In this example, the comparison is static as it relates to populations characterized at the same time, but demography carries so many references to different dimensions of time that the analysis quickly becomes dynamic. In other cases, comparative analysis is dynamic from the start and covers chronological series rather than period data at a given date or a given period (for example, monitoring the current demographic situation).

The resources of demographic analysis are placed at the service of comparative observation in order to disentangle the role of demographic factors in the differences between countries. The evolution of the total fertility rate is thus distributed between what is due to the variation of completed fertility in different cohorts and what is due to changes in age at maternity to explain the role of one or the other in the differences between countries. "*The progressive reduction in age at maternity observed everywhere, and the provisional and identical reactions of women in all the countries of Europe to historical, economic, and social events leading them to more frequently reject or accept maternity at certain periods, are the source of the similarities in the evolution of annual rates. But this homogeneity of behavior, over the years, has not resulted in a correspondingly homogeneous evolution in the final number of children per woman in all the countries: the trend of completed fertility seems more linked to the starting level, declining in the countries where it was high, and increasing or stabilizing in countries where it was low*" (Blayo and Festy, 1975). The differences in life expectancy at birth from one country to another are also distributed between risks of death at various ages. "[From 1981 to 1996,] *most of the increases in life expectancy came as a result of decreasing mortality at older ages (ages over 60). This pattern was observed for males and for females, in East Germany as well as in West Germany. The pattern was more pronounced for females than for males, and it was more obvious in East than in West Germany*" (Gjonça *et al*., 2000, p. 8).

The compared countries are generally all the more numerous when the indicators are simple and their calculation easy. The differences finally exposed and measured by the demographic analysis are linked to an intuitive explanatory model which, at this stage, will not be subjected to a statistical verification in the real meaning of the term. The differences between eastern and western Europe have been a major subject during the decades following the Second World War, with the demographic patterns revealing the confrontation of the political regimes (Bourgeois-Pichat, 1985; Monnier and Rychtarikova, 1991; Meslé and Vallin, 2002). More recently fashionable, the difference in behavior between the countries of southern and northern Europe can lead to the suggestion, according to the case, of differences between Catholic and Protestant traditions, between systems of family and social interdependence, or between cooking with oil and cooking with butter (Ferrera, 1996). The measurement of differences used, whether relative or absolute, is based on descriptive statistics, not the modelling which would support a model of an *a priori* gap which reality would verify.

b. Comparative Analysis in Search of Universal Laws

A major objective of political arithmeticians, and of Adolphe Quetelet after them, was to search within the diversity of empirical observations for the set of *universal laws*, which were sometimes even conceived as revealing the *divine order* (Süssmilch, 1741). International comparison was one of the preferred methods for this. Many years after them, this state of mind is found in research into statistical regularities which reveals the action of underlying extrademographic variables.

During the 1950s, in order to establish the concept of natural fertility, Louis Henry (1961) grouped the curves of the age-specific legitimate fertility rate (number of legitimate births per married woman) established in very different countries; he showed their similar convex form at variable levels (an initially barely perceptible decline in rate when age increases between 20 and 30 years, then a progressive acceleration up to zero rates from 40 or 50 years); he deduced from this the role of physiological (fecundity, fecundability) and behavioral (breast-feeding) factors independent of the number of children already born, without being able to measure and specify the impact of each of these factors.[11] During the same period, the first *model life tables* were developed by compiling national tables in as diverse a group as possible of countries and periods and by exposing the relations between risk of mortality at successive ages which revealed a general trend of increase in mortality with age (beyond early childhood) and a typical rate of the progress in combating mortality over time. "*We had for a long time noticed that when mortality was high at certain ages, it was also high at other ages. In other words, knowledge of mortality at certain ages certainly does not make it possible to determine mortality at other ages, but it provides an idea of what it could be. To express the same idea differently, experts came to the conclusion that, for the same health situation, the model life tables observed were similar and their average could therefore be considered as a good means of spotting this health situation*" (Bourgeois-Pichat, 1994). That Jean Bourgeois-Pichat partly relied on international comparisons to demonstrate the rule of division of mortality in endogenous and exogenous components during the first year of life shows the importance of comparative reasoning during this period (Bourgeois-Pichat, 1951).

Here we will notice that, in particular in the example of model life tables, the empirical observations made on the various countries enter into a theoretical model of dispersion around central values, an expression of the law sought, which justifies the use of measurements and an analysis of variance, through the set of correlations within the group of points drawn by mortality rates at successive ages in the series of 150 national tables collected by Vasilios Valaoras for the United Nations (United Nations, 1955).[12] The same idea can be found today, although the search for universal laws has lost its appeal and has been replaced by the search for a possible convergence in demographic behavior in a group of temporarily diversified countries. The evolution over fifty years of the linkages between fertility, nuptiality, and fertility outside marriage, subjected to analysis of variance in sixteen countries in Western Europe by Alain Monnier, suggests waves of homogenization and heterogenicities which contradict the hypothesis of a movement towards a common stable and sustainable point. "*The main criticism that can be made of the concept of convergence is that beyond the 'point of convergence' nothing else happens. In reality, new behavior continually modifies the European family landscape and, while it spreads, introduces a differentiation factor. This is what occurred with the spread of modern contraception, induced abortion, births outside of marriage, and definitive voluntary infertility. Tomorrow, other elements of family life, which are still very marginal (for*

[11] "*The arithmetic mean of 13 series*

- 9 relative to European populations, 5 of which before 1800;
- 3 relative to populations in Asia (Formosa, India, Iran);
- 1 relative to a population in Africa (Guinea)

is the following:

[Indicators]	Woman's age					
	20–24 years	25–29 years	30–34 years	35–39 years	40–44 years	45–49 years
Legitimate fertility rate	0.435	0.407	0.371	0.298	0.152	0.022
Rate (base 100, average 20–24 and 25–29 years)	103.5	96.5	88.0	71.0	36.0	5.0

The shape of the curves, which represent the 13 series, obviously differs from this average but only very slightly; the average indicator for the 30- to 34-year and 35- to 39-year age groups is only from 67 to 89, 9 out of the 13 values are between 76 and 81; the average rate for the 40–44 and 45–49 age groups is more dispersed (14.5 to 28.5); but part of this dispersion is doubtless due to the influence, likely in these groups, of errors in age declaration" (Henry, 1961, p. 629–630).

[12] The validity of the statistical model used by Vasilios Valaoras (United Nations, 1955) to construct the United Nations' first model tables was seriously disputed thereafter, particularly by K.R. Gabriel and Ilana Ronen (1958, p. 164–169) then by Sully Ledermann (1969). The following model tables were created according to the same intellectual principle, but the statistical method had been vastly improved (Valaoras, 1973).

example couples who 'live apart together' as it is called by Catherine Villeneuve), will act as a new element of diversification. The history of the family is an endless story" (Monnier, 1998).

By building a single network of model life tables based on information contained in around 150 national tables, Vasilios Valaoras took advantage of the similarities revealed by comparisons between countries and periods to establish a reference, a norm which would make it possible to compensate for any shortcomings in statistical information on other populations. By supposing that the mortality model of these fits the form of reference, we can deduce the probable risks of mortality where it is not possible to calculate the actual risks.

Paradoxically, the reference can also be used for shedding light on the specificity of those countries for which good quality statistical information is available, and thus revealing their originality compared to the rest of the world. This type of comparative analysis is used as much for an improved understanding of the situation in each country as it is for questioning the norm on which the comparison is based.

For example, Ansley Coale and Paul Demeny (1966), took up the work of Vasilios Valaoras a few years later, systematically comparing each national life table to the corresponding model life table of the United Nations with the same life expectancy at birth and observed three typical ways of deviating from the reference. These three ways were dominated by the countries of southern, northern, and eastern Europe, respectively, they diversified the United Nations' single network of model tables by adding three others, denominated from the same three cardinal points, and giving the title of West to models for countries that could not be included in one of the other three groups. The single and universal norm was thus separated into four regional norms, through a new comparative analysis.

Furthermore, Jean Bourgeois-Pichat's adjustment, which places on to a line the points representing the cumulated number of deaths during the first twelve months of the life of infants, quickly revealed itself to be inappropriate for countries in which infant mortality is dominated by deaths due to respiratory causes (the deaths accumulated more quickly than is foreseen by the adjustment, at the beginning of the first year of life), or deaths due to digestive causes (the deaths accumulated more quickly than is foreseen by the adjustment, during the second semester of life). But by reversing the logic which initially dominated Bourgois-Pichat's research, a country's difference from the classic adjustment can be used to suggest a frequency of respiratory mortality (cold countries) or digestive mortality (hot countries) that is above the norm.[13] A country's distance from the reference becomes a means of revealing a country's specificity.

Lastly, the gap between the curve of the age-specific legitimate fertility rates in a country and the natural fertility curve drawn by Louis Henry for the populations where no Malthusian behavior reduces fertility after the birth of a certain number of children, was used by Ansley Coale and James Trussell (1974) in order to evaluate the impact of contraceptive behavior, without however measuring the latter's prevalence.

Thus, within the same intellectual context, comparative analysis is successively used to reveal the *laws of behavior*, which unite a large number of countries, as well as the obstacles to the universal adherence to these laws through a set of national specificities, like the many impediments to convergence towards a single model.

c. Identifying Explanatory Factors

Although the search for universal laws is no longer on its agenda, demographic research continues to wonder about the factors which are at the origin of changes in fertility, mortality and migration. Their association through ecological correlations taking the countries as a connection point is a common method here.

For example, when Jacques Vallin (1968) questions whether the decline in mortality is linked to economic development in Third World countries, he plots a graph for life expectancy at birth with the average income per inhabitant of those countries for which these rates are known. He states "*that the group of points fits entirely into the angle made by the y-axis with the first bisector. In other words, for an annual income per capita of 50 to 150 dollars there is a wide range of possible life expectancies ranging from 27 years (Guinea) to over 63 years (Formosa) and includes no fewer than 23 countries occupying the intermediate positions. On the other hand, when the income per capita increases, the likelihood of high mortality seems less. Beyond 300 dollars per capita per year, life expectancy does not fall below 55 years. Therefore, if there is a link between mortality and standard of living in Third World countries, this link seems to be one-way: a very low standard of living is not an insurmountable obstacle to an increase in lifespan but a relatively high income cannot be imagined without high longevity*" (Vallin, 1968).

[13] For example, "*The exceptions to the Bourgeois-Pichat Law, in many countries or regions in Southern Europe, North Africa, and no doubt the Middle East, come from excess mortality due to dietary risks, which only appears at the age of a few months and which mainly occurs during the hot season and increases the excess mortality that these countries may also have due to other causes*" (Biraben, 1957, p. 641–644).

Émile Durkheim, at the end of the 19th century, acted in the same manner when he established the existence of "*concomitant variations*" between the suicide rate and the proportion of Protestants in the thirteen Prussian provinces. "*The simple parallelism of values through which both phenomena pass, provided that it has been established on a sufficient number of sufficiently varied cases, is the proof that a relation exists between them. The method owes this privilege to the fact that it attains causal relations*" (Durkheim, 1937).

In any example, the interpretative model associates a demographic characteristic of a country (in Vallin's case, the survival conditions measured by life expectancy at birth, in Durkheim's case, the risk of suicide) with an economic or social characteristic of the same country (the level of economic development measured by income per capita in one case, the "*degree of integration into religious society*" measured by the proportion of Protestants in the other case). It is a holistic system of general explanation in which the whole of the interpretative model is situated at the macrosocial level.

The risk would be of transposing the interpretation to the individual level and deducing through the example of the relation established between the countries that individuals suffering from a low standard of living are barely affected in their health conditions (this could be either mediocre or medium) while those with a better income are not exposed to the risks of very bad health. This transposition can lead to an error known as *ecological*, when the measured relationship between macrosocial characteristics are not verified at the individual level. Daniel Courgeau (2004) thus demonstrated that migration from Norwegian regions is higher when the proportion of farmers is high, but that the mobility of farmers is lower than that of the rest of the population. Moreover, in Western Europe, current fertility is lower in countries which still have high proportions of practicing Catholics (Italy, Spain), while fertility remains higher among practicing Catholic couples than among those who are not. Furthermore, Daniel Courgeau emphasizes that the risk is more likely in comparative analyzes at the aggregated level when the initial hypotheses relate to the determinants of individual behavior rather than reasoning about social groups: "*Although the approach is macrogeographic, the characteristics selected in the models are according to our concept of individual migratory behavior . . . Moreover, the interpretation of the effects is often given in personal terms: it will be said that it is the attraction of a higher salary which encourages an individual to migrate, whereas the model only shows a relationship between the number of migrants and the average salary in the destination zone*" (Courgeau, 1994).

With this warning in mind, it is possible to introduce into the analysis of the differences in demographic behavior between countries a plurality of social interpretative factors, which progresses from a simple regression, for which Vallin and Durkheim provide illustrations, to multiple regressions. For example, in a recent article, Lee Budget (2003) links the recognition of homosexual couples by industrialized countries to the expression of social norms concerning marriage, religiosity, and tolerance with regard to male-female relationships on the one hand, and sexuality on the other hand. These norms are measured from the results of opinion polls with individual responses being aggregated to the national level. In order to guarantee the comparability of variables, she uses the single harmonized source of the *World Values Survey*, in which the 24 countries of the OECD (as well as South Africa) participated.

Two of the main problems of comparative analysis based on aggregated data are highlighted here. Data comparability, by using a group of social dimensions as explanatory variables causes problems which, as we mentioned earlier, have been tackled far more recently than the comparability of demographic events themselves. The field of comparison is thus restricted to those countries which have data taken from a single standardized source, resulting from a particularly delicate procedure as they concern questions on opinion and attitude which are always difficult to transpose from one culture to another. The issue of the number of statistical units therefore becomes a crucial one. In the most common case of a detailed analysis referring to such a relatively homogeneous group, at least in procedures for statistical production, there will be twenty or so countries at a given point in time which will oblige the analysis to be sparing in the number of predictive behavior variables studied (*parsimonious specification*).

In all cases, the geographic diversity of demographic facts is explained by other demographic, economic, or social facts, at the same level of aggregation. These were implicit in the previous section, they are explicit in this section. International comparisons play an important role in revealing these relations. Did Émile Durkheim (1937) not say that "*the comparative method is the only one that suits sociology*"?

2. Comparative Analysis of Individual Data

With the development of individual data we move away from explaining behavior indicators with social variables where both are at the same level. For interpreting individual behavior, we initially use the determinants, themselves individual, and the national dimension then has problems in defining its role.

In the lessons drawn from the *Fertility and Family Surveys* (FFS) which will serve as a common theme to illustrate the possibilities provided by individual data, Daniel Courgeau (2002) emphasizes "*The tendency is to consider individual behavior as being influenced only by individual characteristics. The danger here is of committing the atomic error, that is, of ignoring the context in which human behavior occurs. In reality of course, individual behavior is influenced by context, and it seems fallacious to consider individuals in isolation from the constraints imposed by the society and milieu in which they live*".

FFS was built on the biographical model, where complex life histories with marital, family, educational, occupational, and residential dimensions were collected with the aim of associating these through a methodology specific to individual data, such as the *event history technique* (see Chapters 23 and 24, as well as Chapter 134). But at the same time it is a comparative program which aims at placing the results within an international context through the introduction of a macrosocial dimension into the analyzes.

The research carried out based on the *FFS* provides numerous examples of the differences between countries in the use of individual data. These attempts differ significantly in their geographic coverage, their objectives, and their ambitions. This diversity clearly finds its place in the classification of three posts proposed by Anne Gauthier (2002), based on her experience of various comparative research programs:

- a descriptive analysis of the difference of situations in a large number of countries
- parallel analyses in several countries of the relationships between variables at the individual level, in order to test the generality of observations made in a country, referred to by John Hobcraft (1980, 2002) as *replicated models*
- the introduction of national level variables in relationships at the individual level, in order to explain the share of international variations which are not due to individual factors.

a. Descriptive Comparative Analysis

Descriptive studies generally cover a large number of countries spread all over Europe, so that the diversity of situations is automatically represented or constructed by a rational choice. Remaining at a descriptive level, these studies highlight the similarities which unite or separate the countries, most frequently on the basis of geographical vicinity (for example, Scandinavian and Mediterranean countries, eastern and western countries); in some cases an interpretation is put forward, based on common sense qualitative assertions, beyond geography but without a thorough demonstration (Scandinavian and Mediterranean countries become known as welfare and "familialistic" states; eastern and western countries become known for socialism and market economy).

For example, in order to observe all the countries of the European Union, Kathleen Kiernan (2002) uses the *Eurobarometer* of 1996 for simple indicators, before using the data of *FFS* for a more thorough analysis in thirteen countries (plus Great Britain, thanks to the *British Household Panel Survey*). Groups are formed *a priori*: "*the countries have been subdivided into four sets: the northern set includes Norway, Finland and Sweden; the western European set includes Austria, Switzerland, France, Great Britain and Germany (we further subdivide Germany into East and West given its different history for much of the postwar period); the southern European set includes Italy and Spain; and the eastern European set includes Poland, Hungary, Latvia and Lithuania*" (Kiernan, 2002, p. 59).

In most of the countries, some results are disaggregated into subgroups, according to three important individual characteristics: level of education, religious affiliation, and experience of their parents' separation.

Union formation is analyzed according to a sophisticated but traditional model: the proportion of individuals who have formed couples, type of union, age at first union and its duration, conversion of consensual unions into marriage, proportion of couples without coresidence. Then it is operated for the dissolution of unions, by duration since the formation of the couple, according to the type of union (with or without premarital cohabitation, with or without marriage).

Differences and similarities between countries are highlighted. The homogeneity and heterogeneity of geographic groups are stressed, but in no way does the text go beyond "*an overview and description of changes in union behavior*" (Kierman, 2002, p. 57).

The political aspects of the trends are the subject of a paragraph on the development of new legislation in the countries of northwestern Europe, concerning informal unions, including homosexual unions. The adoption of these laws is clearly presented as a political response to changes in families, and not as a possible factor of these transformations.

Here is a conclusion that is typical of this prudence: "*In this study, we have seen that in southern European and in some eastern European countries, marriage is still the pre-eminent marker for entry into first union. However, in most western and northern European countries,* *cohabitation has eclipsed marriage as the marker for first partnership, and in the northern countries and France there is evidence that long-term cohabitation has become more prevalent. But whether most countries are on the same trajectory to an ultimate destination where marriage and cohabitation are largely indistinguishable, or even where cohabitation overtakes marriage as the dominant form of union, awaits the future*" (Kierman, 2002, p. 75).

This type of analysis hardly differs from that which was traditionally carried out on civil registration data and censuses, based on aggregated data. It is revealing that several other studies have been based, not on survey data, but on the compilation of tables published in series of standardized national reports.[14] The specificity of *FFS* in this type of study is not so much the possibility of using data at the individual level, but rather that of providing information that other sources cannot. In Kathleen Kiernan's study, for example, marital status including informal unions and their dates of formation and dissolution can only be discovered by a survey, but its use is no different to that which is made of marital status limited to its legal categories in nuptiality analyses.

b. Replicated Models of Individual Data Analysis

It is a different matter in the second type of comparative study, where the individual characteristics of the respondents are preserved during analysis of the data. The diversity of behavior (for example in *FFS*, leaving the parental home, couple formation or fertility, etc.) is explained by the personal characteristics of the respondents through a regression-type modelization, in order to reveal the weight of these determinants on the heterogeneity of behavior. Replicated for several countries, the same operation makes it possible to show the international similarities or diversities of the explanatory capacities of different individual factors. Diversity includes national specificities which need to be explained by factors at national level.

A large number of comparative studies are based on this type of methodology. We will illustrate this with a study by Martine Corijn and Erik Klijzing (2001), who compiled ten national case studies built on the same methodology and placed the results into the context of international comparison. The study concludes with the observation of an international diversity in the action of certain individual factors and the necessity of developing an analytical framework which would make it possible to introduce the national dimension into the comparison of individual data.

Martine Corijn and Erik Klijzing studied the various stages of the transition to adulthood, particularly from school to the labour market, from the family home to the first union, and from this to fertility. They linked these stages to one another and to the characteristics of the individuals according to a well organized set of hypotheses on (1) the destandardization of pathways (which brings an end to the normative sequence of different events), (2) the role of school attendance, (3) the influence of level of education, (4) the effect of employment and unemployment, (5) the influence of religion. Each of these five points is detailed by age and sex, and each time a hypothesis deals with national specificities.

Ten countries took part, seven of which did so with FFS data. "*The countries that were selected in the end cover rather well the different parts of Europe and give a diversified, although necessarily incomplete, picture of the different pathways to adulthood in Europe. The data cover parts of northern Europe (Norway), Southern Europe (Spain, Italy), Eastern Europe (Poland), Central Europe (Austria, Germany), Western Europe (Belgium, The Netherlands, France) and noncontinental Europe (Britain). These countries present a wide array of social, economic and political systems*" (Corijn and Klijzing, 2001, p. 16).

The main conclusions on the role of individual factors go in several directions. In some cases, the influence of the factor on the stages of the transition to adulthood is the same in different countries, in other cases, it varies according to events, in yet other cases, it varies according to country.

As expected, school attendance delays the various stages of the transition to adulthood everywhere. "*An incompatibility between being enrolled in school and starting a union and/or family seems indeed to be the rule: the effect of enrollment is strongly negative in all countries (no appropriate data were available for Austria, France, Poland and Norway). The (longer) enrollment seems to discourage young adults from starting a first union: studying and living with a partner—even without a marriage and parenthood commitment—remains incompatible in terms of time, energy and money*" (Corijn and Klijzing, 2001, p. 330).

At the other extreme, the effect of employment and unemployment varies considerably, in particular among women, whose marriage is brought forward or delayed and whose fertility is favored or discouraged according to country. "*Female unemployment affects the marriage and parenthood transitions in two distinct ways. In West Germany,*[15] *Spain and Norway female unemployment leads to a postponement of marriage. In contrast, in*

[14] See, for example: Schoenmaeckers and Lodewijckx, 1999; Prioux, 2001.

[15] Martine Corijn and Erik Klijzing incorrectly named East Germany instead of West Germany (Hullen, 2001, p. 166).

Flanders and Italy female unemployment speeds up the marriage timing, although this effect changes across age. Female unemployment has no impact on the marriage timing in Poland (Fratczak and Liefbroer, 1996) and in East Germany (Hullen, this volume). In West and East Germany and in Poland female unemployment has no impact on the motherhood timing either. On the other hand, in Flanders, Italy and Spain unemployment has a strong positive impact on the motherhood timing: it doubles the rate" (Corijn and Klijzing, 2001, p. 335).

Halfway, the effect of the level of education is variable and depends on the stage of transition to adulthood. It is systematic on union formation and fertility, which are delayed by the most qualified individuals; it differs according to country with regard to departure from the parental home. "*In general, the educational level does have a negative effect on the union formation, marriage and parenthood timing. The only exception is among East German men. The more these events are disconnected from each other, the stronger the negative effect is for the parenthood timing. The impact of the educational attainment seems to be event-specific. The postulated positive effect of educational attainment on the timing of leaving the parental home was confirmed for the Netherlands, Britain, Austria, Spain, Poland, East Germany (both sexes), and for Norway and France (men only). The higher the educational level of young adults in these countries, the earlier they leave the parental home . . . In contrast, the effect of the educational level on the timing of leaving the parental home was negative at young ages in Flanders, Italy and among French women. The higher the educational level of young adults in these countries, the less inclined they are to leave the parental home before age 25*" (Corijn and Klijzing, 2001, p. 332–333).

One of the fundamental hypotheses was that the destandardization of the stages signified both the decline in the norms concerning the sequence of events from the departure from the parental home, and the increased influence of individual determinants of behavior. In a comparative perspective, it implies a reduction in the differences between countries, i.e. a convergence in trends all over Europe.

Certain facts do not confirm this implied hypothesis. There are considerable differences between countries in the action of various factors on different events, in particular (but not only) female unemployment on nuptiality and fertility. Beyond education, employment, or religion, the country remains an element of differentiation which must be taken into account. The authors and reviewers of the study have rightly emphasized this point. Explaining the differences between countries by introducing national level variables into a model based on individual data remains an unresolved problem.

"*The diversity in the results of some of the individual factors points to the importance of structural factors, including social policy factors. These factors are not investigated directly in this project, but they often surfaced as indirect explanatory factors: the institutionalization of the school-work transition, the unemployment rate, the housing market, the prevailing family model . . . Further specification of the theory must clarify the role of structural and social policy factors. Such a theory will need an appropriate data gathering design and appropriate multilevel analyzes. This could be one of the challenges of a possible second FFS round*" (Corijn and Klijzing, 2001, p. 338–339).[16]

Through this example, it appears that taking into account factors at the national level within an analytical framework focused on individual determinants should be explicit and requires a specific interpretative model. The repeated application of the same analytical model to individual data from different countries paves the way for hypotheses on the specific role of countries in the diversity of individual behavior. But in the absence of an appropriate analytical framework and a methodology which enables the variables at national level to be integrated in models based on individual data, comparative analysis remains rudimentary in the explanation of demographic phenomena.

c. Multilevel Comparative Analysis

Contrary to the first two types of analysis, which have both been illustrated by a study drawn from a large group of research based either on descriptive analysis or on replicated models, multilevel analysis has only a very limited number of examples. The idea of combining micro- and macrodata in the same model is still rare in the social sciences, with the possible exception of the educational sciences, and it is even more rare if it is a question of turning country into one of the macrolevels in order to undertake an international comparative analysis.

Alessandra De Rose and Filomena Racioppi (2001) started from the fact that individual behavior is not only explained by the relationships observed at the individual level, but also by the environment of which the individual is a part. The relationships are therefore defined in a contextual framework. The main question is therefore: how does the causal relationship between the variables studied (individual behavior, choices, attitudes, decisions, etc.) and their determinants vary according to the context? The analytical framework is unavoidably complex, as it must expose the importance of the environment (social, geographic, cultural,

[16] See also the review of the study published by Katrin Golsch (2002).

economic, political, etc.) and its effects on the attitudes, decision processes, and individual behavior.

The authors use a single statistical model which integrates the micro- and macrodeterminants of individual behavior into a multilevel regression. International comparative analysis appears to be a more inclusive method, which could include contextual dimensions other than the nation.

Alessandra De Rose and Filomena Racioppi studied the frequency of very low fertility in 13 European countries, using FFS data. They measured it from the proportion of women who are planning on having no children or one child only. The proportion is below 30% in German cohorts born in 1950–1954 and below 5% in the Norwegian and Swedish cohorts born in 1965–1969. In general the number of infertile or subfertile women declines over time.

The authors selected 13 countries on a technical criterion: they are those who have the greatest number of common comparable variables in the survey files.

National models were initially developed, taking into account selected individual determinants, among those that were revealed as being the most effective in explaining the frequency of voluntary infertility in the past. The general model is indeed replicated 13 times, according to the previously described methodology. The marital status, year of birth, employment status, education, number of brothers and sisters, religiosity, and the size of the place of residence are used as explanatory factors.

The majority of the variables, *when they exert a significant effect*, do so in the expected direction and in the same manner in the various countries of Europe. Low fertility is associated with the fact that women are unmarried, are employed, live in large cities, have no religious feelings, and are themselves only-children. The only unexpected and variable effect is that of educational level, which limits low fertility in the west and supports it in the east of the continent.

From one country to another, the main difference is not in the direction of the effect of the significant variables but in their number. This raises the question of the existence of "*a* [unique] *'European' model enabling us to identify a single typology of women who voluntarily decide to remain childless or stop at the first*" (De Rose and Racioppi, 2001, p. 23). The answer is negative: the application of a single model, which groups all women in all countries, highlights the same effects as previously, but with a mediocre adjustment quality, which implies a very weak explanatory capacity.

"*The bad performance of the 'European model' leads us to suspect that another dimension than the individual one plays an important role, probably the context. We can expect that contextual variables (i.e. characteristics of the environment) may affect individual behavior in terms of reproductive planning. Women with selected characteristics and belonging to one country have different fertility expectations from those with the same characteristics but belonging to another country, just because their respective countries differ as demographic, social, economic or cultural conditions, as well as gender system*" (De Rose and Racioppi, 2001, p. 24).

In order to characterize the countries, Alessandra De Rose and Filomena Racioppi collected a set of 15 demographic and socioeconomic indicators (four of which relate to the status of women) in each of the thirteen countries, and they defined the best combination of these statistics according to a small number of simple and relevant dimensions (through principal component analysis), thus retaining the majority of the information while simplifying it in anticipation of future stages in the research. They showed the differences, on an axis of economic and social development, between countries that are "*traditional*" and those that are "*innovative*" (countries of the east and countries of the north, respectively), and the differences between "*egalitarian*" and "*non-egalitarian*" countries (central and southern Europe, respectively), on an axis where the mechanism of the gender system is expressed. The position of the countries on these two axes represents national level information used in a multilevel analysis.

In the multilevel model, the individuals of each country are characterized by a small number of common variables (date of birth, marital status, and size of place of residence) and the same is true for the countries themselves which are characterized by their level of development and equality between men and women. The role of individual determinants is confirmed and that of the contextual factors is revealed: low fertility is more frequent in countries with a high level of traditionalism with regard to economic and social development, and in countries with a less egalitarian gender system. The first relation is barely significant, but the second one is entirely significant.

It is thanks to an astute innovation that the authors thus obtained a metrics of the national level effects on low fertility: the characterization of countries by their position on a continuous and supposed unidimensional stepladder, instead of a nonhierarchical typology like those of the regimes of welfare states, which group countries in small dispersed clusters.

On the other side of the coin, what is missing from this study is the analytical framework which would have postulated linkages between the individual and national levels and which would have guided their choice of factors and indicators at both levels: "*As to the definition of the context and its links with the* *individual dimension, there is still the problem of formulating a theoretical framework to provide a preliminary basis and to identify suitable indicators, and to clarify the various context roles*" (De Rose and Racioppi, 2001, p. 29). In the future, it would be good that the development of such a framework and its testing by the most appropriate quantitative methods associates sociologists and statisticians in an ambitious common adventure.

The scarcity of other examples in the field of demography where individual data are placed in an international comparative perspective in a tool which integrates the individual and national analytical levels underlines the exploratory and innovative nature of the work of Alessandra De Rose and Filomena Racioppi.[17] Despite the creation of very large comparative data bases that group individual data from coordinated international surveys (not only *FFS* but also the *European Community Household Panel*, the *Luxembourg Income Study*, etc.), the knowledge that enables these materials to be used is still being developed. The introduction of multilevel models in the field of comparative research is one of the most promising methods for the immediate future (Courgeau, 2004).

CONCLUSION

International comparisons have for a long time been the preferred method for making sense of statistical observations of human behavior, whether the observation is made from the point of view of the country or that of the individual. Let us take two emblematic examples.

Using censuses carried out all over Europe during the 1930s, John Hajnal notes that the proportion of people who marry during their lives and the age at which they do so are significantly different in the countries to the east and to the west of the Saint Petersburg to Trieste line: with marriage being more frequent and earlier in the east, less common and later in the west. He associates a second statement with it, that of a difference in household composition: in the east households are more complex, often combining several families, in particular families of successive generations (families of elderly parents and families of adult children); in the West, they are simpler and generally one-family. He concludes that in the east the family system envisages, facilitates, and organizes the coexistence of generations on the same piece of land for common cultivation which makes it possible to keep at least one son with his father and to encourage his marriage. In the same time in the West, the son only succeeds his father when the latter has left him his place; he only envisages marriage later, sometimes too late for a union to be concluded. Comparative analysis thus links nuptiality to the family system and to the system of agricultural cultivation, both of which differ in the east and the west of the continent: "*In [Western] Europe it has been necessary for a man to defer marriage until he could establish an independent livelihood adequate to support a family; in other societies [i.e. Eastern Europe] the young couple could be incorporated in a larger economic unit, such as a joint family*" (Hajnal, 1965).

Based on the first fertility surveys carried out in Europe during the second half of the 1960's, the team of demographers at the United Nations in Geneva showed that the completed fertility of couples is as low as their education level is high in eastern and southern European countries (Poland and Yugoslavia, for example), while the relation is more or less nonexistent and even the reverse in northern and western European countries (Denmark and France). The difference persists even after the inclusion of other individual variables likely to explain the diversity in the behavior of couples. The authors concluded that the level of economic development of countries is associated with the form of the individual relationship between educational level and fertility: "*If the countries were to be arranged by level of GNP per capita, the strength of the association between education and ultimate family size would show an inverse relation*" (United Nations, 1976).

There we have the two main types of comparative analysis, that which is based on indicators linked through the intermediary of the statistical unit of *country* for the first, while the second type associates individual behavior with their individual determinants before placing the relationships within a multinational comparative context.

The two categories of operation can only be carried out if statistical data speaking the same language are available for each country, despite the diversity of systems in which the data were developed. But this language itself can only be created by a series of compromises between national requirements which must satisfy the statistics of one country, and the similarity of concepts, definitions, tools, etc. necessary for international comparison.

The most natural path in this direction seems to be that which makes each stage of the process from the demonstration of the situation and behavior to their statistical interpretation as similar as possible: obviously not only using the same concepts and definitions but also the same questionnaires, sample designs, and collection procedures, etc. But the road is extremely

[17] See also, from *Fertility and Family Surveys* data: Dourlein and Liefbroer, 2004.

long as statistical production is a complex procedure, of which each element is essential in obtaining good quality results. And each stage is an opportunity for reflecting on similarities and differences. How is it possible to translate into two languages questions managing to give an equivalent stimulus for the respondents of two countries? What is the best way to take into account the availability of different sources in order to identify and reach the respondents? etc. Moreover, the result is uncertain, as similar tools are not always enough to guarantee equivalent results: for example, the use of a standard nomenclature of educational levels does not always make it possible to overcome the diversity of national education systems.

Also is it possible to prefer the harmonization of results to the harmonization of procedures by leaving the countries free to choose the best method to integrate the harmonized concepts into their statistical translation?

There is a certain correspondence between the type of harmonization to which the process of data production is submitted and the type of comparative analysis which would then be carried out on it. The census tables published by Eurostat for all the countries of the European Union adapt well to differences in collection methods (wording of questions, questionnaire administration), and even to the heterogeneity of sources (inclusion of results from population registers in some countries). Inversely, the most personal, and even intimate, information often used in the analysis of individual data (opinions and attitudes, religious or sexual orientation, ethnic origin, etc.) can be extremely sensitive to the method through which they are obtained and their comparability can only result from a process which is standardized from beginning to end. But there are intermediate situations and varied combinations between procedures that create comparability and the methods of comparative analysis.

The discussion of methods for international comparative analysis leads to a more general consideration on the evolution of statistical methods of analysis in the social sciences. In the initial phase, the community was in the forefront, research was built on aggregated data at this level and it found its method for explaining social behavior in comparisons of national communities. Today, the individual has come to the forefront and basic statistical data are at this level; analysis must therefore make an extra effort to also introduce a national level in order to supplement the study of individual determinants, as they are not enough to explain the diversity of behavior and situations.

Behind these developments are factors which are due to the transformation of societies, particularly Western societies, but it is not appropriate here to develop considerations on the trend of individualization. But there is also the effect of technical conditions such as the data processing and statistical resources now available to researchers to deal with large files of individual data. However, the vertiginous development of these means and the speed with which these files are created and then analyzed run the risk of making the analyst forget the length and the rigor of the process which leads to the production of data and which determines their quality. By placing problems of international comparability of data at the center of its reflection, comparative analysis has the great merit of refocusing attention on all the linkages of this chain, and on the primacy of the role of concepts, definitions, and tools, etc. in the significance which can be given to the result of analysis, regardless of its level of sophistication.

Acknowledgment

This chapter was translated from French by Accenta Ltd.

References

Badgett M. V. Lee, 2003. *Variations on an equitable theme: explaining international same-sex partner recognition laws.* Amherst, University of Massachusetts, 18 p. (Paper presented at the conference on "*Same-sex couples, same-sex partnerships and homosexual marriages*", Stockholm, September 2003).

Bauman Kurt, 1997. *Shifting family definitions: The effect of cohabitation and other nonfamily household relationships on measures of poverty.* Madison, University of Wisconsin Institute for Research on Poverty, 26 p. (Discussion Paper n. 1123).

Bien Walter and Quellenberg Holger, 2003. How to measure household and family, *in:* Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*, p. 279–293. New York, Kluwer Academic/Plenum Publishers, IX + 413 p.

Biraben Jean-Noël and Henry Louis, 1957. La mortalité des jeunes enfants dans les pays méditerranéens, *Population*, vol. 11(4), p. 615–645.

Blayo Chantal and Festy Patrick, 1975. La fécondité à l'Est et à l'Ouest de l'Europe, *Population*, vol. 30(4–5), p. 875–876.

Bourgeois-Pichat Jean, 1951. La mesure de la mortalité infantile: II. Les causes de décès, *Population*, vol. 6(3), p. 459–480.

Bourgeois-Pichat Jean, 1985. L'évolution récente de la mortalité dans les pays industrialisés, *in:* Jacques Vallin, Alan Lopez and Hugo Behm (eds.), *La lutte contre la mort*, p. 489–522. Paris, INED, PUF, 542 p. (Travaux et Documents, Cahier no. 108).

Bourgeois-Pichat Jean, 1985. Recent changes in mortality in developed countries, *in*: Jacques Vallin, Alan Lopez and Hugo Behm (eds.), *Health policy, social policy and mortality prospects*, p. 507–540. Liège, Ordina Édition, 558 p. (Proceedings of an IUSSP Seminar in Paris, February 28–March 4, 1983).

Bourgeois-Pichat Jean, 1994. *La dynamique des populations. Populations stables, semi-stables, quasi-stables.* Paris, INED and PUF, 296 p. (Travaux et documents, Cahier no. 133).

Brislin Richard W., Lonner Walter J. and Thorndike Robert M., 1973. *Cross-cultural research methods.* New York, John Wiley, 361 p.

CépiDc and INSERM, 2001. *Comparability and quality improvement of European causes of death statistics. Final report.*, European Commission, DGV/F3 and INSERM, SC8, CépiDc, 190 p. (DGV/F3 SOC 98 20108).

Coale Ansley J. and Demeny Paul, 1966. *Regional model life tables and stable populations.* Princeton (New Jersey), Princeton University Press, 871 p.

Coale Ansley J. and Trussell T. James, 1974. Model Fertility Schedule: variations in the age structure of childbearing in human populations, *Population Index*, vol. 40(2), p. 185–258.

Corijn Martine and Klijzing Erik (eds.), 2001. *Transitions to adulthood in Europe.* Dordrecht, Boston and London, Kluwer, Kluwer and Academic Publisher, 340 p.

Courgeau Daniel, 1994. Du groupe à l'individu: l'exemple des comportements migratoires, *Population*, vol. 49(1), p. 7–26. (Also published in English: From the group to the individual: what can be learned from migratory behaviour), *Population. An English Selection*, vol. 7, 1995(1), p. 145–162.

Courgeau Daniel, 2002. New approaches and methodological innovations, *in*: Miroslav Macura and Gijs Beets (eds.), Dynamics of fertility and partnership in Europe. Insights and lessons from comparative research, Volume I, p. 99–114. New York and Geneva, United Nations, 159 p.

Courgeau Daniel, 2004. *Du groupe à l'individu: Synthèse multiniveau.* Paris, INED, 250 p. (Manuals).

De Bruin Agnes, Picavet H. Susan and Nosikov Anatoliy, 1996. *Health interview surveys: Towards international harmonization of methods and instruments.* Copenhagen, Voorburg, World Health Organisation, Regional Office for Europe and Statistics Netherlands, 162 p. (WHO Regional Publications. European Series no. 58).

De Rose Alessandra and Racioppi Filomena, 2001. Explaining voluntary low fertility in Europe: a multilevel approach, *Genus*, vol. 57(1), p. 13–32.

Dourlein Edith and Liefbroer Aart C., 2004. *Unmarried cohabitation and union stability: testing the role of diffusion using data from 16 European countries.* Paris, INED, 23 p. (Paper presented at the seminar on "Comparaisons européennes", Paris, INED, 5 January 2004; http://www-les-lundis.ined.fr/textes/cohabdivorce.pdf).

Duchêne Josianne and Wunsch Guillaume (eds.), 1995. *Collecte et comparabilité des données démographiques et sociales en Europe.* Louvain-la-Neuve, Academia and L'Harmattan, 671 p. (Chaire Quetelet 1991).

Durkheim Emile, 1937. *Les règles de la méthode sociologique.* Paris, PUF, 128 p.

Egidi Viviana and Buratta Vittoria, 1997. Comparability of definitions and data on morbidity, *in*: Guillaume Wunsch and Attila Hancioglu (eds.), *Morbidity and mortality data: problems of comparability. Proceedings of the European Association for Population Studies and the Hacettepe Institute of Population Studies workshop, Ürgüp (Turkey), 18–20 October 1995*, p. 11–29. Ankara (Turkey), Hacettepe Institute of Population Studies, 236 p.

Ehling Manfred, 2003. Harmonising data in official statistics, *in*: Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*, p. 17–32. New York, Kluwer Academic/Plenum Publishers, 413 p.

Eurostat, 1996a. *The European Community Household Panel (ECHP): Methods, Survey questionnaire: Waves 1–3.* Luxembourg, Office for Official Publications of the European Communities.

Eurostat, 1996b. *The European Community Household Panel (ECHP): Survey methodology and implementation. Vol. 1.* Luxembourg, Office for Official Publications of the European Communities.

Eurostat, 1999a. *Community census programme guidelines and table programme*. Luxembourg, Eurostat.

Eurostat, 1999b. *The future of European social statistics, Harmonisation of social statistics and quality. The fourth Mondorf Seminar, 26–27 March 1998.* Luxembourg, Eurostat.

Eurostat, 2003a. *Household budget surveys in the EU. Methodology and recommendations for harmonisation.* Luxembourg, Eurostat, 26 p.

Eurostat, 2003b. *The European Union Labour Force Survey. Methods and definitions 2001 Eurostat, Methods and Nomenclatures.* Luxembourg, Eurostat, 90 p.

Ferrera Maurizio, 1996. The 'Southern' model of welfare in social Europe, *Journal of European Social Policy*, vol. 6(1), p. 17–37.

Gabriel K. R. and Ronen Ilana, 1958. Estimates of mortality from infant mortality rates, *Population Studies*, vol. 12(2), p. 164–169.

Gauthier Anne-Hélène, 2002. The promises of comparative research, *Journal of Applied Social Sciences Studies*, vol. 122(1), p. 5–30.

Gjonça Arjan, Brockmann Hilke and Maier Heiner, 2000. Old-age mortality in Germany prior to and after reunification, *Demographic Research*, vol. 3, art. 1, 26 p. (http://www.demographic-research.org/volumes/vol3/1).

Golsch Katrin, 2002. Martine Corijn and Erik Klijzing (eds.), Transitions to adulthood in Europe, Dordrecht, Kluwer Academic Publishers, 2001, 340 p., *European Journal of Demography*, vol. 18(1), p. 85–87. (Book reviews).

Grais Bernard, 1998. *Harmonisation in social statistics and quality.* Luxembourg, Eurostat, 72p. (Report prepared for the Fourth Mondorf Meeting, 26–27 March 1998).

Guyatt Gordon H., 1993. The philosophy of health related quality of life translation, *Quality of life research* no. 2, p. 461–465.

Hajnal John, 1965. European marriage pattern in perspective, *in*: David V. Glass and David E. C. Eversley (eds.), *Population in history. Essays in historical demography*, p. 101–143. Chicago & London, Aldine Publishing Company, 682 p.

Harkness Janet, Pennell Beth-Ellen, Schoua-Glusberg Alisaú, forthcoming. Questionnaire translation and assessment, *in*: Stanley Presser, Jennifer Rothgeb, Michael Couper, Judith Lessler, Elizabeth Martin and Eleanor Singer (eds.), *Methods for testing and evaluating survey questionnaires*, p. 453–473. New Jersey, John Wiley & Sons Inc., 624 p. (Wiley Series in Survey Methodology).

Harkness Janet, 2003. Questionnaire translation, *in*: Janet A. Harkness, Fons J. R. van de Vijver and Peter Ph. Mohler (eds.), *Cross-cultural survey methods*, p. 19–34. New Jersey, John Wiley & Sons Inc, x + 420 p. (Wiley Series in Survey Methodology).

Henry Louis, 1961. La fécondité naturelle. Observations, théorie, résultats, *Population*, vol. 16(4), p. 625–636.

Hobcraft John, 1980. Strategies for comparative analysis of WFS data, *in*: ISI/IIS (ed.), *World Fertility Survey conference 1980: a record of proceedings.* Vol. 3, p. 509–574. Voorburg (Pays-Bas), ISI/IIS, 574 p.

Hobcraft John, 2002. Moving beyond elaborate description: towards understanding choices about parenthood, *in*: Miroslav Macura and Gijs Beets (eds.), *Dynamics of fertility and partnership in Europe. Insights and lessons from comparative research, Vol. I*, p. 131–144. New York and Geneva, United Nations, 159 p.

Hoffmeyer-Zlotnik Jürgen H. P., 2003. How to measure race and ethnicity, *in*: Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*, p. 267–277. New York, Kluwer Academic/Plenum Publishers, 413 p.

Hullen Gert, 2001. Transitions to adulthood in Germany, *in*: Martine Corijn and Erik Klijzing (eds.), *Transitions to adulthood in Europe*, p. 153–172. Dordrecht, Boston and London, Kluwer, Kluwer and Academic Publisher, 340 p.

ISTAT, 2004. *Death certificate and certification practices: an international comparison.* Rome, ISTAT, 44 p.

KIERNAN Kathleen, 2002. The state of European unions: an analysis of partnership formation and dissolution, *in*: Miroslav Macura and Gijs Beets (eds.), Dynamics of fertility and partnership in Europe. Insights and lessons from comparative research, Vol. I, p. 57–76. New York and Geneva, United Nations, 159 p.

KRIEGER Nancy, 2003. Genders, sexes, and health: what are the connections and why does it matter?, *International Journal of Epidemiology*, vol. 32(3), p. 652–657.

LEDERMANN Sully, 1969. *Nouvelles tables-types de mortalité.* Paris, INED and Presses universitaires de France, 260 p. (Travaux et documents, Cahier no. 53).

MEJER Lene, 2003. Harmonisation of socio-economic variables in EU statistics, *in*: Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*, p. 67–95. New York, Kluwer Academic/Plenum Publishers, 413 p.

MESLÉ France, 1995. L'enregistrement des causes de décès en Europe, *in*: Josianne Duchêne and Guillaume Wunsch (eds.), *Collecte et comparabilité des données démographiques et sociales en Europe*, p. 401–431. Louvain-la-Neuve, Academia/L'Harmattan, 671 p. (Actes de la Chaire Quetelet 1991).

MESLÉ France and VALLIN Jacques, 2002. Mortalité en Europe: la divergence Est-Ouest, *Population-F*, vol. 57(1), p. 171–212. (Also in English: Mortality in Europe: the divergence between East and West, *Population-E*, vol. 57(1), p. 157–198).

MONNIER Alain, 2000. La convergence: réalité et limites d'une notion, *in*: AIDELF (ed.), *Régimes démographiques et territoires: les frontières en question*, p. 629–636. Paris, Association internationale des démographes de langue française, 636 p. (Colloque international de La Rochelle, 22–26 September 1998, no. 9).

MONNIER Alain, 2004. L'union européenne à l'heure de l'élargissement, *Population*, vol. 59(2), p. 379–380.

MONNIER Alain and RYCHTARIKOVA Jitka, 1991. Comment l'Europe s'est divisée entre l'Est et l'Ouest, *Population*, vol. 46(6), p. 1617–1650.

NIZARD Alfred and VALLIN Jacques, 1980. Les causes de décès en France. II. Couplage des 7e et 8e révisions de la Classification internationale. Application à la période 1958–1974, *Population*, vol. 35(6), p. 1057–1100.

NOSIKOV Anatoliy and GUDEX Claire (eds.), 2003a. *EUROHIS: developing common instruments for health surveys.* Amsterdam, (Pays-Bas), IOS Press, 208 p. (Biomedical and Health Research, no. 57).

NOSIKOV Anatoliy and GUDEX Claire, 2003b. Overview of the EUROHIS Project, *in*: Anatoliy Nosikov and Claire Gudex (eds.), *EUROHIS: developing common instruments for health surveys*, p. 13–20. Amsterdam, (Netherlands), IOS Press, 208 p. (Biomedical and Health Research, no. 57).

OMS/WHO, 1995. *Classification statistique internationale des maladies et des problèmes de santé connexes. Dixième révision*, *Vol. 2.* Geneva, World Health Organization, 167 p.

PACE Monica, BRUZZONE Silvia, FROVA Luisa and PAPPAGALLO Marilena, 2004. *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe.* Rome, ISTAT, 104 p. (Contribution 14).

PACE Monica, BRUZZONE Silvia, MARCHETTI Stefano, GRIPPO Francesco, CINQUE Simona, DI FRAIA Gennaro, PAPPAGALLO Marilena, PENNAZZA Simona, SINDONI Giuseppe and FROVA Luisa, 2004. Preparation of an EU training package on certification of causes of death. *in*: Federal Statistical Office Germany (ed.), *Proceedings. European conference on quality and methodology in official statistics (Q2004), Mainz, Germany, Castle of the Prince Elector, 24–26 May 2004.* Wiesbaden, Statistisches Bundesamt, CD-ROM.

PERACCHI Franco, 2002. The European Community household panel: a review, *Empirical Economics*, vol. 27(1), p. 63–90.

PRIOUX France, 2001. *Evolution récente et comparaison des modes de formation des familles en Europe: que nous révèlent les enquêtes FFS?* Paris, INED, 17 p. (Paper presented at the 24th General Population Conference, Salvador de Bahia, Brazil, August 18–24, 2001).

ROBINE Jean-Marie, ROMIEU Isabelle, JAGGER Carol and EGIDI Viviana, 1998. *Les espérances de santé dans l'Union Européenne. Analyse des données du Panel des Ménages de la Communauté Européenne.* Montpellier, EuroREVES, 62 p. (REVES paper no. 320).

SCHOENMAECKERS Ronald C. and LODEWIJCKX Edith, 1999. Demographic behaviour in Europe: some results from FFS country reports and suggestions for further research, *European Population Journal/Revue européenne de démographie*, vol. 15(3), p. 207–240.

STATISTISK SENTRALBYRÀ6, 2004. *Use of Administrative Records in GGS.* Oslo, Statistisk sentralbyrà6 (Statistics Norway), 10 p. (http://www.unece.org/ead/pau/ggp/tmiwg/ard_report_spetses5.pdf).

SÜSSMILCH Johann Peter, 1741. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortplflanzung desselben erwiesen.* Berlin, Kultur-Verlag, 360 p. (Reprint of the original edition by J.C. Spener, Berlin, 1741).

UNITED NATIONS, 1955. *Age and sex patterns of mortality. Model life-tables for underdeveloped countries.* New York, United Nations, no. 22, 38 p. (Population Studies, no. 22).

UNITED NATIONS, 1976. *Fertility and family planning in Europe around 1970: a comparative study of twelve national surveys.* New York, United Nations, 180 p.

VALAORAS Vasilios, 1973. Model life tables as a measure of mortality, *in*: UIESP/IUSSP (ed.), *International Population Conference, Liège 1973. Proceedings/Congrès international de la population, Liège 1973. Actes. Volume 3*, p. 133–142. Liège, IUSSP/UIESP, 470 p.

VALLIN Jacques, 1968. La mortalité dans les pays du tiers monde: évolution et perspectives, *Population*, vol. 23(5), p. 843–868.

VALLIN Jacques and MESLÉ France, 1988. *Les causes de décès en France de 1925 à 1978.* Paris, INED, PUF, 608 p. (Travaux et Documents, Cahier no. 115. Publication accompanied by 7 appending volumes).

VALLIN Jacques and MESLÉ France, 1998. Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925 à 1993, *in*: Gérard Pavillon (ed.), *Enjeux des classifications internationales en santé*, p. 113–156. Paris, Éditions INSERM, 220 p. (Questions en santé publique).

VALLIN Jacques and NIZARD Alfred, 1978. Les causes de décès en France. I. Pour une typologie simple et homogène. Application à la période 1968–1974, *Population*, vol. 33(3), p. 547–608.

WHOQOL Group, 1994. The development of the World Health Organization Quality of Life Assessment Instrument (the WHOQOL), *in*: John Orley and Willem Kuyken (eds.), *Quality of Life Assessment: international perspectives*, p. 41–61. Berlin and Heidelberg, Springer, 200 p.

WHOQOL Group, 1998. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties, *Social Science and Medicine*, vol. 46(12), p. 1569–1585.

WOLF Christof and HOFFMEYER-ZLOTNIK Jürgen H.P., 2003. How to measure sex/gender and age, *in*: Jürgen H. P. Hoffmeyer-Zlotnik and Christof Wolf, *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*, p. 259–265. New York, Kluwer Academic/Plenum Publishers, 413 p.

C H A P T E R

126

Methods of Historical Demography

LORENZO DEL PANTA, ROSELLA RETTAROLI, AND PAUL-ANDRÉ ROSENTAL*

Dipartimento di Scienze Statistiche, Università de Bologna, Bologna, Italy

**École des havtes études en sciences sociales and Institut national d'études démographiques (INED), Paris, France*

I. DEMOGRAPHY AND HISTORICAL DEMOGRAPHY

Before attempting to examine the distinctive methods of historical demography and without imposing overly strict limits, let us try to define the discipline or, at least, its specific features within the field of demographic research. As common sense indicates, historical demography is in charge of studying past populations. However, it is almost equally obvious that the boundary between past and present is purely artificial or, at the very least, conventional. Admittedly, some key events in general history—for example, Italy's political and administrative unification in 1860—naturally influence the organization of statistical offices and the formation of demographic information sources; such events serve to draw the frontier between the domain of demographers scrutinizing the past and that of analysts examining the contemporary age. Likewise, the contrast between the "prestatistical" and "statistical" ages—in fact, between the periods before and after the introduction of national population censuses—arguably triggered the birth of modern historical demography after World War II. But it remains no less true that (1) demography, even of the contemporary variety, may be construed as a discipline that studies the changes in populations over time, as well as the determinants and effects of these changes, and (2) the study must be conducted over the *long duration*, on which it inherently *depends*.

Indeed, the time frame of demographic investigations is necessarily a long one, and the measurement of demographic processes often forces researchers to extend their gaze over several generations. Consequently, the statement that historical demography concerns *the study of past populations*, while acceptable in didactic terms, is not of much help in understanding the discipline's specific traits or the need for customized methods. It is surely more accurate to characterize historical demography by (1) emphasizing its links with (rather than distance from) many other social sciences (from economic and social history to human geography, from sociology to anthropology and population genetics) and (2) recalling the big questions that it tries to answer.

II. DATA ASSESSMENT AND CHECKING: THE CORNERSTONE OF RESEARCH IN HISTORICAL DEMOGRAPHY

This brief chapter is obviously not the place to discuss all the issues relating to the assessment of data quality and completeness. Nor can we provide a comprehensive account of the countless methods that historical demography uses to attempt to correct or integrate the imprecise or fragmentary data on which it must rely. More than in many other sectors of demography, researchers studying past populations need to exploit *basic data* that are often very rough. They cannot lose sight of the fact that their *results*, whatever the methods employed, are ultimately constrained by these limitations imposed at the outset.

In other words, one of the primary characteristics of historical demography is its forced reliance on documentary sources that are nearly always incomplete and imprecise. This has fostered among historical demographers a sensitivity and special attention to issues of *data assessment*; it has often stimulated the development of new error-assessment and error-correction methods, in regard to *coverage* as much as *precision*. Current practices now include the re-use of old statistical counts, which historical demography typically regarded with suspicion. Such approaches require a combination of historical demography and the history of social science, i.e., an understanding of the cognitive and social context in which the old data were produced, in order to exploit them for contemporary purposes.[1]

To focus on what long constituted the discipline's sole pursuit—namely, the production of fresh data—and to better understand the major burden that this research phase constantly imposes, suffice it to recall the two large-scale reconstitutions undertaken in recent decades by the *Cambridge Group for the History of Population and Social Structure* to describe changes in the English population. The first, in chronological order of completion, was obtained by means of a *back-projection*[2] based on a time series built up from a thorough statistical analysis of parish registers (Wrigley and Schofield, 1981). The second is based on the demographic analysis of individual biographies reconstructed thanks to the nominal processing of those sources[3] (Wrigley *et al*., 1997).

The first of the two operations—the trial reconstitution of aggregate data on the long-term change in the English population—uses birth and death series taken from the parish registers of a sample of 404 parishes, representative of England as a whole. Wrigley and Schofield (1981) devote five of their eleven chapters to the presentation of basic parish data, their critical assessment, and the methods used to identify biases and gaps so as to determine correct parish values—and, in the final stage, the estimation of national series. The basic technique adopted for defining potential periods of underrecording of events essentially relies on the analysis of monthly series[4] and was initially used on individual data (Wrigley *et al*., 1997). In this case, therefore, the monthly series analyzed are not the ones obtained directly from parish registers, but the series of events (arranged by month and year) concerning the families reconstituted from the family records. Previous efforts were directed at evaluating the completeness of the basic data. Now, the emphasis was on assessing the relevance of the reconstitution of individual and family life histories—hence the reliability of the conclusions subsequently drawn from this type of reconstitution. For this evaluation, the authors performed a series of *a posteriori* checks and *tests* that may usefully be recalled here, very briefly, as an example. This is also a way of acknowledging that the reconstitution of the characteristics of the English demographic system—based on the analysis of individual data from a sample of 26 parishes—is surely one of the most important historical demography investigations ever completed in recent years (its final results are now available[5]).

As source material, the Anglican parish registers pose special problems. In particular, a significant share of the population belonged to denominations other than the Anglican church, and that share grew over time. However, the reconstitution of individual biographies within the family histories somewhat attenuates this type of coverage error, given that non-Anglican families, as a rule, were not included in the reconstructed families.[6] Other types of checks were performed, particularly for the 19th century, a period for which the results[7] can be matched against infor-

[1] See Bourdelais (1993) on aging, Brian (2001) on birth rates under the Ancien Régime, Brian and Jaisson (2005) on the secondary sex ratio, and Szreter (1996) on the demographic transition.

[2] See section 4 for an overview of these special modeling techniques—developed by Jim Oeppen for the research project in question—as well as for some related methods.

[3] Using the method developed by Louis Henry, discussed in Section 3.

[4] Following a critical assessment of the reliability and completeness of the monthly data, based chiefly (but not only) on the results provided by the application of this technique, about a hundred of the 530 initial parishes for which the series had been recorded were eliminated from the final sample (Wrigley and Schofield, 1981, p. 16). This shows the value of modern techniques of statistical analysis of time series even at a preliminary stage of identification of underrecording phenomena, which a more cursory examination would not have revealed. The methods are also useful for filling gaps in incomplete series with estimates. This is especially true when the gaps—as is the case for the English parish series—concern only some of many series of demographic events recorded in the same sample to supply estimates of the components of population dynamics at the overall level (see, for example, Tonini, 1993a).

[5] Begun some thirty years ago in the wake of Louis Henry's studies, the reconstitution of family and individual histories was long passed over in favor of the parallel enterprise of aggregate reconstitution, completed in the early 1980s (Wrigley *et al*., 1997, p. 3ff.).

[6] The main concern will accordingly be to check whether the reconstituted families (presumably Anglican) can be regarded as representative of the total population (Wrigley *et al*., 1997, p. 89).

[7] This situation—with timing differences—is typically found in many European countries.

mation from independent sources. The most interesting feature is, no doubt, the checking of the *plausibility* and *internal consistency* of the results of family history reconstitution. In this field, many options are available. The most commonly adopted (even by British authors) is the identification and correction of recording errors, initiated by Louis Henry (1980). Significant examples include: (1) the procedure for estimating the underrecording of infant deaths based on the observation of baptismal names,[8] (2) the verification of the completeness of birth recordings based on the observation of birth intervals in the event of early child deaths (Wrigley *et al*., 1997, p. 108),[9] and (3) the measurement of the underrecording of baptisms (and therefore of births) due to the length of birth intervals and the underadministration of the sacraments.

We have dwelled on these examples drawn from the *Cambridge Group* surveys primarily for one purpose: to emphasize that these issues are one of the specific traits of historical demography research, as they feature very prominently at every research stage (however, we make no pretense of conducting a thorough assessment of problems and procedures concerning error evaluation and correction). Another emblematic example of the centrality of these issues is the reconstitution of the French population and the components of its dynamics in the 19th century published more recently by Noël Bonneuil (1997). Data quality and completeness are the dominant issues in Bonneuil's work (incidentally, the data are considerably richer and more varied than those used by Anthony Wrigley and Roger Schofield for their aggregate reconstitution of the English population). Bonneuil did not simply conduct a rigorous preliminary assessment of the quality of the basic data. He also proposed protocols for adjusting the biases observed in the population age structure in the censuses and for correcting the underrecording of statistics on population movement. These proposals come at the end of a research sequence involving the application of an original method of model-based reconstitution. Using data from Biraben's Institut national d'études démographiques (INED) survey, Bonneuil is currently undertaking an overall reconstitution of the population for a period that produced only fragmentary data: the sixteenth and seventeenth centuries. To this end, he is relying on several types of econometric comparison between partial time series (Bonneuil, 1998).

[8] We refer here to the custom of reassigning the name of a child who dies in infancy to the following child.

[9] After the end of breastfeeding, the death of a child actually stimulates the return of the menstrual cycle and makes a long interval far less likely.

Admittedly, it is often the preliminary assessment of data quality that guides the researchers' choice of the most suitable methodological instruments (sometimes even beyond what might have been their natural preference given the issues they wanted to address). We should therefore briefly draw attention to these initial choices before moving on to a more detailed discussion of the usefulness and limitations of some of the most common techniques in historical demography.

III. MICRO AND MACRO IN HISTORICAL DEMOGRAPHY: ANTITHETICAL OR COMPLEMENTARY APPROACHES?

The birth of historical demography as an autonomous discipline, even in methodological terms, coincided with Louis Henry's development of a technique known as *family reconstitution*, which relies on the statistical use of nominal sources—the most important being, of course, parish registers (Dupâquier, 1984; Rosental, 2003). Louis Henry's protocol has been used for many years without significant amendments to its original coding. As demographers know, it enables them to make measurements despite the lack of census data and the imprecision of some information items in registers of baptisms (or births), marriages, and burials (or deaths). The method involves *record linkage*, i.e., the collation of information from different sources on a single individual or family nucleus. This allows an accurate retrieval of some characteristics of an event that do not appear in the record for the event—in the case of a birth, for example, linkage may show the age of the mother at marriage and at the birth of her previous children. By grouping individual data and successive events on a *family reconstitution form*, demographers can reconstruct family histories and so measure the length of exposure to the risk of experiencing vital events. This approach ultimately yields precise, detailed measurements of fertility and mortality.

In recent decades, the *microdemographic* approach, linked to the collection of nominal data, has undergone major change. There has been very rapid progress in data estimation and the automation of archiving procedures. As a result, record linkage can now be applied to much larger databases than before. Above all, it has stimulated efforts to extend the scope of application of nominal surveys even to sources that are not strictly demographic. The accumulation of ever-expanding databases[10] has, in turn, driven historical demogra-

[10] See the thematic dossier entitled "Banques de données" (Databases) in *Annales de démographie historique*, no. 2, 1998.

phers to make efforts to implement an array of often fairly sophisticated statistical techniques.[11] These are capable of managing complex data systems and achieving substantial progress, at least in the description and measurement of demographic phenomena. In fact, this is a process that has affected—albeit in different ways—both historical demographers and demographers studying contemporary populations.[12] The main difference is, perhaps, the following. Among the second group, the collection and analysis of individual biographies—a means of going beyond the mere use of aggregate data—are undoubtedly much more recent phenomena. By contrast, in historical demography (as we shall see below), the evidence does not suggest an outright abandonment of the *microdemographic* approach. Instead, there has clearly been a strong impulse toward the parallel development of new analytical instruments—most notably, based on models—using aggregate data.

This design of new modeling instruments applicable to past populations relied only in part on the set of theories and instruments already available to demographers (in particular, Lotka's theory and its later versions) for analyzing incomplete or defective data. In fact, the new instruments took shape in a period, the early 1970s, when the enthusiasm for the opportunities afforded by nominal surveys may have begun to fade. It was not yet possible to foresee—at least realistically—the new momentum that the growth of information technology would impart to this field of historical demographic research. Without minimizing the major contribution of other researchers to the rise of model-based aggregate reconstitutions of past populations,[13] we can state that the most fertile methodological path was opened by Ronald Lee (1974), with what has become known as *inverse projection*. This procedure allows the reconstruction, from a series of hypotheses, of the main parameters of a population and their changes over time. Apart from the initial and final population size, the only data used are the annual time series of births and deaths. In one of the sections below, we use selected details to illustrate the later developments of this type of methodological approach. For the moment, it is interesting to note that the need for a fuller description of the mechanisms of a demographic system was not totally satisfied by the application of the nominal family reconstitution method. Setting aside the previously discussed issue of whether the results can encompass unreconstituted families as well as reconstituted ones, the method does not allow the construction of a *net maternity function*. There are two reasons for this. First, its analysis of fertility is confined to fertility within marriage; second, the method faces many obstacles in measuring adult mortality.

The techniques of model-based reconstitution range from Ronald Lee's *inverse projection* and its later variants and elaborations to Jim Oeppen's *back projection*, used in Anthony Wrigley and Roger Schofield's first reconstitution of the English population (1981), up to *generalized inverse projection*, in some ways a more flexible tool, developed later by Jim Oeppen (1993a and 1993b) himself (according to the information available, the third method might rely on one of the first two formulations). To ensure that the expected results are sufficiently consistent with the reality under study, all these techniques require the fulfillment of a large set of assumptions concerning the share of information about the demographic system that can be directly deduced from the data. In particular, users need to postulate the form of the age-specific fertility and mortality functions. Such information—often gathered from small samples—can hardly be obtained except by analyzing individual data.

There are several ways of viewing the *complementarity* between the different methods of historical demography—from those using individual data to those allowing aggregate reconstitutions by means of suitable models. Modeling techniques reveal the underlying trends and thus identify the distinctive phases of change in demographic regimes, the turning points, and possible splits along geographic lines; we can then investigate these in greater detail by analyzing individual data. In exchange, the latter may sometimes provide valuable indications about the choice of new *models* best suited to identifying the structural characteristics of the demographic system that might not have been initially spotted through reconstitutions based on aggregate data alone.[14]

Using partly different methods and benchmarks, the large-scale surveys by INED on the French popu-

[11] The difficulties were greater, no doubt, for the older members of the profession and for those who lacked an initial solid training in statistics.

[12] See, for example, on mortality studies, Caselli, 1995, p. 1532.

[13] We are thinking in particular of the reconstitution methods of Jean Bourgeois-Pichat (1951) and Étienne van de Walle (1974).

[14] In this respect, the most convincing example may be drawn from the recent publication of the results of the sample survey conducted in England on individual data (Wrigley *et al*., 1997, p. 515ff.). The results of nominal reconstitutions have actually highlighted major changes, over the long periods examined, both in the fertility tempo (variations in the mean age of childbearing) and in the age distribution of mortality. This has made it possible to revise and adapt the previous assumptions made in the application of *back projection* to the English population.

lation (of which the complete results are now available)[15] and by the *Cambridge Group* on the English population abundantly illustrate the value of a comprehensive, detailed vision of the demographic history of a country (France or England) that can be obtained by combining the results of vast aggregate reconstitutions and those of surveys that cover smaller samples but are based on the analysis of individual biographies.

In the following paragraphs, we shall therefore attempt a brief discussion—with no claim to exhaustiveness—about specific aspects of the techniques for analyzing individual data and aggregate reconstitution models. Our emphasis will be on some of the more interesting recent developments in both approaches. Concerning the analysis of individual data, we shall give a concise description of the introduction into historical demography of *event history analysis* (the analysis of biographies or life courses), which has been applied only recently to individual parish data.

For brevity's sake, however, we shall not address the many techniques of aggregate data analysis routinely employed by historical demographers. As a rule, they are simply adaptations of well-known analytical instruments—commonly used in other branches of demography[16]—to the sources specific to historical demography. They are abundantly discussed elsewhere in this manual.[17]

IV. ISSUES AND NEW DIRECTIONS IN THE ANALYSIS OF INDIVIDUAL DATA

There are several different sources in historical demography that can directly supply individual data. Depending on the context, *record linkage* itself can be applied in different ways, which cannot be fully described here. In our rapid summary, we shall therefore confine ourselves to emphasizing some issues in the application of the particular technique of nominal data analysis known as *nominal reconstitution of families*. As formulated by Louis Henry (Fleury and Henry, 1976; Henry and Blum, 1988),[18] it relies on the extraction of data from marriage, birth, and death registers, in the absence of status data. However, we should note that, in many countries (such as Italy, Sweden, and Belgium), the relative abundance of *status animarum* (literally, "state[s] of souls" or population registers) ought to entail a change in the reconstitution technique—as has actually happened in some cases: individual and family life courses have now been enhanced with status information deduced, at different moments, from nominal lists of inhabitants.

As the method proposed by Louis Henry is now familiar to all historical demographers, we shall not describe the entire procedure for reconstituting families and then obtaining fertility, marriage, and mortality indicators from family and individual histories. Instead, we shall dwell at greater length on some aspects that we find particularly interesting:

- the impossibility of reconstructing in identical detail the different components of the demographic system
- the problem of the representativeness of results and the possible alternatives for attempting to optimize the cost/benefit ratio of the reconstitutions
- the biases that can diminish the credibility of measures obtained from nominal reconstitutions, mainly due to the difficulty of continuously observing individual life courses

1. The Limits of Family Reconstitution

As regards the first point, the aspect worth noting is that the technique turns out to be far better suited to the analysis of fertility (for which it was actually designed) than to the study of mortality and nuptiality. Absent any status data, it is in fact almost impossible to extract reliable information on the never-married rate from family reconstitutions; likewise, it is very difficult to use nominal reconstitution to determine a *net reproduction* function—and thus a rough estimate of the intrinsic growth rate of the population studied. These shortcomings seem hard to overcome within the framework of Louis Henry's method. Often, however, we can add information (even of a sporadic nature) on population size and structure—generally, if the study is at the parish level, a few *status animarum* will suffice. We can then deduce, even with standard

[15] The results of the INED surveys, unlike those of the Cambridge Group on the English population, have never been collected in a single volume. For the nominal survey, see esp. Henry, 1972a and 1972b; Henry and Houdaille, 1973; Houdaille, 1976; Henry 1978. The results of the survey on aggregate data were published in 1975 in a special *Historical demography* issue of the journal *Population*. For a description of the data, see Séguy, 2001.

[16] For a fairly complete overview of these methods, see Del Panta and Rettaroli, 1994.

[17] Particularly in vol. 1 and the methodological chapters of vols. 2, 3, and 4.

[18] The first of these two publications (Fleury and Henry, 1976) is a small manual (1st ed. 1956) that provides a very precise and detailed description of the family-reconstitution protocol. The second (Henry and Blum, 1988) gives the methods for analyzing data from the reconstituted families.

aggregation techniques, a series of indicators that may help to fill out the demographic picture supplied by the nominal reconstitution.

2. Representativeness of Results

Perhaps the most acute problem concerns the representativeness of the results of nominal reconstitutions from parish data. First, it is well known that individual and family mobility—along with a broad set of causes intrinsic to the nature of the data and to the linkage procedures—is often responsible for a low yield on reconstitutions. Second, as a rule, there is no guarantee that the families whose complete life courses can be determined are representative of the socioeconomic conditions and demographic behaviors of the total population to be examined. In addition to this issue of representativeness in the parish or community, there is the problem of whether the communities or parishes analyzed are representative of all the communities or parishes in the region or country.

The large-scale surveys by INED in France and the Cambridge Group in England have been conducted, as we know, on samples of parishes or villages deemed to be potentially representative, to a certain degree, of the total (rural) population of the respective countries in different ways and hence, presumably, in demographic terms as well. Admittedly, we can raise the *yield* of reconstitutions and at least partly neutralize the mobility effect by working on large populations.[19] This approach was made possible by the computer, which has quickened data-archiving procedures and allowed the development of now well-tested *automatic linkage* procedures.[20] In any event, the reconstitution of families in a large demographic set—even if it consists simply of parishes or localities forming a representative sample—does offer a major advantage: the results can later be analyzed in greater detail, even from a *diachronic* standpoint, which is usually impossible when working on a handful of parishes.

Another approach, from a rather different research angle, offers unquestionable benefits at least in conceptual terms: the reconstitution of *genealogical lineages.* These enable demographers to track, from generation to generation, through the successive cohorts, all the demographic events occurring in a group of predefined family stocks.[21]

As a rule, genealogical studies are either descending or ascending. Descending genealogies are regarded as the more productive for historical demography, as their practical consequence is to supply family reconstitutions. Ascending genealogies are mainly used in genetics, for example to identify hereditary characteristics. They are less suitable for estimating demographic indicators because of the difficulty of locating all the individuals in the most distant generations—especially persons without offspring. In any event, methodological instruments have been developed to try to overcome the limitations due to the inadequacy of data sources in both approaches.

The aim is to try to follow each individual from birth to death even beyond the boundaries of his or her parish or community of origin and thus to analyze (1) *geographic mobility* and (2) *social mobility* over time by combining the demographic histories of successive generations. To some extent, the genealogical approach facilitates the identification of differential demographic behaviors (between family stocks, of course, but also between different social groups, as between successive generations). In certain conditions—most notably, when we can assume negligible information loss due to emigration—it also allows us to estimate the size of the observed population at different points in time (Rallu, 1992). This yields indicators that are totally unobtainable with classic nominal reconstitutions, such as empirical birth, marriage, and death rates. The linkage of these information items on genealogical lineages with nondemographic information also offers an interesting variety of applications—particularly, as one can easily imagine, in genetics and biomedical research (Chapter 96). By contrast, the evidence suggests we must be less optimistic about

[19] The leading example of an *exhaustive* nominal survey is undoubtedly the reconstitution of the Francophone population of Quebec by a University of Montreal research group (Charbonneau *et al*., 1987).

[20] Various programs are now available that even provide a complete, automated family reconstitution. The most notable is the one developed by the University of Montreal for reconstituting the Francophone population of Quebec (Charbonneau *et al*., 1987). Note, however, that these programs are generally not easy to adapt to sources whose characteristics differ from those for which they were initially designed. By far the most serious problem to solve is the *standardization* of nominal information (correction of reading and transcription errors, but also, and more important, the variants in name forms).

[21] The sample of 3,000 French families (Dupâquier, 1984, p. 128) is well known in the world of genealogical research. It seeks to reconstruct until the recent period the male genealogical lineages derived from original couples (early nineteenth century), with the aim of describing geographic and social mobility in nineteenth- and twentieth-century France. Many other interesting comments, on the implications of the genealogical approach in historical demography, are to be found in the monograph entitled *Démographie historique et généalogie* published by *Annales de démographie historique* in 1984. Lastly, we should recall the pioneering study by Andrea Schiaffino (1979, reprinted 1993) on fertility in Bologna in the late nineteenth century, based on the longitudinal analysis of the descendants of 532 first marriages registered in the city in 1871.

the real capability of the genealogical approach to solve—in terms of more strictly demographic results—the problems that cannot be solved via nominal reconstitutions covering a small number of parishes or communities. We shall return to these issues later on.

3. Tracking Life Courses

One example will suffice to illustrate the difficulty of fully tracking individual life histories: the obstacles encountered when seeking to measure the risk of dying in adulthood and the different alternatives that have been proposed to achieve this. Usually, the indicator cannot include the never married, except those who died in the parish studied. But, even for the ever married and widow(er)s, it will not do to restrict the risk-of-dying indicator to individuals belonging to the subset of "closed" families.[22] This would drive us toward an *overestimation* of mortality, as emigration introduces a selection in favor of individuals who have died at earlier ages.[23]

Louis Henry drew attention to the problem on several occasions, proposing increasingly refined solutions (Gautier and Henry, 1958; Henry, 1967; Henry, 1976); later, Jacques Dupâquier took up the subject (1978). In substance, we can try to retrieve the individuals—spouses or widow(er)s—whose dates of death are unknown by identifying the likely intervals within which the deaths occurred. However, depending on the type of information used to define the intervals, we may introduce biases, if the probability of observing an individual—and hence the probability of occurrence of a given event that would attest to the individual's presence—is not independent of the age at death (Blum, 1987, p. 40).

Alain Blum (1987) has explored a more "correct" method, at least in conceptual terms. In keeping with Louis Henry's suggestions, it involves the compilation of two different tables showing the upper and lower boundaries of *possible* mortality in the set of individuals studied. The novelty lies in the fact that the upper boundary of the interval in which the death should occur is estimated from information on fertility—more precisely, the mean birth intervals. For an individual who has had n children, the method basically consists in using the analysis of the fertility of the population studied to estimate the likely date at which (s)he would have had an $(n + 1)^{th}$ child.[24]

At the price of various methodological subterfuges, we can thus obtain information on adult mortality from nominal family reconstitutions. Nevertheless, it remains very incomplete. In particular, there will be no observations on older ages. Modeling (Blum and Bringé, 1992) therefore seems the logical path to follow in order to obtain plausible estimates of the total survival function by combining the probabilities of infant death—whose computation involves narrower risk margins—with the probabilities of adult death.

Naturally, when we can supplement the information on life histories gathered from family reconstitutions by information drawn from the nominal use of census sources, we can correspondingly reduce the adult mortality indicator's margins of approximation (Bideau and Brunet, 1993). The effect of exits from observation due to migration can then be estimated with a far smaller margin of error. Section VI of this chapter will deal precisely with how the application of sophisticated statistical techniques creates possibilities for the combined use of nominal status sources and movement sources.

V. DATA SERIES AND MODEL-BASED RECONSTITUTIONS

As noted earlier, there have been major advances, especially since the 1970s, in the methods for reconstituting the history of the main structure and movement indicators for past populations in the absence of full data. The development and application of new reconstitution methods have provided answers to a series of questions on the components of past-population dynamics hitherto visible only in theory. The pioneer in this field is unquestionably Ronald Lee (1974), with his earlier-mentioned formulation of an original reconstitution method subsequently known as *inverse projection* (*IP*).[25] At first, however, *IP* did not have the

[22] "Closed" families are defined as those for which the date of marriage and the dates of death of the spouses are known. To determine the exact ages at death, we also need to know the dates of birth of both spouses.

[23] In fact, it is obvious that the probability of leaving the observation field through migration and thus the frequency with which—for this reason—the family history cannot be "closed" increase as the individuals grow older. In a manner of speaking, the older the age at death, the greater the likelihood that the individual has migrated before dying. The deaths of younger persons are therefore overrepresented.

[24] The method proposed by Alain Blum has been criticized by Steven Ruggles (1992) who, using micro-simulation procedures, has found that it yields reasonably correct estimates of mortality when migration flows are intense, but can lead to significant overestimation of mortality when the migration flows are reduced. The limits of the method, Ruggles contends, stem from the fact that (unlike reality) it does not link the duration of intervals between successive events (used to estimate the upper boundary of the individuals' presence) to the probability of migration (Ruggles, 1992, p. 517).

[25] On the genesis of the method, see Ronald Lee's recent account (2004, esp. p. 1–4).

reception it deserved. Its application potential did not fully materialize until its author developed a second version (Lee, 1985) that made it possible to abandon the overrestrictive hypothesis of a zero net migration over the entire time interval examined. In any event, the debate over the use of models for reconstituting past populations was triggered by the publication of Anthony Wrigley and Roger Schofield's famous book (1981), based on a different method devised by Jim Oeppen: *back projection* (*BP*).[26] In some ways inspired by Ronald Lee's technique, *BP* enabled Wrigley and Schofield to paint a broad picture of demographic change in England spanning more than three centuries. Their work sparked a lively methodological discussion and stimulated other reconstitution projects in different geographic environments.

Space precludes our giving a full account of the lively, fertile debate fueled for many years by the inventors of *IP* and *BP*, and many other specialists in their wake (Lee, 1985; Wachter, 1986; Bonneuil, 1992). But it is this debate that has, among other things, led to the formulation, by Jim Oeppen (1993a) himself, of a new algorithm for reconstituting past populations: *generalized inverse projection* (*GIP*). Its advantage is high flexibility,[27] while enabling demographers to perform a more rigorous analysis using the same information required for *BP*. However, we shall confine ourselves to a few brief remarks on the potential and limitations of *IP*, especially as, in its various versions now available,[28] this reconstitution method has been by far the most commonly used in historical demography in recent years.

In its author's own words (Lee, 1993, p. 8), we can say that *IP* is a very simple demographic method for estimating the *levels* of a population's fertility, mortality, and age structure and their changes over time, from available time series of births and deaths and some data on population size. It is important to specify that the *inversion* is logical and not temporal, in the following sense: it does not project the sequences of births, deaths, and age structures on the basis of distributions of specific fertility and death rates as in normal projections; instead, it uses time series of births and deaths to compute sequences of rates and age structures.

Assuming, for simplicity's sake, zero migration flows, let us therefore try to summarize in a few words the idea underlying Ronald Lee's model. We assume that we know the initial population's size and age structure, and that the age distribution of mortality complies with that of a model (for example, one of Coale and Demeny's families of model life tables[29]). If so, there will be only one distribution of probabilities of dying (derived from the family chosen) that, applied to the population's initial age structure, will give the exact number of deaths recorded that year.[30] If we also know the total births in the first annual interval considered, and if we apply this distribution of probabilities of dying (which complies with the relationship indicated earlier) to the initial age structure, we can deduce the age structure at the start of the following year.

Having chosen a uniparametric family of life tables, we can estimate the probability $q_{x,t}$ of dying between ages x and $x + 1$ in the time interval t, $t + 1$ by means of the following formula:

$$q_{x,t} = {}^{1}q_x + k_t d_x \qquad [1]$$

where $d_x = {}^{2}q_x - {}^{1}q_x$, and ${}^{1}q_x$ and ${}^{2}q_x$ are the probabilities of dying between ages x and $x + 1$ obtained from two tables in that family (called table 1 and table 2, the first reflecting a higher mortality than the second). The two tables should be chosen so as to encompass, as fully as possible, the change in mortality of the population studied during the period examined.

The probabilities ${}^{1}q_x$ and ${}^{2}q_x$ (and hence d_x) depend only on age (meaning that they remain identical for each calendar year), while k_t depends only on time (its value, determined for each year, being identical at all ages).

Equation [1] can be resolved if we know k_t, whose value we can obatain by applying the following equation (demonstration not provided here):

[26] *Inverse projection* could not be applied to the English population for it requires a fairly accurate value for the initial population size—information unavailable for sixteenth-century England.

[27] Theoretically, *GIP* allows the model to incorporate all types of supplementary information. The fundamental difference between *IP* and *GIP* (Barbi and Rosina, 1997) lies in the treatment of migrations. In *IP*, migration rates are estimated from a net migration computed from empirical data (deduced from the comparison between movement data and census data); in *GIP*, by contrast, they are obtained with the reconstitution procedure (consistently with the same logic that informed *BP*). It would be interesting to test the application of *IP* and *GIP* to the same database. However, this is not possible for the moment, as Jim Oeppen's program is very sophisticated and hard to use (except for its author!).

[28] In this area, we can also mention—despite its being based on slightly different logical premises—a program developed by Salvatore Bertino and Eugenio Sonnino (1995): *stochastic inverse projection* (*SIP*). Although it is an original *micro-simulation* model, it basically uses the same *inputs* and produces the same *outputs* as the most common inverse-projection methods.

[29] Coale *et al*., 1983.

[30] The program actually uses two life tables, derived from the same model. They are uni-parametric tables, in the sense that they specify different levels of life expectancy at birth but have the same structure. Between the two tables, the model interpolates (or extrapolates beyond) a third table that belongs to the same family and that, once applied to the age distribution of the population, provides the number of deaths occurring in the year.

$$k_t = \frac{D_t - D_1}{D_2 - D_1}$$

where D_t is the (known) number of deaths recorded in year t, and D_1 and D_2 are respectively the deaths that we would have obtained by applying the probabilities of dying of tables 1 and 2 to the population's age structure in t.

Thus, once we have determined the value of k_t (which we may regard as a *mortality adjustment factor*) for the period $(t, t+1)$, we can use equation [1] to calculate the complete distribution of $q_{x,t}$ values for the same period, since k_t depends not on age but only on time.

Having thus obtained the distribution $(q_{x,t})$ of the probabilities of dying for period $(t, t+1)$, we can use them to project the population's age structure between t and $t + 1$. Simultaneously, we deduce an estimated life expectancy at birth (e_0) for each year.

We can repeat the procedure, in an altogether comparable manner, from year to year,[31] which will enable us to reconstruct not only an estimated life expectancy at birth but also an age distribution for each calendar year. If we further assume a model of age-specific fertility, we will be able to redistribute total births, year by year, as a function of the mother's age and hence determine an overall fertility indicator as well (total fertility rate or crude reproduction rate).

Our highly simplified description of the procedure enables us to emphasize the advantages, but also the limitations, of this reconstitution method. First, we should note that the method, now often tested (Lee, 1985; Breschi, 1990; McCaa and Vaupel, 1992), has displayed outstanding *robustness* with respect to the diversity of the mortality and fertility models used, as well as to the changes in the initial age structure, rather quickly "forgotten" during the projection (Wachter, 1986).

By contrast, *IP* is extremely sensitive to errors on the total size of the initial population—errors that grow over time for decreasing populations. In addition, systematic errors in birth and death series produce heavy estimation biases.

These phenomena confirm the need, even in the present context, of an accurate prior assessment of the reliability of the basic data. We should ponder the fact that the *IP* procedure merely translates the rather scanty information available on the changes in the population studied, these changes being expressed in terms of formally more refined indicators (the translation is achieved with the aid of fairly rigid assumptions, even if they are, at first glance, acceptable). Indeed, if the basic information is not sufficiently accurate, no method, however ingenious and sophisticated, can ever supply indicators—or, rather, estimates—closely reflecting a real situation that we cannot observe directly.

Nevertheless, having mentioned the robustness of the procedure with respect to the chosen model of mortality, we must admit that the hypothesis of a constant age distribution of mortality over time significantly restricts the power of *inverse projection*, as formulated by its inventor. The hypothesis is not very realistic if applied to historical periods characterized by frequent epidemics. Some of these often serious outbreaks could certainly strike all population strata indiscriminately and catastrophically. More commonly, however, they had a heavy selection effect by concentrating their effects on specific categories, most notably children. For the years marked by mortality spikes, we could obviously adopt (within the procedure) a different model for redistributing age-specific death rates, if we had at least summary indications on the nature of the event (typically, an epidemic) and hence on its possible differentiated effects. In dealing with catastrophic events such as the plague, historical demographers have sometimes assumed that mortality was neutral as far as age structure is concerned, impacting the entire population with a roughly identical force. Noël Bonneuil (1993) accordingly proposed to modify the procedure for distributing the age-specific death rates. He suggested adopting the standard model for deaths that would have been observed in the absence of a crisis and distributing the excess proportionally to the population age structure.[32]

In sum, there is ample room for amending and refining the procedure in this area. One alternative to the most commonly used mortality models (Coale *et al*., 1983; Ledermann, 1969) is William Brass's relationship system (1971), already applied by Jim Oeppen to *GIP* (Wrigley *et al*., 1997). The system might make it easier to incorporate any additional information, particularly from independent sources, on the characteristics of the mortality of the period[33] (doing this would also gradually modify the mortality structure over time, or in

[31] Applying an only slightly more complex procedure, we estimate annual net migration (using our knowledge of at least two population data items at the start and end of the reconstitution period) and therefore, by means of a subsequent model, its age distribution. The migrants distributed by age are then incorporated into the reconstitution, in a manner comparable to the procedure used in any ordinary projection.

[32] This implies identical ratios of crisis mortality at all ages. This idea was proposed by Robert McCaa (1993) and introduced into the most recent version of the "*Populate*" program.

[33] Brass's *logit* system is actually very useful when there exists an empirical life table for the population studied (or a table which, in any event, is assumed to be more consistent than a theoretical model with the age distribution of mortality in the population). In these cases, both tables used in the *IP* procedure to establish the death distribution, year after year, can be drawn, precisely, from the empirical table by keeping its structure (β parameter) constant and increasing or decreasing the α parameter, which determines the survival level.

specific years).

In cases where we have information on the age distribution of deaths in addition to the time series of deaths (a rather frequent occurrence, especially from the 19th century onward), another alternative is available: a modified version of *IP*, which its author describes (Rosina, 1994, 1996) as *differentiated inverse projection* or *DIP*. In many cases, *DIP* improves the estimate of survival levels and also yields indications on the typology of mortality.[34]

We have very little room here for a fuller discussion of this highly fertile and stimulating field of varied methodological proposals, although they do fall within the scope of reconstitution models operating on aggregate data. Earlier, we mentioned Noël Bonneuil's original method for reconstituting the 19th-century French population. Such a method appears, among other things, to have a structure closely tied to the data available in a particular historical and geographic context. It thus seems hard to generalize to other situations without some basic adjustments.[35]

For a diametrically opposite solution, let us take the case where our information is supposedly confined to a time series for births and our goal is to obtain an estimate of the population's size and, possibly, its age structure. We can cite two different methodological proposals. They are described in two publications: Noël Bonneuil's book (1993) and a paper by Massimo Livi Bacci and David Reher (1993). Despite their differences, both solutions involve (1) the reconstitution of cohort-specific mortality starting from birth (they therefore rely on vague general information on the characteristics of mortality in the period and geographic area studied) and (2) an assessment of the intensity and structure of net migration—a more difficult task. These situations seem, to us, representative of many others where data are almost totally lacking. Clearly, the priority here is to estimate the margins of approximation and to determine whether their size (usually rather large) is at least compatible with the research goals pursued.

Let us return to the most robust reconstitution techniques already described at length (chiefly *IP* and *GIP*) for one brief comment: the production of series of demographic indicators that we can easily obtain thanks to them has already made it possible, in many cases, to offer an extremely useful frame of reference for the characteristics and dynamics of past demographic systems.

Such results take on a special value when we can link the demographic series to comparable series of indicators reflecting economic or other conditions (for example, climatic) that may influence the demographic regime. In historical demography, as in the other branches of demography, the vast field of causal research is one of the most fascinating, but also one of the most arduous to explore—for two reasons: first, the chronic shortage of quantitative information needed; second, the difficulty of formulating and hence verifying the plausible models describing the relationships between the demographic system, the natural-environment system, and the socioeconomic system. The massive study by Anthony Wrigley and Roger Schofield on England is, in this respect as well, an essential benchmark;[36] however, it is also a model that may be hard to imitate, owing to the difficulty of obtaining the same quantity of historical economic data in other contexts.

The techniques most widely used today for analyzing the short-term relationships between economic (or meteorological) indicators and demographic indicators—but also to interlink demographic series—are common to econometric research and presuppose a consolidated theory on which we can rely to define the model's specifications (Tonini, 1993b, p. 16). In short, we need a very specific hypothesis on the meaning of the possible relationships between the variables studied. The most commonly used solutions are the *distributed-lags* models, applied by Ronald Lee (1985) in his study on English series, and by Patrick Galloway (1994) in a large-scale research project on the demographic system's responses to environmental and economic conditions, spanning four countries of northwestern Europe, between the mid-fifteenth and early 20th centuries. In distributed-lags models, the analysis rests on the following principle: the effects of a change in, say, the nutrition level (flagged by price observations) will not be confined to the demographic indicator examined (number of deaths, mortality rate, etc.) in the year in which the variation occurs; it may also influence a certain number of later years—or months, if the observation is monthly. The existence of such short-term relationships may therefore be effec-

[34] Taking, for example, two broad age groups (0 to 9 years and 10+ years), the program will follow the procedure that allows total deaths to be distributed into annual age classes separately for each group, by identifying for each group the survival level compatible with the total number of deaths. In the final stage, the complete results allow a comparative assessment of mortality trends in the two different age groups.

[35] The method in question, at least for the 1856–1906 period, is based on the age distribution of deaths and so requires data typically unavailable for the oldest periods (Bonneuil, 1997).

[36] The volume (Wrigley and Schofield, 1981) contains, among other things, an important study by Ronald Lee entitled "Short-term variation: vital rates, prices, weather." It is a fundamental contribution to the analysis of relationships between the demographic system, the natural environment, and the economic system.

tively identified by the models discussed above, which begin by eliminating the trend component, i.e., by reducing the series to stationarity conditions.

More recently, to study the short-term relationships between demographic and economic series, several researchers have used far more sophisticated statistical methods for analyzing time series, such as *Vector AutoRegressive Moving Average* (*VARMA*) models. Their main difference with econometric models based on distributed-lag regressions is that they are less structured (Tonini, 1993b): they posit a theory, at the most, for the choice of explanatory factors, and let the data suggest the form of the model. There is, for now, a fair amount of disagreement about whether these analytical methods can address the questions raised by historical demography research in a satisfactory manner and thus, ultimately, about whether they can clarify the mechanisms used by the demographic system to respond to stimuli from the economic and environmental system. Ronald Lee's highly critical verdict (1993) is particularly worth noting. He argues that structured econometric models remain a far more prudent means of attaining these objectives. In fact, all of the studies based on the use of modern statistical methods are rather explicitly informed by a different research philosophy that may be summed up as follows: we should not establish *a priori* the direction of relationships between the variables studied (see, for example, Entorf and Zimmermann, 1990); the set of potential linkages between the data should thus be determined on a purely empirical basis.

VI. STATISTICAL METHODOLOGY AND HISTORICAL DEMOGRAPHY

While historical demography has developed specific techniques for studying past populations—often needed to offset the chronic shortage of data—statistical methods have always been crucial to an understanding of past demographic dynamics.

Let us therefore review very briefly some of the statistical applications proposed in the literature and the new research approaches offered by available methods, most notably the analysis of individual data.

1. Analysis of Aggregate Data

Thanks to the results of observations and studies that have allowed the construction of demographic indicators at a fairly detailed level, the number of investigations of geographic variations—of mortality in particular—has been growing steadily since the 1970s. The approach known as *ecological* is chiefly based on factor analysis methods or, in their broadest sense, multidimensional analyses. The goal has been to link mortality indicators with contextual external variables, such as socioeconomic or health conditions, assumed to have a potential influence on the risk of death. Many valuable studies concentrate on this specific approach.[37] Researchers have also started to apply the ecological approach, albeit more belatedly, to historical periods, in particular the ones closest to us: the 19th and early 20th centuries. In addition to mortality, these studies cover fertility, marriage, and the identification of differential demographic regimes. The approach here is purely descriptive. It is largely focused on the identification of specific profiles, typologies or models of demographic behavior at the aggregate level.

This type of analysis, as noted earlier, rests primarily on the availability of aggregate statistics drawn from status data (such as censuses) and movement data or special surveys: in the 19th century, such material started to become available on a reasonably regular basis in many regions, in Italy and elsewhere in Europe. From these sources, we can often calculate or enumerate not only the standard demographic indicators, but also other variables in areas such as socioeconomic conditions, farming systems, occupational classification, some information on health status, educational attainment, and so on. While displayed as aggregate data, such inputs are well suited to multidimensional analysis. The basic scientific goal is always to determine geographic differences in behavior. Thanks to the links between variables that can be established with these multidimensional analyses, we can, to some extent, begin to examine the causal mechanisms underlying the differences observed.

The choice of the methods mentioned above is guided not only by technical criteria concerning the form of data availability—the data are aggregated, although the aggregation sometimes involves a very refined level of territorial division, even down to basic communities—but also by conceptual motives. The contextual indicators computable from status data, for example, are seen as measuring variables that influence the demographic parameter studied: death rate or probability of dying, total fertility rate, mean age at marriage, never-married rate, etc.

The ecological approach, however, has well-known limitations. While offering a rich and interesting description of phenomena, it does little to identify causal relationships. But its main shortcoming is inherent in the type of data employed and the time scale of the indicators. Because the approach uses aggregate

[37] For a selection, see: Castiglioni *et al*., 1991; Pozzi, 1990, 2000; Lesthaeghe, 1992.

data on entire populations or subgroups of individuals, the subject of the investigation is not the specific risk factors for a given behavior, but their *proxy* or, more precisely, an approximation of their proxy consisting of indicators of the group's average behavior (Caselli, 1995). Also recall that one of the major unresolved problems of ecological analysis is its inability to allow for the time that a context variable may require to influence the demographic process leading to the event studied.

Another area of methodological exploration that also concerns historical demography is the interpretation of time series for rates or probabilities in three dimensions: age, period, and cohort.

The reading of cohort indicators and their relationships with period measures have consistently interested historical demographers ever since Louis Henry stressed their potential via *nominal reconstitution*. Indeed, historical demography—most notably in the area of fertility analysis—was the first discipline to emphasize the importance of a combined reading of behavioral changes by cohort group and of period variations. This approach has long captured the attention of researchers seeking an overall interpretation of fertility trends but also, and above all, an identification of the start of the transition process for reproductive behaviors and their control.

The analysis of fertility through the family reconstitution technique thus clearly requires putting individual data in perspective. By contrast, the cohort analysis of aggregate data is a more recent development in historical demography.

The demographic use of specific statistical models to measure and identify the three different components of a matrix of rates or probabilities by age, period, and cohort (APC) is recent. In any event, the APC approach (see Volume I, Chapter 18) remains a descriptive approach that seeks to define a time frame of reference for the processes but cannot replace a causal analysis. As we know, demographers are applying APC models in an effort to distinguish the weights of the age, period, and cohort components in the explanation of the internal variability of a matrix of rates or probabilities.[38] An understanding of the different dimensions of the phenomenon analyzed can only help to move through the later stages of causal analysis.

The application of this type of statistical techniques can have highly stimulating effects in historical demography. If we are trying to reconstruct matrixes of probabilities of dying for sufficiently long periods, it is certainly worth describing the change in the age distribution of general mortality or of mortality in specific age groups by identifying the effects of periods or cohorts whose mortality diverges from the standard underlying level. We can thus specifically identify (1) period effects, due, for example, to epidemic crises or cycles, which were extremely frequent in the past, and (2) cohort effects, understood as the delayed consequences of events affecting particular cohorts, such as cohorts born in an unusually adverse period. With this technique, we can therefore begin to study the selection effect of a given epidemic—such as the plague—on certain cohorts by identifying its impact on the population's age structure (Rosina *et al*., 1999).

In this area, we should mention the application of the *age-period-cohort* approach to the life tables of Italian cohorts born between 1882 and 1953 (Caselli and Capocaccia, 1989). One of the study's main aims was to determine, by means of a log linear function, the relationship between child and adolescent mortality and adult mortality. The theoretical frame of reference is precisely the one just mentioned, i.e., the potential consequences, in survival terms, of unusually adverse conditions experienced by certain cohorts in childhood. Such conditions may exert a positive selection effect by eliminating the weakest individuals at early ages. As a result, they may increase the statistical weight of the most robust individuals and diminish the cohort's mean mortality at older ages. If, instead, the main consequence of the event is to weaken the survivors, the cohort's adult mortality will be higher. Without going into the details of the modeling results, when we can take account of age cohort interactions in the mortality variations, we can pinpoint the specific cohort effect by underlining the linkage between (1) the probability of dying in adulthood and (2) child and adolescent mortality. The model also allows a fairly specific analysis of intercohort differences in mortality by clearly identifying the most disadvantaged cohorts.

2. Analysis of Individual Data

Our knowledge of the differences between demographic regimes and of the changes in processes over time is steadily advancing. Hence it is ever more relevant and necessary to seek methods and forms of information analysis with which we can identify the true relationships underlying the data. To do this, we

[38] As we now know, APC models face identification problems (see vol. 1, chap. 18). In fact, it is impossible to estimate all the parameters simultaneously because of the linear dependence between the three variables involved. As a result, the models linking the dependent variables (rates or probabilities) to the three components (age, period, cohort) do not have a single solution, given that period = age + cohort. Therefore, we can obtain identical descriptions of data from different *sets* of parameter values; equally well, non-different groups of parameters may lead to very different interpretations. The only solution to the identification problem is to impose mathematical constraints on the model coefficients (Martelli, 1997).

must search for causes.

Such analysis makes it more vital than ever to take account of individual behaviors and the sequence of phases and choices that determine them. Concretely, this means that analysis, even in historical demography, now tends to reappraise the individual approach in terms of *life courses*.

The limits of the classic method of nominal family reconstitution have been recalled earlier; as we stressed, the yield of reconstitutions—measured by the numbers of individuals or couples under observation at different moments—can be significantly improved through the joint use of status data and movement data (Bideau and Brunet, 1993; Ge Rondi, 1988; Manfredini, 1996). This improvement is, as we know, perfectly feasible and rich with potential for many countries, thanks to the existence of records updated on an annual, multiannual, or even continuous basis: *status animarum* (nominal registers of religious origin) and *anagrafe* (civil registration) in Italy, "population books" in Belgium and Germany, parish registers tracking household changes in Scandinavia, and so on.

The aspect to stress here, however, is not so much the possibility of improving the yield from reconstitutions. The far more important task, taking Louis Henry's basic intuitions as a starting point, is to amplify the concepts underpinning such techniques in order to arrive at the statistical analysis of individual data.

As emphasized earlier, the powerful growth of information technology, especially since the 1980s, has had two consequences. First, it has allowed the collection of enormous databases and facilitated their management. Second, thanks to ever more effective algorithms, it has solved many of the problems involved in the *linkage* of documents pertaining to a single individual taken from different sources (status data, movement data, cadastral data, etc.). Today, it is far easier than in a still recent past to achieve a rapid reconstitution of life histories for a large number of persons who were active in a given period of variable length and were registered in nominal sources during that time.

Such possibilities may actually enable demographers to overcome many of the limitations specific to classic nominal reconstitutions by altering the perspective from which reconstituted event histories are viewed. We can thus switch from an often selective analysis—such as that of nominal reconstitution—to a study that treats individual histories as a sequence of transitions between different statuses. The transitions, which may involve different time intervals or durations, are marked by the events. In this approach, we can read a life course as a sequence of events and statuses along a time line (individual age or the time elapsed since a previous event), between a given instant that serves as the starting point and another that marks the exit from observation. To these reconstituted biographies, we can apply the many statistical techniques of regression analysis grouped under the name of *event-history analysis* (Alter, 1998).

Since our focus is on the advantages of moving from old forms of demographic analysis to new ones, it is interesting to note that we are not witnessing a conceptual break, particularly in historical demography: the fundamental principles of the statistical analysis of event histories were already present in Louis Henry's system and spelled out in the rules of family reconstitution. Indeed, in his famous monograph on the town of Crulai in Normandy, Henry had explicitly argued that the tracking of life histories would have been a valuable approach. The only reason that kept him from pursuing it was the lack of suitable statistical tools at the time.

In an interesting work on the links between *nominal reconstitution* of families and *event-history analysis*, Myron Gutmann and George Alter (1993) have stressed the resemblances and discontinuities between the two techniques by summing up the main limitations of the first in three points: rigidity of analytical rules for arriving at the computation of demographic indicators,[39] impossibility of conducting multivariate analysis, and difficulty of analyzing event sequences.

In reality, the notion that the understanding of individual demographic processes implies a continuous observation over time—already present in *nominal reconstitution*—is also the basic premise in *event-history analysis*.

In nominal reconstitutions, however, the rigidity of the rules for calculating demographic indicators entails a sizable loss of information by excluding all the data for which a truncation boundary cannot be determined. The analysis is thus confined to a subset of the universe analyzed—with all the problems this implies for assessing their representativeness. Actually, this limitation may be partly overcome, as noted earlier, by using supplementary sources of information, such as nominal status sources (e.g., censuses or tax lists), that may provide a clear end point for the observation.

Determining a specific moment in which the individual enters the observation field means that, if we know the starting point, we can identify with certainty the period of exposure to the risk of experiencing certain events. That said, if we know the sequence of events on the time scale, we can apply a series of statistical methods to analyze the durations and occur-

[39] This rigidity underlies the selectivity of the data used with respect to the data gathered initially.

rence rates of the events for all individuals exposed to the risk (Yamaguchi, 1991). Of course, there are different ways to describe the event histories: number of persons in a given status, mean time spent in a given status, rate of transition between statuses, distributions of durations between transitions, and so on (Gutmann and Alter, 1993).

The analysis of *event histories* also allows a fuller, multivariate analysis. Whether we choose continuous observation or discrete observations, regression techniques enable us to measure the extent to which certain individual or family characteristics can influence the likelihood of a transition between statuses, interstatus durations, or the rate of occurrence of a given event. Event-history analysis also allows us to take account of the succession of events in an individual life and of how yesterday's events can influence the likelihood of a later event (Courgeau and Lelièvre, 2001).

For historical demographers, *life histories* are certainly nothing new. However, the collection of ever more massive amounts of data and the enormous financial and labor costs needed to create event histories via *linkage* may find a practical justification if these data are then properly analyzed. The aim of the exercise must be to identify and describe the true relationships between the data. One can then move from the description and its attendant hypotheses to the confirmation and explanation of the relationships.

As noted earlier, the data from which we can reconstruct life courses may consist of movement data and status data from nominal lists or censuses. Movement data give us the dates of occurrence that tell us about the main stages in a person's life (birth, death, marriage, and, with greater difficulty, migration). Status data provide information on individual characteristics and attributes at a given instant or at successive moments.

When information abounds—for example, in Italy, where there are annual nominal status sources going back several centuries—we can reconstitute long portions of life histories in two steps. First, we establish a nominal link between successive lists. Second, we introduce the event dates from the continuous recordings into the life histories on an annual basis. In that case, the observation measurement unit is the year, and the biography naturally becomes a reconstitution in discrete durations. Where population registers (vital statistics) are available, the identification of attributes and the dating of events are very often combined in the same source; the life history is automatically generated by the reconstitutions performed on a routine basis. In both cases, the biographical reconstitutions are amenable to *event-history analysis* (Lagazio and Pagani, 1996).

The life histories thus identified are not always complete, however: they are beset by well-known *censoring* problems.[40] Both right-censoring and left-censoring are common in life histories reconstituted from historical sources. Observations often end before the analyzed event occurs (right-censoring). In certain cases, on the contrary, we can determine that an individual has already experienced the terminal event without our being able to date its occurrence (left-censoring).

Censoring must not be linked to the analyzed event, nor must the mechanism that causes the censoring enable us to predict the transition studied in any way. This essential notion was already explicit in Louis Henry's nominal reconstitution; Henry duly underscored its importance in his description of the possibility of "closing" families (Fleury and Henry, 1976). Both types of censoring can be resolved by applying suitable techniques of statistical analysis. Likewise, solutions have been proposed to overcome the difficulties involved in the use of fragmentary data, in particular the discontinuous data obtained by reconstituting life courses from civil registration records (Courgeau and Najim, 1995; Bonneuil *et al*., 2004).

Two other aspects of historical individual data should be emphasized (Diamond *et al*., 1993). Such data may be subject to recording errors that involve omission or to reporting and linkage errors. Such errors weigh on the computation of transition lengths and thus on the final estimate of the transition.

In addition, unobserved and hence unidentifiable factors may influence the processes studied. We will have to make allowance for these latent variables by looking out for them in the statistical analysis and inserting them into the models, so that we can accurately quantify the effect of the measurable variables.

In demography, the techniques of *event-history analysis* were first applied to the study of mortality (Hobcraft and Murphy, 1986). Today, demographers also tend to use them in the analysis of other demographic phenomena such as fertility, nuptiality, and migration. In historical demography, the first attempts at application concerned data from population registers, which are particularly well suited to the study of event histories (Gutmann and Watkins, 1990; Kertzer and Hogan, 1989).

The technique has now been rapidly disseminated via applications covering a wide range of biographical data. One of the most interesting applications concerns

[40] For a clear definition of censoring and truncation, see Trussell and Guinnane (1993); Yamaguchi, 1991.

the analysis of infant mortality, which has consistently played a leading role in the quest for a unifying theory of demographic transition (Lynch and Greenhouse, 1994; Lee *et al*., 1994; Breschi *et al*., 1997). Thanks to the methods suitable for aggregate data, our knowledge of the changes in infant mortality is making great progress; but the determination of the underlying conditions of such changes is important as well. For example, demographers have empirically observed massive variations in infant mortality levels between social classes, ethnic groups, and cultural groups. As we explore in ever greater detail the variables linked to infant mortality, it is interesting to study, for example, the potential influences of family factors on the risk of death, such as the family's general health conditions (which we can approximate, among other ways, via the infant and child mortality of the same mother's earlier children), its socio-occupational status, or its display of protective behavior toward children. To these variables operating at the family level we can add others concerning the environment and the community of residence, such as season of birth, crisis experience, and economic conditions.

If we have long-term data and can link them to the environment, we can also measure the lagged effect of childhood living conditions on adult mortality (Alter and Oris, 2002; Bengtsson and Lindstrom, 2000). This causal investigation even extends to the prenatal period. A recent study used multivariate regression analysis to track the birth cohorts of a sample of 16,000 European aristocrats born in the years 1800–1880. It found significantly shorter adult life expectancies among daughters conceived by fathers over 45. The method identified a strictly gender-specific effect: the authors found no evidence of a negative relationship between the father's age at conception and life expectancy for male children (Gavrilova *et al*., 2003).

Fertility studies are equally of interest in methodological terms (Mroz and Weir, 1990). We can use *event histories* to study if a woman is likely to give birth again and in what time frame the event may occur. The risk of giving birth and of parity progression can also be measured as a function not only of the woman's demographic characteristics (age, length of marriage, birth cohort, etc.), but also of factors directly linked to reproductive choices, such as the effect of replacing a child who died in infancy, or the physiological impossibility of conception due to breast-feeding. A number of the attributes included in the model will not remain fixed for the entire observation period but will vary over time, and their treatment will require the use of multivariate regression that will take proper account of time-dependent variables.

In conclusion, event-history analysis is a promising tool for causal research even regarding past populations. A final point concerns the great variety of relationships that can be analyzed with these methods. The occurrence exposure rate for a given event or the probability of its happening can be influenced by many independent factors. Let us assume that the dependent variable is adult mortality. An *event-history* regression model will enable us to establish relationships with equal ease between (1) the variables that pertain to the personal sphere (such as sex or year of birth) or identify family status (socio-occupational category, migration status, family composition) and (2) covariates acting at the community level (economic fluctuations proxied by prices, temperature, crises). Some of these factors may vary in the course of an individual life; we can then take such variations into account when determining exposure to the risk of dying. The new feature here is that the researcher can check the effect of macroscopic variations—such as economic or climatic fluctuations—at the individual level and the individual response to them, even on small populations. Such versatility makes the event-history method a key instrument for analyzing living conditions in pre-industrial societies by once again combining the *micro* and *macro* approaches into a single structure.

VII. KINSHIP AND ENVIRONMENT

Another significant feature of contemporary historical demography is the growth in studies based on genealogical data. To some extent, this development carries on an old tradition. In the early 1950s, Louis Henry began his quest for sources that might provide information on "natural" fertility levels. He first considered published genealogies. But he soon realized that using such sources could create social-selection effects that might bias the demographic conclusions. Henry soon opted instead for parish registers, to which he had been introduced by the Canadian Jacques Henripin, then an intern at INED.[41] Yet he did not entirely abandon the use of genealogical documents: they enabled him to produce his first major monograph (Henry, 1956), and he continued to explore their biases.

This early use of genealogies to supply basic information for demographic reconstitutions was hardly confined to France (Hollingsworth, 1957). In the past twenty years or so, it has enjoyed a revival thanks to the

[41] On the history of this trial-and-error process that led Louis Henry to "invent" historical demography, see Rosental (2003, p. 221ff).

powerful momentum of historical demographic studies on China. The country has a wealth of kinship data on descendants of a common ancestor, often spanning many generations (Lee and Campbell, 1997). They therefore constitute, if not the sole source, at least a major documentary series for reconstructing the main demographic variables of ancient China over the long term (Peng and Hou, 1996; Zhao 1997). However, there has been considerable debate over the risks of bias induced by this type of document. In a recent summary, Zhongwei Zhao (2001) focuses on the demographic-selection effect of the sources, whose presentation implies that the lineages under study were reproduced from generation to generation. The author concludes that the effect of this selection bias is concentrated in the initial generations, and that its intensity is inversely proportional to the genealogical depth available. Zhao relies on microsimulation, one of the techniques now commonly used to measure the potential biases entailed by the use of genealogies, whether preconstructed or reconstituted by the demographer: it allows to measure the gap between the observed data and the data that would presumably result from the interaction of what are regarded as plausible demographic variables (Smith and Oeppen, 1993).

A second use of genealogical data combines historical demography and genetics. It too has a long tradition, as it intersected the study of human settlement through the tracking of the spatial diffusion of genetic malformations (Jacquard, 1970). The compilation of large bases of nominal data by historical demographers has given a new impulse to this field of study, which can simultaneously (1) identify the immigrant couple(s) initially responsible for the spread of a genetic disease in their host country and (2) trace settlement processes at the regional level (Tremblay *et al*., 2003; Zeegers *et al*., 2004). Using its own specific sources, this approach thus bolsters geneticists' efforts (Cavalli-Sforza *et al*., 1994) to reconstitute the dissemination of human settlements across the planet. Another use of genealogical databases derived from historical demography is the analysis of potential demographic effects of consanguinity, particularly on infant mortality (Eberst Dorsten *et al*., 1999).

But, if we look in a totally different direction, the growth of genealogical databases also illustrates how studies on nominal data are providing historical demography with new approaches. At the outset, it will be recalled, Louis Henry's establishment of historical demography went hand in hand with a non-nominal use of parish register data. Family reconstitutions served to produce aggregate data, offsetting the lack of series comparable to those produced by administrative statistics beginning around 1800. Since the 1980s, the methods of life course statistical analysis have enhanced the potential for investigations, particularly of a causal nature. But the compilation of genealogical databases goes even further by calling into question the observation unit of reference. Between the population, defined as a macroscopic aggregate, and the individual, captured at the microscopic level, we are witnessing the emergence of a mesoscopic level that may radically reshape the priorities and explanatory frameworks of historical demography. Its observation unit is the interpersonal configuration in which each individual operates.

This shift has occurred gradually, starting with a question at the intersection between historical demography and family history: What was the actual availability of kinship in traditional societies? How did the number of kinship ties in the widest sense (mother and father, grandparents, siblings, etc.) change during a life cycle? This question, which, for a long time, could be addressed only via microsimulation (Le Bras, 1973), can now be tackled more directly—and more precisely—thanks to the advent of large databases with a family component. Post *et al.* (1997) do this for the Netherlands using information on 160,000 individuals for the period 1830 to 1990. But this wholly demographic approach to the issue is clearly incapable of transcending a purely formal vision of kinship, in which ties are defined by a degree of genealogical proximity, but not by their effective degree of activation or by their effects on ego. To sum up, we could say that the approach is useful for defining the widest boundaries of kinship in a given sociohistorical context, and that we can then measure the intensity of family mobilization actually observed against this theoretical denominator.

How can we break out of the confines of the household—a unit that long enjoyed the more or less explicit preference of historical demography—without overstating the importance of the extended family? How can we conceptualize a kinship network whose boundaries are intrinsically variable? To find the proper compromise and identify the boundaries of the family ties actually mobilized, many authors use indirect measures. In his study on the protective effect of the extended family on the mortality of illegitimate children, Anders Brändström (1997) looks at the geographic distance between related individuals. He relies on the database compiled in Umea on the Sundsvall region of Sweden in the mid-19th century: this involved a comprehensive enumeration and genealogical linkage of the population of several contiguous parishes. Brändström observes that an unmarried mother whose parents resided in the parish in which she gave birth was significantly more likely to see her children survive. In addition to Derosas (2002), a recent review study by Frans van Poppel (2000) suggests the degree of generality of this result at the European level for infant mortality in the aggregate.

The study of family proximity also enables us to reassess the relative influence of kinship networks on the geographic direction of migrations, as Bonneuil *et al*. (2004) show using the database of the "3,000 Families" survey (see footnote 22). The authors of the study have created individual files for each person in which the demographic, geographic, and social characteristics of ego are correlated with those of all the person's kin. We can thus observe the impact of each individual movement on the other kin: does it take ego away from, or bring ego closer to, his/her parents, siblings, aunts and uncles, etc.? The study breaks with the increasingly widespread notion that migration is chiefly determined by interpersonal networks, and in particular by kinship networks. The authors show that most family constellations (if we use the term for a given kinship tie) do not "explain" the destinations chosen by migrants. Migration does not bring related persons closer—or, if it does, the effect is weak. From a complementary viewpoint and using the same data, Kesztenbaum (2004) observes that, on the whole, internal migrants seldom choose as their destinations the localities in which their relatives resided. These counterintuitive results reveal the contribution of a genealogical unit of observation: it enables us to analyze a representative national sample, chosen independently of the phenomenon studied; by contrast, studies based on the observation of migrant populations may sometimes overrepresent networked individuals to the detriment of those who have dispersed toward seldom-chosen destinations.

These studies also show that extended kinship is not an intrinsically relevant unit for observing individual behavior (Plakans and Wetherell, 2003). To move out of the confines of the household, we must objectify the nature of the actual relations between members of a kinship group. With this aim in mind, historical demography is beginning to share the social sciences' growing interest in the study of interpersonal relationships (Lemercier, 2005). The practical result is a new approach to database design. It is no longer enough to collect biographical or intergenerational information (i.e., on individual life courses); we must now combine this with relational data, such as the choice of godparents at baptism and marriage witnesses.

This information can be used in different ways. One solution is to objectify it in the form of indicators—an approach well suited to samples of lineages that do not share the same environment, such as national samples. We can thus observe the extent to which members of a given lineage choose kin as marriage witnesses. This indicator arguably enables us to distinguish between *autocentric* lineages, in which the lineage and reference group coincide, and *exocentric* lineages, whose members do not share personal ties (Rosental, 1999). We can then incorporate the variable into logistic regressions measuring its relative power to explain sociodemographic behaviors, as against the classical attributes of ego such as age, gender, and social status (Bourdieu *et al*., 2000).

A second solution consists in the use of measuring instruments derived from *network analysis* (see Chapter 137, this volume). It is suitable for analyzing lineages that all function in the same social space. One example is the pioneering study by Cristina Munno (2005, 2006), which offers a genealogical and relational reconstruction of the entire population of a wool-producing town in Venetia in the 19th century. By combining event-history analysis and network analysis, the author seeks to determine the extent to which the integration of households into the social fabric of the parish determines their demographic characteristics, such as mortality and geographic mobility.

Besides their intrinsic contribution, studies of this kind are important because they clearly show that the analysis of genealogical databases calls for a sociohistorical approach—in the sense in which, for example, the nature of interpersonal networks depends on social position. This stands in contrast to the drift toward sociobiological approaches, which, through the use of contemporary statistical tools, revive the quest for a so-called genetic heredity of behavior. Geneticists are often the first to warn demographers and historical demographers against the simplistic nature of such determinism, and the interpretative reductionism that informs it. To take just one example, Westendorp and Kirkwood (2001), using a genealogical database on the English nobility for the period 740 to 1995, claim to demonstrate "influences" of parental longevity on offspring lifespan. But they follow this up immediately with an assertion that the result does not, in itself, imply a hereditary mechanism. To show the limits of the finding, the authors introduce an environmental variable easily available in their database: the age at death of ego's spouse. They see this as a convenient indicator of the influence of the environment on ego's mortality, and effectively illustrate its importance.

The scope of Westendorp and Kirkwood's study does not require them to define in greater detail what constitutes the "environment." But, as it happens, the notion is now under intense investigation. The development of historical demography in the 1950s was informed by an organic and broadly biological approach to population, which still permeates some contemporary work. However, the renewal of statistical methods coincided with a deeper change. Alongside historical demography—which, as its name indicates, is a demographic enclave in the land of history—we are witnessing the emergence of a social and political history that sees population as the product of social and political dynamics.[42] If we confine our analysis to the most quantitative aspects, we can observe this change in all the studies that examine the direct and indirect effects of social arrangements on demographic behaviors—instead of treating demographic dynamics as the outcome of more or less endogenous variables (such as fertility and mortality), or as the mechanical product of economic conditions. The link between population and social policies becomes the focus of the analysis.

The United Kingdom has proved a particularly fertile terrain for such explorations: a growing number of historians are combining the study of populations with that of the Poor Law system. Given the wide variations in poor relief from one parish to another, and the fact that it could involve a large share of the population, we may legitimately ask whether it might be an explanatory factor in accounting for local demographic differences. Andrew Blaikie (1998), for example, has studied the effects of local social protection (as well as the effects of family configurations) on infant mortality in two Scottish parishes. But the scholar who has unquestionably gone furthest in using the Poor Law system as the key to understanding all of 19th-century English demographics is Steve King. In addition to the classical sources of historical demography, he examines all the documentation produced by Poor Law administration—most notably the "pauper letters" sent to local authorities to request and justify relief. King uses these sources in two ways. First, the quantitative analysis of the content of these documents—of which tens of thousands survive—gives clues about the boundaries of kinship regarded by the "overseers" of poor relief as relevant and capable of being called upon. Admittedly, King is not the first to qualify the assumption, introduced by Peter Laslett, of "nuclear hardship"—i.e., that English households could not count on support from their kinship group.[43] Rather, King's originality is his incorporation of quasinarrative sources, far removed from the tradition of historical demography, into the social history of population.

A second use of these sources consists precisely in matching them against households' sociodemographic trajectories. King begins by noting the existence of demographic "sub-groups," i.e., households displaying a propensity for a specific behavior such as high infant mortality, illegitimacy, early marriage, or high fertility.[44] These concentration phenomena have been known for a long time, and not only in England. As they are often reproduced by successive generations, they are readily interpreted in genetic or culturalist terms.[45] True, they do not overlap conventional sociological variables such as occupation or land-tenure status. But King re-examines the issue in terms of the Poor Law criteria, distinguishing between the lineages that never require relief, those that fail to emancipate themselves from need, and those that seek relief either frequently or occasionally. These categories, drawn from sources that are unconventional in historical demography, turn out to provide far more effective explanations of behaviors. King attributes this to the different paths to marriage followed by the poor, and to the impact of these differences on their later life cycle.

On another scale, many demographic differences between parishes become easier to understand when we analyze the differences in the volume of poor relief distributed and in the procedures for doing so—such as cash payments, medical examinations, and handouts of food or clothing—whose practical effectiveness could vary. These differences, in turn, depend on the balance of power within the elites and between the elites and the rest of the population, as well as on pressure from neighboring parishes that did not want to see an influx of paupers ill-protected by their native villages or towns. In another area, the study by Frans van Poppel *et al.* (2005) on three provinces of the Netherlands reveals the contemporary spread of this type of approach, which is profoundly reshaping the picture of the population drawn by historical demographers. Using an excerpt of the "Historical Sample of the Population of the Netherlands," which seeks to document 76,700 life courses between 1812 and 1922, the authors show that the provincial environment—principally, health conditions—has an even stronger long-term effect on infant mortality than social status. Moreover, far from displaying a linear uptrend or downtrend, social inequality in the face of infant mortality fluctuates over time, precisely as a function of the environment. This finding has a major methodological implication: it demonstrates the low relevance of

[42] On this point, see Rosental (2006).

[43] See a formulation of this hypothesis by Laslett (1988) and its critique by Reay (1996).

[44] On this and subsequent points, see Hudson and King (2000), and King (1997, 2005, 2006).

[45] Cf. the notion of a culture of romantic relationships producing illegitimate children (Laslett, 1980), or Bengtsson and Lindstrom's "genetic" interpretation (2004).

studies covering a brief time period.

In sum, contemporary demographic research no longer sees the population as an organic whole or as an entity automatically influenced by economic constraints. The approach that is gaining ground views the population as the product of sociopolitical power plays. In methodological terms, this vision requires demographers to widen the range of sources used, and to develop, if need be, statistical methods combining paths (intra- or intergenerational), social networks (both inside and outside the family), and environmental variables.[46] If these steps are taken, the demographic approach will help us to interpret, in a broader perspective, the ways in which societies are continuously shaping themselves.

Acknowledgment

This chapter was translated by Jonathan Mandelbaum from Jacques Vallin's French version of the original Italian.

References

Alter, George, 1998. L'*event history analysis* en démographie historique. Difficultés et perspectives, *Annales de démographie historique*, no. 2, p. 25–35.

Alter, George and Oris, Michel, 2002. Early life conditions and the decline of adult mortality in the Belgian Ardennes, 1812–1890, University of Indiana, ELC Working Paper Series 2002, no. 2, 28 p.

"Banques de données," thematic dossier, *Annales de démographie historique*, 1998, no. 2.

Barbi Elisabetta and Rosina Alessandro, 1997. *Generalized inverse projection: caratteristiche metodologiche e possibilità applicative ai dati italiani.* Rome, Gruppo di Coordinamento per la Demografia della Società Italiana di Statistica, 11 p. (paper given at "Giornate di studio sulla popolazione," Roma, January 7–9, 1997)

Barbi Elisabetta, Bertino Salvatore and Sonnino Eugenio (eds.), *Inverse Projection Techniques. Old and New Approaches*, Demographic Research Monographs, Heidelberg, Springer, 2004, 155 p.

Bengtsson Tommy and Broström Göran, 2004. Effects of Mothers on Children's Mortality in Older Ages, Southern Sweden, 1829–1894, "New History of Kinship" Conference, UIESP-INED-EHESS, Paris, October 1–2, 2004, mimeo.

Bengtsson Tommy and Lindstrom Martin, 2000. Childhood Misery and Disease in Later Life: The Effects on Mortality in Old Age of Hazards Experienced in Early Life, Southern Sweden, 1760–1894, *Population Studies*, vol. 54(3), p. 263–277.

Bertino Salvatore and Sonnino Eugenio, 1995. La proiezione inversa stocastica: tecnica e applicazione, in: Marco Breschi and Lorenzo Del Panta (eds.), *Le Italie demografiche. Saggi di demografia storica*, p. 99–122. Udine, Dipartimento di Scienze Statistiche, Università degli Studi, 391 p.

Bideau Alain and Brunet Guy, 1993. The construction of individual life histories. Application to the study of geographical mobility in the Valserine valley (French Jura) in the nineteenth and twentieth centuries, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 111–124. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Blaikie Andrew, 1998. Infant Survival Chances, Unmarried Motherhood and Domestic Arrangements in Rural Scotland, 1845–1945, *Local Population Studies*, vol. 60, p. 34–46.

Blum Alain, 1987. Estimation de la mortalité locale des adultes à partir des fiches de familles, *Population*, vol. 42(1), p. 39–56.

Blum Alain and Bringé Arnaud, 1992. Mortalité locale et générale en France: 1670–1829, in: Alain Blum, Noël Bonneuil, and Didier Blanchet (eds.), *Modèles de la démographie historique*, p. 11–43. Paris, Institut National d'Études Démographiques ("Congrès et colloques" series, no. 11), 370 p.

Bonneuil Noël, 1992. Non identifiabilité et cohérence démographique de la rétro-projection, in: Alain Blum, Noël Bonneuil, and Didier Blanchet (eds.), *Modèles de la démographie historique*, p. 99–107. Paris, Institut National d'Études Démographiques ("Congrès et colloques" series, no. 11), 370 p.

Bonneuil Noël, 1993. The trend method applied to English data, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 57–65. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Bonneuil Noël, 1997. *Transformation of the French demographic landscape 1806–1906.* Oxford, Clarendon Press, 218 p.

Bonneuil Noël, 1998. Traitement des données manquantes dans les séries issues des registres paroissiaux, *Population*, "Histoire" special issue, no. 1–2, p. 249–72.

Bonneuil Noël, Bringé Arnaud, and Rosental Paul-André, 2004. *Familial Components of First Migration after Marriage in nineteenth century France*, "New History of Kinship" Conference, UIESP-INED-EHESS, Paris, October 1–2, 2004, mimeo.

Bourdelais, Patrice, 1993. *Le nouvel âge de la vieillesse. Histoire du vieillissement de la population*. Paris, Odile Jacob, 441 p.

Bourdieu Jérôme, Postel-Vinay Gilles, Rosental Paul-André and Suwa-Eisenmann Akiko, 2000. Migrations et transmissions inter-générationnelles dans la France du XIX[e] et du début du XX[e] siècle, *Annales. Histoire, Sciences Sociales*, vol. 55(4), p. 749–790.

Bourgeois-Pichat Jean, 1951. Evolution générale de la population française depuis le XVIII[e] siècle, *Population*, vol. 6(4), p. 635–662.

Brändström, Anders, 1997. Kinship and Mortality: Illegitimate Children in Sundsvall, Sweden, 1800–1890, Social Science History Association, Washington, DC, October 16–19.

Brass William, 1971. On the scale of mortality, in: William Brass (ed.), *Biological aspects of demography*, p. 69–110. London, Taylor & Francis (Symposia of the Society for the Study of Human Biology, vol. X), 167 p.

Breschi Marco, 1990. *La popolazione della Toscana dal 1640 al 1940. Un'ipotesi di ricostruzione.* Florence, Dipartimento Statistico, Università degli Studi di Firenze, 283 p.

Breschi Marco, Derosas Renzo and Manfredini Matteo, 1997. *Infant mortality in historical Italy: interaction between ecology and society.* Udine, Dipartimento di Scienze statistiche, 23 p. (Paper given at Annual Meeting of the Population Association of America, Washington D.C. March 27–29, 1997)

Brian Éric, 2001. Nouvel essai pour connaître la population du royaume: histoire des sciences, calcul des probabilités et population de la France vers 1780, *Annales de démographie historique*, no. 2, p. 173–222.

Brian Éric and Jaisson Marie, 2005. Le critère du fait social. Retour sur le taux de masculinité à la naissance, in: Maurice Halbwachs, Alfred Sauvy et al. (eds.), *Le point de vue du Nombre*, p. 169–197.

[46] On this point, see Courgeau's presentation (2004) of multilevel approaches, which are specifically designed to determine the causal share of each scale—and hence of each institutional construct, including kinship.

Paris, Éditions de l'INED, reprint, Marie Jaisson and Éric Brian (eds.), 470 p., 1st. ed. 1936.

Caselli Graziella, 1995. Parcours et étapes des études sur la mortalité dans la dynamique de la recherche démographique, *Population*, vol. 50(6), p. 1509–1544.

Caselli Graziella and Capocaccia Riccardo, 1989. Age, period, cohort and early mortality: an analysis of adult mortality in Italy, *Population Studies*, vol. 43(1), p. 133–153.

Castiglioni Maria, Dalla Zuanna Giampiero, and La Mendola Salvatore, 1991. Differenze di fecondità tra i distretti del Veneto attorno al 1881. Analisi descrittiva e ipotesi intepretative, in: Fiorenzo Rossi (ed.), *La transizione demografica nel Veneto. Alcuni spunti di ricerca*, p. 73–149. Venice, Fondazione G. Corazzin Editrice, 196 p.

Cavalli-Sforza Luigi Luca, Menozzi Paolo and Piazza Alberto, 1994. *The History and Geography of Human Genes*, Princeton (New Jersey), Princeton University Press, 1088 p.

Charbonneau Hubert, Desjardins Bertrand, Guillemette André, Landry Yves, Legaré Jacques and Nault François, 1987. *Naissance d'une population. Les Français établis au Canada au XVII^e siècle.* Paris and Montreal, Institut National d'Études Démographiques and Presses de l'Université de Montréal, 232 p.

Coale Ansley J., 1972. *The growth and structure of human populations: a mathematical investigation.* Princeton (New Jersey), Princeton University Press, 227 p.

Coale Ansley J., Demeny Paul and Vaughan Barbara, 1983. *Regional model life tables and stable populations*. New York and London, Academic Press, 496 p. (2nd ed.)

Courgeau Daniel and Najim Jamal, 1995. Analyse de biographies fragmentaires, *Population*, vol. 50(1), p. 149–168.

Courgeau Daniel and Lelièvre Eva, 2001. L'analyse démographique des biographies, in: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Démographie: analyse et synthèse, Volume 1, La dynamique des populations*, Paris, INED, chap. 23, p. 503–517.

Courgeau Daniel, 2004. *Du groupe à l'individu: synthèse multiniveau*, Paris, INED, 242 p.

Del Panta Lorenzo and Rettaroli Rosella, 1994. *Introduzione alla demografia storica.* Bari and Rome, Editori Laterza, 314 p.

Derosas Renzo, 2002. Fatherless families in nineteenth-century Venice, in: Renzo Derosas and Michel Oris (eds.), *When dad died: individuals and families coping with distress in past societies*, p. 433–64. Bern: Peter Lang, 502 p.

Diamond Ian, Davies Rhodri and Egger Peter, 1993. Some applications of recent developments in event-history analysis for historical demography, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 223–238. Oxford, Clarendon Press, 426 p. (International Studies in Demography).

Dupâquier Jacques, 1978. Réflexion sur la mortalité du passé: mesure de la mortalité des adultes d'après les fiches de famille, *Annales de démographie historique*, p. 31–48.

Dupâquier Jacques, 1984. *Pour la démographie historique.* Paris, Presses universitaires de France, 188 p.

Eberst Dorsten Linda, Hotchkiss Lawrence and King Terri M., 1999. The effect of inbreeding on early childhood mortality: twelve generations of an Amish settlement, *Demography*, vol. 36(2), p. 263–271.

Entorf Horst and Zimmermann Klaus, 1990. Interrelationships between mortality and fertility in Germany: rural and urban Prussia and modern Germany, *Genus*, vol. 46(1–2), p. 133–146.

Fleury Michel and Henry Louis, 1976. *Nouveau manuel de dépouillement et d'exploitation de l'état civil ancien.* Paris, Éditions de l'Institut National d'Études Démographiques, 182 p. (2nd ed.)

Galloway Patrick, 1994. Secular changes in the short-term preventive, positive, and temperature checks to population growth in Europe, 1460 to 1909, *Climatic Change*, vol. 26, p. 3–63.

Gautier Étienne and Henry Louis, 1958. *La population de Crulai, paroisse normande*. Paris, Presses universitaires de France ("Travaux et documents" series, Cahier no. 33), 272 p.

Gavrilova Natalia S., Gavrilov Leonid A., Evdokushkina Galina N., and Semyonova Victoria G, 2003. Early-life predictors of human longevity: analysis of the XIX^th century birth cohorts, *Annales de démographie historique*, no. 2, p. 177–198.

Ge Rondi Carla, 1988. *L'analisi nominativa in demografia storica: metodi e problemi. Il caso di una parrocchia.* Milan, Giuffrè, 122 p.

Gutmann Myron and Alter George, 1993. Family reconstitution as event-history analysis, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 159–177. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Gutmann Myron and Watkins Susan C., 1990. Socio-economic differences in fertility control: is there an early warning system at the village level? *European Journal of Population/Revue européenne de démographie*, vol. 6, p. 69–101.

Henry, Louis, 1956. Anciennes Families Genevoises. Étude Démographique: XV^e–XX^e siècles. Paris, INED-Presses Universitaires de France.

Henry Louis, 1967. *Manuel de démographie historique.* Geneva and Paris, Droz, 146 p.

Henry Louis, 1972a. Fécondité des mariages dans le quart sud-ouest de la France de 1720 à 1829, *Annales ESC*, vol. 27(3), p. 612–640.

Henry Louis, 1972b. Fécondité des mariages dans le quart sud-ouest de la France de 1720 à 1829, *Annales ESC*, vol. 27(4–5), p. 977–1023.

Henry Louis, 1976. Étude de la mortalité à partir de la reconstitution des familles, *Bulletin d'information de la Société de démographie historique*, no. 18, p. 4–20.

Henry Louis, 1978. Fécondité des mariages dans le quart sud-est de la France de 1670 à 1829, *Population*, vol. 33(4–5), p. 855–83.

Henry Louis, 1980. *Techniques d'analyse en démographie historique.* Paris, Éditions de l'Institut National d'Études Démographiques, 177 p.

Henry Louis and Blum Alain, 1988. *Techniques d'analyse en démographie historique.* Paris, Éditions de l'Institut National d'Études Démographiques, 180 p. (2nd ed.)

Henry Louis and Houdaille Jacques, 1973. Fécondité des mariages dans le quart nord-ouest de la France de 1670 à 1829, *Population*, vol. 28(4–5), p. 873–922.

Hobcraft John and Murphy Michael, 1986. Demographic event history analysis: a selective review, *Population Index*, vol. 52(1), p. 3–27.

Hollingsworth T. H., 1957. A demographic study of the British Ducal families, *Population Studies*, vol. 11(1), p. 4–26.

Houdaille Jacques, 1976. Fécondité des mariages de 1670 à 1829 dans le quart nord-est de la France, *Annales de démographie historique*, p. 341–391.

Hudson Pat and King Steve, 2000. Two textile townships: a comparative demographic analysis, *Economic History Review*, vol. 54(4), p. 396–429.

Jacquard Albert (ed), 1970. *Etude des isolats. Espoirs et limites.* Paris, INED, 334 p.

Kertzer David I. and Hogan Dennis P., 1989. *Family, political economy and demographic change.* Madison, University of Wisconsin Press, 270 p.

Kesztenbaum Lionel, 2004. Places of life events as bequestable wealth. Familial territory and migration in France, XIX^th and XX^th centuries, "New History of Kinship" Conference, UIESP-INED-

EHESS, Paris, October 1–2, 2004, mimeo.

King Steve, 1997. Dying with style: infant death and its context in a rural industrial community, *Social History of Medicine*, vol. 10, p. 3–24.

King Steve, 2005. Back to Malthus? The existence and characteristics of demographic sub-groups in English towns and villages 1650–1837, mimeo.

King Steve, 2006. La pauvreté, la protection sociale et la politique locale de la démographie dans l'Angleterre du XIX[e] siècle, *Annales. Histoire, Sciences Sociales*, in press.

Lagazio Corrado and Pagani Laura, 1996. L'applicazione della "event history analysis" a studi di carattere socio-demografico, *Bollettino di Demografia Storica*, no. 24–25, p. 97–112.

Laslett Peter, 1980. The bastardy-prone sub-society, in: Peter Laslett, Karla Oosterveen, and Richard Smith (eds.), *Bastardy and its comparative history: studies in the history of illegitimacy and marital nonconformism in Britain, France, Germany, Sweden, North America, Jamaica and Japan*, p. 217–240. London, Edward Arnold, 431 p.

Laslett Peter, 1988. Family, kinship and collectivity as systems of support in pre-industrial Europe, *Continuity and Change*, vol. 3, p. 153–175.

Le Bras Hervé, 1973. Parents, grands-parents, bisaïeux, *Population*, vol. 28(1), p. 9–37.

Ledermann Sully, 1969. *Nouvelles tables-types de mortalité.* Paris, INED and Presses Universitaires de France ("Travaux et documents" series, Cahier no. 53), 260 p.

Lee James and Campbell Cameron, 1997. *Fate and fortune in rural China: social organization and population behavior in Liaoning, 1774–1873.* Cambridge, Cambridge University Press, 280 p.

Lee James, Feng Wang and Campbell Cameron, 1994. Infant and child mortality among the Qing nobility: Implications for two types of positive check, *Population Studies*, vol. 48(3), p. 395–411.

Lee Ronald D., 1974. Estimating series of vital rates and age structures from baptisms and burials: a new technique, with applications to pre-Industrial England, *Population Studies*, vol. 28(3), p. 495–512.

Lee Ronald D., 1981. Short-term variation: vital rates, prices and weather, in: Edward A. Wrigley and Roger S. Schofield, *The population history of England 1541–1871*, p. 356–401. London, Edward Arnold, 781 p.

Lee Ronald D., 1985. Inverse projection and back projection: a critical appraisal, and comparative results for England, 1539 to 1871, *Population Studies*, vol. 39(2), p. 233–248.

Lee Ronald D., 1993. Inverse projection and demographic fluctuations, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 7–28. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Lee Ronald D., 2004. Reflections on inverse projection: its origins, development, extensions, and relation to forecasting, in: Elisabetta Barbi, Salvatore Bertino, and Eugenio Sonnino (eds.), *Inverse projection techniques.*, op. cit., p. 1–9.

Lemercier Claire, 2005. Analyse de réseaux et histoire de la famille: une rencontre encore à venir?, *Annales de démographie historique*, no. 2 (in press).

Lesthaeghe Ron, 1992. Motivation et légitimation: conditions de vie et régimes de la fécondité en Belgique et en France du XVI[e] au XVIII[e] siècle, in: Alain Blum, Noël Bonneuil, and Didier Blanchet (eds.), *Modèles de la démographie historique*, p. 275–317. Paris, Institut National d'Études Démographiques ("Congrès et colloques" series, no. 11), 370 p.

Livi Bacci Massimo and Reher David, 1993. Other paths to the past: from vital series to population patterns, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 66–83. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Lynch Katherine A. and Greenhouse Joel B., 1994. Risk Factors for Infant Mortality in Nineteenth-Century Sweden, *Population Studies*, vol. 48(1), p. 117–33.

Manfredini Matteo, 1996. L'utilizzo degli status animarum nelle ricostruzioni nominative: miglioramenti informatici, qualitativi e quantitativi. Il caso di Madregolo (1629–1914), *Bollettino Demografia Storica*, no. 24–25, p. 113–30.

Martelli Cristina, 1997. I modelli età-periodo-coorte in demografia: applicazioni prospettive e problemi, in: Piero Giorgi and Salvatore Strozza (eds.), *Studi di popolazione. Temi di ricerca nuova* (*Secondo convegno dei giovani studiosi dei problemi di popolazione, Rome, June 25–27 1996*), p. 677–690. Rome, Università degli Studi di Roma "La Sapienza," Dipartimento di Scienze Demografiche, 727 p.

McCaa Robert, 1993. Benchmarks from a new inverse population projection program: England, Sweden, and a standard demographic transition, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 40–56. Oxford, Clarendon Press (International Studies in Demography), 426 p.

McCaa Robert and Vaupel James, 1992. Comment la projection inverse se comporte-t-elle sur des données simulées?, in: Alain Blum, Noël Bonneuil, and Didier Blanchet (eds.), *Modèles de la démographie historique*, p. 129–146. Paris, Institut National d'Études Démographiques ("Congrès et colloques" series, no. 11), 370 p.

Mroz Thomas A. and Weir David, 1990. Structural change in lifecycle fertility during the fertility transition: France before and after the Revolution of 1789, *Population Studies*, vol. 44(1), p. 61–87.

Munno Cristina, 2005. Prestige, intégration, parentèle: réseaux de parrainage dans une communauté ouvrière vénitienne (1834–1854), *Annales de démographie historique*, no. 2 (in press).

Munno Cristina, 2006. *Réseaux de parenté et dynamiques sociales: essai d'analyse et de mesure. Follina XIX[e] siècle*, in press. (Doctoral dissertation, EHESS, Paris; supervised by Patrice Bourdelais).

Oeppen Jim, 1993a. Back projection and inverse projection: members of a wider class of constrained projection models, *Population Studies*, vol. 47(2), p. 245–267.

Oeppen Jim, 1993b. Generalized inverse projection, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 29–39. Oxford, Clarendon Press (International Studies in Demography), 426 p.

Peng Xizhe and Hou Yangfang, 1996. Demographic and social change in Jiangsu and Zhejiang between 1370 and 1900: a study of the genealogy of the Fan family, *Chinese Journal of Population Science*, vol. 8(4), p. 361–372.

Plakans Andrejs and Wetherell Charles, 2003. Households and kinship networks: the costs and benefits of Contextualization, *Continuity and Change*, vol. 18, p. 49–76.

Post Wendy, van Poppel Frans, van Imhoff Evert and Kruse Ellen 1997. Reconstructing the extended kin-network in the Netherlands with genealogical data: methods, problems, and results, *Population Studies*, vol. 51(3), p. 263–78.

Pozzi Lucia, 1990. *La mortalità per cause nelle province italiane dal 1890 al 1950: differenziazioni geografiche e fattori esplicativi.* Florence, Dipartimento Statistico, Università degli Studi di Firenze, 250 p. (Doctoral dissertation in demography)

Pozzi Lucia, 2000. *La lotta per la vita. Evoluzione e geografia della sopravvivenza in Italia fra '800 e '900*. Udine, Forum, 128 p.

Rallu Jean-Louis, 1992. Note sur l'utilisation des généalogies, in: Alain Blum, Noël Bonneuil, and Didier Blanchet (eds.), *Modèles de la démographie historique*, p. 45–53. Paris, Institut National d'Études Démographiques ("Congrès et colloques" series, no. 11), 370 p.

Reay Barry, 1996. Kinship and the neighbourhood in nineteenth-century rural England: the myth of the autonomous nuclear family, *Journal of Family History*, vol. 21(1), p. 87–104.

Rosental Paul-André, 1999. *Les sentiers invisibles. Espace, familles et migrations dans la France du XIX^e siècle*. Paris, Éditions de l'E.H.E.S.S., 255 p.

Rosental Paul-André, 2003. *L'Intelligence démographique. Sciences et politiques des populations en France* (*1930–1960*). Paris, Odile Jacob, 369 p.

Rosental Paul-André, 2006. Vers une histoire non démographique des populations?, *Annales. Histoire, Sciences Sociales* (in press).

Rosina Alessandro, 1994. Una estensione dell'Inverse Projection con mortalità differenziata per età, in: Alessandro Rosina and Fiorenzo Rossi (eds.), *Ricostruzioni aggregate dei processi evolutivi delle popolazioni*, p. 39–57. Padua, CLEUP Editore, 104 p.

Rosina Alessandro, 1996. *Applicazione automatica dell'inverse projection differenziata* (*passo annuale e quinquennale*). Padua, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, 13 p.

Rosina Alessandro, Etonti Mirto and Billari Francesco, 1997. *La popolazione di Chioggia nel Seicento.* Florence, SIDES. (paper given at Congress on "La popolazione italiana nel Seicento," Florence, November 28–30, 1996).

Rosina Alessandro, Etonti Mirto and Billari Francesco, 1999. La popolazione di Chioggia nel Seicento, in: Società Italiana di Demografia Storica (ed.), *La popolazione italiana nel Seicento*, p. 607–641. Bologna, CLUEB, 796 p.

Ruggles Steven, 1992. Migration, marriage, and mortality: correcting sources of bias in English family reconstitutions, *Population Studies*, vol. 46(3), p. 507–522.

Schiaffino Andrea, 1993. Fecondità e vita coniugale a Bologna nell'ultimo secolo: analisi longitudinale delle discendenze. Il ciclo di vita coniugale, in: Andrea Schiaffino, *Scritti di demografia storica*, p. 161–247. Florence, Centro Editoriale Toscano, 248 p. (Reprint of 1st ed., Università di Bologna, Istituto di Statistica, 1979, Cooperativa Libraria Universitaria Editrice, Bologna).

Séguy, Isabelle *et al.*, 2001. *La population de la France de 1670 à 1829: l'enquête Louis Henry et ses données*. Paris, INED, Classiques de l'économie et de la population. Études & enquêtes historiques, 208 p.

Smith James and Oeppen Jim, 1993. Estimating numbers of kin in historical England using demographic micro-simulation, in: David Reher and Roger Schofield (eds), *Old and New Methods in Historical Demography*, Oxford, Clarendon Press, p. 281–317.

Szreter Simon, 1996. *Fertility, Class and Gender in Britain 1860–1940*. Cambridge, Cambridge University Press, 704 p.

Tonini Giovanni, 1993a. *L'analisi delle serie temporali in demografia secondo l'approccio moderno.* Trento, Istituto di Statistica e Ricerca Operativa, Facoltà di Economia e Commercio, Università degli Studi di Trento, 26 p.

Tonini Giovanni, 1993b. Le applicazione demografiche dei modelli Arima: casi, problemi e prospettive, in: Luciano Ciucci and Filomena Racioppi (eds.), *Studi di popolazione. Nuovi approcci per la descrizione e l'interpretazione*, p. 411–432. Rome, Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," 742 p.

Tremblay Marc, Arsenault Julie and Heyer Évelyne, 2003. Les probabilités de transmission des gènes fondateurs dans cinq populations régionales du Québec, *Population*, vol. 58(3), p. 403–423.

Trussell James and Guinnane Timothy, 1993. Tecniques of event-history analysis, in: David Reher and Roger Schofield (eds.), *Old and new methods in historical demography*, p. 181–205. Oxford, Clarendon Press (International Studies in Demography), 426 p.

van de Walle Étienne, 1974. *The female population of France in the 19^th Century: a reconstruction of 82 départements*. Princeton (New Jersey), Princeton University Press, 483 p.

van Poppel Frans, 2000. Children in one-parent families: survival as an indicator of the role of the parents, *Journal of Family History*, vol. 25(3), p. 269–290.

van Poppel Frans, Jonker Marianne and Mandemakers Kees, 2005. Differential infant and child mortality in three Dutch regions, 1812–1909, *Economic History Review*, vol. 58(2), p. 272–309.

Wachter Kenneth, 1986. Ergodicity and inverse projection, *Population Studies*, vol. 40(2), p. 275–287.

Westendorp Rudi and Kirkwood Thomas, 2001. La transmission héréditaire de la longévité en lignes maternelle et paternelle, *Population*, vol. 56(1–2), p. 253–268. English version, 2001: Maternal and Paternal Lines of Familial Longevity, *Population: An English Selection*, vol. 13(1), p. 223–235.

Wrigley Anthony, Davies Ros, Oeppen Jim and Schofield Roger, 1997. *English population history from family reconstitution 1580–1837.* Cambridge, Cambridge University Press, 657 p.

Wrigley Edward A. and Schofield Roger S., 1981. *The population history of England 1541–1871.* London, Edward Arnold, 781 p.

Yamaguchi Kazuo, 1991. *Event history analysis.* Newbury Park, Sage Publications, Inc., 182 p.

Zeegers Maurice, van Poppel Frans, Vlietinck Robert, Spruijt Liesbeth and Ostrer Harry, 2004. Founder mutations among the Dutch, *European Journal of Human Genetics* vol. 12(7), p. 591–600.

Zhao Zhongwei, 1997. Long-term mortality patterns in Chinese history: evidence from a recorded clan population, *Population Studies*, vol. 51(2), p. 117–127.

Zhao Zhongwei, 2001. Chinese genealogies as a source for demographic research: a further assessment of their reliability and biases, *Population Studies*, vol. 55(2), p. 181–193.

C H A P T E R

127

Indirect Estimation Methods

KENNETH HILL

Department of Population and Family Health Sciences, Johns Hopkins University, Maryland, United States

Direct demographic measures are typically derived from counts of events and corresponding measures of exposure times. The resulting rates and their derivatives will be accurate as long as all events and exposures are reported and correctly classified, for example, by age. In many countries lacking fully developed administrative systems and with populations that include substantial proportions with limited education, accuracy of reporting and classification cannot be assumed. As a general rule, these countries come mainly from the ranks of developing countries, with developed countries enjoying accurate data. However, it should be remembered that being a developed country is neither a necessary nor a sufficient condition for accurate data: death registration in the United States has only been regarded as adequately complete for measurement purposes since the mid 1930s, and births were adjusted for underregistration until the late 1950s; similarly, data on international migration are of dubious quality almost everywhere, developed and developing alike, partly because administrative systems are not well developed in this area.

Indirect estimation techniques have been developed over the last 50 years to provide adequate measures of demographic processes in situations in which direct measures were too inaccurate to be informative. The basic philosophy underlying indirect estimation, conceived and largely developed by William Brass, was to take advantage of the most reliably reported portions of data, and discard the rest. The techniques can be divided into two broad groups. One group uses consistency checks between different measures of the same underlying phenomenon to explore likely accuracy, and then applies simple models of data errors to arrive at plausible corrections. The second group uses robust indicators of an underlying phenomenon, even though the indicator may be affected by other factors. An analytical method is then used to obtain an estimate of the underlying phenomenon itself by allowing for the effects of these other factors on the indicator.

A key characteristic of indirect techniques is the use of simple questions that can be included in data collection exercises conducted by interviewers with limited education and short training. In almost all cases it is possible to include the questions needed in population censuses. Several of the early methods were developed by Brass expressly to make use of information that was already often collected by demographic inquiries, but for which no systematic analytical methods existed for converting the information into standard demographic measures. In the last couple of decades, the need for indirect estimation methods for national measures of two key demographic phenomena, fertility and child mortality, has been greatly reduced by the development of standardized though complex event history questionnaires for use in sample surveys. Indirect estimation methods have remained important, however, in the measurement of adult mortality and international migration, and for small-area estimates of fertility and child mortality for which the small samples necessitated by complex data collection instruments cannot provide adequate disaggregation.

The two types of indirect estimation will be illustrated in the next section. Subsequent sections will then discuss the more widely used methods according to the demographic process measured: fertility, child mortality, adult mortality and migration. Within each section, one or more methods will be discussed and an application illustrated; modifications of the main methods will be discussed briefly.

I. TWO TYPES OF INDIRECT ESTIMATION

Estimation methods can be divided into two groups. An example of each group, both originally developed by Brass, will be given. The first group consists of those methods that can be literally called *indirect*. A data collection exercise collects information on a population characteristic that can be reported on reasonably accurately. The characteristic in question must be strongly influenced by the demographic process of interest, but will also be affected by other factors not of direct interest. Analytical methods are then used to transform the characteristic into a pure demographic measure, by allowing for the influences of other, extraneous factors. The second group of methods consists of those methods that are primarily *consistency checks*. Two (or more) different measures of the same demographic process are compared. Adjustment (if necessary) is made on the basis of prior knowledge of the error processes that would result in the inconsistencies observed. Reliance is placed on the portions of the alternative pieces of information that are expected to be most accurate.

The classic example of an *indirect* method is the procedure for estimating childhood mortality from summary birth history data, in which women report the number of children born alive and the number that were born alive but have since died. These data permit the calculation of the proportion of dead children according to the age group of mother. The proportion of dead children for women of a particular age group is primarily determined by the overall level of childhood mortality, but it is also affected by the age pattern of fertility. In an early-childbearing population, children of women of a particular age have been exposed to the risk of dying for longer than in a late-childbearing population. By allowing for the age pattern of childbearing (through ratios of average parities in successive age groups), proportions dead of children given birth by women classified by standard 5-year age groups can be used to estimate conventional probabilities of dying between birth and specified exact ages of childhood.

A good example of a *consistency check* is what is commonly referred to as the P/F ratio method for measuring fertility. Information is available on two indicators of the same process: the number of children ever born by women classified by age group, and the number of children born in a defined time period (usually 1 year) by age group of mother. These data permit the calculation of the average number of children ever born by women by age group and of age-specific fertility rates. If fertility is not changing and there are no selection effects at work, cumulated age-specific fertility rates will provide measures of lifetime fertility equivalents that can be compared to the recorded levels of lifetime fertility. The reported lifetime fertility is denoted by P for parity, the lifetime fertility equivalents obtained by cumulating age-specific fertility rates are denoted by F. The ratios of P to F, age group by age group, provide indicators of the consistency of the two types of information.

Once the comparison for consistency has been made, there is the possibility for adjustment. If the P and F values are consistent, there is no indication of error and no need for adjustment, but if they are not, a plausible adjustment can be made in some circumstances if there is some prior knowledge about the likely nature of data errors.

II. ESTIMATION METHODS FOR FERTILITY

1. Indirect Estimates

Indirect estimates of fertility are mostly based upon age distributions, and represent different systems of reverse projection. In a closed population, the number of children of a given age represents the survivors of the number of births that occurred that number of years earlier. The primary influence on the number of survivors is the earlier births, and the secondary influence is childhood mortality. If allowance can be made for the deaths, by reverse projection, for example, the number of births can be estimated. If the entire population can also be reverse-projected to the same time point, the birth rate can be estimated, and if the female population of childbearing age can be estimated, the general fertility rate (GFR) can be estimated. Reverse projection in this simple form requires only data on the age distribution of the population and knowledge of mortality rates over the preceding few years.

This approach to fertility estimation dates back many years. In practice, it is not very useful: it only provides estimates of (at best) the GFR, rather than estimates of more standard measures such as the Total Fertility Rate; age distributions are affected by typical age reporting errors (generally reducing the number of children reported as being under age 5 years), introducing a (downward) bias into fertility estimates for recent periods; child mortality (and especially trends in childhood mortality) may not be accurately known; and the age distribution may be affected by migration.

The *Own Children* method (Cho *et al*., 1986) is a variant of the reverse-projection methodology that provides standard age-specific fertility rates. Information within a data source (such as relationship to head of household) is used to match children in a particular household to a likely mother in the same household. The information on age of child provides information on when the births occurred, while information on age of mother minus age of child provides an indication of the age of the mother at the time of the birth, and thus a basis for estimating age-specific fertility. The method has been widely used with census data, particularly for East and South East Asia, where age distributions are of good quality.

The use of stable and quasistable population analysis techniques (Coale and Demeny, 1968) can also be thought of as a form of reverse projection. Standard age distributions reflecting known, fixed fertility and mortality schedules are used as reference patterns against which an observed age distribution is compared. Some single characteristic of the age distribution, such as the proportion under age 15, which is believed to be fairly accurately observed even in situations in which age reporting is of poor quality, is generally used as the starting point. Adjustment for mortality is then made either directly, if an estimate of child mortality is available, by matching the proportion under age 15 against stable populations which have the same childhood mortality as that of the population being studied, or indirectly, if no estimate of childhood mortality is available, by matching against stable populations with the same rate of growth as the population being studied.

The development of mathematical expressions to represent unstable age distributions gave rise to a method of reverse projection proposed by Preston (1983). This method is appealing because it does not require estimates of mortality *per se*, but rather an estimate of the general age pattern of mortality. It is also appealing because it makes use of the entire age distribution, and linear methods for fitting. In an unstable but closed population, the proportion of the population of a particular age, $c(a)$, is a function of the current birth rate, b, the current probability of surviving to age a, $p(a)$, and the sum of age-specific growth rates below a, $R(a)$. Thus

$$c(a) = b\ p(a)\ exp(-R(a))$$

where

$$R(a) = \int_0^a r(x)dx$$

Preston expressed the odds ratio of dying by age a as equal to a constant fraction of the odds ratio of dying by age a in a standard mortality schedule or model life table. Thus

$$(1 - p(a))/p(a) = g\ \{(1 - p^s(a))/p^s(a)\}$$

Substituting in for $p(a)$ and rearranging gives the relationship

$$1/\{c(a)\ exp(R(a))\} = 1/b + (g/b)\{(1 - p^s(a))/p^s(a)\}$$

Thus the reciprocal of the proportion age a adjusted for the sum of growth rates from age 0 to age a is linearly related to the odds ratio of dying by age a in the standard life table, with intercept equal to the reciprocal of the birth rate and slope equal to the ratio of the proportional factor linking the actual and model odds ratios to the birth rate. Even better, if child mortality is known, from some independent source, the expression can be rewritten

$$p(5)/\{c(a)\ exp(R(a))\} = 1/b + (g/b)\ \{(p^s(5) - p^s(a))/p^s(a)\}$$

where $p(5)$ and $p^s(5)$ are the actual and standard probabilities of surviving to age 5. This formulation is to be preferred because much of the variability between human mortality schedules and between different families of model life tables is the relationship between mortality before and after age 5.

Although the Preston method appears to make very effective use of both data and theory in a combination that should be robust to error, in practice the method has performed rather poorly. At best, the only fertility estimate it produces is of the birth rate, and this estimate turns out to be sensitive to selection of points to fit. Age misreporting is the most likely explanation of the problem.

2. Consistency Checks

The principal consistency check method in fertility estimation is the P/F method outlined earlier, whereby fertility rates for a recent period (for example from vital registration) are compared with lifetime fertility measures (for example from a census question on children ever born). Age-specific fertility rates cumulated across age groups provide cumulative fertility measures $F(a)$ that are equivalent to lifetime fertility measures at exact ages. For data tabulated by 5-year age groups, the cumulative fertility to the beginning of a given age group is just the sum of age-specific fertility rates for younger age groups multiplied by 5 (to allow for the fact that each woman spends 5 years in a 5-year age group). Lifetime fertility measures *P*(*i*), however, are averages for women within a 5-year age group. A method is therefore needed to interpolate between the cumulative fertility measures at exact ages to approximate lifetime fertility measures for an age group. Various ways of carrying out this interpolation have been developed. The original method, proposed by Brass (1975), used a very simple fertility model (a third degree polynomial) to interpolate between cumulative fertility to the beginning of an age group and cumulative fertility at the beginning of the next age group in order to arrive at an estimate of average cumulative fertility *F*(*i*) in the age group. The polynomial was fitted using either the ratio of fertility in the age group 15 to 19 years to that in the age group of 20 to 24 years, or the mean age of childbearing.

Once the *P* and *F* values for each age group have been calculated, *P*/*F* ratios can be calculated for each age group. The ratios indicate the consistency of the current and lifetime fertility information, but may, if certain assumptions about data errors can be made, also provide a basis for adjustment. Say that the data on children ever born come from a census question. Experience suggests that in some cultures older women may omit some of the children they have given birth to, if for example the children have died or moved away. In these circumstances, the P values for older women will be biased downwards, but those for younger women will be more accurate. If the data on births by age of mother come from a vital registration system that only covers 80% of the national population, there will be too few births registered, but the undercoverage of births may be roughly constant at all ages of mother. If so, the age-specific fertility rates and the *F* values derived from them will all be too small by a factor of 0.8. Under these circumstances, the *P*/*F* ratios for younger women should be an approximately constant 1.25, reflecting the underregistration of births and reasonable accuracy of the *P*s, but should drop for older women as the *P*s become biased downwards. An estimate of fertility better than either source on its own can then be obtained by adjusting the age-specific fertility rates by the *P*/*F* ratios for younger women. In practice, of course, the assumptions will only approximate reality, or may not even do that if fertility is changing rapidly, so adjustment may not be possible. The consistency check and the pattern of *P*/*F* ratios by age will always be informative, however.

The original method was slightly modified in the United Nations (Hill *et al*., 1983) *Manual X*. The interpolation procedure between successive cumulative fertility observations was based on a regression equation. Model fertility schedules were generated by the Coale-Trussell (1974) fertility model for a range of model parameters and combined with likely age distributions of women within 5-year age groups given different levels of fertility. Table 127–1 shows an application of the method to data from the 1992 census of Zimbabwe. The *P*/*F* ratios (column 5) rise steeply with age of mother, from below unity in the first two age groups to nearly 1.5 for the women over 40. Clearly, the two sources of fertility data are far from consistent, but as they are the ratios give little clue as to which source of data should be regarded as more accurate. Knowledge of the specific situation suggests that changing fertility may explain much of the pattern.

A further modification of the interpolation procedure is proposed by Brass (1981), making use of his Relational Gompertz fertility model, which is particu-

TABLE 127–1 Application of P/F Ratio Method. Zimbabwe, 1987 and 1992

Age Group	Age-Specific Fertility Rate (1992) f(i)	Average No. of Children Born (1992) P(i)	Parity Equivalents From f(i)s F(i)	P(i)/F(i)	Average No. of Children Born (1987) P′(i)	Hypothetical Cohort (1987–1992) $P^h(i)$	$P^h(i)/F(i)$
15–19	0.099	0.189	0.220	0.86	0.170	0.189	0.86
20–24	0.223	1.119	1.158	0.97	1.311	1.138	0.98
25–29	0.204	2.537	2.234	1.14	2.983	2.364	1.06
30–34	0.177	4.021	3.174	1.27	4.533	3.402	1.07
35–39	0.141	5.278	3.957	1.33	5.913	4.147	1.05
40–44	0.074	6.262	4.441	1.41	6.770	4.496	1.01
45–49	0.026	6.738	4.690	1.44	7.262	4.464	0.95
Total average fertility rate	4.72						

Source: United Nations (Hill *et al.*, 1983) Manual X.

larly convenient for fitting to lifetime fertility data or their equivalents. The Relational Gompertz model can be fitted to the cumulated age-specific fertility rates to age a as a proportion of total fertility, $F(a)/TF$, or to ratios of successive cumulative values, $F(a)/F(a + 5)$, and the model parameters estimated. The model parameters can then be used with the standard average parity distribution values of the model, $P^s(i)/TF$, to estimate the expected average parity in each age group as a proportion of total fertility, $P^e(i)/TF$. An estimate of total fertility can then be obtained for each age group i as the ratio of observed parity, $P(i)$, to the expected proportion of total fertility represented by average parity at that age group, based on the model fitted to the *Fs*.

The exact method used to apply the *P/F* Ratio method makes little difference to the results obtained, since all that the methods are doing is using different ways of interpolating between observed cumulative fertility levels. The factors that do make a difference to the results are fertility trends and data errors that do not fit the very simple models specified. Results for women in the age group 15–19 are often unreliable, perhaps because of more complex error patterns affecting very young women. Fertility trends result in *P/F* ratios that deviate from unity, as noted above with regard to the Zimbabwe example, but a basis for adjustment can sometimes be obtained if data are available on children ever born from two successive surveys. If the two surveys are separated in time by approximately 5 years, the parity distribution for a hypothetical intersurvey cohort can be created by chaining together the parity increments for true cohorts aging from one survey to the next. This parity distribution for the hypothetical cohort, with an explicit time reference, can then be compared with the cumulated age-specific fertility rates from the second survey. Table 127–1 (columns 6 to 8) shows the application of this technique to the Zimbabwe data, combining the 1992 census data with data from a 1987 intercensus survey. The hypothetical cohort *P/F* ratios still start below unity for the youngest women, rise moderately to values slightly over 1.0 for the central ages, and then fall again. This application confirms that fertility trends were largely responsible for the pattern of *P/F* ratios in column 5, and also suggests that young women may underreport children ever born, while women over 40 may also omit some of their lifetime births (note that the average parity of 45- to 49-year-old women in 1992 is *lower* than that of 40- to 44-year-old women in 1987). The application suggests that an upward adjustment of about 5 percent to the reported age specific fertility rates would be appropriate.

The *P/F* ratio methodology can also be applied to data specific by birth order. For example, cumulated first birth rates can be compared with proportions of women who report having had at least one child. The assumptions about error are then that women are less likely to underreport ever having had a child than underreporting the number, and that first births are as well reported as all births. In practice, these assumptions about error are rather strong, and the order specific methodology has not been widely used.

3. Combined Indirect Estimates and Consistency Checks

Parity progression ratios are useful for studying long-term trends in family building behavior, and the techniques of indirect estimation and consistency check can be used in the process. Distributions of women after the end of childbearing by number of children ever born permit the calculation of parity progression ratios, and performing the calculations by age group provides an approximate time reference for the measures of fertility. Such ratios may be affected by age misreporting and omission of children from women's reports, but consistency checks of the measures for similar time points from successive surveys can demonstrate whether or not data errors are important (Feeney, 1988). Indirect estimation can then be invoked to fill in the incomplete fertility records of women who have not yet reached the end of childbearing in order to use these incomplete records as a basis for calculating parity progression ratios. Current age-parity–specific fertility rates are used to project future births of each parity for each age group of women, in order to fill in the truncated birth histories. This procedure will of course damp trends, because current levels are being assumed to apply to future childbearing of all cohorts.

III. ESTIMATION METHODS FOR CHILD MORTALITY

1. Indirect Estimates

Indirect estimates of child mortality, using data from population censuses and large-scale surveys, have contributed enormously to our knowledge of child mortality levels and trends in the developing world. The recent widespread use of birth history surveys, in programs such as the World Fertility Survey and the Demographic and Health Surveys, has provided alternative, direct estimates of child mortality at the national level in a large number of countries, and has also contributed to our knowledge of the factors associated with child mortality differentials. However, the indirect estimates included in censuses still provide the only viable way of examining small-area variations in child mortality in much of the developing world.

a. Method Based on Proportions of Children Dead

The method is based on proportions dead of children ever born by women, as classified by age. The basic questions are very simple: *How many children have you given birth to?* and *How many of them are still alive?* It is claimed that better results can be obtained by breaking the question down into three: children still alive and living at home, children still alive but living elsewhere, and children that have died. The questions can be further refined by asking about boys and girls separately. The analysis uses age of mother as a proxy for exposure to risk of the children. Other things being equal, older women have children exposed to the risk of dying for longer, and thus have higher proportions of dead children for a given level of childhood mortality, than do younger women. The proportion dead within a given age group is also affected by the age pattern of fertility. In an early childbearing population, average exposure time of the children of mothers of a given age group will be longer than in a late childbearing population, so the proportion dead will be higher for the same overall level of mortality. For simplicity, assume that women in a population bear all their children at one age. For women of exact age 30 who bore all their children at exact age 23 years, the proportion dead will measure (other things being equal) the probability of dying by age 7 years, ${}_7q_0$. For women of exact age 30 years who bore all their children at exact age 25 years, the proportion dead would measure the probability of dying by age 5, ${}_5q_0$.

Brass (1964) developed the first analytical method for such data by modelling proportions dead for standard 5-year age groups using a single model life table and a simple polynomial fertility model with varying age locations. He related the model proportions dead, age group by age group, to probabilities of dying by exact ages of childhood. A particular application used an indicator of the age location of childbearing, the ratio of the average numbers of children given birth to women in the age groups 20 to 24 years and 25 to 29 years, $P(2)/P(3)$, to select appropriate factors for converting proportion of dead children into probabilities of dying between birth and exact ages of childhood.

The method has been subject to a number of modifications over the years. Sullivan (1972) used a wider range of model life tables to model proportions dead, and proposed a regression-based method to obtain conversion factors based on observed values of $P(2)/P(3)$. Trussell (1975) proposed a set of regression equations based on a range of fertility models derived from the Coale-Trussell (1974) fertility model and a wide range of model life tables. Preston and Palloni (1977) proposed a method for using the ages of each mother's surviving children to infer the distribution of children by duration of exposure. The United Nations *Manual X* (Hill *et al*., 1983) presented a rather simplified version of the Trussell procedure.

The precise method of allowing for exposure time makes little difference to the results obtained relative to other data errors and to assumptions about another factor that affects the relationship between the proportion dead and probabilities of dying by exact ages of childhood, namely the age pattern of child mortality. Both model simulations and extensive experience have shown that that the probability of dying by age 5, ${}_5q_0$, is robustly estimated by the proportion dead among children given birth by women aged 30 to 34 years, over a wide range of age patterns of early child mortality. However, infant mortality (based on proportions dead among women aged 15 to 19 years) and the probability of dying by age 2 years (based on the proportion dead among children ever born by women aged 20 to 24 years) are not robustly estimated, partly because they are sensitive to the assumed pattern of childhood mortality and partly because of selection biases.

The second major breakthrough in the use of data on children ever born and surviving was the recognition that estimates of the time reference of the estimate based on each age group of women could be derived if the rate of change of mortality could be assumed to be approximately constant. Feeney (1980) developed a method for using the proportion dead for each age group of women to derive an estimate of the infant mortality rate and the years to which it referred. The basic idea is that both the average exposure time of children and the average time ago to which that exposure refers increase with age of mother. Under conditions of constant child mortality change, the pace of change does not materially affect the time to which each estimate refers. The United Nations *Manual X* provides conversion factors based on parity ratios for estimating the reference dates of successive estimates. One of the important results of being able to locate estimates in time is that results from different surveys can be compared for consistency. Each estimate of ${}_nq_0$ can be converted into a common indicator (${}_5q_0$ is normally

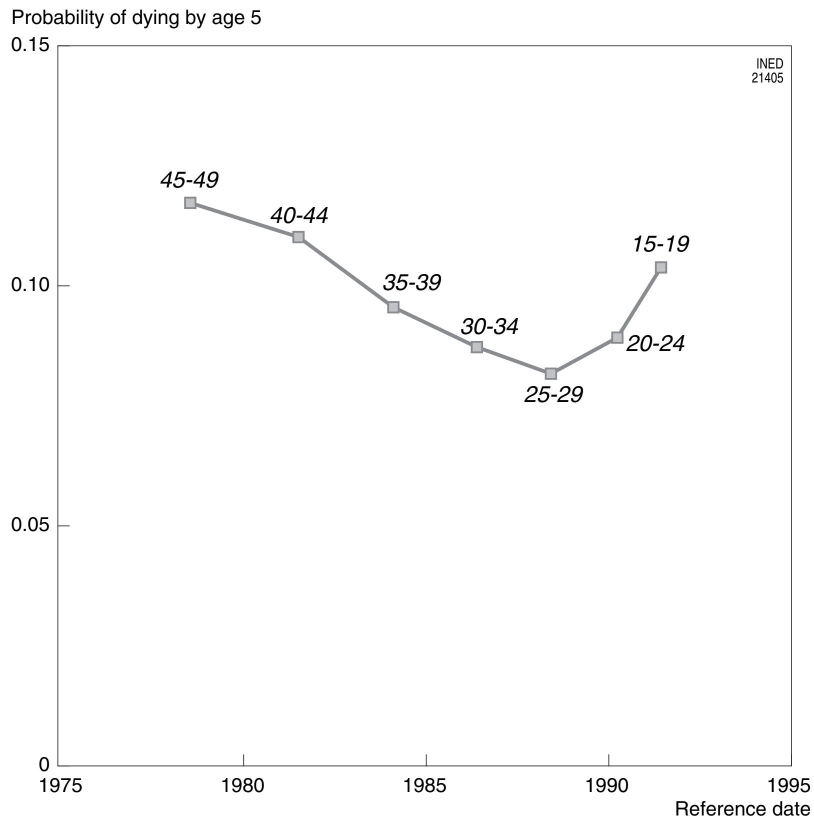


FIGURE 127–1 Indirect estimation of child mortality: Zimbabwe, 1992. Version applied: United Nations (Hill *et al*., 1983) *Manual X* using Coale-Demeny (1983) *West* Family model life tables.

used) using a system of model life tables. The estimates of $_5q_0$ can then be plotted against time and the consistency of successive surveys assessed. Results of the application of the *Manual X* method to data from the 1992 Zimbabwe census are shown graphically in Figure 127–1.

Further developments of the method have explored alternative controls for exposure to risk. One of the problems with the age based method is that early child bearers tend to be selected for low socioeconomic status and thus higher than average child mortality. This effect is clearly visible in Figure 127–1. Thus estimates based on proportions dead of women under age 25 tend to be above the population average. Sullivan (1972), Trussell (1975) and the United Nations *Manual X* provide factors for converting proportions dead by 5-year duration of marriage groups into standard risks of dying in childhood. The advantage of this approach is that the estimate for the first duration group, which gives an estimate with a reference period closest to the survey date, avoids the selection bias that affects the young age groups, since all socioeconomic groups are represented by children in the first duration of marriage group. The problem is that age at marriage may not be closely related to age at initiation of childbearing, particularly in populations with significant levels of premarital fertility. To avoid this problem, Hill and Figueroa (2001) have proposed a method based on time since first birth, but the method has not yet been tested widely enough to draw conclusions about its usefulness.

b. The Previous Birth Technique

A second indirect approach to the estimation of child mortality is the previous birth technique (Brass and Macrae, 1984). This approach is intended for application in settings in which registration or notification of births occurs, but recording of child deaths may be incomplete. At the time of a delivery, the mother-to-be is asked whether she has had a previous birth, and, if so, whether the previous birth is still alive. The proportion dead of the previous births closely approximates the probability of dying by age 2, $_2q_0$, approximately two years earlier. The simple questions can be added to a registration system (if a high proportion of births are registered) or to clinic record-keeping systems (if a high proportion of births occur in clinics). Even if proportions of births covered by the record-keeping system are not very high, it may still be argued that trends in ${}_2q_0$ will approximate population trends. However, concerns about selection bias or changing compositions of births recorded over time (particularly if the data are to be used for evaluating a program that aims to change that composition) may weaken possible interpretation.

2. Consistency Checks

Consistency checks have not played an important role in the development of measurement methods for child mortality (except for comparisons of consistency of alternative indirect estimates for similar time periods, noted above). Although in theory it is possible to compare a time sequence of births and child mortality estimates with the surviving population of young ages, in practice misreporting of ages of young children and uncertainties about true fertility levels make it impossible to draw conclusions about the child mortality element.

IV. ESTIMATION METHODS FOR ADULT MORTALITY

Large numbers of methods have been developed for measuring adult mortality, but the performance of the methods is often less satisfactory than in the areas of fertility and child mortality. As a result, firm information about current patterns, levels and trends of adult mortality is lacking for much of the developing world. One of the main reasons for the uncertain performance of the available methods is age misreporting. Exaggeration of age is an important error in many developing countries, and it affects all the available estimation methods to a greater or lesser extent.

1. Indirect Estimates

Indirect estimates are based on data on survival of relatives. Methods have been developed for three types of relative: parents, spouses and siblings. However, data on parental survival (and particularly survival of mother) are much more widely available than data on spouses or siblings, and more effort has been put into the development of analytical methods for parental survival than into methods for other types of relative. The presentation here therefore will focus mainly on parental survival methods.

a. Parental Survival Method

The basic questions needed to apply the parental survival methods are very simple: *Is your father alive?* and *Is your mother alive?*. Given that the respondent's mother must have been alive at the respondent's birth (and the father at the respondent's conception), the respondent's age provides a measure of the duration of exposure to risk of the parent. If all births occurred to mothers at exact age 25 years, the proportion of mothers alive among respondents aged exactly 20 years would approximate the probability of surviving between ages 25 and 45, $l(45)/l(25)$. If all fathers were exact age 30 years at the conception of their children, the proportion of fathers alive would approximate $l(50.75)/l(30)$, assuming the gestation period to be three quarters of a year. In practice, of course, women bear children at a range of ages from about 15 to 50 (and men beget children at an even wider range of ages). The proportion of those exact age 20 with a surviving mother (or father) is thus a weighted average of a range of 20-year survival probabilities.

The proportion of respondents in a given age group with a surviving parent depends primarily on the level of adult mortality, but also depends on the age pattern of childbearing. The analytical methods for parental survival data (Brass and Hill, 1973; Hill and Trussell, 1977; Timaeus, 1992) have all used model fertility and mortality schedules to model relationships between proportions with a surviving mother (or father) in a given age group and a standard life table conditional survivorship probability. The method is applied by using a standard indicator of the age location of childbearing (the average age of parent for a cohort of births) to control for differences in proportions with surviving parents that result from differences in age patterns of childbearing in order to convert each proportion into a life table survivorship probability. Thus in the United Nations *Manual X* version of the survival of mother method,

$$l^f(25 + N)/l^f(25) = a(N) + b(N)\,M + c(N)\,{}_5PSM_{N-5}$$

where $l^f(25 + N)/l^f(25)$ is the probability of female survival from age 25 to age 25 + N, M is the mean age of mother for a cohort of births (usually for data availability reasons a recent cohort, for example births occurring during the last 12 months), ${}_5PSM_{N-5}$ is the proportion of respondents in the age group (N – 5, N) with a surviving mother, and $a(N)$, $b(N)$ and $c(N)$ are constants estimated by regression analysis of model simulations.

The survivorship indicator estimated from parental survival is not a current estimate but rather an estimate of an average of mortality conditions over the exposure period (determined by the age group of respondents). Brass and Bamgboye (1981) have proposed a simple method for deriving an estimate of time location in the past of each estimate. Because mortality risks increase with age, the time reference is more recent than the midpoint of the exposure period, but it is still the case that estimates of female mortality based on survival of mother data for respondents aged between 15 and 50 years generally refer to time periods from 7 to 15 years before the interviews. As with indirect estimates of child mortality, the ability to locate adult mortality estimates in time makes it possible to compare consistency of estimates between surveys as well as track trends in mortality.

Parental survivorship methods have been widely applied, but with mixed results (for a review, see Timaeus, 1991a). There are some theoretical reasons for cautious interpretation. It is only persons with surviving children whose survival gets reported, and the survival is then reported in proportion to the number of surviving children; any correlation between survivorship and numbers of surviving children (whether through a fertility or mortality differential) will mean that parental survival proportions are not representative of the population at large. An attempt to limit responses to one per parent by restricting analysis to firstborn or eldest surviving children was not successful (Hill, 1977). There are also well-established data accuracy problems, most notably age misreporting (exaggeration of age of respondent increases apparent survival time and biases mortality estimates downwards) and what has been dubbed the *adoption* effect (young children with a dead parent appear to be reported as for a surviving adoptive parent, biasing upwards proportions with a surviving parent). The adoption effect is particularly important for respondents under 15, but age exaggeration may become a problem at almost any age thereafter. Finally, adults with high fertility are overrepresented.

A modification of the basic technique has been developed by Timaeus (1991b) to avoid problems with adoption. In some surveys, data have been collected on parental survival at the time of the survey as well as at some point in the past such as first marriage. The proportions with a surviving parent at first marriage whose parent has died by the time of the survey reflect parental mortality after the respondent has passed the stage of life during which adoption is possible, thus avoiding the adoption effect, and also reflect quite recent mortality risks, reducing the time ago of the reference periods of the estimates. Timaeus developed a method for analyzing such data using fertility, mortality and marriage models. Unfortunately, the necessary data have not been widely collected, so despite the method's promise, it cannot be widely applied. The only survey vehicles in which it has been applied are complex fertility surveys of the Demographic and Health Surveys program phase 1, so experience is lacking on how the method would perform in, say, a population census.

b. Spouse Survival Technique

Attempts have also been made to use information on survival of spouses as a basis for indirect estimates of adult mortality (Hill, 1977). In theory, such methods should work quite well: the problems with parental survival of adoption and multiple responses are removed, the spread of ages at marriage is narrow, and the mortality experience reflected is quite recent. However, in practice, the problems with data errors have been found to be formidable. To avoid effects of remarriage, it is necessary to specify a first spouse, but then misreporting current spouse as first spouse, or adopting a different standard of marriage for surviving partners versus those that may have died, become problems. Also, in populations affected by HIV/AIDS, correlations between spouse mortality may be quite high, resulting in selection biases leading to an underestimation of adult mortality.

c. Sibling Survival Technique

The third type of relative used for indirect estimates of adult mortality is siblings. On average, for respondents age 15 years or older, siblings born to the same mother are distributed by age approximately equally on either side of the respondent's age. Thus if mortality risk is constant with age or changes linearly with age, the proportion surviving of a respondent's siblings should closely approximate the probability of surviving to the age of the respondent. Hill and Trussell (1977) first suggested the use of sibling survival data for estimating mortality, and used fertility and mortality models to examine the relationship between proportions with surviving siblings and standard life table survivorship values. They confirmed that the relation between the proportion of surviving siblings and survivorship to the age of the respondent was extremely strong. Timaeus *et al*. (2001) have since modified Hill and Trussell's approach to reflect the fact that the dispersion of childbearing within families is smaller than the dispersion of fertility at the population level.

Despite its theoretical appeal, sibling survival in its indirect form has not been widely used to estimate mortality. The questions needed are not as simple as they at first appear: *How many brothers/sisters (siblings) did your mother give birth to in total?* and *How many brothers/sisters (siblings) born by your mother are still alive now?*. Note that asking separately about brothers and sisters provides estimates of mortality by sex. One problem with the questions is how to make clear that the respondent should not include him or her self. A second problem is that the respondent simply may not know about children born and died before the respondent reached an age to be aware of such matters.

A modification of the sibling method has been developed to measure maternal mortality (Graham *et al.*, 1989). The questions are changed to focus only on numbers of sisters who survived to reach some life milestone such as marriage or age 15, the number of those who have since died, and, for the dead sisters, whether she was pregnant, delivering a child, or had delivered within the 2 months before death. The risk of maternal death was modeled as a modified fertility function, and for each age group the relation between the proportion of sisters dead of maternal causes and the lifetime risk of maternal death was modeled. The approach has mainly been used in small surveys, and results across age groups have been pooled into a single estimate. Unfortunately, this single estimate has a reference date about 12 years before the survey, reducing the value of the method for monitoring purposes.

The development of the indirect sibling method for estimating maternal mortality led to the inclusion of full sibling histories in a number of DHS surveys (Stanton *et al.*, 1997). The data collection required is much more complex (for each sibling: name, sex, whether still alive, how old if still alive, and if not, year of death, age at death, and, for dead sisters, whether the death met a time of death criterion to be regarded as a maternal death). These data provide a basis for the direct calculation of mortality risks and maternal mortality risks, but initial assessments of data quality suggest substantial omissions of sibling deaths more than 6 years before interview (Stanton *et al.*, 1997).

2. Consistency Checks

Consistency checks in the estimation of adult mortality consist primarily of comparisons of the age pattern of deaths with the age pattern of the population. Brass (1975) was the first to propose the use of consistency checks to estimate death registration coverage. In a closed population, the birth (or entry) rate is equal to the sum of the death rate and the growth rate. Such a relation is necessarily true for any age segment of a population. Thus

$$b(a+) = r(a+) + d(a+)$$

where $b(a+)$ is the "birthday" rate (or entry rate) for the age segment a and over (equivalent to the birth rate for the total population), $r(a+)$ is the growth rate of the population over age a, and $d(a+)$ is the death rate over age a. In a stable population, the growth rate is constant at all ages. If the "birthday" rate (i.e., the rate of entries) is calculated only from the age distribution, but the death rate is calculated from recorded deaths of completeness c (constant at all ages) relative to the population coverage, the relationship then becomes

$$b(a+) = r + (1/c)d^o(a+)$$

where $d^o(a+)$ is the death rate based on reported deaths. Plotting the observed death rate against the observed entry rate for a range of ages a thus permits (if the assumptions of stability, and constant death coverage, hold) the estimation of both the stable growth rate r and the reciprocal of the coverage of the death records c.

If two successive population counts are available, the assumption of stability can be relaxed (Hill, 1987). Assuming that census coverage is constant by age, but may change from count to count, the relationship becomes

$$b(a+) - r^o(a+) = k + (1/c)d^o(a+)$$

where $r^o(a+)$ is the observed growth rate over age a, k is the natural logarithm of the change in coverage from the first to the second count divided by the length of the intercensus interval in years, and c is the coverage of death recording relative to the average coverage of the two counts. Figure 127–2 shows the application of this method to data for males from the 1982 and 1992 censuses of Zimbabwe, using deaths by age recorded in the form of household deaths by age and sex in the 12 months before the 1992 census.

Equations for stable and closed general populations have also been exploited for comparisons between age distributions and distributions of deaths by age. In a stable population, the population age a can be expressed in terms of the deaths over age a and the stable rate of growth (Preston *et al.*, 1980):

$$N(a) = \Sigma D(x)\ exp(r^*(x - a))$$

If deaths are reported with completeness c relative to the population count, the ratio of the population age a based on the above expansion of deaths at ages above a to the reported population age a estimates the value of c. If the stable growth rate is not known, the use of a range of growth rates can be used, with the one that gives the closest approximation to a consistent set of estimates of c across age groups being selected as the best single estimate.

Bennett and Horiuchi (1981) generalized the above relationship to apply to all closed populations, replacing the product of the stable growth rate and the age difference between the population and the deaths

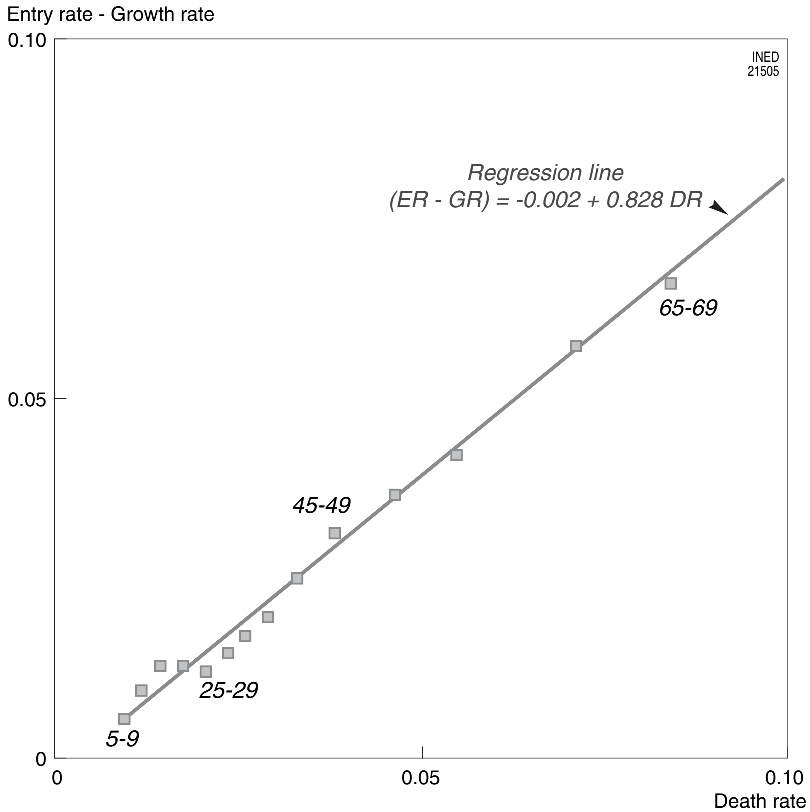


FIGURE 127–2 Application of consistency check to Zimbabwe males, 1982 to 1992.

$(x - a)$ in the above relationship by the sum of the observed growth rates between a and x.

These consistency checks between age distributions and age at death distributions can be thought of as extensions of conventional intercensus survival (whereby intercensus cohort mortality is estimated from cohort decrements between one census and the next) with the intercensus survival (or population age structure) estimates of adult mortality being compared with the estimates based on deaths by age. Assumptions about the nature of the errors (typically that errors are invariant with age) then permit adjustments of the original data and improved estimates of mortality.

These consistency checks have been very widely applied to death distributions obtained from registered deaths and from census or survey questions on recent household deaths. Experience suggests that the methods that assume stability are becoming increasingly inappropriate as populations in the developing world destabilize. The methods that do not assume stability appear to work very well in populations with good quality age reporting (Chinese populations, for example) as long as the coverage of deaths is high enough to be representative of the true age pattern of deaths; coverage less than 60% of those deaths that should be included is likely to give rise to selection bias and lead to unsatisfactory results. In populations with substantial amounts of age misreporting of a nonrandom nature, results can be disappointing.

V. ESTIMATION METHODS FOR MIGRATION

Direct measures of migration are often weak even in developed societies, so it might be expected that indirect estimation procedures would be well developed. In fact, however, the development of estimation procedures for migration has also lagged those for fertility and mortality. Even those methods that have been developed have not been widely applied: utilization of available data has been no faster than the development of methods.

1. Indirect Estimates

Indirect estimates have been largely limited to the use of information on residence of children or siblings to estimate emigration (Zaba, 1986). Mothers are asked how many surviving children they have, and how many of those live in the same country. For the sibling approach, respondents are asked how many surviving brothers and sisters (born of the same mother) they have, and how many live in the same country as the respondent and how many live elsewhere. Models of fertility and emigration are used to develop methods for deriving estimates of migration from the data, usually limiting analysis to the native-born population.

2. Consistency Checks

Consistency checks have followed lines similar to those used for adult mortality. For example, intercensus estimates of migration computed from cohort changes in numbers of migrants can be compared with direct measures of numbers of migrants available from administrative records (Hill, 1987). Use of an appropriate study population (for example, the nonnative population to study immigration) concentrates the numbers of events and improves robustness. As always, to use consistency checks as a basis for adjustment requires assumptions about patterns of data error. These assumptions (constant coverage by age, for example) may often be more problematic in the case of migrants than in the case of deaths or births.

CONCLUSIONS

Techniques for estimating fertility, mortality and migration in developing countries from less than optimal data have been developed over the last 40 years, and have contributed enormously to our knowledge of demographic conditions in the developing world. Over that same period, conventional data on event frequencies—registration of births and deaths—have not improved greatly, though population censuses have become much more an established part of routine and have in many cases improved in quality also. The other major change in demographic measurement has been the widespread use of complex data collection instruments, particularly birth histories, that permit direct estimation of a number of demographic processes. However, because these instruments are elaborate, they are typically only used for small samples, which in turn provide limited geographic or other detail. Indirect estimation procedures, including both indirect methods and consistency checks, remain important as ways of producing estimates for small population subgroups and for tracking trends. The procedures have become increasingly sophisticated and robust over time, but assumptions about true demographic patterns or error patterns still have to be made to interpret results. Purists sometimes find this indirectness distressing, whereas pragmatists accept what they can get.

References

Bennett Neil G. and Horiuchi Shiro, 1981. Estimating the completeness of death registration in a closed population. *Population Index*, vol. 42(2).

Brass William, 1964. Uses of census or survey data for the estimation of vital rates. Paper prepared for the African Seminar on Vital Statistics, Addis Ababa, 14–19 December 1964.

Brass William, 1975. *Methods for estimating fertility and mortality form limited and defective data*. Occasional publications of the laboratory for population statistics, University of North Carolina. Chapel Hill, North Carolina, University of North Carolina.

Brass William, 1981. The use of the relational Gompertz model to estimate fertility. *Proceedings of the IUSSP Conference, Manila*. Liège, Belgium, IUSSP.

Brass William, Coale Ansley J., Demeny Paul, Heisel Don F., Lorimer Frank, Romaniuk Anatole and van de Walle Étienne, 1968. *The Demography of Tropical Africa*. Princeton, New Jersey, Princeton University Press.

Brass W. and Hill K., 1973. Estimating adult mortality from orphanhood. *Proceedings of the International Population Conference, Liège*, p. 111–123. Liège, Belgium, IUSSP.

Brass William and Bambgoye Elijah A., 1981. The time location of reports of survivorship. Estimates for maternal and paternal orphanhood and the ever-widowed. Research Paper 81-1, Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Brass William and Macrae Sheila, 1984. Childhood mortality estimates from reports on previous births given by mothers at the time of a maternity. I. Preceding births technique. *Asian and Pacific Census Forum*, vol. 11, p. 5–8.

Cho Lee-Jay., Retherford Robert D. and Choe Minjak K., 1986. *The "Own Children" method of fertility estimation*. Honolulu, Hawaii, East-West Center.

Coale Ansley J. and Demeny Paul, 1968. *Methods of evaluating basic demographic measures from limited and defective data*. New York, United Nations.

Coale Ansley J., Demeny Paul and Vaughan Barbara, 1983. *Regional model life tables and stable populations, second edition.* New York and London, Academic Press, 496 p. (Second edition)

Coale Ansley J. and Trussell T. James, 1974. Model fertility schedules: variations in the age structure of childbearing in human populations. *Population Index*, vol. 40(2), p. 185–258.

Feeney Griffeth, 1980. Estimating fertility trends from child survivorship data; a demographic approach. *Population Studies*, vol. 34, p. 109–128.

Graham Wendy, Brass William and Snow Robert W., 1989. Estimating maternal mortality. The sisterhood method. *Studies in Family Planning*, vol. 20, p. 125–135.

Hill Kenneth, 1977. Estimating adult mortality levels from data on widowhood. *Population Studies*, vol. 31, p. 75–84.

Hill Kenneth, 1987. Estimating census and death registration completeness. *Asian and Pacific Population Forum*, vol. 1(3), p. 8–13.

Hill Kenneth and Figueroa Maria-Elena, 2001. Child mortality estimation by time since first birth, *in*: Basia Zaba and John Blacker (eds.), *Brass tacks. Essays in medical demography. A tribute to the memory of Professor William Brass*, p. 9–19. London and New York, The Athlone Press, 301 p.

Hill Kenneth and Trussell James, 1977. Further developments in indirect mortality estimation. *Population Studies*, vol. 31, p. 313–334.

Hill Kenneth, Zlotnik Hania and Trussell T. James, 1983. *Manual X. Indirect techniques for demographic estimation.* New York, United Nations/Nations unies, 304 p.

Preston Samuel H., 1983. An integrated system for demographic estimation from two age distributions, *Demography*, vol. 20(2), p. 213–226.

Preston Samuel H. and Palloni Alberto, 1977. Fine-tuning Brass-type mortality estimates with data on ages of surviving children, *Population Bulletin of the United Nations*, no.10, p. 72–87.

Stanton Cynthia, Abderrahim Noureddine and Hill Kenneth, 1997. *DHS maternal mortality indicators: an assessment of data quality and implications for data use*. Analytical Report no. 4, Demographic and Health Surveys. Calverton, Maryland, Macro International Inc.

Sullivan Jeremiah M., 1972. Models for the estimation of probabilities of dying between birth and exact ages of early childhood. *Population Studies*, vol. 26, p. 79–97.

Timaeus Ian, 1990. Advances in the measurement of adult mortality from data on orphanhood. Research Paper 90-1, Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Timaeus Ian, 1991a. Measurement of adult mortality in less-developed countries: A comparative review. *Population Index*, vol. 57(4), p. 552–568.

Timaeus Ian, 1991b. Estimation of adult mortality form orphanhood before and since marriage. *Population Studies*, vol. 45(3), p. 455–472.

Timaeus Ian, 1992. Estimation of adult mortality from paternal orphanhood: A reassessment and a new approach. *Population Bulletin of the United Nations*, vol. 33, p. 47–63.

Timaeus Ian, Zaba Basia and Ali Mohamed, 2001. Estimation of adult mortality from data on adult siblings, *in*: Basia Zaba and John Blacker (eds.), *Brass tacks. Essays in medical demography. A tribute to the memory of Professor William Brass*, p. 43–66. London and New York, The Athlone Press, 301 p.

Trussell T. James, 1975. A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. *Population Studies*, vol. 29(1), p. 97–108.

Zaba Basia, 1986. *Measurement of emigration using indirect techniques*. Manual for the collection and analysis of data on residence of relatives. Liége, Belgium, Ordina Editions.

CHAPTER

128

Model Life Tables

JOSIANNE DUCHÊNE

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

One alternative to the analytical approach based on mortality curves by age, sex and possibly cause, is to use the life tables produced in some countries to create what are known as model life tables—tables defined numerically, or by the analytical relationships between the "parameters" and other mortality functions, and intended to represent average mortality scenarios at different levels. Broadly speaking, model life tables are used whenever a life table calculated for a given period or place is applied to other populations, either in order to estimate missing data or to predict mortality.

I. MODEL LIFE TABLES BASED ON REGRESSION METHODS AND ON DATA FROM DEVELOPED COUNTRIES

The first attempt to construct a model life table summarizing the mortality patterns of several countries was published in 1944 by Frank Notestein *et al.*, in *The Future Population of Europe and the Soviet Union*, a study whose aim was to project the population of Europe and the Soviet Union. The authors calculated the regression coefficients relating ${}_5q_x$ and e_{10} for $5 \leq x \leq 85$ on the basis of European life tables used for the projections. Using regression techniques, values for e_{10} were obtained from the observed ${}_5q_x$ from $x = 5$ to $x = 20$, and these estimated average life expectancies at ten years were then used to correct the ${}_5q_x$ for $x \geq 25$.

1. The United Nations Tables

Most sets of model life tables were developed through applying regression methods to an array of life tables estimated from relatively reliable data. Vasilios Valaoras elaborated the first collection of such model tables for the United Nations in around 1955, using data from 158 tables essentially from developed countries; some of them had been smoothed using analytical or graphical techniques (United Nations, 1955, 1956; United Nations, 1956). A chain regression method of mortality probabilities (both sexes combined) linked successively ${}_5q_0$ to ${}_1q_0$, ${}_5q_x$ to ${}_5q_{x-5}$, etc., according to the quadratic relation:

$${}_5q_x = A + B\ {}_5q_{x-5} + C\ {}_5q^2_{x-5}$$

where the least squares method is used to estimate parameters A, B and C for each value (5, 10, . . . , 80) of x. These are called *one entry* tables (${}_1q_0$). Forty tables were created in this way for values of ${}_1q_0$ from 20‰ to 330‰ (‰ = per 1,000), corresponding to life expectancies at birth varying from 19 to 72 years. From these tables, model tables for each sex were then established for the total population, by applying standard (fixed) models of age differences in mortality by sex. Ruben Gabriel and Ilana Ronen (1958) criticized this chain method for estimating mortality rates with second degree terms, showing that this approach introduced a systematic bias and, in particular, overestimated expectation of life by approximately two years on average. With linear regression, they then calculated the probabilities ${}_4q_1$, ${}_5q_5$, . . . , ${}_5q_{80}$ from infant mortality ${}_1q_0$:

$${}_nq_x = A + B\ {}_1q_0$$

nonetheless drawing attention to the fact that a quadratic relation would be preferable for estimating ${}_4q_1$, at least for extreme values of ${}_1q_0$. Their aim was to estimate missing data in countries with incomplete—even quasi-nonexistent—or defective statistics.

2. The Princeton Tables

It quickly became evident, as shown notably by Sully Ledermann, Jean Bréas and Jean Bourgeois-Pichat (Ledermann, 1955; Ledermann and Bréas, 1956; United Nations, 1962; Bourgeois-Pichat, 1963), and Paolo De Sandre (1973, 1974), that a mortality schedule could not be adequately characterized by a single parameter. In 1964, Ramakrishna Kurup (cited by Valaoras, 1973) produced five groups or families of model tables, each corresponding to a level of socioeconomic development. In 1966, Ansley Coale and Paul Demeny published four families of model tables constructed from 193 male and female life tables, the great majority from European populations or those of European origin:

- the *East* family, based on 31 tables (5 from Austria, 13 from Germany, 4 from Czechoslovakia, 6 from Central or Northern Italy, and 3 from Poland), characterized by relatively high mortality during childhood and at ages above 50 years compared with other adult ages;
- the *North* family, summarizing 9 tables (4 from Norway, 4 from Sweden and 1 from Iceland), and characterized by relatively low mortality levels below the age of 1 year and above 45 or 50 years.
- the *South* family, based on 22 tables (5 for Spain, 8 for Portugal, 5 for all Italy, 1 for Sicily and 3 for Southern Italy), characterized by relatively high mortality below the age of 5 years and over 65 years, but relatively low between 40 and 60 years.
- a residual group, called the *West* family, was constructed from 130 other national tables for countries with relatively good statistics (7 from Australia, 4 from Belgium, 7 from Canada, 11 from Denmark, 11 from England and Wales, 1 from Estonia, 4 from Finland, 16 from France, 5 from Ireland, 3 from Israel, 6 from Japan, 1 from Latvia, 1 from Luxembourg, 10 from the Netherlands, 12 from New Zealand, 2 from Northern Ireland, 7 from Scotland, 5 from Sweden, 3 from Taiwan, 4 for the white population of South Africa and 10 from the United States).

A matrix of correlation coefficients was estimated for each family, and separately for men and women, between $\log({}_1q_0)$, $\log({}_4q_1)$, $\log({}_5q_5), \ldots, \log({}_5q_{75})$ and both e_0 and e_{10}. As the male mortality indicators were not as well correlated as female mortality indicators, the authors first created a set of female model tables, using least squares linear regression on life expectancy at 10 years (e_{10}) of ${}_nq_x$, for low values of e_{10}, and of $\log({}_nq_x)$ for high values of e_{10}, and an average of the values of ${}_nq_x$ obtained through the two regressions for the values of e_{10} in between. Each family of models included 24 mortality levels defined by female life expectancies at birth from 20 to 77.5 years, by steps of 2.5 years. Ansley Coale and Paul Demeny then derived the corresponding male model life tables, based on standardized differences in sex-specific mortality. The tables were closed using a life expectancy at 80 years estimated by a regression on the number of survivors at 80 years.

In 1983, Ansley Coale, Paul Demeny and Barbara Vaughan completed this set of model life tables by calculating a 25th level, corresponding to a life expectancy at birth of 80 years for women, and extending table estimates to the age of 100 years. Above 80 years, however, the Gompertz formula, based on the hypothesis that instantaneous death rates, $\mu(x)$, grow exponentially with age, was used to calculate the survival curve l_x.

In 1989, Ansley Coale and Guang Guo revised these model tables for very low mortality levels (e_0(F) over 75 years). They modified the tables at three levels:

1. for ages above 80 years, they replaced the exponential growth hypothesis of $\mu(x)$ by a hypothesis of a linear decrease of the ratio $\ln \frac{{}_5m_x}{{}_5m_{x-5}}$.
2. they constructed new model tables, from ages 0 to 100 years, for female life expectancy levels of 77.5 and 80 years for the West family, from 72.5 to 80 years for the East and from 65 to 80 years for the South family of models.
3. they estimated new model tables, common to the various family types, for female life expectancies of 82.5 and 85 years.

Their creators believed that these tables could prove useful for projections in 1989; in fact, from this point of view they very quickly became obsolete as, ten years later, women in some countries had already reached this level.

3. The Ledermann Tables

After demonstrating that entering a model table through a parameter other than the one used to construct it introduced a risk of bias, Sully Ledermann and Jean Bréas (1956) analyzed the principal components of 157 tables which enabled them to identify three vital components of age-specific mortality:

1. the first component close to e_0 depends especially on mortality between 10 and 45 years
2. the second component is defined by mortality over 40 years
3. the third component summarizes mortality over the age of 70 years.

Based on this analysis, Sully Ledermann (1969) then defined seven sets of one-parameter model tables and three two-parameter sets:

1. **One-parameter Sets**

For an entry X, the set is constructed on the relation:

$$\log(q_j) = a_{j0} + a_{j1}\log(X) + s_j\varepsilon$$

where a_{j0}, a_{j1} and s_j are three parameters, ε being a normal variable with a mean of zero and a standard variation of 1.

The seven parameters included are:

set 100: $100 - e_0(MF)$
set 101: ${}_5q_0(MF)$
set 102: ${}_1q_0(MF)$
set 103: ${}_{15}q_0(MF)$
set 104: ${}_{20}q_{30}(MF)$
set 105: ${}_{20}q_{45}(F)$
set 106: $T_{50+} = 44{,}93 + 0{,}5046t_{50+} - 0{,}5587f$

where t_{50+} is the observed death rate over 50 years and f is the proportion (%) of those aged over 65 years among the population aged over 50 years.

2. **Two Parameter Sets**

For two parameters X_1 and X_2, the set is constructed on the relation:

$$\log(q_j) = b_{j0} + b_{j1}\log(X_1) + b_{j2}\log(X_2) + s_j\varepsilon$$

where b_{j0}, b_{j1}, b_{j2} and s_j are four parameters and ε a normal variable with a mean of zero and a standard variation of 1.

The three pairs of parameters entered are:

set 1: ${}_5q_0(MF)$ and ${}_{20}q_{45}(F)$
set 2: ${}_{15}q_0(MF)$ and ${}_{20}q_{30}(MF)$
set 3: ${}_{15}q_0(MF)$ and T_{50+}

In addition to these ten sets of single and double-entry tables, Sully Ledermann published tables of regression coefficients as well as indicators of variation around the estimated central value. The probability (‰) at 85 to 89 years is calculated from the probability at 80 to 84 years, according to the relation:

$${}_5q_{85} = 337.8 + 0.698\,{}_5q_{80}$$

Henri Leridon and Laurent Toulemon (1997) demonstrated that "*the splitting of the single parameter $e_0(MF)$ into two tables* [. . .] *holds several surprises: life expectancies re-estimated for each sex can both be inferior (or superior) to the parameter's value for both sexes combined.*"[1] The same authors suggest a formula to correct the observed life expectancy (EV) and provide the life expectancy to be used (EVU):

$$EVU = 0.009EV^2 + 0.11EV + 17.0 \text{ for } 75 \le EV \le 90$$

The probability (‰) at 85–89 years should be corrected as follows:

$${}_5q_{85} = 200 + 0.9\,{}_5q_{80}$$

II. THE LOGIT SYSTEM

Brass's logit system does not provide a set of model tables but rather the relationships making it possible to create a life table for a given population from a known table, taken as the reference and dubbed "standard" (Brass and Coale, 1968). William Brass developed a method introduced by William Kermack, Anderson McKendrick and P. McKinlay in 1934 (cited by Brass, 1969) to establish longitudinal mortality perspectives, and based on the assumption of a constant relationship between the instantaneous mortality rates at age x of two succeeding generations:

$$\frac{\mu^*(x)}{\mu(x)} = \beta$$

or, since $\mu(x) = -\dfrac{d\ln l(x)}{dx}$.

$$\frac{\ln(l^*(x))}{\ln(l(x))} = \beta$$

William Brass (1971a) demonstrated that the association $\dfrac{\mu^*(x)}{\mu(x)}$ is not constant and that a preferable hypothesis is:

$$\frac{\dfrac{\mu^*(x)}{1-l^*(x)}}{\dfrac{\mu(x)}{1-l(x)}} = \beta$$

or

$$\frac{1}{2}\frac{d\ln l^*(x)}{(1-l^*(x))dx} = \beta\frac{1}{2}\frac{d\ln l(x)}{(1-l(x))dx}$$

[1] "l'éclatement de l'entrée unique $e_0(MF)$ en deux tables [. . .] réserve quelques surprises: les espérances de vie recalculées pour chaque sexe peuvent être toutes deux inférieures (ou supérieures) à la valeur entrée pour les deux sexes."

and, after integration,

$$\frac{1}{2}\ln\left(\frac{l^*(x)}{(1-l^*(x))}\right) = \alpha + \beta\frac{1}{2}\ln\left(\frac{l(x)}{(1-l(x))}\right)$$

or, by definition of logit$(1 - l(x))$,

$$\frac{1}{2}\text{logit}(1-l^*(x)) = \alpha + \beta\frac{1}{2}\text{logit}(1-l(x))$$

In practice, the standard series of survival probabilities from birth to age x, noted $l_s(x)$ by the function logit$(1 - l_s(x))$ is thus transformed, and the probability $l(x)$ for a given population is inferred from the linear relation:

$$\text{logit}(1 - l(x)) = \alpha + \beta\,\text{logit}(1 - l_s(x)),$$

which amounts to the assumption that:

$$\frac{1-l(x)}{l(x)} = e^{2\alpha}\left(\frac{1-l_s(x)}{l_s(x)}\right)^{\beta}$$

The parameter α sets the level of mortality in the sense that, if m is the median age at death in the standard table, then

$$l_s(m) = 0.5$$

$$\text{logit}(1 - l_s(m)) = 0$$

and thus

$$\text{logit}(1 - l(m)) = \alpha$$

Consequently,

- if the parameter α is positive, the probability $l(m)$ is superior to 0.5 and the median age will be higher than m;
- if the parameter α is negative, the probability $l(m)$ is inferior to 0.5 and the median age will be lower than m;

The parameter β translates differences in mortality's age structure in the sense that it controls the relationship between childhood and adult mortality since

$$\text{logit}(1 - l(x)) - \alpha = \beta\,\text{logit}(1 - l_s(x))$$

and thus

- if parameter β is superior to 1, logit$(1 - l(x)) - \alpha$ will be lower than logit$(1 - l_s(x))$ and x will be lower than the median age m whereas logit$(1 - l(x)) - \alpha$ will be superior to logit$(1 - l_s(x))$ if x is higher than m;
- if parameter β is inferior to 1, logit$(1 - l(x)) - \alpha$ will be higher than logit$(1 - l_s(x))$ and x will be lower than the median age m whereas logit$(1 - l(x)) - \alpha$ will be inferior to logit$(1 - l_s(x))$ if x is higher than m;

William Brass uses the logit system for the essentially pragmatic reason that it satisfactorily describes most mortality schedules. To give meaning to this system, William Brass (1975) assumes that the relation

$$y = \frac{1}{2}\left(\text{sech}\left(\frac{x-\alpha}{\beta}\right)\right)^2 = \frac{2}{\left(\exp\left(\frac{x-\alpha}{\beta}\right)+\exp\left(\frac{\alpha-x}{\beta}\right)\right)^2}$$

is used to define a symmetrical curve of mean α and standard deviation β, and that any other curve with the same form but a mean of zero and a standard deviation of 1 is defined by

$$y^* = \frac{1}{2}(\text{sech}(x))^2 = \frac{2}{(e^x + e^{-x})^2}$$

For each value of x, $1 - l(x)$ and $1 - 1^*(x)$ are defined as the proportion of the area under curves y and y^* between the origin and the abscissa x; the following relationship exists between $l(x)$ and $l^*(x)$:

$$\text{logit}(1 - l(x)) - \alpha = \beta\,\text{logit}(1 - l^*(x))$$

the quantity $1 - l(x)$ represents the proportion of deaths occurring before age x, and y is the mortality curve.

This model is good for populations with mortality close to the standard, that is for values of α not too far from zero and values of β close to 1. For a well-chosen standard, the parameter α varies between –1.5 and +1.0 whereas values of the β parameter fall between 0.5 and 1.5. Guillaume Wunsch (1978) demonstrated that β values close to 1 can, however, be as easily associated with very flawed adjustments as very good ones; the value of coefficient β is not therefore a good indicator of the quality of fit. William Brass (1971b), Norman Carrier and Thuang-Jig Goh (1972) and Guillaume Wunsch (1978) all emphasized the importance of the choice of standard. The general standard established by William Brass summarizes the patterns observed in Europe and, as a result, approaches the United Nations model tables or Coale and Demeny's West family of models. William Brass also constructed an African standard to correct data from countries with relatively high mortality in early childhood, and André Lambert (1973) summarized mortality patterns in Latin-American countries during the twentieth century into a standard for each sex. Alongside this diversification of standard tables, Basia Zaba (1979) developed William Brass's approach for generating a set of standards from a general standard in the following way:

$$l_n(x) = l_s(x) + \psi k(x) + \chi t(x)$$

where $k(x)$ and $t(x)$ should distort the mortality schedule at over 70 years and in infancy. The WHO (Murray et al., 2000; Lopez et al., 2000; Murray et al., 2003) elaborated new standards from data from developed and developing countries. Five strata were defined (from A to E) varying according to the ratio of mortality at 0–5 years (${}_5q_0$) to mortality between 15 and 60 years (${}_{45}q_{15}$). These tables were constructed with the aim of taking AIDS into account. Joseph Kamara (1988, 1989) created a three-parameter logit system which he applied to Peru in particular (Kamara and Lansana, 2001).

III. MODEL TABLES BASED ON REGRESSION METHODS AND DATA FROM DEVELOPING COUNTRIES

1. The OECD Tables

From 1977, the OECD Development Centre, with the support of the Department of Demography at the *Université catholique de Louvain*, the project's instigator, gathered raw data on population and mortality structures in developing countries; these were evaluated and adjusted to create mortality tables for these countries. Subsequently, model tables were put together (Clairin *et al.*, 1980) on the basis of data from developing countries whose mortality levels were still high, and including only the 125 cases where it was at least possible to estimate rates at 0, 1–4 years, 5–9 years, . . . , 60–64 years, for each sex. The tables were calculated from unsmoothed age and sex-specific rates and constructed from parameters adapted to the requirements of mortality analysis in developing countries.

Several measures were applied to the 125 tables to evaluate the death registration, and the 21 tables with an undercount of over 50% were rejected. The 104 remaining tables were compared with the Coale and Demeny West model on the basis of a median life expectancy at birth corresponding to different mortality probabilities. Four age groups were compared: 0–4 years (children), 5–19 years (adolescents), 20–44 years (young adults) and 45–59 years (adults). This made it possible to define 3 subgroups (A, B and C) with systematic deviations, and one residual group (D) with no specific characteristics:

- group A has high child, adolescent and young adult mortality, but low adult mortality,
- group B has low mortality among children and the elderly, but high adult mortality,
- group C shows high child mortality, low adolescent mortality, average mortality among young adults and high adult mortality.

From the 4 groups of tables, 4 linear and quadratic regressions were applied:

$$\hat{q}_x = a + b(100 - e_1)$$

$$\hat{q}_x = a + b(100 - e_1) + c(100 - e_1)^2$$

$$\ln(\hat{q}_x) = a + b(100 - e_1)$$

$$\ln(\hat{q}_x) = a + b(100 - e_1) + c(100 - e_1)^2$$

As information on infant mortality is often imprecise, e_1 was preferred to e_0 and the equation regressing $\ln(\hat{q}_x)$ on $(100 - e_1)$ was used to calculate $\ln(\hat{q}_x)$ for each group, and only tables with probabilities falling within the 95% confidence interval defined by $\ln(\hat{q}_x) \pm 2\sigma$ were included. Nine tables that could not be attached to any of the four groups were excluded, and some tables from the residual group (D) were incorporated into one of the first three groups (A, B or C). After these reclassifications, 95 tables remained, distributed between the four groups, known as regions, as indicated in Table 128–1. Region A is characterized by slightly above-average infant mortality, and a very high excess mortality between ages 1 to 9 years which declines until 25 years. Above this age, mortality is below average and becomes increasingly so with age. Region B's mortality pattern is almost the inverse of region A's. In region C, mortality deviates from the mean less markedly than in regions A and B. The mortality structure in region D (residual group containing 40% of the 95 tables) has no relative deviation of over 5%. Not only is there no obvious regional homogeneity but, for a same country, the tables can be classified into different regions according to the period and context. Substantial changes in health conditions affecting these countries have at times transformed radically age structures and mortality levels.

Two models were created: the first used regression and the second was based on median regional standards.

The model based on regression was founded on the relations:

$$\ln({}_n\hat{q}_x) = a_x \ln(q(1)) + b_x \text{ for one parameter}$$

$$\ln({}_n\hat{q}_x) = a_x \ln(q(1)) + b_x n(q(2)) + c_x \text{ for two parameters}$$

The publication contains all the regression coefficients that make it possible to calculate the models, for men and women separately, for the five regions (A, B, C, D and the four groups combined, termed E), and for the 7 one-parameter sets (${}_1q_0$, ${}_5q_0$, ${}_{15}q_{15}$, ${}_{20}q_{25}$, ${}_{15}q_{35}$, $100 - e_0$ and $100 - e_1$) and the 6 two-parameter sets (${}_1q_0$ and ${}_{15}q_{15}$ or ${}_{20}q_{25}$ or ${}_{15}q_{35}$, as well as ${}_5q_0$ and ${}_{15}q_{15}$ or ${}_{20}q_{25}$ or ${}_{15}q_{35}$) that combine information on both juvenile and adult mortality. The most precise single entry sets are those in which ${}_{20}q_{25}$, $100 - e_0$ or $100 - e_1$ are used as the entry parameter. For two-entry sets, ${}_5q_0$ is the best first entry associated with one of three parameters (${}_{15}q_{15}$, ${}_{20}q_{25}$ or ${}_{15}q_{35}$).

TABLE 128–1 Tables Used to Construct the Organization for Economic Co-operation and Development Models

Region A	Mexico (1959/1961)
(23 tables)	Mexico (1968/1971)
Senegal Paos Koto (1963/1965)	Barbados (1959/1961)
Tanzania, rural (1967)	Barbados (1969/1971)
Jordan, East Bank (1960/1962)	Puerto Rico 1949/51
Jordan, West Bank (1975)	Puerto Rico 1959/61
Jordan (1960/1962)	Region D
Kuwait (1970)	(39 tables)
Tunisia, total population (1966)	Ivory Coast, Bouake (1961/1962)
Tunisia (1965/1967)	Guinea (1954/1955)
Algeria residents (1965/1967)	Upper Volta (1960/1961)
Greater Algiers (1965/1967)	Liberia PGS, urban (1971)
Algeria, urban (1966)	Reunion (1960/1963)
Paraguay (1971/1973)	Seychelles (1960)
Peru (1960/1962)	Tanzania, urban (1967)
Philippines (1958/1961)	Jordan, East Bank (1971/1973)
Philippines (1969/1971)	Morocco Cered 1st pas. (1972)
Sri Lanka (1951/1954)	Tunisia, urban (1965/1967)
Thailand (1959/1961)	Tunisia, rural (1969)
Costa Rica (1962/1964)	Tunisia, urban (1969)
El Salvador (1949/1951)	Argentina (1970)
Guatemala (1963/1965)	Colombia (1950/1952)
Panama (1959/1962)	Colombia (1963/1965)
Panama (1969/1971)	Afghanistan (1973)
Belize (1969/1971)	Afghanistan, rural (1973)
Region B	Afghanistan, urban (1973)
(16 tables)	Bangladesh (1963)
Mauritius (1951/1953)	Bangladesh (1965)
Reunion (1966/1969)	Bangladesh, rural (1963)
China Taiwan (1939/1942)	Brunei (1970/1972)
China Taiwan (1955/1957)	Republic of Korea (1966/1967)
China Taiwan (1970/1971)	Hong Kong (1959/1962)
Malaysia, west (1969/1971)	Iran (1975)
Malaysia, rural west (1969/1971)	Pakistan (1963)
Malaysia, urban west (1969/1971)	Pakistan (1965)
Sabah (1969/1971)	Pakistan (1968)
Singapore (1956/1958)	Thailand, sample (1965)
Singapore (1969/1971)	Thailand (1969/1971)
Surinam (1950/1951)	El Salvador (1960/1962)
Guiana (1959/1961)	Honduras (1960/1962)
French Guiana (1960/1963)	Mexico (1938/1941)
Trinidad and Tobago (1945/1947)	Nicaragua (1939/1941)
Trinidad and Tobago (1950/1961)	Nicaragua (1962/1964)
Region C	Belize (1945/1947)
(17 tables)	Belize (1959/1961)
Cameroon—Bamileke (1965)	Puerto Rico (1969/1971)
Mauritius (1961/1963)	Dominican Republic (1968/1971)
Mauritius (1971/1973)	
Kuwait (1975)	
Argentina (1959/1961)	
Chile (1929/1932)	
Chile (1939/1942)	
Chile (1951/1953)	
Uruguay (1962/1965)	
Costa Rica (1972/1974)	
El Salvador (1970/1972)	

The OECD model tables stop at 65 years as most life tables end at 65 or 70 years; in cases where they are taken further, death rates tend to take on unrealistic values above this age. For probabilities above 65 years, the authors recommend linear extrapolation of their logarithms. Guillaume Wunsch (1984) considers it preferable to adjust a 3rd degree polynomial to the probabilities ${}_5q_{50}$, ${}_5q_{55}$, ${}_5q_{60}$ and ${}_5q_{95}$ (assumed equal to 1), and to use it to interpolate the probabilities between 65 and 95 years.

For each region, an $e_0(M)$ can be associated with an $e_0(F)$, and vice versa, using regression equations, although this may be imprecise if there is substantial variation around the central values. The document provides the regression coefficients, calculated for life expectancies at birth between 30 and 70 years. Region D has excess female mortality (women's life expectancy lower than men's) for life expectancies lower than 37.1 years.

The OECD publication includes separate standard tables for men and women, and for all families (A, B, C, D and E). For each family and each sex, these standard tables use median probabilities at every age. Using these standards is identical to Brass' logit approach.

The OECD model tables are the first to be created from mortality data from developing countries whose poor quality affects the precision of the estimates provided. The families do not always correspond to precise geographical areas. It is preferable to use regression coefficients as some tables in annex are abnormal and mortality at high ages is sometimes badly estimated. The current state of computing techniques means that the quantity of calculations needed is no longer an obstacle.

2. The United Nation's New Tables

The United Nations Population Division (1982) applied stricter selection criteria to data gathered by the OECD Development Centre. It included only 36 tables, some of which had been smoothed or corrected. Gaining with respect to data quality, however, entailed losses in terms of representation. Mortality was divided into four sets or "patterns": the Latin American, Chilean, South East Asian and Far Eastern patterns, the average of the four groups known as the "overall average pattern." The model tables were elaborated on the basis of principal components analysis.

The Latin-American pattern is characterized by excess childhood mortality, the Chilean pattern by excess infant mortality, the Southern Asian pattern by excess juvenile and elderly mortality and, finally, the Far Eastern pattern, by excess mortality at older ages. The overall pattern is close to Coale and Demeny's West Model.

IV. APPLYING MODEL TABLES TO DATA FROM DEVELOPING NATIONS

In order to apply model tables to data from the developing world, not already smoothed by some model (including one of the model tables), life tables were compiled from data gathered at the demographic surveillance sites at Agincourt (South Africa), Bandafassi (Senegal) and Matlab (Bangladesh), and published by the International Development Research Centre (IDRC, 2003, taken from http://web.idrc.ca, January 22, 2005). These INDEPTH sites (International Network of field sites with continuous Demographic Evaluation of Populations and their Health in developing countries) possess data on the number of deaths by age, as well as on the person-years at risk during the period 1995–1999. These are recent data. The estimated tables are similar to those presented by the IDRC http://web.idrc.ca/fr/ev-43000-201-1-DO_TOPIC.html, 22nd January 2005). For the Agincourt (South Africa) and Matlab (Bangaldesh) sites, the IDRC used values of ${}_na_x$ corresponding to high mortality countries (over 0.107), whereas we estimate ${}_na_x$ in line with the observed level of infant mortality (under 0.107). On the Agincourt site (South Africa), mortality under five years is low (15‰ for infant mortality and 18‰ for juvenile mortality) in relation to the national average estimate (infant mortality estimated at 45‰, http://www.measuredhs.com, January 22, 2005). On the Bandafassi site, infant and juvenile mortality is high (above 107‰). As for the Matlab (Bangaldesh) site, mortality among the under-5s lies between the high level (above 100‰) and low level (less than 30‰), with infant mortality higher than juvenile mortality. This mortality pattern falls between the Bandafassi (Senegal) pattern of high infant and juvenile mortality, and that of Agincourt (South Africa), characterized by low infant and juvenile mortality. The two graphs below present the death rates observed in each of the three populations (Fig. 128–1).

The three empirical tables were then estimated using the different systems of model life tables. The main results are summarized in the graphs in Figure 128–2.

At Agincourt (South Africa), most models overestimate mortality among children aged under one year, under five years and even under fifteen years, more for boys than girls. Male mortality between 20 and 25 years is precise enough whereas female mortality is estimated in diverse fashions. Mortality between 30 and 50 years is underestimated substantially. As for the two summary mortality measures—life expectancy at birth and at ten years—their estimates carries a maximal error of 10%.

The model life tables make a much more accurate estimation of Bandafassi (Senegal) mortality schedules, even if there is still a 30% gap between the observed and estimated mortality for some model tables.

Matlab's (Bangladesh) mortality is less well modelled than Bandafassi's. With respect to mortality among the under-15s, the Coale and Demeny model tables underestimate by close to 50% mortality probabilities under one, five or fifteen years. On the other hand, mortality between 20 and 25 years is clearly overestimated (up to five times for men by the Ledermann set).

In conclusion, these applications show that the model tables remain adequate for estimating global mortality levels, such as life expectancy at birth or at 10 years. Their ability to represent the age structure of mortality, on the other hand, varies enormously, and no system of model tables stand out as clearly better than the others. The new United Nations model tables or the OECD tables, based on data from the developing world, have no advantage over other, older tables constructed from developed countries' data.

CONCLUSION

Henri Leridon and Laurent Toulemon (1997) compared the different model life tables (United Nations, 1956; Princeton, 1983; United Nations, 1982; OECD, 1980; and Ledermann, 1969) for female mortality corresponding to a life expectancy at birth of forty years. There is clearly (see Table 13.6, page 233) enormous variation in the probabilities ${}_1q_0$ (between 126 and 216‰), ${}_4q_1$ (between 89‰ and 253‰), ${}_5q_{40}$ (between 43‰ and 94‰), and ${}_5q_{70}$ (between 200‰ and 395‰), which reflects the fact that many mortality patterns may exist for the same life expectancy at birth.

Although one might have expected model tables based on data from developing nations to have an advantage, this is in no way born out by the three applications presented. Only life expectancies at birth or at ten years are well reproduced by the model tables.

Almost twenty years after the publication of the OECD and United Nations model life tables, it would

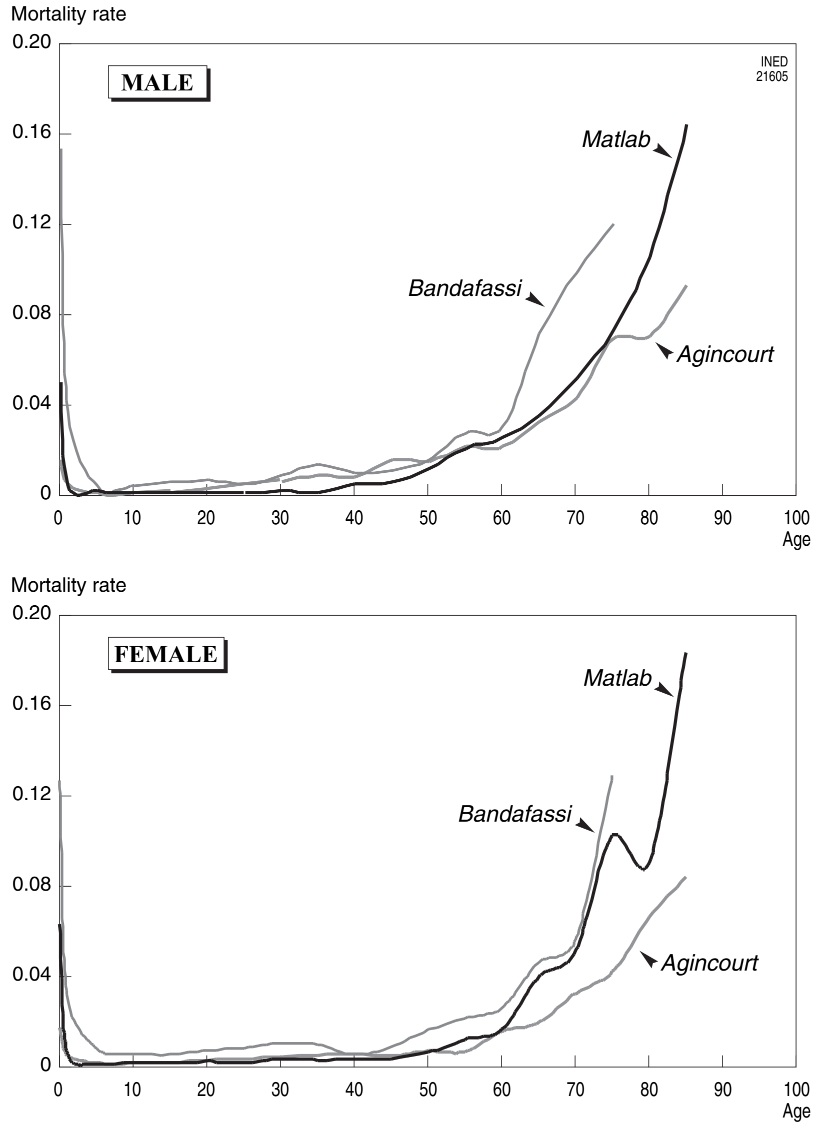


FIGURE 128–1 Age-specific death rates observed at Agincourt, Bandafassi and Matlab.

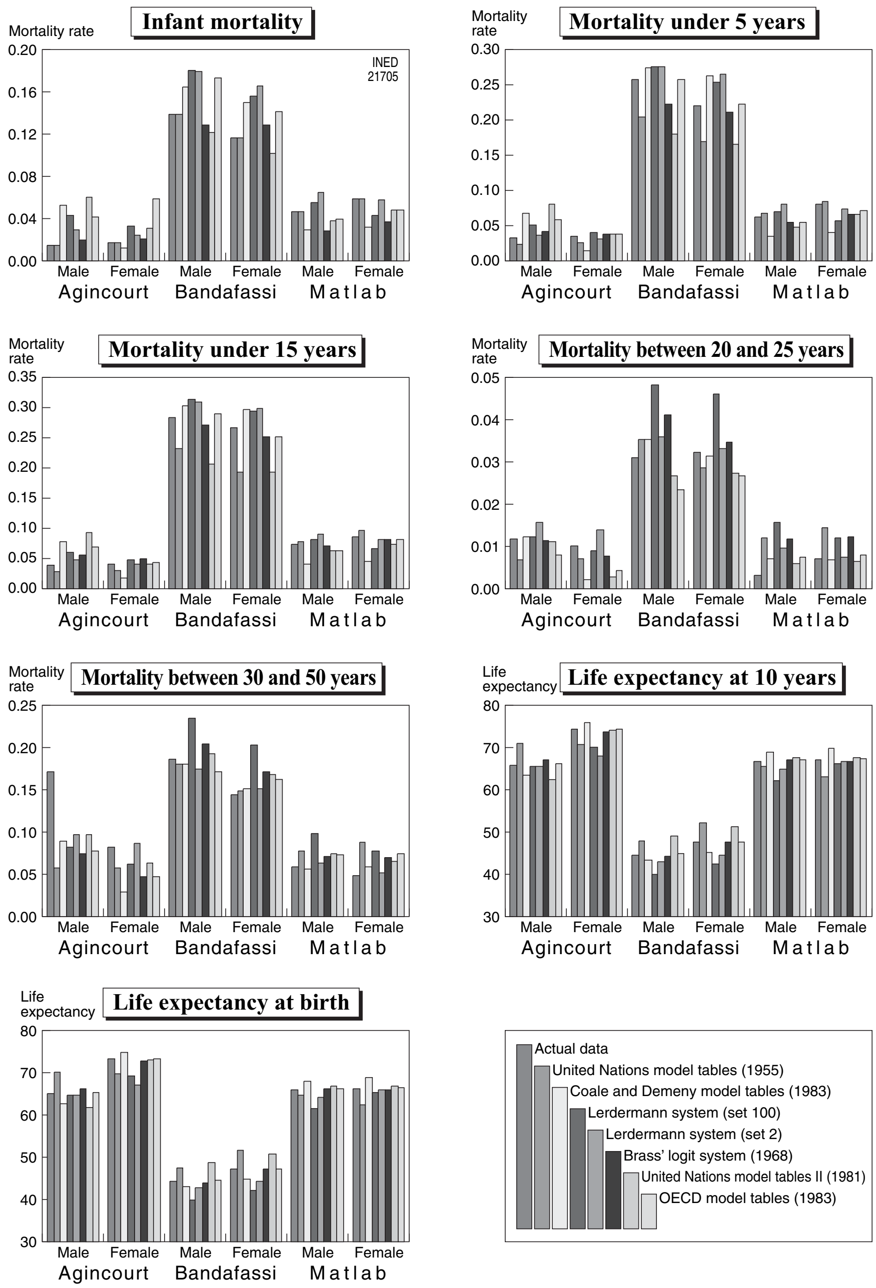


FIGURE 128–2 Results obtained with the different model life tables systems.

undoubtedly be useful to make an inventory of the new tables available for developing countries, and to repeat the exercise to establish new model tables or complete the existing ones, as the WHO did in the year 2000 when it wished to take AIDS into account.

Acknowledgment

The author would like to thank Jacques Emina Be-Ofuriyua for collecting and analyzing the data.

References

Bourgeois-Pichat Jean, 1963. Application of factor analysis to the study of mortality, *in*: Milbank Memorial Fund (ed.), *Emerging techniques in population research*, p. 194–229. New York, Milbank Memorial Fund, 307 p. (Proceedings of a Round Table at the Thirty-Ninth Annual Conference of the Milbank Memorial Fund, September 18–19, 1962).

Brass William, 1969. A generation method for projecting death rates, *in*: Frank Bechhofer (ed.), *Population growth and the brain drain*, p. 75–91. Edinburgh, Edinburgh University Press, 236 p.

Brass William, 1971a. Disciplining demographic data, *in*: IUSSP/UIESP (ed.), *International population conference of the IUSSP, 1969, London, vol. 1*, p. 183–204. Dorchester, Henry Ling Ltd, 843 p.

Brass William, 1971b. On the scale of mortality, *in*: William Brass (ed.), *Biological aspects of demography*, p. 69–110. London, Taylor & Francis, 167 p. (Symposia of the Society for the Study of Human Biology, vol. 10).

Brass William, 1974. Perspectives in population prediction. Illustrated by the statistics of England and Wales, *Journal of the Royal Statistical Society, Series A*, vol. 137(4), p. 532–583.

Brass William, 1975. *Methods for estimating fertility and mortality from limited and defective data*. Chapel Hill (North Carolina), Laboratory for Population Statistics, University of North Carolina, 160 p. (Occasional Publications of the Laboratory for Population Statistics).

Brass William and Coale Ansley J., 1968. Methods of analysis and estimation, *in*: William Brass, Ansley J. Coale, Paul Demeny, Don F. Heisel, Frank Lorimer, Anatole Romaniuk and Étienne van de Walle (ed.), *The demography of tropical Africa*, p. 88–139. Princeton, Princeton University Press, 539 p.

Carrier Norman H. and Goh Thuan-Jig, 1972. The validation of Brass's model life-table system, *Population studies*, vol. 26(1), p. 29–51.

Clairin Rémy, Condé Julien, Fleury-Brousse Michèle, Waltisperger Dominique and Wunsch Guillaume, 1980. *La mortalité dans les pays en développement/Mortality in developing countries, Tome III Nouvelles tables types de mortalité à l'usage des pays en développement*. Paris, Centre de développement de l'OCDE, 250 p.

Coale Ansley J. and Demeny Paul, 1966. *Regional model life tables and stable populations*. Princeton (New Jersey), Princeton University Press, xiii+ 871 p.

Coale Ansley J., Demeny Paul and Vaughan Barbara, 1983. *Regional model life tables and stable populations, second edition*. New York and London, Academic Press, 496 p. (Second edition).

Coale Ansley J. and Guo Guang, 1989. Revised regional model life tables at very low levels of mortality, *Population Index*, vol. 55(3), p. 367–609.

Coale Ansley J. and Guo Guang, 1991. The use of new model life tables at very low mortality in population projection, *Population Bulletin of the United Nations*, no. 30, p. 1–22.

De Sandre Paolo, 1973. Primi risultati di un'applicazione dell'analisi dei fattori a tavole di mortalità italiane (1881–1966), *Atti e memorie dell'Academia Padavina di Scienze Lettere ed Arti*, vol. 86(3), p. 251–277.

De Sandre Paolo, 1974. *Introduzione ai modelli demografici*. Padoue, CLEUP, 281 p. (Serie di Statistica, no. 14).

Duchêne Josianne, 1980. *Un essai de modélisation de la répartition des décès selon l'age et la cause dans les pays industrialisés*. Louvain la Neuve, Cabay, 182 p.

Gabriel K. Ruben and Ronen Ilana, 1958. Estimates of mortality from infant mortality rates, *Population Studies*, vol. 12(2), p. 164–169.

Kamara Joseph A. L., 1988. The use of model life tables in the estimation of mortality trends for African countries, *in*: IUSSP/UIESP (ed.), *African Population Conference Proceedings/Congrès africain de population, Dakar, 1988, Volume 3*, p. 7.31–7.46. Liège, IUSSP/UIESP.

Kamara Joseph A. L., 1989. *The use of model life tables in indirect estimation*. London, University of London, 350 p. (PhD thesis).

Kamara Joseph A. L. and Lansana Antonia H., 2001. *The use of model systems to link child and adult mortality levels: Peru data*. Paris, IUSSP/UIESP, 16 p. (Paper presented at the meeting of the International Union for the Scientific Study of Population, IUSSP, Salvador de Bahia, Brazil, 2001).

Kermack William G., McKendrick Anderson G. and McKinley P. L., 1934. Death rates in Great Britain and Sweden: some regularities and their significance, *Lancet*, March 31, p. 698–703.

Kurup C. K. Ramakrishna, 1964. *Recent trends in world mortality and application for a revised system of model life tables*. Chicago, University of Chicago.

Lambert André, 1973. *Un logit standard pour l'Amérique Latine*. Louvain, Département de démographie, Université Catholique de Louvain. (Masters dissertation in demography).

Ledermann Sully, 1955. Application de l'analyse factorielle à l'étude de la mortalité, *in*: CNRS (ed.), *L'analyse factorielle et ses applications*, p. 263–293. Paris, Centre national de la recherche scientifique, 431 p.

Ledermann Sully, 1969. *Nouvelles tables-types de mortalité*. Paris, INED and Presses Universitaires de France, 260 p. (Travaux et documents, Cahier no. 53).

Ledermann Sully and Bréas Jean, 1956. Les dimensions de la mortalité, *Population*, vol. 14(4), p. 637–682.

Leridon Henri and Toulemon Laurent, 1997. Les "dimensions" de la mortalité. Notion de table type, *in*: Henri Leridon and Laurent Toulemon, *Démographie. Approche statistique et dynamique des populations*, p. 221–234 (Chapter 14). Paris, Economica, 440 p. (Économie et statistiques avancées).

Lopez Alan D., Salomon Joshua A., Ahmad Omar B., Murray Christopher J. L. and Mafat Doris, 2000. *Life tables for 191 countries for 2000: data, methods, results*. Geneva, World Health Organization, 2000, 512 p. (Global Program on Evidence for Health Policy, Discussion Paper no. 9).

Murray Christopher J. L., Ahmad Omar B., Lopez Alan D. and Salomon Joshua A., 2000. *WHO system of model life tables*. Geneva, World Health Organization, 2000. (Global Program on Evidence for Health Policy, Discussion Paper no. 8).

Murray Christopher J. L., Ferguson B. D., Lopez Alan D., Guillot Michel, Salomon Joshua A. and Ahmad Omar B., 2003. Modified logit system: principles, empirical validation, and application, *Population studies*, vol. 57(2), p. 165–182.

Newell Colin, 1988. *Methods and models in demography*. London, Belhaven Press, 217 p.

Notestein Frank W., Taeuber Irene B., Coale Ansley J., Kirk Dudley and Kiser Louise K., 1944. The future population of Europe and the Soviet Union: population projections 1940–1970. Geneva, League of Nations, 332 p.

Tabutin Dominique, 1995. *Analyse des données démographiques déficientes. Méthodes d'évaluation et d'analyse des données démographiques pour les pays en développement*. Louvain-la-Neuve, Diffusion universitaire, CIACO, 236 p.

United Nations, 1955. *Age and sex patterns of mortality. Model life-tables for underdeveloped countries*. New York, United Nations/Nations unies, 38 p. (Population Studies, no. 22, ST/SOA/Ser.A/22).

United Nations, 1956. *Methods for population projections by sex and age*. New York, United Nations/Nations unies, 81 p. (Population Studies, no. 25, ST/SOA/Ser.A/25).

United Nations, 1956. Schémas de variations de la mortalité selon l'âge et le sexe. Tables types de mortalité pour les pays sous-développés. New York, United Nations/Nations unies, 30 p. + annexes. (Études démographiques, no. 22).

United Nations, 1962. Factor analysis of age-sex specific death rates. A contribution to the study of dimensions in mortality, *Population bulletin*, no. 6, p. 155–210.

United Nations, 1982. Model life tables for developing countries. New York, United Nations, 351 p. (Population studies no. 77, ST/ESA/Ser.A/77).

Valaoras Vasilios, 1973. Model life tables as a measure of mortality, *in*: IUSSP (ed.), *International Population Conference, Liège 1973. Proceedings. Volume 3*, p. 133–42. Liège, IUSSP/UIESP, 470 p.

Waltisperger Dominique, 1985. Les tables de mortalité, *in*: Francis Gendreau, Françoise Gubry, Louis Lohlè-Tart, Étienne van de Walle and Dominique Waltisperger, *Manuel de Yaoundé. Estimations indirectes en démographie africaine*, p. 47–135. Liège, UIESP and Ordina Editions, 290 p.

Wunsch Guillaume, 1978. *Les tables logit à deux paramètres. Un test de la qualité de l'ajustement. Démographie Africaine*, Bulletin de liaison, 27, 59–67.

Wunsch Guillaume, 1984. Tables-types de mortalité, *in*: Guillaume Wunsch, *Techniques d'analyse des données démographiques déficientes*, p. 15–48. Liège, Ordina Editions, 221 p.

Zaba Basia, 1979. The four-parameter logit life table system, *Population Studies*, vol. 33(1), p. 79–150. http://web.idrc.ca/fr/ev-9435-201-1-DO_TOPIC.html. Accessed January 22, 2005.

APPENDIX 128–1 Model life table estimates of various indicators from three life tables

1. Male mortality indicators

	$_1q_0$	$_5q_0$	$_{15}q_0$	$_5q_{20}$	$_{20}q_{30}$	e_0	e_{10}
Agincourt site (South Africa): male sex							
Actual data	0.01501	0.03229	0.03856	0.01135	0.17014	66.12	58.46
United Nations model tables (1955)	0.01501	0.02275	0.02747	0.00673	0.05671	71.12	62.91
Coale and Demeny model tables[2] (1983)	0.05276	0.06712	0.07779	0.01211	0.08896	63.49	58.40
Ledermann system[1] (1966)	0.04268	0.05148	0.06052	0.01205	0.08074	65.77	59.66
Ledermann system (1966)	0.02914	0.03674	0.04650	0.01535	0.09639	65.68	58.49
Brass's logit system[4] (1968)	0.01918	0.04114	0.05548	0.01106	0.07417	67.07	60.55
United Nations model tables II[3] (1981)	0.06059	0.08045	0.09134	0.01069	0.09664	62.48	58.30
OECD model tables (1983)	0.04210	0.05891	0.06774	0.00764	0.07650	66.17	60.59
Bandafassi site (Senegal): Male sex							
Actual data	0.13860	0.25462	0.28137	0.03016	0.18520	44.76	50.53
United Nations model tables (1955)	0.13860	0.20139	0.23060	0.03429	0.17879	47.84	50.78
Coale and Demeny model tables (1983)	0.16426	0.27062	0.30175	0.03445	0.17879	43.35	50.47
Ledermann system[1] (1966)	0.18006	0.27262	0.31187	0.04697	0.23345	39.93	46.47
Ledermann system (1966)	0.17868	0.27308	0.30699	0.03497	0.17289	43.03	50.63
Brass's logit system (1968)	0.12920	0.22073	0.26837	0.04023	0.20263	44.44	51.67
United Nations model tables II(1981)	0.12199	0.17743	0.20397	0.02601	0.19092	49.09	50.51
OECD model tables (1983)	0.17295	0.25525	0.28737	0.02263	0.17009	44.81	51.23
Matlab site (Bangladesh): Male sex							
Actual data	0.04703	0.06222	0.07150	0.00305	0.05798	66.95	61.79
United Nations model tables (1955)	0.04703	0.06657	0.07604	0.01178	0.07696	65.52	60.49
Coale and Demeny model tables (1983)	0.02894	0.03436	0.03998	0.00687	0.05607	69.10	61.74
Ledermann system[1] (1966)	0.05534	0.06900	0.08112	0.01540	0.09809	62.35	57.40
Ledermann system (1966)	0.06446	0.08049	0.08987	0.00922	0.06359	64.98	61.03
Brass's logit system (1968)	0.02870	0.05442	0.06987	0.01143	0.07024	67.23	61.77
United Nations model tables II(1981)	0.03781	0.04719	0.06087	0.00576	0.07433	67.76	61.61
OECD model tables (1983)	0.03874	0.05397	0.06209	0.00717	0.07293	67.06	61.15

Notes:
[1] Life expectancy at birth, both sexes (entry)
[2] Coale and Demeny network
Agincourt and Matlab = western network
Bandafassi = southern network (high mortality under 5 years)
[3] United Nations network
Overall model for all the sites
[4] Brass system
African standard

(continues)

APPENDIX 128–1 *(continued)*

2. Female mortality indicators

	$_{1}q_{0}$	$_{5}q_{0}$	$_{15}q_{0}$	$_{5}q_{20}$	$_{20}q_{30}$	e_{0}	e_{10}
Agincourt site (South Africa): Female sex							
Actual data	0.01663	0.03369	0.04056	0.00989	0.08070	74.43	67.26
United Nations model tables (1955)	0.01663	0.02493	0.02985	0.00695	0.05756	70.85	62.81
Coale and Demeny model tables[2] (1983)	0.01258	0.01438	0.01619	0.00203	0.02838	76.04	67.21
Ledermann system[1] (1966)	0.03269	0.04032	0.04719	0.00864	0.06231	70.23	63.41
Ledermann system (1966)	0.02471	0.03179	0.04089	0.01358	0.08522	68.11	60.62
Brass's logit system[4] (1968)	0.02102	0.03878	0.04934	0.00761	0.04682	73.88	67.35
United Nations model tables II[3] (1981)	0.03011	0.03771	0.04122	0.00267	0.06284	74.16	67.35
OECD model tables (1983)	0.05944	0.03734	0.04259	0.00409	0.04726	74.41	67.51
Bandafassi site (Senegal): Female sex							
Actual data	0.11639	0.21740	0.26434	0.03146	0.14292	47.56	52.35
United Nations model tables (1955)	0.11639	0.16738	0.19113	0.02785	0.14768	52.20	53.42
Coale and Demeny model tables (1983)	0.14965	0.25978	0.29396	0.03051	0.15036	45.25	52.30
Ledermann system[1] (1966)	0.15561	0.25069	0.29199	0.04496	0.20138	42.61	48.41
Ledermann system (1966)	0.16574	0.26175	0.29640	0.03228	0.15106	44.72	52.04
Brass's logit system (1968)	0.12840	0.20930	0.25070	0.03388	0.17066	47.54	51.86
United Nations model tables II(1981)	0.10238	0.16409	0.19195	0.02655	0.16739	51.33	52.32
OECD model tables (1983)	0.14166	0.21987	0.24945	0.02615	0.16136	47.71	52.03
Matlab site (Bangladesh): Female sex							
Actual data	0.05944	0.07910	0.08537	0.00681	0.04876	67.08	62.93
United Nations model tables (1955)	0.05944	0.08400	0.09566	0.01415	0.08679	63.25	59.40
Coale and Demeny model tables (1983)	0.03176	0.03963	0.04572	0.00649	0.05937	69.81	62.89
Ledermann system[1] (1966)	0.04332	0.05552	0.06530	0.01164	0.07712	66.33	60.56
Ledermann system (1966)	0.05725	0.07248	0.08047	0.00711	0.05077	67.03	62.58
Brass's logit system (1968)	0.03671	0.06523	0.08161	0.01190	0.06980	66.97	62.36
United Nations model tables II(1981)	0.04821	0.06534	0.07322	0.00637	0.06441	67.85	62.87
OECD model tables (1983)	0.04731	0.07005	0.08007	0.00785	0.07308	67.58	63.03

OECD = Organisation for Economic Co-operation and Development

Notes:

[1] Life expectancy at birth, both sexes (entry)

[2] Coale and Demeny network
Agincourt and Matlab = western network
Bandafassi = southern network (high mortality under 5 years)

[3] United Nations network (2)
Overall model for all the sites

[4] Brass system
African standard

PART

II

STATISTICAL ANALYSES, SPECIFIC TOOLS AND QUALITATIVE APPROACHES

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

This second part brings together a number of methods that are auxiliary or supplementary to demography and are either simply borrowed from other disciplines, or the result of recent developments in particular methods which, while not entirely unconnected with demography, evolved out of (or are being developed from) exposure to approaches prompted by other disciplines. They cover a wide range of different standpoints from the philosophy of science to the most practical statistical methods. There are three main focal points to this section: the application of specific statistical methods to demography, how individual studies can inform demographic analysis, and the value of qualitative approaches in demography.

First, however, Hubert Gerard looks in Chapter 129 at the constraints on (in theory, at least) any attempt at theorization, especially verifying a theory in demography. As in any other social science, it is often impossible to satisfy all these constraints, and it would certainly have been extremely discouraging to commence this treatise on that basis, however logical that might be. But this vexed issue had to be addressed, even if only to agree on the overarching need to know one's limitations: the advancing of any particular theory should always be subject to a note of caution as to its universal applicability.

No less if not more demanding is the search for proof. Albeit one of the most quantitative social sciences, demography like all its fellow disciplines is all-but unable to furnish certain and incontrovertible proof of any causal link. In most cases, the best that can be established is some probability. And even that requires the basis of real causality to be established. A number of causal models can and must be defined so as not to stray from the objective. This is the topic addressed by Josianne Duchêne and Guillaume Wunsch in Chapter 130.

In the following two chapters, Laurent Toulemon reviews various general statistical methods that are particularly useful in demography: multidimensional exploratory analysis (Chapter 131) and regression analysis (Chapter 132). In both cases, the author highlights the value of these methods to demography, exemplified by particularly apposite cases in point from the literature.

Two Chapters then set out to look further into the issue already raised at several points in preceding volumes in relation to particular topics—that of the individual level. While the quantitative analysis of collective macro data is central to the demographic approach, the main aim of which is to understand the driving forces of population development, the question arises as to whether the micro individual behavior component of the mechanisms observed may undermine the undertaking at the final explanatory stage. The further the discipline advances, the more acutely its initial approach is felt to be approaching its limits. The issue then is about going back to the individual level. But how and using what methods? There are many possible answers to this. An outline of just three alternatives is given here.

In Chapter 133, firstly, Francesco Billari puts the question of whether there is really an unbridgeable gap between the micro and macro approaches, or whether an improved dialogue between the proponents of both might be mutually informative and advance matters somewhat more or somewhat better towards the ultimate explanation; although that may still be as unachievable a quest as finding the Holy Grail.

In Chapter 134, Dominique Tabutin addresses the less far-reaching but also more practical topic of developing individual demographic indicators that can in many cases, if not provide firm answers, then at least dispel some ambiguities, by verifying whether the ostensible findings from aggregate indicators are borne out at this more forensic level of study.

In Chapter 135, Alberto Palloni and Paolo De Sandre return to more traditional concepts of demographic analysis, but also come at explanatory research from a different, more advanced and detailed angle through further mathematical analysis of all-order and order-specific fertility.

The two concluding chapters move into an entirely different field, showing how areas which demographers would certainly be misguided in regarding as the preserve of sociology can inform demography. More specifically, Michel Bozon demonstrates in Chapter 136 the benefit that demographic research could derive from a more frequent and purposive use of qualitative methods, while Alexis Ferrand and Ainhoa de Federico de la Rúa offer insights into the potential that an emerging field of sociological study—network analysis—offers for many issues in demography.

C H A P T E R

129

Theory Building in Demography

HUBERT GÉRARD

Université catholique de Louvain, Louvain-la-Neuve, Belgium

In demography, as in social sciences generally, *theory* is a polysemous term whose different meanings have their own relevance and value. Proof of this can be found in scientific periodicals, in the chapters of this book, or in systematic studies, such as those of Peter Swanborn (1990) or Robert Merton (1945). The latter, published 60 years ago, has become a classic but remains completely current. A number of these meanings can be arranged along a continuum, progressing from a precisely defined and elegantly rigorous theory[1] to one that is flexible enough to encompass virtually anything.

This chapter does not pontificate on what is a "real" theory, reducing all other types to more or less close approximations; it seeks instead to present an approach—already shared by many demographers and social science researchers—that engages the problem of theory building and of defining theory in a less doctrinal fashion, one closer to research as practiced. Rather than defining theory at the outset and then presenting ways of constructing it, I prefer to start with the process of theory building, the process that leads to the construction of one or more theories, that compares theory with reality, and that incorporates these findings into a new theory, which itself is submitted to the process of testing and synthesis. Although I find it useful to present this process in a sequential manner, starting with formulation of the initial question and progressing through theory revision, in actual research, the process has no mandatory path. At any time, it may set off in a number of directions, according to the needs of theory building.[2]

I. THEORIZATION

Without becoming mired in discussions about the definition of science while nonetheless aligning myself with Jean Ladrière (1996), who characterizes the scientific process as both reflexive and prospective, I postulate that science can be viewed simultaneously as the controlled production and communication of knowledge.[3] I also attempt to demonstrate that theory building is the suitable procedure to ensure this control. For a demographer to maintain a critical distance from his or her research, to open it up to new avenues of thinking, and to better communicate the findings, it is essential to be aware of all the twists and turns of its layout, its rationale and foundations, its explicit premises, its strengths and weaknesses, its potential and limits.

[1] Such as Johan Galtung's definition of theory (1973, p. 451), "as a set of hypotheses structured by the relation of implication or deducibility," or that of Paul-Marie Boulanger (1980, p. 11), "a theory is a system of hypotheses relating to a determined factual object, such that each element in the system is either an initial proposition (axiom, postulate, datum), or a logical consequence of one or many of these initial propositions." As Hubert Blalock posited (1971b, p. 156), such theories are far from frequent in social sciences.

[2] This was also emphasized by Hox and Mellenbergh (1990, p. 133).

[3] This material is largely based on my contribution at the Chaire Quetelet, 1987 (Gérard, 1989).

1. Definition

I define *theory building* as "a process of critical systematization of acquired knowledge in a designated corpus of specified propositions generating hypotheses that can be tested against empirical evidence" (Gérard, 1989, p. 271).

First, the process is continuous throughout the research, ideally moving along four fundamental axes. The first axis refers to the scope of the burgeoning theory; this scope is limited to a single situation or specific cases or is more broadly applicable, even claiming universal validity. The second axis concerns the theory's formulation; it ranges from theories that are formulated in ordinary language and are relatively imprecise in the definition of concepts and of the relationships linking them to highly formalized theories. The third axis refers to the qualities of the theoretical framework organized in this way, moving from the simple juxtaposition of propositions to their integration into a hypothetical-deductive corpus.[4] The set of statements organized to create testable hypotheses are tested against empirical evidence, and the extent of testing forms the fourth axis, which ranges from sets that have never been empirically tested to sets that have already been confirmed or refuted by means of relatively substantial testing.

This process begins as soon as the initial question is formulated, and its progression along the four axes is far from linear because of opposing movements of expansion and contraction. Expansion broadens the developing theory's scope, for example, to include a greater number of concepts or to identify new dimensions of existing concepts and give greater complexity to their relationships. Contraction encourages a more rigorous definition of concepts and of their interrelationships, even to the point of reducing their scope, eliminating concepts that escape precise definition, or spelling out the theory's scope more rigorously. This dual movement condemns theory building to be forever incomplete and theory to be no more than an instrument of ephemeral knowledge, constantly to be questioned and improved.

Second, the process is one of critical systematization, aiming to build a set or a designated corpus of propositions in a critical way with clear awareness, explicit formulation, elucidation, justification, and evaluation of what is being constructed, of the manner in which it was achieved, and of its underlying rationale. This type of systematization should make it easier to move ahead on the four axes and should indicate the direction, the detours, and the furthest recesses of theorization. To do this, the concepts and their interrelationships, which together make up the various propositions integrated in this corpus, must have been defined beforehand in a critical manner.

Third, systematization is related to *acquired knowledge*, a generic expression designating the theories and findings of research carried out in identical or neighboring fields—whose relevance for the current research should be specified—and all the *a priori* beliefs, attitudes, value judgments, and stereotypes, which in subtle and hidden ways inevitably introduce all kinds of biases into the research. Rather than ignoring them, these biases need to be spelled out, criticized, and selected; those that seem relevant for the current theorization should be incorporated, and the others should be controlled for as far as possible.

Fourth, the acquired knowledge, thoroughly scrutinized and critically assessed in this way, is then transformed into propositions integrated into a designated corpus—a set of propositions whose characteristics and claims have been made explicit and evaluated. In demographic areas of study, this corpus rarely consists of a deductive set of propositions clearly defining its basic[5] and derived concepts and of its postulates[6] and the propositions inferred from them. Instead, it is a composite set, with contours in need of definition, and one integrating, in a relatively precarious manner, empirical generalizations or propositions founded on analogy, plausibility, or common sense, whose quality needs to be evaluated.

Fifth, identifying and understanding the characteristics and claims of this set, or designated corpus, depend on the manner in which its constituting statements and their relations are designated. Each proposition and every relation between propositions should be assessed in terms of their theoretical foundations and how they stand up in the face of empirical testing.

Sixth, from this set, which is defined, evaluated, and justified in the manner previously described, should follow *hypotheses that can be tested against reality*. Theory

[4] These are logically deduced from postulates that are not necessarily all true or sure (Lalande, 1988, p. 430).

[5] The basic concepts, also known as *primitive concepts*, are stated and often defined from the very beginning of theory building, during the course of which they continue to be added to. They are based on other theories, earlier research, premises stated from the outset, or empirical observation. Their meaning should be stated explicitly without requiring any justification from the elements of the theory being developed.

[6] Like primitive concepts, *postulates* are assertions posited as though they should be accepted as such, that do not need to be derived from the theory being constructed nor tested by it. This should not prevent us from supporting them up as much as possible and justifying their selection for the relevant theory.

built in this way operates as a tool for the production of knowledge, rather than being a mere statement summarizing the end results of research. "Taking note of new knowledge, theory acts above all to display, or ruthlessly expose, everything that contributes to formulate new questions and put forth new hypotheses to answer them, hypotheses elaborated in such a way as to be themselves testable" (Gérard, 1989, p. 172–273).

2. Basic Principles of Theory Building

If, like Mario Bunge (1975, p. 197), we admit that "no technique exists that makes it possible to construct a theory *ab initio*," it is nonetheless possible to suggest some principles that ease progress along the four axes and preserve the theory's status as an instrument of knowledge. I discuss four of the principles: transparency, clarity and precision, empirical testing, and relativity of the theory.

The whole process of theorization, from before the initial question through the results of empirical testing and their integration into a theory, should be as *transparent* as possible to the researcher and to those to whom the research will be communicated. Flaws, weaknesses, and badly constructed concepts or interrelations between concepts should not be disguised, nor should the *a priori* ideologic positions or preconceptions be concealed. Everything should be exposed to the critical evaluation of the author and his or her readers. The researcher needs to know himself and to make that knowledge known; he needs to openly state *a priori* his personal choices, even his feelings about his research object. He should be aware of his discipline's presuppositions and of the theories, methods, and techniques he is using. He should always know and explain what he is doing, how he is doing it, and why.

The principle of *clarity and precision* is fundamental for theory building to progress, even though it tends to lead to contraction rather than expansion. It should therefore be used with care to avoid hindering the type of creativity that thrives in blurred areas, a more fertile soil for developing new concepts or new dimensions of already established concepts, and for adding to the analytical tightness to the relations linking them. This principle leads to tracking down all implicit postulates, to understanding their relevance, and to defining concepts and the chosen interrelations with greater precision, while making explicit their necessary characteristics. It helps to explain the relations established between the different propositions integrated by the theory.

It may seem astonishing to consider *empirical testing* a principle of theory building. Testing often is relegated to the end of the process of theorization, and Hubert Blalock (1969, p. 5) even recommends a division of labor between those who build theories and those who test them. However, I consider theory and empirical evidence as obligate partners, whose relations are admittedly rather complicated. Theorizing can as easily take off from an abstract question, such as from one or more empirical observations. Theory and reality are in constant dialogue throughout the entire process to confirm or refute what one or the other suggests, to open up new avenues, and to refer to those that have already been traveled. As it progresses, theory will attempt to demand more from empirical evidence than mere illustrations or cursory analogies. For its part, empirical evidence will demand well-structured concepts and relations between these concepts to be able to elaborate relevant indicators for these very concepts and their associations, thereby answering theory's demands. To make this dialogue more relaxed and effective, it is also necessary to rely on such artificial constructs as the main and auxiliary theories proposed by Hubert Blalock (1990; 1971a; 1969, p. 5, 151–154).

The principle of the *theory's relativity* highlights the fact that theory is a tool for producing knowledge, not an end in itself. It claims no truth for its statements but presents itself as an ephemeral structure that can always be shaken by empirical testing and whose premises can be questioned at any time.

3. Markers for a Course with No Compulsory Route

Each research process is necessarily unique and is designed according to numerous elements arising simultaneously from the question studied, from the number and quality of studies already carried out on the topic, from the researcher's own characteristics, and from the objectives and conditions of research. Nevertheless, the course that research follows can usefully be marked by key points, or *moments*, when the decisions to be made involve the whole of the theorization process. I identify six moments: the initial question, the main theory, the auxiliary theory, the manner of testing them against empirical evidence, the techniques used for testing, and the results.[7] Even if, logically, they are set down in sequence, in practice, they are articulated in various ways that do not have to follow a linear direction; at any point, they often come back to earlier moments.

[7] These distinctions refer to Hubert Blalock's studies and to the work of Josianne Duchêne and Guillaume Wunsch (1985), who added the *statistical model.*

The *initial question* (i.e., object of the research) supposedly precedes all the other moments, but the latter may lead the researcher to specify, direct, and transform the question in various ways, depending on the case, or even to replace it with another one.

The process of theory building begins as soon as the initial question is formulated; so do elaboration of the main theory—the set of propositions constructed theoretically as well as empirically—and generating hypotheses to be tested empirically. In most cases, scholars present this main theory in ordinary language, even if they take particular care to construct clearly and explicitly the concepts to be used and their interrelations.

In general, the main theory cannot be directly tested against empirical evidence; it needs to be adapted to the concrete conditions of the testing procedure and to the real possibility of collecting and analyzing empirical data to test the hypotheses. The concrete conditions are determined by the objective of the research undertaken (e.g., responding to a request from a ministry or another organization, a term paper, longer-term academic research) and by the researcher's own ability, the time available, and the financial and technical resources at his or her disposal. Moreover, gathering and analyzing empirical data generally involves translating concepts and their connections in terms of the indicators required by the testing method and techniques selected. Adapting and translating the main theory in this way gives rise to one or more *auxiliary theories*, which should be used to assess if these necessary adjustments to the main theory, together with the new propositions that had to be postulated to do this, do not excessively reduce the possibility of empirical testing.

Equally important in the process is selecting a *testing procedure*. This choice also influences the formulation of the auxiliary theory or theories and indirectly influences the main theory that will actually be tested. Without assessing the different procedures available, I discuss quantitative and qualitative procedures and the procedures that use data that is or is not specific to a given research.

Despite the fact that some authors[8] have asserted the primacy or even the exclusivity of quantitative methods for conclusive hypothesis testing, others[9] have attempted to give rigor and quality to qualitative methods in this field. It appears completely legitimate to choose between the two types of procedures or to combine the two. Because these procedures have very different requirements in terms of operationalizing concepts and their associations and of the type of data needed, their auxiliary theory or theories are also different.

To test a theory against empirical evidence on the basis of data collected expressly for this purpose will undoubtedly meet fewer constraints than to do so using existing data gathered for other study purposes. The choice between these two types of data has a bearing on the auxiliary theory and, in some cases, on the main one.

The same applies to the *techniques of observation and analysis* used for testing. Far from neutral for the research, they are hedged in by various postulates and constraints that must be taken into consideration in the auxiliary theory and often in the main one.

After the *results* are in, they need to be assessed with respect to the main theory to ensure that all the hypotheses that the main theory required to be tested have been. If they have not, it is imperative to specify those parts of the main theory that have effectively been tested and, if need be, to decide on how the other results can be used. When the research is over, its findings should make it possible to appreciate the extent to which the main theory answered the initial question and to outline new directions for future research.

In practice, the researcher simultaneously works on the main theory, auxiliary theories, empirical testing procedure and techniques, and to some extent, the results themselves. This is generally the case in qualitative approaches and frequently so in quantitative ones, even in research based on extensive and specific observations that require the first draft of main and auxiliary theories to have been formulated and the testing procedure to have been chosen before starting to collect the data.

The research's key points or moments presented earlier are useful points of reference, but they are in no way a recipe. Each investigator must make the best possible use of them for the research in progress.

II. MAIN THEORY

Developed from an initial question about a given set of facts about the research, the main theory is a structured set of statements, from which are inferred hypotheses to be tested. This main theory and its derived hypotheses claim to reconstruct, partially but meaningfully, the initial set of facts. Usually, the latter

[8] Even Pau1 Lazarsfeld (1970), who was open to qualitative approaches, believed that the existence of a relation between two variables can be truly tested only by statistical analysis or experimentation and that qualitative studies lead only to confirmation or corroboration.

[9] Examples include Michael Huberman (1981), Michael Huberman, and Matthew Miles (1991), as well as the work of Barney Glaser and Anselm Strauss (1970).

is already partially constructed from other theories or from other types of knowledge, such as common sense.

To speak of a partial reconstruction is to recognize that theorizing is not about explaining the data's tiniest details or all the earlier constructions but is concerned with reconstructing the data set in a partial and simplified manner and according to a certain number of reasoned choices explained and assessed by the theory.

Although partial, this reconstruction is nonetheless meaningful; it makes sense in relation to the set of facts (or gives them meaning) and highlights from that set the elements that are essential and sufficient for that theory's project. The essential and crucial elements are those deemed necessary for the reality itself of the facts being reconstructed; they are sufficient from the moment they allow its complete reconstruction, at least relative to the theorization project. A theory aiming to account for high fertility in Africa reconstructs this fact differently when read from the perspective of demographic theory, economic theory, or sociologic theory.

When elaborating the main theory, three elements are particularly important: conceptualizing or constructing concepts, establishing relations between them to build premises and hypotheses, and organizing these assertions into a structured and coherent whole, which was previously called the *designated corpus*.

1. Conceptualization

a. An Absolute Necessity

Although often neglected in general research methodology textbooks in demography and other social sciences, the question of conceptualization is all the more important that social scientists generally use ordinary language rather than neologisms or formal terms or terms sufficiently formal not to be open to interpretations other than the intended one.[10] One exception is demographic analysis *stricto sensu*; its key concepts are rigorously defined, leave little room for various interpretations, and are usually directly measurable.

Everyday language enables us to communicate well in daily life. Up to a point, it makes it possible to label objects, express our moods, think, and communicate. Given the semantic and emotional load it has accumulated over time, it also works as a lens that distorts our perception. Words have an interpretative role that creeps into speech and affects our views. For instance, in the animated and sometimes violent debates at scientific meetings, speakers can use the same words, such as demographic policy, but give them different meanings without ever being aware of it. In the demographic literature, anyone can see the ambiguous nature of much-used concepts such as development, elementary needs, family, family head, household, education, and the desired, expected, or ideal number of children. It is therefore essential to break with the polysemy that words carry and to use words from everyday language in such a way as to make them rigorously and clearly express the only meaning we wish to give them.

An important postulate of our process is that concepts do not exist in themselves; they exist only through creation, use, and convention. In a given discipline or even more in its schools of thought, some concepts are created in a similar way and even codified in specialized dictionaries.[11] It is useful to standardize concepts in such a way as to make possible the communication of knowledge or its accumulation through time. Nonetheless, variations whose scope fluctuates according to the concepts, the discipline, the schools of thought, and other elements, including the context or evolution of the facts themselves, are also acceptable.

For several decades, civil marriage or its equivalent seemed to be the incontrovertible original event for legitimate childbearing, the main source of fertility in most Western countries. Evolution of the types of union since the 1970s has altered the demographic and social meaning of marriage radically enough to make it necessary to modify the standard concepts associated with it.

All investigators claiming to do scientific research must have a firm grasp on the conceptualization of all the terms they use, and these should be similarly understood by those to whom the research is presented. This level of understanding is produced through a new, well-explicated definition of concepts, or by adopting a particular conceptualization recognized by the researchers and those interested in reading the research's results, or by accepting, most often implicitly, the conceptualization generally held by the research's actual and potential readers and shared by the author. As the audience widens, their understanding of concepts risks becoming that of common sense, with all the imprecision and semantic variability that goes with it.

[10] This section of the chapter is mainly based on the work of Giovanni Sartori (1984).

[11] For instance, in demography, Louis Henry (1981) or Roland Pressat (1979).

b. A Complex and Finely Shaded Process

Concept building or conceptualization means giving a word a precise meaning in a manner that is relevant and grounded in sound reasoning and crafting it such that it designates the very objects related to that meaning. It also means breaking with the other meanings and emotional values attached to the word; this is often the most difficult task, so much so that some words need to be treated with extreme caution or even completely avoided. To different degrees, this is the case for ethnic groups, ethnic cleansing, race, and assimilation of foreigners.

The conceptualization should be relevant and built on solid foundations. Relevance is assessed against the aims of the given theorization and research. An exploratory study on the link between infant mortality and mothers' educational levels, based on national aggregate data, may rely on concepts defined in a relatively crude manner; this would be unacceptable, however, for a study of the influence mothers' education has on infant mortality in a particular country.

The foundations of conceptualization can arise from three main registers:

1. From theory, when concepts are constructed from an earlier theory or from an original use of diverse theoretical considerations
2. From empirical observation, when priority is given to individuals' life experiences and conceptualizations or to name and conceptually abstract new empirical observations
3. From usage, if we seek to explain how the concept is generally constructed within a particular discipline (as is generally the case with dictionaries), within one of its schools, or even by a particular author and, in some cases, by common sense

The process of conceptualizing can be grounded in the interaction between these three registers and be so intimately interconnected that it is often difficult to find out which one is at work in a particular instance.

Examples are provided by two definitions of population policy, one rooted mainly in use and the other in more theoretical considerations (Gérard, 1983). Definitions, as we shall see, are important but not sufficient parts of conceptualizing.

Deborah Oakley (1977) carried out a content analysis of 34 definitions published between 1940 and 1975 by well-known demographers or political scientists such as David Glass, Joseph Spengler, Alfred Sauvy, Kingsley Davis, and Hope Eldridge. Oakley observed widespread agreement on the fact that population policy is a governmental action (mentioned in 30 of 34 definitions) that influences or acts on (33 definitions) demographic variables (25 definitions), directly or indirectly (27 definitions). Opinions are more divided on whether nongovernmental organizations should be included (11 in favor, 17 against) and whether the action should be intentional (13 in favor). From this analysis, the Oakley discerns a definition of population policy based on those used by the scientists studied: "Governmental population policies are those actions of government that affect or attempt to affect the balance between births, deaths, and migration of human beings" (Oakley, 1977, p.18).

I wish to define population policy with respect to a question: To what extent can a society voluntarily control its demographic processes? On the basis of mainly theoretical considerations, I suggest the following definition: "A population policy is a system of integrated programs created and implemented with the explicit aim of guiding or changing population dynamics in a particular direction" (Gérard, 1983, p. 25). Such a definition makes sense only with respect to the question asked and in no way represents the definition of a population policy in itself.

Conceptualization should give words precise meanings to remove all uncertainty or avoid possible mistakes. If a definition that states and gives meaning (i.e., a *declarative definition*[12]) can largely reduce or even eliminate all ambiguity, it often remains inadequate for giving the desired meaning with precision and for identifying all the objects that it should designate. As a consequence, it is essential to identify the concept's every characteristic by distinguishing the characteristics that are necessary, without which the intended meaning cannot be given and no object can therefore correspond to this concept, i.e., from those that are contingent, whose presence or absence in no way affects the concept's meaning. A concept's whole set of characteristics and their division into necessary or conditional subsets constitute the concept's *comprehension.*[13] The process of constructing a concept to which a word is attached requires that its comprehension be delineated with care. After this designation established, the selected word no longer refers to anything else, except in a purely contingent manner, and no other word should be used to give this meaning.

In the text (Gérard, 1983, p. 13–26) from which the declarative definition of population policy has been taken, the concept's comprehension consists of the fol-

[12] This is Giovanni Sartori's term (1984, p. 29).

[13] Some authors also talk of the concept's *intention* or *connotation*. According to André Lalande (1988), these do not exactly mean the same thing; in the context of this chapter, however, they are used synonymously.

lowing: (1) to fall under the authority of an organization with powers of decision and implementation in the population sphere; (2) to be action oriented, made up of a set of integrated programs; (3) to have been deliberately created and implemented; (4) to guide or change population dynamics as such, not only one of its components; (5) to have a scientific and ideologic basis for setting the desired population targets, the means to reach these, and the ultimate objectives that these population goals should make it possible to achieve, such as development or well-being, for example. Of these characteristics and given the purpose of the suggested conceptualization—to know the extent to which a society can voluntarily control its population dynamics—the first four are necessary, and the fifth is contingent.

Conceptualization should make it possible to *specify flawlessly all the objects encompassed* by the established meaning and to exclude all others with equal rigor. In this manner, the concept's *extension* (some authors also speak of *denotation*) *is established*, which consists in setting up the criteria according to which every object can be evaluated without risk of error. These criteria are necessarily intimately linked to all the concept's necessary characteristics, because the latter, taken together, are the only ones to establish the concept's meaning. To make things easier, Giovanni Sartori (1984) suggests constructing an *operational definition*, a definition in extension, restricted to observable or even measurable characteristics and providing criteria better adapted to locating and, if necessary, to measuring the objects, even at the expense of reducing the concept's comprehension,[14] albeit only up to a level considered acceptable.

Using the population policy example, a concrete policy could be viewed as falling under the compass of the suggested definition if (1) it comes within the province of national government or of a political body answerable to it, insofar as it has the power to decide and implement the policy; (2) it constitutes a system of programs, not a series of measures taken independently; (3) its demographic objectives are explicitly stated; and (4) its relation to population dynamics as such is clearly explained.

Close to definitions in extension is the usual practice of associating a given word with a specific combination of observations—to indicate, for example, the factors in a factor analysis or some other constructions of empirical data reached through various statistical method, or through thinking and trial and error. The choice of words often seems left to chance or to the researcher's creativity. From the perspective taken here on conceptualizing and to break with the range of meanings and emotions attached to everyday language, it would be best to consider factors or other sets of empirical data as indicators of necessary characteristics of concepts to be constructed "in comprehension." The gap is often wide between what is proposed here and what is done in practice.

Concepts deserve particular attention and effort. Concepts are constructed to serve as tools of controlled knowledge, specifying single meanings for terms. The tools of controlled knowledge should make it possible to understand concepts in the same way when used in the context of the theory. Evidently, all other concepts—the secondary concepts used to construct the theory's concepts—need to be understood in the same unequivocal manner.

Should we therefore define all the secondary, tertiary, and higher-order concepts? Would there be no end to it? A healthy approach would be to construct all the concepts used, whatever their order, unless there are good reasons to believe that these concepts have a single, obvious meaning for everybody or that the possible range of semantic variation remains within the accepted limits to keep the same comprehension of key or primary concepts or because their precise meaning is of little relevance.

2. Establishing Relations between Concepts

The concepts constructed in this way are then related to one another to elaborate the theory's propositions and the hypotheses to be tested. As with the concepts themselves, these relations should be constructed in an appropriate and well-founded manner, to give them precise content. Their *appropriateness* is assessed with respect to the theory's objectives (i.e., to explain, predict, or act) and to the proposition's place within the theory being constructed.

The *foundations* or bases of the relation to be constructed can be (1) a system of postulates deemed satisfactory, when the proposition has been derived in a deductive manner; (2) findings from earlier studies that have held up to critical re-examination and that may not relate to the society under observation; (3) results of exploratory research; (4) a generalization from singular observations, if its grounds are explicitly stated and appraised; (5) the *a priori* argument, whose foundations the investigator attempts to spell out and critically examine; or (6) other elements as long as they are explained and stand up to critical evaluation.

[14] Giovanni Sartori (1984, p. 31) states, "The operationalization of a concept often entails a drastic and eventually distorting curtailment of its connotation."

These foundations are not equivalent and do not endow the constructed relations with the same validity; hence, the importance of identifying them and submitting them to fine critical analysis.

The *exact content* of these relations, except during an exploratory stage, cannot be elucidated merely by juxtaposing statements or through empirical observations. For example, since 1964, fertility has been falling in European countries, and divorce rates started to rise, slowly at first and then much more rapidly. In another example, Algeria's total fertility rate was 5.4 in 1991, with infant mortality at 74 per 1000; 10 years later, in 2001, the rates stood at 3.1 and 55 per 1000, respectively.[15]

As with concepts, relations should be defined "in comprehension," and their necessary characteristics should be identified so as to operationalize the relations, because such operationalization is essential to testing. How can these relations be given an exact content, or how can accurate relations be constructed between events (e.g., , the number of births and infant deaths) or between some variables (e.g., education) and some observed behaviors at the individual or aggregate leve (e.g., breast-feeding)?

For a first approximation, let us look at two main types of relations or models of relations: causal and systemic. In causal models, the constructed associations aim to clarify how effects are produced by earlier or at least concurrent causes (see Chapter 130). In the systemic model, the constructed relations form the framework of a system whose elements are in constant interaction. In the first case, we talk about causes, factors, favorable or unfavorable conditions, direct or mediated determinants, dependence, function, influence, or impact.[16] In the second, we find terms such as information flows, their transformation into action, regulation, balance, retroactions, positive or negative circle, and so on, although without completely rejecting any causal perspective.

Elucidating a relation implies *laying bare its mechanisms* and assessing the validity of our knowledge about them. The relationship between women's education and lower fertility levels in a high-fertility African country is a case in point. We may believe that education provides women with the necessary information for reducing their fertility. However, we need to ascertain that this is effectively the case, that classes are taught on this specific topic in the school program taken by the women and that the women have sufficiently assimilated this information to put it into practice when they need to. We may instead assume that education introduces women to social groups characterized by a cultural model[17] of lower fertility or by a greater openness to cultural models competing with the dominant one and more oriented toward reduced fertility. In the latter case, we would need serious analytic grounds to put forth such a thesis and, if possible, have substantial empirical evidence to sustain it.

Spelling out the mechanisms of this relation makes it possible to specify its content, which may prove much more complex than initially expected, and to identify the necessary characteristics. It also highlights possible intermediate terms between the related concepts and other elements that may be related to these intermediate terms. It therefore provides a clearer insight into the other characteristics of the relationship.

A standard set of questions[18] provides useful guidelines to help specify the relation's mechanisms. Originally devised for the causal approach, these questions seem to have wider application.

1. What model does the relation refer to: causal model, systemic model, or other types, such as relations of belonging, implication, or reference?
2. Does the relationship go in a particular direction and, if so, which? Does it start from one of the two concepts involved (i.e., irreversible relationship), or can it start from one or other of the two (i.e., reversible relationship)? In the latter case, we could also specify the possible conditions, such as time constraints on reversibility.
3. Is the direction positive or negative? Is it associated with an increase or a decrease of one of the given concept's characteristics?
4. With what level of certainty does the association occur? If it does so with complete certainty, the relation is deterministic; if not, it is stochastic.
5. Does it occur immediately (i.e., concurrent relationship), or does it involve a time lapse (i.e., sequential relationship)?
6. Are the two concepts involved sufficient to establish a relationship (i.e., sufficient relationship),

[15] According to data from *Population et sociétés*, 1991, no. 225, and 2001, no. 370.

[16] These terms are not all synonymous, and authors often prefer one or another without explaining the reasons for their choices or the exact meanings they give the terms. It is time to change this practice, which makes empirical testing impossible.

[17] "Cultural models crystallize and interpret the influence that different elements of the sociocultural system may have, even by chance and in an ephemeral way, on procreation, mortality or special mobility; in so doing, they form the type of mental orientations and behavior that society exudes, suggests and values with respect to each risk and to which the subjects adhere undoubtedly more unconsciously than consciously" (Gérard, 1995, p. 48).

[18] Inspired by Hans Zetterberg (1965, p. 69–74).

or is another concept essential (i.e., conditional relationship)?

7. Can the primary concept be replaced by another one, giving the same result for the other concept (i.e., contingent relationship), or is it indispensable (i.e., necessary relationship)?

To illustrate, let us take the example of the association between women's education and reduced fertility in an African country. Education would introduce women to a social world characterized by a cultural model of reduced fertility, one competing with the dominant model. As a consequence, it leads them to identify with this alternative model, to wish for fewer children, and to behave accordingly. This relation belongs primarily to a causal model; it can be broken down, however, into several relations of a different type.

One is a *relation of belonging* that educated women have with a particular social universe. This relation has a direction, starting from education, and it is not reversible.[19] It is neither positive nor negative, but it is possible to foresee a minimum education threshold necessary to belong to this world. Even without much information, a stochastic relationship can be assumed, although it would be necessary to estimate its probability. The relation is sequential, except perhaps for women with educated mothers; it is not necessary if education is not the only condition for belonging to this group and if it can be replaced by economic status. Neither is it sufficient if schooling does not prove to be the only characteristic on which this group is founded and perhaps not even the main one.

Another type is a *relation of reference* to a particular cultural model This relationship has a direction, moving from the group to which an individual belongs toward the cultural model; it is not oriented positively or negatively. It is stochastic if we assume that some of the group's women do not share this particular cultural model for various reasons, which would have to be studied in another research. It is concurrent even if we admit that the reference does not occur immediately on entering the group.[20] It is not sufficient, given the assumption that some women in the group may not adhere to this cultural model. It is necessary from first principles, because reference to a cultural model is posited as inescapable.

A *determinant relation* starts from the cultural model and ends in specific mental, behavioral, and physiologic elements that characterize the individual with respect to procreation.[21] This determining relation implies that the cultural model defines the limits of possible variations of the specific mental (e.g., representations, knowledge, opinions, attitudes, projects), behavioral (e.g., sexual relations, contraception, abortion), and physiologic (e.g., fecundability, infecund periods, intrauterine mortality, likelihood of twins) elements. This determinant relation is irreversible, and it is neither positive nor negative. It is also stochastic but highly probable. It involves a time lapse, one difficult to define, between the model's assimilation through socialization and its realization under social control; it is necessary and sufficient.

It is more problematic to try to describe *causal relations* between specific individual elements in this case, lacking any reference to a well-defined population. A first causal relationship starts from mental elements, proceeds through behavioral ones, and ends up with physiologic elements. Moreover, all kinds of causal relations exist between specific elements, including mental (e.g., effect of knowledge on attitudes, effect of attitudes on projects), behavioral (e.g., effect of frequency of sexual relations on contraceptive use, effect of contraceptive failure on abortion), and physiologic (e.g., effect of intrauterine mortality on temporary infecundability) ones. Relations can be observed that have their roots in physiologic elements and that influence mental and behavioral elements (e.g., lower fecundity brings about a greater lack of precision in the projected birth spacing and number of children or leads to lower contraceptive use). Despite the problems involved, it also seems useful to try to assess whether the cultural model is more closely related to aspirations, motivations, and projects than behavior or whether it affects behavior first.

This effort to spell out relations is rather onerous and involves many postulates that may hinder theory building. However, it makes for much more precise empirical testing.

3. Organizing a Set of Propositions

After they are made explicit on solid and rational grounds, the propositions are related in such a way as to form a coherent set or a set with at least some degree of coherence. This set defines the theory and gives rise to the hypotheses to be tested. Some of these propositions making up the theory's theoretical underpinnings or acting to link with other propositions are postulated without having to be tested within the

[19] It can be assumed that belonging to these groups through an educated mother implies that the young girl is more likely to be educated herself.

[20] We could also assume this reference to have developed before entering the group if a person strongly wishes to join the group.

[21] More detail on specific individual elements can be found in Gérard (1995, p. 46–47, 55–57).

current research because they already have been confirmed empirically or because they have been introduced through speculations deemed sound. Some propositions are undoubtedly better founded and more clearly explicated or relevant than others. When aware of this, it becomes possible to identify the theory's weak points, which may need to be strengthened at times.

In principle, the theory's propositions should be articulated into a hypothetical-deductive whole. Even though greater flexibility may be called for in demography and the social sciences in general, it seems essential to spell out how the different propositions fit together by identifying all the postulates—whatever their soundness and plausibility (assuming that we can evaluate them) and whatever their number (often quite high)—and to explicate as much as is possible and without forgetting to mention the manner in which the other propositions and hypotheses are linked to these postulates and themselves connected. This view on the organization of propositions and hypotheses follows the style of what Lee Freese (1980, 1981) calls *systematic theory*.[22]

III. AUXILIARY THEORY

To test the main theory empirically, it needs to be adapted to the concrete conditions of this testing, to the real possibility of observing the hypotheses to be tested, and to the testing models and techniques to be used. This adaptation requires a language other than that of the main theory: the operational language (Blalock, 1971a, p. 23–24). As Hubert Blalock points out, these two languages are interconnected neither naturally nor by definition, and it is essential to bind them to each other by stating the necessary postulates that, if not grounded in solid reasoning, can imperil the testing process.

Adapting the main theory and translating the hypotheses into operational language are the two tasks of the auxiliary theory. As a result, the auxiliary theory gains from being closely linked to the construction of the main theory.

1. Adapting the Main Theory

The *concrete conditions* to which the main theory needs to be adapted depend on the researcher and the research environment. First, it depends on the research's main purpose, such as work within the framework of a teaching program, a conference paper, an article or book, a study commissioned by a ministry or an international organization, or research within a long-term contract. Also important are the time available to carry out the research, the financial and technical resources at the researcher's disposal, and the idiosyncrasies of his or her training and preferred work methods (i.e., quantitative methods or specific statistical techniques, interdisciplinary comparative approaches, and qualitative approaches). The conditions also relate to the object of study, depending on the degree to which it has already been studied and on the wealth of data available.

The auxiliary theory must take account of the *real possibilities of observing* the hypotheses to be tested through observation of the concepts in all their wealth and many dimensions or through the relations established between them. In this regard, the connection between the main theory's language and the operational one to be developed by the auxiliary theory raises a considerable problem. This is discussed later with respect to the study of indicators. Usually, the auxiliary theory enforces greater precision in the definition of concepts and their relations and reduces the scope of their claims. In some cases, the main theory must be scaled down to remain the theory to be tested empirically.

The *testing models* and associated *techniques* also impose constraints, postulates, and predetermined *forms for receiving* the observed data on the concepts and their relations.

Several auxiliary theories are often needed, such as when combining testing models, when working with several populations for which the same indicators are not available, or when we wish to use parallel indicators within the same population. If, *a priori*, several auxiliary theories seem to better corroborate or refute the hypotheses to be tested, if at least the results converge, it remains to know how to integrate the findings, especially if they diverge, including what procedures to use, which techniques, at what conditions, and at what cost in terms of new premises to be posited.

The auxiliary theory makes it possible to specify the main theory's real scope, to construct the necessary

[22] According to Lee Freese, systematic theory is expressed in everyday language without needing to use mathematical formulations, although these are not precluded. It defines and gives foundation to concepts and to the relations between them, and it dissociates basic propositions from derived ones. Lee Freese also writes, "Systematic theories resemble hierarchies of propositions in which the more specific, ordinary propositions may be deduced from the more general theoretical propositions according to the deductions rules of ordinary language" (Freese, 1980, p. 196–197). This systematic theory, according to Freese, belongs to the type of theories that Hans Zetterberg (1965, p. 94–100) called axiomatic, and it has been used in the work of Wesley Burr and colleagues (1979) in family sociology.

indicators for some concepts and relationships, and to assess whether the adjustments made to the main theory reduces its compass too much. The *precision of the main theory's scope* is achieved (1) by appreciating how much is lost through the constraints imposed by the researcher's and the research environment's concrete conditions; (2) by specifying its precise subject matter, which may be more restricted than the one initially envisaged given these constraints; (3) if it has not already been done in the main theory, by indicating the population to be studied and the parts of it which will be observed.

If necessary, *indicators are constructed*[23] for concepts and their relations emerging from the hypotheses to be tested by identifying and appraising the postulates needed to accept that the selected indicators properly represent the concepts and their relations and to accept not to observe this or that concept, or part of a concept, or this or that relation, or part of a relation that, for reasons to be elucidated and understood, would not be observable.

The loss involved in all these adjustments should to be assessed. When the loss is too great, it may be preferable to discard the main theory and elaborate another one that is more open to empirical testing.

2. Indicators

The scientific use of the term *indicator* is flexible enough to be conceptualized in various ways. Here, an indicator denotes something observable and potentially measurable that encapsulates a given concept or relation.

Definitions "in comprehension" specify the characteristics of concepts and the links between them and classify them as necessary or contingent. For instance, the concept C is defined by C1, C2, and C3 as necessary characteristics and by C4, C5, and C6 as contingent characteristics. To observe C, the indicator should grasp each of the necessary characteristics, C1, C2, and C3, in as much as the following ideas are true:

1. There is no necessary connection between some of these characteristics. If such were the case, observing the relation's starting characteristic would be enough: if C1 leads necessarily to C2, observing C1 and C3 suffices; if the relation between C1 and C2 is both necessary and reversible, it is enough to observe one or the other, as well as C3.
2. No other concept, C′, necessarily characterized by C8 and C9 and necessarily or contingently characterized by C1, C2, and C3, exists. In this case, to observe C, an indicator is needed which ensures both that C1, C2 and C3 are present, and C8 and C9 are absent. For instance, let us suppose fertility planning to be characterized necessarily by (a) a project concerning the number of births, (b) the behavior to achieve this, (c) a project to space births, for each birth, and (d) the behavior to achieve this. If I wish to look at planning only from the point of view of the number of births among couples no longer procreating because the woman has reached menopause, I need to make sure that a and b are present and that c and d are absent.

[23] In demographic analysis, some concepts have already been operationalized, such as age, sex, total fertility rate, and infant mortality rate.

When some necessary characteristics of a concept or a relation cannot be observed, we can attempt to observe instead one or more instrumental variables, preferably necessary ones, or one or more consequences, preferably necessary ones. Alternatively, we can give up any attempt to observe them. In any case, it is important to be aware of the biases this introduces and to evaluate them as well as possible.

To study couples who plan only the number of births—who are necessarily characterized by the two elements a and b and by the absence of elements c and d, we could do the following:

1. We can keep as the indicator an almost necessary consequence in our societies: the number of two births. This means including couples who also plan birth spacing and those who do not plan births because their fecundity level is low and excluding couples who plan only family size but have fewer or more than two births.
2. Use an instrumental variable very common in the society studied, such as including sterilized couples only, if we already know that sterilization is commonly used by couples not wishing to have more children. This means including couples planning only the number of births this indicator adds, those who plan birth spacing, and those who, although not planning numbers of births, were sterilized for other reasons; it excludes couples who plan numbers of births but do not use sterilization.

Whatever the problems encountered and the solution chosen, the procedure must always control and critically scrutinize our research practices. We must be aware of what we are doing but keep a critical distance from it to appreciate the consequences for hypothesis testing.

Whatever the case, an indicator should do the following:

1. Be observable and, if necessary, measurable
2. Be highly predictive or have a good grasp of the particular concept or the relation studied

3. Be connected exactly to the concept or relation, with respect to the observed content (i.e., all, some, or only one of the necessary characteristics or one of the necessary consequences, the very likely consequences, the necessary instrumental variables, or the highly probable instrumental variables) and with respect to the observation's accuracy (i.e., the concept's or relation's reality or reality itself and, in the case of a gradient, the level of the concept or relation).

4. Be valid by representing the concept or relation studied properly, which can be recognized from the information collected in item 3.

5. Be faithful or reliable by always relating to the same content of the concept or relation

The indicators' validity and reliability determines their scope. In principle, to have universal application, the indicator must apprehend all the necessary characteristics. If not, its range is relative and must be reconfirmed every time.

The question of indicators changes when the research begins with data already collected in its raw form or after statistical analysis and when we reflect on the concepts or relations to which these data may correspond. This problem is far from simple, and its solution seems to involve much coming and going between the data and the construction of concepts and relations. For instance, (1) at the start, one or several pieces of information suggest a possible concept; (2) we then attempt to construct this concept and identify its characteristics, separating them into necessary or contingent types; (3) subsequently, we go back to the facts to find data for necessary characteristics for which we have none; and (4) we return to concept building to adjust it more closely to the available data. Evidently, these steps should be well founded and explicated.

CONCLUSIONS

The whole theorization process may appear highly persnickety, uselessly rigid, and at risk of producing meticulous, even obsessive, minds rather than particularly creative ones. What has happened to freedom of thought, the freedom to combine words into sentences that are striking, that arouse debate, and that make clear what the real stakes are, how rich society is, and so on? Nothing in the theorization procedure seems to bridle this freedom. In elaborating a main theory, the researcher can proceed freely and creatively at first. However, if he or she wishes to go beyond preaching to the converted to produce and communicate knowledge that everyone can understand and master, theorization must be transparent and precise, and empirical testing can strengthen these desiderata.

In practical research, theory building follows no mandatory course. To facilitate the presentation, however, I have followed a linear sequence to highlight the key moments of critical reflection, the elements to include, and the postulates to accept where and when they are essential for the process to continue. This does not seem to me more inflexible, persnickety, or obsessional than some statistical technique or content analysis. However, the latter methods are often assisted by computer programs that enable the researcher to avoid appreciating the degree of labor and meticulousness necessary to apply such techniques.

Undoubtedly, theorization as presented in this chapter will need to be refined and perhaps even simplified. It nonetheless seems to be adequate to ensure the controlled production and communication of knowledge needed for all scientific study.

Acknowledgment

This chapter was translated from the French version by Heather Juby.

References

Blalock Hubert M. Jr., 1969. *Theory construction: from verbal to mathematical formulation.* Englewood Cliffs, NJ, Prentice-Hall, 180 p.

Blalock Hubert M. Jr., 1971a. The measurement problem: a gap between the languages of theory and research, *in*: Hubert M. Jr. Blalock and Ann Blalock (eds.), *Methodology in social research*, p. 5–27. London, McGraw-Hill, International Student Edition, 493 p.

Blalock Hubert M. Jr., 1971b. Theory building and causal inferences, *in*: Hubert M. Jr. Blalock and Ann Blalock (eds.), *Methodology in social research*, p. 155–198. London, MacGraw-Hill, International Student Edition, 493 p.

Blalock Hubert M. Jr., 1990. Auxiliary measurement theories revisited, *in*: Joop J. Hox and Jenny de Jong-Gierveld (eds.), *Operationalization and research strategy*, p. 33–48. Amsterdam, Lisse, Swets & Zeitlinger, 283 p.

Blalock Hubert M. Jr. and Blalock Ann (eds.), 1971. *Methodology in social research.* London, McGraw-Hill, International Student Edition, 493 p.

Boulanger Paul-Marie, 1980. *Introduction à l'usage de quelques concepts épistémologiques dans le champ de la démographie.* Louvain-la-Neuve, Département de démographie, Université catholique de Louvain, 19 p. (Working Paper, no. 82.)

Bunge Mario, 1975. *Philosophie de la physique.* Paris, Le Seuil, 299 p. (Translated from English.)

Burr Wesley R., Hill Reuben, Nye F. Ivan, and Reiss Ira L., 1979. *Contemporary theories about the family*. New York, Free Press, 2 vols. 668 + 226 p.

Duchêne Josianne and Wunsch Guillaume, 1985. From theory to statistical model, *in*: IUSSP/UIESP (eds.), *International population conference, Florence 1985,* vol. 2, p. 209–224. Liège, IUSSP/UIESP, 456 p.

FREESE Lee, 1980. Formal theorizing, *Annual Review of Sociology*, vol. 6, p. 187–212.

FREESE Lee, 1981. The formalization of theory and method, *American Behavioral Scientist*, vol. 24(3), p. 345–363.

GALTUNG Johan, 1973. *Theory and methods of social research.* London, Allen & Unwin, 534 p.

GÉRARD Hubert, 1983. Politique de population: à la recherche du sens perdu . . . ou à donner, *Politiques de population. Études et Documents*, vol. 1(1), p. 9–27, 63–82.

GÉRARD Hubert, 1989. Théories et théorisation, *in*: Josianne Duchêne and Guillaume Wunsch (eds.), *L'explication en sciences sociales. La recherche des causes en démographie, Chaire Quetelet 1987*, p. 267–281. Louvain-la-Neuve, CIACO, 478 p.

GÉRARD Hubert, 1995. Pour une reconstruction sociologique des faits de population, *in*: Hubert Gérard and Victor Piché (eds.), *La sociologie des populations*, p. 39–60. Montreal, PUM and AUPELF-UREF, 508 p.

GLASER Barney G. and STRAUSS Anselm L., 1970. Discovery of substantive theory: a basic strategy underlying qualitative research, *in*: William J. Filstead (ed.), *Qualitative methodology: firsthand involvement with social world*, p. 288–304. Chicago, Makham Publishing, 352 p.

HENRY Louis, 1981. *Dictionnaire démographique multilingue, 2nd ed.* Liège, UIESP/IUSSP, Ordina édition, 179 p.

HOX Joop J. and DE JONG-GIERVELD Jenny (eds.), 1990. *Operationalization and research strategy.* Amsterdam, Lisse, Swets & Zeitlinger, 283 p.

HOX Joop J. and MELLENBERG Gideon J, 1990. The interplay of substantive and auxiliary theory, *in*: Joop J. Hox and Jenny de Jong-Gierveld (eds.), *Operationalization and research strategy*, p. 123–136. Amsterdam, Lisse, Swets & Zeitlinger, 283 p.

HUBERMAN Michael, 1981. Splendeurs, misères et promesses de la recherche qualitative, *Éducation et recherche*, vol. 3(3), p. 238–242.

HUBERMAN Michael and MILES Matthew B., 1991. *Analyses des données qualitatives. Recueil de nouvelles méthodes.* Bruxelles, De Boeck Université, 506 p. (First edition translated from English.)

LADRIÈRE Jean, 1996. Sciences et discours rationnel, *in*: *Encyclopaedia universalis*, vol. 20, p. 721–725. Paris, Encyclopaedia Universalis, 1103 p.

LALANDE André, 1988. *Vocabulaire technique et critique de la philosophie, 16th ed.* Paris, PUF, 1323 p.

LAZARSFELD Paul, 1970. *Philosophie des sciences.* Paris, Gallimard, 506 p.

MERTON Robert K., 1945. Sociological theory, *The American Journal of Sociology*, vol. 50(May), p. 462–473.

OAKLEY Deborah J.H., 1977. *The development of population policy in Japan, 1945–1952, and American participation.* Ann Arbor, The University of Michigan, 2 volumes, 528 p. (Ph.D. thesis, Xerox University Microfilms.)

PRESSAT Roland, 1979. *Dictionnaire de démographie.* Paris, PUF, 295 p.

SARTORI Giovanni, 1984. Guidelines for concept analysis, *in*: Giovanni Sartori (ed.), *Social science concepts. A systematic analysis*, p. 15–85. Beverly Hills, CA, Sage Publications, 455 p.

SWANBORN Peter G., 1990. The changing concept of theory, *in*: Joop J. Hox and Jenny de Jong-Gierveld (eds.), *Operationalization and research strategy*, p. 17–32. Amsterdam, Lisse, Swets & Zeitlinger, 283 p.

ZETTERBERG Hans L., 1965. *On theory and verification in sociology, 3rd ed.* Totawa, NJ, The Bedminster Press, 177 p.

C H A P T E R

130

Causality and Causal Models

JOSIANNE DUCHÊNE AND GUILLAUME WUNSCH

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

One of the main objectives of science is to explain how things work. If some events are interdependent, explanation aims at disentangling the mechanisms that cause these events to occur as they do. Although the scientific approach is not necessarily causal, because science is also interested in functional dependences and in noncausal interdependencies (Robert Franck, 1994), the explanation objective is often achieved through causal analysis. We can therefore refer to causal explanation as the objective of basic research in social science. According to Edgard Morin (1985), "Explanation is a path which makes it possible to acquire objective knowledge by attempting to understand the causal relationships between objects as opposed to understanding based on inter-subjective knowledge, on self-identification with others."[1] Most philosophers and scientists acknowledge that explanation is basically investigation into causes, if we take into account a broad definition of *cause*.

In social science, causal modeling has become an important research tool in recent decades. The Simon-Blalock method for analyzing simple causal structures and the Sewall-Wright path analysis are at the root of many methods that have been developed to test causal hypotheses in social science. In his original article, Sewall Wright (1921, cited by Wright, 1968, p. 325) emphasized the fact that path analysis was developed "as a method of measuring the direct influence along each separate path in a system and thus of finding the degree to which variation of a given effect is determined by each particular cause."

Causal modeling is one of the main objectives of econometrics. The terminology of the *dependent* and *independent* variables used, for example, in linear regression models conveys the idea that a group of variables has an effect on the dependent variable. Consequently, the essential role of econometrics is to estimate and test the explanatory models derived from economic theories. In the preface of his book *Causality in Economics*, Sir John Hicks (1979) indicates two important tasks for economic theory; the first is descriptive, and the other is analytic in the sense that it is used as a means to explain: "We ask not merely what happened, but why it happened. That is causation."

Epidemiology is a field in which the analysis of causal relations has been significantly developed. In their publication on epidemiology, Milos Jenicek and Robert Cléroux (1982) affirm that research into the causes of a disease and its persistence within a human population is one of the main objectives of epidemiologic research. Causal modeling of the process of morbidity has become a basic aspect of epidemiology and a good example of the experimentation model that applies to human populations. However, Daniel Schwartz (1989) concluded his presentation to the Chaire Quetelet 1987 by pointing out "that it is always possible to act without understanding."[2]

These examples show that causal explanation seems to be one of the main objectives of the sciences that

[1] "L'explication est une voie qui permet d'atteindre la connaissance objective en tentant de saisir les relations causales entre objets en contra-distinction avec la compréhension qui est basée sur la connaissance intersubjective, sur l'auto-identification avec l'autre."

[2] "Qu'il reste toujours possible d'agir sans comprendre."

deal with human populations. However, this view is contested by some researchers in social sciences. According to Carol Gould (1983), for example, causality applies to the subject's relationship with the object but not to the subject's relationship with another subject, because causal explanations of human actions are incompatible with the freedom of the individual. Even if we accept Gould's ideas regarding the fact that human actions are guided by a goal, these views do not necessarily lead to the rejection of causality in social science. Carol Gould's argument, according to which objective conditions are necessary but not sufficient conditions of human actions, suggests only that strict determinism is not applicable in this field. If weaker definitions of causality are adopted, the question "what is the cause of human behavior?" remains a relevant one. Another criticism leveled against the Durkheim-like causal explanation in sociology was highlighted by Claude Javeau in his book, *Leçons de Sociologie* (1986). For this author, causes alone do not give meaning to social processes such as revolutions. We must also discover their reasons and motives. According to Javeau, the *cause* of a revolution is the ruling classes' abandonment of power, the *motive* is the will of the dominated classes to take power; the *reason* is their desire to relieve their misery. This view does not seem to be in contradiction with a causal explanation. Claude Javeau's example could be interpreted by a *causal chain* or a *causal process*,[3] which would coincide with current views on causality. Consequently, there is no reason to distinguish between causal explanation and the search for meaning, because the causal theory that we are testing can take into account (under *causes*) the reasons and motives of social actions.

In this context, we would like to develop two points: the concept of cause and effect in social science and the statistical modeling of causality.

I. CAUSE AND EFFECT IN SOCIAL SCIENCE

There is no unanimously accepted definition of causality. A large part of the debate on causality is rooted in the fact that scientists use the term *cause* with different meanings.[4] In epidemiology (Jenicek, 2003), a classic distinction relates to the difference between risk *factors* (i.e., modifiable causes such as smoking) and risk *markers* (e.g., age or sex), which cannot be modified. The same distinction is used by Paul Holland (1986) in statistics; he distinguishes causes that can be manipulated from attributes (e.g., age) that cannot. In social sciences, we speak of determinants or factors rather than causes, without generally making a distinction between risk markers and risk factors.

To what extent does an effect follow its cause? In addressing this point, we separate the deterministic theories of causality from the nondeterministic theories (Karhausen, 1996). Although causality in social science is primarily stochastic because the multiple causes of social phenomena cannot all be known, examining the deterministic theories of causality is a useful frame of reference. Even in social sciences, many researchers still refer implicitly to the deterministic nature of causality. Several concepts examined within this context, such as the disjunction or the conjunction of causes or the asymmetry between causes and effects, have their counterparts in the probabilistic approach.

A well-known supporter of the deterministic theory of causality was John Stuart Mill, and his ideas were outlined in *A System of Logic* (1843, republished in 1889). For John Stuart Mill, the cause of a phenomenon is a combination of all its conditions. When all the necessary conditions are present, they make up the sufficient condition. "The cause, then, philosophically speaking, is the sum total of the conditions positive and negative taken together; the whole of the contingencies of every description, which being realized, the consequent invariably follows" (Mill, 1889). More recently, the "necessary and sufficient" approach of causality has been supported primarily by Richard Taylor. For him, "X is the cause of Y" means that X and Y are two conditions (or combinations of conditions) that occur and that each one is (given all the other conditions that occur) necessary and sufficient for the other one to occur (Taylor, 1966).

A complex version of John Stuart Mill's "combination of all the conditions" was developed by John Mackie (1974). Let $AB\overline{C}$, for example, be the conjunction of sufficient factors for Y to occur. For illustration, Y is the start of a fire, A is the occurrence of a short-circuit, B is the presence of flammable materials, and $\overline{C}$ is the lack of a smoke detector and extinguisher (i.e., the combination of positive and negative conditions in Mill's terms). According to John Mackie, $AB\overline{C}$ is a minimal sufficient condition for Y. There are, however, other possible minimal sufficient conditions, such as $LB\overline{C}\overline{D}$, in which L represents lightning striking the house, D represents the absence of a lightning conductor, and so forth. The disjunctive plurality of the

[3] Other components of the process could exist and could also be taken into account.

[4] For an overall view, see Wunsch (1986); for a detailed discussion, see three papers presented to the Chaire Quetelet 1987 on explanation in social science: Callebaut (1989), Nowak (1989), and Swain (1989). On the historical evolution of the concept, Sosa and Tooley (1993) and Laudisa (1999).

finite number of conjunctions $AB\bar{C} \cup LB\overline{CD} \cup \ldots \cup KMG$ is a necessary and sufficient condition for the effect (i.e., the "complete cause" of Y). Each element (e.g., A) of one of the conjunctions (e.g., $AB\bar{C}$) is an *i*nsufficient but *n*ecessary part of a condition, which is *u*nnecessary but *s*ufficient for the effect (i.e., A is an *INUS* condition for Y).

1. Temporal Anteriority of the Cause over the Effect

An important aspect is the symmetry of the relationship between the cause X and the effect Y. If X is a necessary and sufficient condition for Y, then logically Y is a necessary and sufficient condition for X. To guarantee asymmetry between the cause and the effect, David Hume (1739, republished 1968) imposed the temporal precedence of the cause over the effect. For this author, simultaneous causality does not exist because "if one cause were co-temporary with its effect, and this effect with its effect, and so on, it is plain there would be no such thing as succession." All the objects would coexist, which would lead to the annihilation of time itself. David Kline supported the point of view of David Hume on the basis of the special theory of relativity (Kline, 1980). He starts with the counter-example given by Richard Taylor and taken up in a similar counter-example by Myles Brand (1976): If a force is exerted on a locomotive L at time t_1, it is simultaneously exerted over the wagon C. This implicitly presupposes that *causal information* is propagated from L toward C at a speed equivalent to or greater than the speed of light. This would produce an infinite relative mass that would require the deployment of an infinite force to move it, which is impossible. Kline concluded that for cases of simultaneous causality to have a chance of being verified, the events concerned should be distinct, linked causally, and not separated spatially.[5]

While acknowledging that no real cases of retrocausality have been observed, Michel Ghins (1989) argues in favor of the conceptual possibility of such a retrocausality in physics. He concludes, however, by stating that the possibility of retrocausality does not imply the possibility of influencing the future. Huw Price (1996) also examined the issue of the possible reversibility of the arrow of time and supported the idea of temporal symmetry in physics. James Griesmer (1991) insists that the asymmetry of the relationship is a necessary but not sufficient condition for establishing causality: All asymmetric relationships are obviously not causal.

[5] However, according to Salmon, the *intersection* of two processes can be simultaneous.

The condition of temporal precedence of cause over effect has been adopted by most causal theories because it enables differentiation of causal relations from false associations (Pearl, 2000). However, if several causes contribute to the effect, it is also frequently necessary to take into account the temporal sequence on which these causes occur. This is the reason why we have put forward a causal model where the effect (in this case, the death of an individual) would be the result of different *states* in which this individual had lived during his or her life, with the state defined as the conjunction of relevant variables for the effect, the *time* lived in each state, and the *order* of temporal succession of these states (Wunsch *et al*., 1996). The causal sequence can also be introduced into a path analysis. The main problem with these approaches is the lack of adequate data, except in cases of detailed individual longitudinal observation.

2. Probabilistic Theories of Causality

In many cases, purely deterministic causality is difficult to defend. If we opted for a combination of causes, and if on the basis of this combination, we predicted the occurrence of an event, this event should occur in all of the predicted cases. If the event occurs only in one-half of the cases, something must be incorrect in the deterministic theory. There are two possible explanations. According to the first explanation, the phenomena are by nature deterministic, but the theory is badly specified considering the limits of our knowledge, or our method of observation is prone to random errors. Within this hypothesis, there are "hidden variables" that, if they were discovered, would enable a deterministic description. This was the position adopted by Albert Einstein about the nondeterministic nature of quantum mechanics. In the second case, we are faced with the presence of intrinsically nondeterministic phenomena; in this situation, no additional information can lead to forecasts devoid of error. In the field of human actions, we are often faced with nondeterminism, even if choices are made with certainty (Wunsch, 1988). Consequently, there is a fundamental source of *fuzzy causality* in social sciences, because the events are the results of human actions, and these can be considered to a great extent as nondeterministic. This justifies a probabilistic approach to causality, including statistical modeling.

Several probabilistic theories[6] of causality have been developed. Among the most interesting approaches is

[6] Because of a lack of space, we do not develop the different concepts of probability or discuss their application in scientific explanation (see Jaynes, 2003).

that of Wesley Salmon, which was based on the concept of *statistical relevance* (i.e., *S-R* model) (Salmon, 1971). Later, Salmon (1980, 1984) developed a qualitative theory of probabilistic causality based on the concept of process, interaction, and conjunction of processes, meant to complete the *S-R* model based on conditional probabilities. According to Salmon, a *causal process* is that which transmits a marker, or a signal, of information, energy, or a causal influence from one region of space-time to another. The *S-R* model is therefore only the *basis* of a causal explanation, which needs to explain the ongoing causal process. According to the *S-R* model, let a property C divide a population A (a reference class) into two subgroups AC and $A\bar{C}$. The property C is said to be statistically relevant for an attribute B in A if $P(B/C) \geq P(B/\bar{C})$. For example, $P(B/C)$ is the probability of cure (B) for a reference class of patients suffering from lung cancer who are taking a medicine (C). For those who are not taking the medicine, the probability of a cure is equal to $P(B/\bar{C})$. We observe (Holland, 1986) that the determination of a possible causal impact necessitates at least two causes, C and $\bar{C}$ in the present case. If the medicine is an effective treatment, we should observe a significant increase in the probability of cure in the cases in which the medicine is taken compared with the probability in cases in which the medicine is not taken. For example, if $P(B/C) = 0.20$ and $P(B/\bar{C}) = 0.10$, the probability ratio (or *relative risk*) is equal to $P(B/C)/P(B/\bar{C}) = 2$ (i.e., there is twice a likelihood of cure for those who are administered the medicine compared with the control group subjects who have not been given the medicine). C can be considered in this case as a cause of B. The example shows that, if we follow Salmon, a correct probabilistic explanation does not require high probabilities. What counts is the statistical relevance of the factor (i.e. a significant difference between the *a priori* and *a posteriori* probabilities, or between the results of the target group and the results in the control group).

The reference class should be selected as the greatest *homogeneous* class to which the event belongs. Unfortunately, in most cases, we do not know how to partition the population into the largest group of homogeneous cells (i.e., we do not know all the relevant factors). In the example, if the treatment and the control groups do not differ other than in the fact that one takes and the other does not take the medicine, the difference between the two groups would be due (except for random errors) to the factor that separates the population—in this case, the fact of taking or not taking the medicine. In reality, the exposed and unexposed groups frequently differ in other factors. For example, the treatment can be administered to patients suffering from cancer in hospital M and the results compared with those obtained from hospital N, where this treatment is not used. In this case, the sampling is not perfect because the patients who attend hospital N may have characteristics (e.g., age, social status) different from those of the patients attending hospital M, and the difference between M and N can result from the treatment itself and from the other confounding variables that were not controlled for.

The preceding approach is a specific case of the *counterfactual* theory of causality, whose origin goes back to David Hume. A contemporary version, adapted to epidemiology, is presented by George Maldonado and Sander Greenland (2002). A target population, B, is receiving preventative treatment against an illness (e.g., influenza). During a specified period, let A_1 be the incidence of new cases of influenza observed within this exposed population. What would have been the incidence if this same population B had not had the treatment (i.e., counterfactual situation)? Let A_0 be this incidence. In this case, the *causal contrast* is written in terms of relative risks: $RR_{causal} = (A_1/B)/(A_0/B)$. The problem lies in the fact that we do not know A_0 as it is a counterfactual and therefore hypothetical situation. We must therefore find a substitute for the frequency (A_0/B); let this be, for example, (C_0/D), where C_0 is the incidence of new cases in another population D, which this time does not take the medication. In the case of an imperfect substitution, the causal contrast could be subject to bias as a result of the presence of confounding variables (LaLonde, 1986) (see Chapter 15). Consequently, the problem consists of choosing the correct substitute. *Double-blind randomization* would very probably be the best solution in this example. However, this approach is experimental and frequently impossible for ethical or practical reasons. In a nonexperimental approach that, for example, compares a target group receiving treatment with a control group that is not receiving treatment, for which *a priori* randomization cannot be carried out, we can use the approach proposed by Paul Rosenbaum and Donald Rubin (1983) using *propensity scores*. It is a matter of pairing the observations that have the same propensity score to see whether the treatment is effective or not. The propensity score is defined as the conditional probability of receiving the treatment, taking into account the co-variables observed. This method makes it possible to adequately control the explicit variables (Rubin and Neal, 1992), but it cannot guarantee the avoidance of potential errors introduced by the presence of implicit or hidden confounding variables.

II. CAUSALITY AND STATISTICAL MODELING

Although this chapter is devoted to methods of statistical analysis, epistemologically, there is no reason to limit the determination of causal impact in social science to multivariate statistical methods. In some cases, the theory of information, simulation, or acyclic directed graphs, for example, can also describe the causal mechanism; there are close analogies between the acyclic directed graphs and some statistical methods that are examined later. In other cases, qualitative approaches can be used to shed new light on causal relations that are difficult to understand through other methods. However, most of the work carried out in social science is based on statistical modeling, because many studies are carried out in nonexperimental contexts,[7] in which the researcher is not necessarily aware of all the relevant variables. In this case, the absence of checks in the experiment should be compensated by a series of conditions that should be set (and fulfilled) if statistical methods are used in a causal perspective. However, a statistical model, such as logistic regression, can be used for explanatory ends without explicitly referring to the causal structure of variables.

1. Causal Precedence and Longitudinal Analysis

Any cause needs a period of time, however short, in which to exercise its effects. The importance of the effect can vary over the period between the appearance of a cause and the measurement of its effect. This fact is illustrated by Harry Gollob and Charles Reichardt (1987) by means of a situation that is easy to understand for everyone. If two aspirins are taken to get rid of a migraine, the effect can be nil 2 minutes after ingestion, substantial after 30 minutes, close to maximum after 2 or 3 hours, very reduced after 5 hours, and once more nil after 24 hours.

If the only information available is cross-sectional, the causal effects are generally estimated on the basis of models that are also cross-sectional and that consider only the relationships between the contemporary values of the different variables. These models do not make it possible to specify the relationships between the value at a given time of one or more variables and the value at a previous period of some of these variables and other variables. Among the conditions that justify the use of a confirmatory analysis to evaluate the scientific usefulness of causal hypotheses, Lawrence James, Stanley Mulaik, and Jeanne Brett (1982) called on the condition of stability of the structural model (i.e., the invariance) over a certain period of the parameters of the equations. Harry Gollob and Charles Reichardt (1985) demonstrated that this hypothesis is neither a sufficient condition nor a necessary condition for the cross-sectional model to produce nonbiased estimates of causal effects. For example, if the variables X and Y are standardized in two simple models (Fig. 130–1) and if R is the correlation coefficient between X_{t-1} and Y_{t-1}, we can write $\beta^* = \beta + R\ \varphi_Y$, where φ_Y is the path coefficient between Y_{t-1} and Y_t. Consequently, β^* is a biased estimator of β as long as $R\ \varphi_Y$ is different from zero. Moreover, if the structural stability condition is not respected, the fact that X is the cause of Y is translated by a longitudinal causal system (Fig. 130–2). In this case, if the variances of X and Y and the correlation between X and Y are the same in $t-1$ and in t, $\beta^* = \beta[\varphi_X / (1 - \varphi_X \varphi_Y)]$ will be a nonbiased estimate of β if $\varphi_X = 1 / (1 + \varphi_Y)$, where φ_X is the path coefficient between X_{t-1} and X_t. This relationship is nevertheless as difficult to satisfy as when $R\ \varphi_Y = 0$.

The best method to estimate causal effects is to adjust longitudinal models to individual longitudinal data, while taking into account the presence of contextual effects if it is justified (Courgeau, 2003). If we only have cross-sectional data, Harry Gollob and Charles Reichardt (1985 and 1987) suggest applying longitudinal models with latent variables where unobserved variables are substituted for unavailable variables. To identify this model, additional hypotheses must be set; for example, if the variables are standardized, the equality of the correlations between the periods t and $t-1$ and $\varphi_Y = R(Y_{t-1}) - \beta R(Y_{t-1}, X_{t-1})$. Longitudinal models with latent variables are preferable to cross-sectional models because they oblige researchers

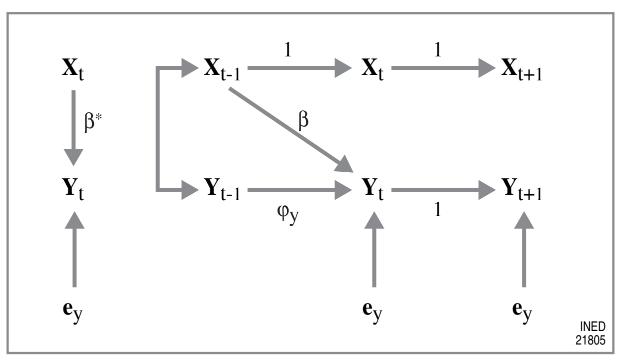


FIGURE 130–1. Causal System A.

[7] Double-blind experimentation is ethically unacceptable or impossible to implement in many cases.

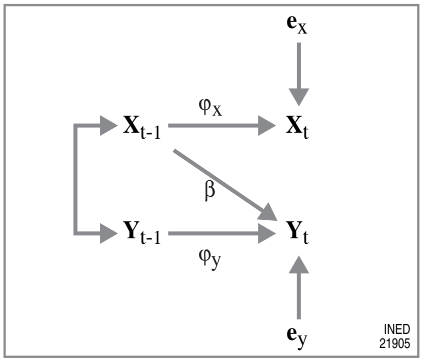


FIGURE 130–2. Causal System B.

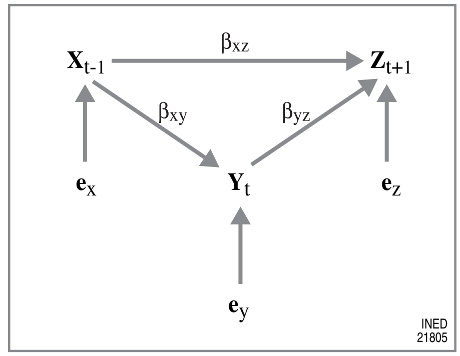


FIGURE 130–3. Longitudinal Causal Model.

to clarify their hypotheses. However, estimation problems can be avoided only by collecting longitudinal data and using longitudinal models, such as life-course analysis (see Chapter 23). However, every population is heterogeneous (see Chapters 10 and 21), and prospective longitudinal data can be affected by selection bias (e.g., the most vulnerable individuals die before others, the strongest individuals migrate and therefore leave the field of observation). Moreover, retrospective longitudinal data depend on memory recall and are therefore not always reliable due to recall lapses.

2. Correlations and Causal Relations

Paul Games (1990) draws attention to the fact that methods based on correlations, such as multiple regression, partial correlation, structural equation models, path analysis, and every method applied to data collected without experimental control, do not enable us to draw conclusions concerning any causal effect. They can only suggest hypotheses on causal relations that must be tested later; association does not necessarily mean causation. These methods are too often used incorrectly to discover causal relations, whereas they were developed to test hypotheses regarding these relations. Sewall Wright, the initiator of path analysis, very clearly defined the objectives of this method in 1934: "It has been emphasized that the method of path coefficients is not intended to accomplish the impossible task of deducing causal relations from the values of the correlation coefficients. It is intended to combine the quantitative information given by the correlations with such qualitative information as may be at hand on causal relations to give a quantitative interpretation" (Wright, 1934, p. 193).

When all the variables are observed on a longitudinal basis, they can be integrated into a linear and recursive causal model, which fulfills the following four conditions: (1) it includes neither causal loops nor reciprocal causal relations; (2) all the relations between variables can be expressed as linear equations; (3) the error terms are not correlated among themselves; and (4) the independent variables of each equation are not correlated with any error terms. Such a model, including three variables for example (Fig. 130–3), can be translated by the following set of equations:

$$X_{t-1} = e_X$$

$$Y_t = \beta_{XY}\, X_{t-1} + e_Y$$

$$Z_{t+1} = \beta_{XZ}\, X_{t-1} + \beta_{YZ}\, Y_t + e_Z$$

The path coefficients β_{XY}, β_{YZ}, and β_{XZ} are estimated by least squares regression methods such that

$$\beta_{XY} = (\sigma_Y/\sigma_X) \,/\, R_{XY}$$

$$\beta_{YZ} = (\sigma_Z/\sigma_Y)\, (R_{YZ} - R_{XY}\, R_{XZ})/(1 - R^2_{XY})$$

$$\beta_{XZ} = (\sigma_Z/\sigma_X)\, (R_{XZ} - R_{XY}\, R_{YZ})/(1 - R^2_{XY})$$

In a recursive longitudinal model, the direct and indirect effects can be estimated through path analysis, which applies to continuous data (Kendall and O'Muircheartaigh, 1977; Herbert Asher 1976). Christopher Winship and Robert Mare (1983) extended its use to discrete data, and Thomas Wilson and William Bielby (1983) applied it to polytomic data.

In the case of probabilistic causal relations between dichotomous variables, Frederick Ellett and David Ericson (1986a, 1986b) show that the β coefficients and three other measurements of association:

$$cov_{XY} = P(X \cap Y) - P(X)\, P(Y)$$

$$cov_{XY}/P(X) = P(Y/X) - P(Y)$$

$$cov_{XY}/\sigma_X\, \sigma_Y = R_{XY}$$

can give only a *qualitative* measurement of the different causal impacts. The signs of these indices indicate whether the causes are positive or negative. They cannot provide quantitative measurements of the causal impact of one dichotomous variable on another dichotomous variable. Ellett and Ericson (1984) also developed an approach based on conditional probabilities to analyze probabilistic causal systems with dichotomous variables. According to Gürol Irzik (1986), Wesley Salmon's approach (1984), which is also based on conditional probabilities, provides the same numeric results as those of causal modeling applied to dichotomous variables. The debate remains open with regard to the adequacy of methods based on correlation coefficients for measuring the causal impact of dichotomous variables.

3. Reciprocal Causal Relations and Temporal Precedence of Cause over Effect

The most frequently mentioned limitation of recursive models is the fact that they postulate the absence of a reciprocal causal relation. These reciprocal causal relations between, for example, X and Y are usually represented in graphs as $X \leftrightarrow Y$ without any reference (explicit or implicit) to the temporal precedence of the cause over the effect. To eliminate ambiguity on this point, it would be useful to employ any one of the following representations:

$$X_t \rightarrow Y_{t+1} \text{ and } Y_t \rightarrow X_{t+1}$$

$$X_t \rightarrow Y_{t+1} \rightarrow X_{t+2}$$

$$Y_{t-1} \rightarrow X_t \rightarrow Y_{t+1}$$

If the system is not in equilibrium (Strotz and Wold, 1960), it is not certain that different formulations will lead to the same results. With these notations, the non-recursive model can be substituted with a model that, of the four possible models, requires the fewest observations. For the model to be recursive, the error terms must not be correlated among themselves, and the independent variables of each equation must not be correlated with any error term. If the systematic measurement errors of variable Y in time $(t - 1)$ and $(t + 2)$ are avoided and if the possible common causes of the error terms are explicitly introduced into the model, most models can be made recursive, and as a result, we can avoid the maximum likelihood or the least squares method of estimation at two or three levels. David Sanders (1980) applied recursive (ordinary least squares) and non-recursive (two-stage least squares and maximum likelihood) models to the same data. Each of these methods provides results that partially contradict those obtained using the other methods.

4. Covariance Structure Analysis

The recursive linear model is sometimes presented as a specific case of covariance structure analysis,[8] combining the path analysis approach with regard to causal structure and the factor analysis approach for the relations between latent variables and their indicators. The use of this method became widespread due to the development of software such as LISREL (Jöreskog and Sörbom, 1996) and EQS (Bentler, 1989), among others. The ordinary least squares used by the correct recursive linear model exactly identified the result, and the maximum likelihood used by the software LISREL provide the same results as long as the matrix of correlations not include any random components, which amounts to presuming that large enough samples are being used. Bruce Biddle and Marjorie Marlin (1987) advised against using LISREL with small samples. Moreover, Peter Bentler and Paul Dudgeon (1996) examined the consequences of not respecting the hypotheses of the model in the case of nonexperimental studies, notably the hypothesis of distribution normality. Various solutions have, however, been suggested to take into account the non-normality of variables (Shipley, 2000). If we construct latent variables from several indicators using the LISREL or EQS software, we should be aware that this procedure is not equivalent to a regression by ordinary least squares, which would take into account the factor scores resulting beforehand from factor analysis applied to the indicators. Some studies have extended the approach to latent dichotomous variables (Eshima *et al*., 1996) and to multinomial variables (Moors, 2003, provides an example of the application). As is the case for path analysis, we must emphasize the fact that the sole analysis of covariance structures with latent variables cannot confirm a causal relation; causation needs modeling, background knowledge, and explanatory theory. A detailed discussion of this issue is provided by Judea Pearl (2000). However, it enables certain structures to be disconfirmed (or falsified) in favor of others that are better suited to the available data.

5. The Concept of Functional Relation

Stanley Mulaik (1986 and 1987) suggested using the functional relation concept with a view to synthesizing the deterministic and probabilistic formulations of causal relations between *continuous* variables. This concept makes it possible, according to Mulaik, to unify different models such as analysis of variance, the generalized linear model, structural linear equations, and log-linear models. It is a case of setting hypotheses whose *cause* variable values determine not the specific *effect* variable values but rather the conditional

[8] For an example, see Lopez Rios and colleagues (1992).

probabilities with which the different values of the effect variable are observed.

6. Correlation and Causal Link between Dichotomous Variables

Frederick Ellett and David Ericson (1986b) located, in social science research and in the philosophy of science, four conceptions of the correlation between two dichotomous variables X and Y.

First, the social science researchers Patricia Kendall and Paul Lazarsfeld (1950) and the science philosopher Ernest Nagel (1961) use the conception according to which two dichotomous variables X and Y are positively (or negatively) correlated if, and only if, the probability of X and Y minus the probability of X multiplied by the probability of Y is greater (or less) than zero. It can be shown that

$$P(X \cap Y) - P(X)\,P(Y) = P(X \cap Y)\,P(\bar{X} \cap \bar{Y}) - P(X \cap \bar{Y})\,P(\bar{X} \cap Y) = cov_{XY}$$

Second, for the science philosophers Hans Reichenbach (1956) and Patrick Suppes (1970), two dichotomous variables X and Y are positively (or negatively) correlated if, and only if, the (conditional) probability of Y if X minus the probability of Y is greater (or less) than zero. If $P(X) > 0$, it can be shown that

$$P(Y/X) - P(Y) = [P(X \cap Y)\,P(\bar{X} \cap \bar{Y}) - P(X \cap \bar{Y})\,P(\bar{X} \cap Y)]/P(X) = cov_{XY}/P(X)$$

Third, according to a third conception used by Wesley Salmon (1980), two dichotomous variables, X and Y, are positively (or negatively) correlated if, and only if, the (conditional) probability of Y if X minus the (conditional) probability of Y if $\bar{X}$ is greater (or less) than zero. If $1 > P(X) > 0$, one can show that

$$\begin{aligned} P(Y/X) - P(Y/\bar{X}) &= [P(X \cap Y)P(\bar{X} \cap \bar{Y}) - P(X \cap \bar{Y})P(\bar{X} \cap Y)]/[P(X)\,P(\bar{X})] \\ &= cov_{XY}/\sigma_X^2 \\ &= b_{XY} \end{aligned}$$

Fourth, social science researchers such as Herbert Asher (1976), Hubert Blalock (1964), and Herbert Simon (1957) use the fourth conception, according to which two dichotomous variables X and Y are positively (or negatively) correlated if, and only if, the coefficient ϕ is positive (or negative). If $1 > P(X) > 0$ and $1 > P(Y) > 0$, the coefficient ϕ is defined as

$$\begin{aligned} \phi_{XY} &= [P(X \cap Y)P(\bar{X} \cap \bar{Y}) - P(X \cap \bar{Y})P(\bar{X} \cap Y)]/[P(X)P(\bar{X})P(Y)\,P(\bar{Y})]^{1/2} \\ &= cov_{XY}/(\sigma_X\,\sigma_Y) \\ &= r_{XY} \\ &= b_{XY}\,\sigma_X/\sigma_Y \end{aligned}$$

These four measurements of association have the same sign (i.e., the cov_{XY} sign). Consequently, all the measurements lead to the same *qualitative* conclusion that X is not a cause of Y if they are equal to zero or that X is a positive (or negative) cause of Y if they are greater (or less) than zero.

A *quantitative* approach to causality attempts to establish whether X is a cause of Y and to measure the degree of causal influence that X has over Y. However, from a quantitative point of view, it is possible to show that if $1 > P(X) > 0$ and $1 > P(Y) > 0$,

$$cov_{XY} < cov_{XY} / P(X) < b_{XY} \qquad (1)$$

$$cov_{XY} < \phi_{XY} \qquad (2)$$

and if $\phi_{XY} > 0$, it satisfies one of the following three inequalities:

$$cov_{XY} < \phi_{XY} < cov_{XY}/P(X) \qquad (3.1)$$

$$cov_{XY}/P(X) < \phi_{XY} < b_{XY} \qquad (3.2)$$

$$b_{XY} < \phi_{XY} \qquad (3.3)$$

Frederick Ellett and David Ericson (1986b) stated the following:

> If some effect B has two causes, A and C, and if A is a sufficient cause of B, or if A is a probabilistic cause of B (with probability 1), in general, the correlation between A and B must be nonnegative but it need not equal 1. If A is a probabilistic cause of B (with probability p < 1), then the sign and the magnitude of the correlation will depend on background assumptions involving certain features of the other causes of B. . . . If A and the other causal factor are uncorrelated, the correlation between A and B will be non-negative, but in general, it will not equal p. If the expectation of A and the other causal factor is zero, or if A and the other factor are maximally negatively correlated, the correlation between A and B can be negative. We also claimed that the INUS conception of causation and the INUP[9] conception of causation produce results which parallel those obtained for the probabilistic conception.

III. TOWARD AN EVIDENCE-BASED DEMOGRAPHY

Causal relations cannot be highlighted by using statistical methods only. A statistical model can be suitably adapted to available data, and the results can be statistically significant. However, this does not mean that it alone can result in causal inference. Like medicine, in which evidence-based scientific conclusions (see Chapter 52) prove essential for patient survival, we can conclude this chapter by emphasizing some conditions that must be fulfilled if we wish to respond to causal-type questions.

[9] A probabilistic generalisation of the INUS cause.

According to Mark Elwood (1988), a causal influence necessitates the following questions. What hypothesis is being tested in this study? How has the observation been carried out (e.g., is it an experimental study or not; is it cross-sectional or longitudinal)? Which population has been observed? What was the main result? At the level of internal validity, have the results been influenced by observation errors, by confounding variables, or by randomness? The following questions could also be added. If a statistical model has been used in a nonexperimental approach, does the model fit well with the data?[10] If several models fit equally well with the data, have we used the simplest one, and if not, why not? With regard to causality, are the observed variables in the correct temporal order? Is the relation between presumed cause and effect a strong one? How does the effect differ according to the variation of the cause? Is the presumed causal relation observed within different subpopulations? At the level of external validity, can the results be extended to cover the whole population? For example, in the case of a longitudinal study, to what extent does loss from observation affect the extrapolation of the survey results to the general population? Can the results be extrapolated to other populations? Are the results consistent with those of other studies? Are they plausible with regard to what we already know about the subject and according to the underlying theory of the study? These questions are applied at various degrees according to whether the research is descriptive, explanatory, or intervention based. It is nevertheless useful to ask them, before coming (often too quickly) to any conclusions on the cause-effect relationship.

Acknowledgments

The authors are particularly grateful to Alberto Palloni and Federica Russo, as well as an anonymous referee, for their relevant suggestions about a previous version of this chapter. This chapter was translated from the French version by Accenta Ltd.

References

Asher Herbert B., 1976. *Causal modeling*. Beverly Hills, CA, Sage, 80 p. (Sage Series on Quantitative Applications in the Social Sciences, no. 07-003.)

Bentler Peter M., 1989. *EQS structural equations program manual*. Los Angeles (California), BMDP Statistical Software, 254 p.

Bentler Peter M. and Dudgeon Paul, 1996. Covariance structure analysis: statistical practice, theory, and directions, *Annual Review of Psychology*, vol. 47, p. 563–592.

[10] This can be achieved, for example, by verifying the consistency of results from the various tests of fit. Inconsistency between the results of these tests probably reflects a bad choice of model.

Biddle Bruce J. and Marlin Marjorie M., 1987. Causality, confirmation, credulity, and structural equation modeling, *Child Development*, vol. 58(1), p. 4–17.

Blalock Hubert M., 1964. *Causal inference in nonexperimental research.* Chapel Hill, University of North Carolina Press, 200 p.

Brand Myles, 1976. *The nature of causation*. Urbana, University of Illinois Press, 372 p.

Callebaut Werner, 1989. Post-positivist views of scientific explanation, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales. La recherche des causes en démographie*, p. 141–196. Louvain-la-Neuve, CIACO éditeur and ARTEL, 478 p. (Chaire Quetelet 1987.)

Courgeau Daniel (ed.), 2003. *Methodology and epistemology of multilevel analysis*. Dordrecht, Kluwer, 235 p.

Ellett Frederick S. and Ericson David P., 1984. Probabilistic causal systems and the conditional probability approach to causal analysis, *Quality and Quantity*, vol. 18(3), p. 247–259.

Ellett Frederick S. and Ericson David P., 1986a. An analysis of probabilistic causation in dichotomous structures, *Synthese*, vol. 67(2), p. 175–193.

Ellett Frederick S. and Ericson David P., 1986b. Correlation, partial correlation, and causation, *Synthese*, vol. 67(2), p. 157–173.

Elwood J. Mark, 1988. *Causal relationships in medicine*. Oxford, UK, Oxford University Press, 332 p.

Eshima Nobuoki, Asano Chooichiro, and Tabata Minoru, 1996. A developmental path model and causal analysis of latent dichotomous variables, *British Journal of Mathematical and Statistical Psychology*, vol. 49(1), p. 43–56.

Franck Robert (ed.), 1994. *Faut-il chercher aux causes une raison*? Paris, Vrin, 447 p.

Games Paul A., 1990. Correlation and causation: a logical snafu, *Journal of Experimental Education*, vol. 58(3), p. 239–246.

Ghins Michel, 1989. Causal precedence: logical or temporal order? *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales. La recherche des causes en démographie*, p. 215–223. Louvain-la-Neuve, CIACO éditeur and ARTEL, 478 p. (Chaire Quetelet 1987.)

Gollob Harry F. and Reichardt Charles S., 1985. Building time lags into causal models of cross-sectional data, *in*: ASA (ed.) *Proceedings of the Statistical Section of the American Statistical Association*, p. 165–170. Washington, DC, American Statistical Association, 442p.

Gollob Harry F. and Reichardt Charles S., 1987. Taking account of time lags in causal models, *Child Development*, vol. 58(1), p. 80–92.

Gould Carol C., 1983. Beyond causality in the social sciences, *in*: Robert S. Cohen and Marx W. Wartofsky (eds.), *Epistemology, methodology and the social sciences*, p. 53–88. Dordrecht, Reidel, 270 p.

Griesmer James R., 1991. Must scientific diagrams be eliminated? The case of path analysis, *Biology and Philosophy*, vol. 6(1), p. 155–180.

Hicks John, 1979. *Causality in economics*. Oxford, UK, Basil Blackwell, 124 p.

Holland Paul W., 1986. Statistics and causal inference, *Journal of the American Statistical Association*, vol. 81(396), p. 945–960.

Hume David, 1968. *A treatise of human nature*. London, Dent, 258 p.

Irzik Gürol, 1986. Causal modeling and the statistical analysis of causation, *Proceedings of the Biennal Meetings of the Philosophy of Science Association (PSA)*, no. 1, p. 12–23.

James Lawrence R., Mulaik Stanley A. and Brett Jeanne M., 1982. *Causal analysis: assumptions, models, and data*. Beverly Hills, CA, Sage, 176 p.

Javeau Claude, 1986. *Leçons de sociologie.* Paris, Meridiens Klincksieck, 281.

Jaynes Edwin, 2003. *Probability theory: the logic of science*, Cambridge, UK, Cambridge University Press, 758 p.

Jenicek Milos, 2003. *Foundations of evidence-based medicine*. New York, Parthenon, 392 p.

Jenicek Milos and Cléroux Robert, 1982. *Épidémiologie. Principes, techniques, applications*. St Hyacinthe, Quebec, Edisem, Maloine, 454 p.

Jöreskog Karl G. and Sörbom Dag, 1996. *LISREL 8 user's reference guide*. Moresville, Scientific Software, 374 p.

Karhausen Lucien R., 1996. The logic of causation in epidemiology, *Scandinavian Journal of Social Medicine*, vol. 24(1), p. 8–13.

Kendall Maurice G. and O'muircheartaigh Colm A., 1977. *Path analysis and model building*. Voorburg, ISI, 27 p. (World Fertility Survey technical bulletin N2/TECH.414.)

Kendall Patricia L. and Lazarsfeld Paul F., 1950. Problems of survey analysis, *in*: Robert K. Merton and Paul F. Lazarsfeld (eds.), *Continuities in social research: studies in the scope and methods of "the American soldier,"* p. 133–196. Glencoe, IL, Free Press, 255 p.

Kline A. David, 1980. Are there cases of simultaneous causation? *Proceedings of the Biennal Meetings of the Philosophy of Science Association (PSA)*, no. 1, p. 292–301.

LaLonde Robert J., 1986. Evaluating the econometric evaluations of training programs with experimental data, *American Economic Review*, vol. 76(4), p. 604–620.

Laudisa Federico, 1999. *Causalità. Storia di un modello di conoscenza.* Rome, Carocci, 120 p.

Lopez-Rios Olga, Mompart Ana and Wunsch Guillaume, 1992. Système de soins et mortalité régionale: une analyze causale, *European Journal of Population/Revue européenne de démographie*, vol. 8(4), p. 363–379.

Mackie John L., 1974. *The cement of the universe: a study of causation*. Oxford, UK, Clarendon Press, 329 p.

Maldonado George and Greenland Sander, 2002. Estimating causal effects, *International Journal of Epidemiology*, vol. 31(2), p. 422–429.

Mill John Stuart, 1889. *A system of logic*. London, Longmans, Green and Co., 622 p.

Moors Guy, 2003. Estimating the reciprocal effect of gender role attitudes and family formation: a log-linear path model with latent variables, *European Journal of Population*, vol. 19(2), p. 199–221.

Morin Edgar, 1985. Sur la sociologie, *Projet*, no. 191, p. 7–19.

Mulaik Stanley A., 1986. Toward a synthesis of deterministic and probabilistic formulations of causal relations by the functional relation concept, *Philosophy of Science*, vol. 53(3), p. 313–332.

Mulaik Stanley A., 1987. Toward a conception of causality applicable to experimentation and causal modeling, *Child Development*, vol. 58(1), p. 18–32.

Nagel Ernest, 1961. Explanation and understanding in the social sciences, *in*: Ernest Nagel (ed.), *The structure of science: problems in the logic of scientific explanation*, p. 474–546. New York, Harcourt, Brace & World, 618 p.

Nowak Stefan, 1989. Causality and determinism in the social sciences, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales. La recherche des causes en démographie*, p. 225–265. Louvain-la-Neuve, CIACO éditeur and ARTEL, 478 p. (Chaire Quetelet 1987.)

Pearl Judea, 2000. *Causality. Models, reasoning, and inference*. Cambridge, UK, Cambridge University Press, 384 p.

Price Huw, 1996. *Time's arrow and Archimedes' point. New directions for the physics of time*. Oxford, UK, Oxford University Press, 320 p.

Reichenbach Hans, 1956. *The direction of time*. Berkeley, CA, University of California, 280 p.

Rosenbaum Paul R. and Rubin Donald B., 1983. The central role of the propensity score in observational studies for causal effects, *Biometrika*, vol. 70(1), p. 41–55.

Rubin Donald B. and Neal Thomas, 1992. Characterizing the effect of matching using linear propensity score methods with normal distributions, *Biometrika*, vol. 79(4), p. 797–809.

Salmon Wesley C., 1971. *Statistical explanation and statistical relevance*. Pittsburgh, PA, University of Pittsburgh Press, 117 p.

Salmon Wesley C., 1980. Probabilistic causality, *Pacific Philosophical Quarterly*, vol. 61(1–2), p. 50–74.

Salmon Wesley C., 1984. *Scientific explanation and the causal structure of the world*. Princeton, NJ, Princeton University Press, 305 p.

Sanders David, 1980. Path analysis/causal modeling, *Quality and Quantity*, vol. 14(1), p. 181–204.

Schwartz Daniel, 1989. L'explication en épidémiologie, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.). *L'explication en sciences sociales: la recherche des causes en démographie*, p. 127–140. Brussels, CIACO éditeur and ARTEL, 478p. (Chaire Quetelet 1987.)

Shipley Bill, 2000. *Cause and correlation in biology*. Cambridge, UK, Cambridge University Press, 317 p.

Simon Herbert A., 1957. Spurious correlation: a causal interpretation, *in*: Herbert A Simon (ed.),*Models of man*, p. 37–49. New York, Wiley, 279 p.

Sosa Ernest and Tooley Michael (eds.), 1993. *Causation*. Oxford, Oxford University Press, 252 p.

Strotz Robert H. and Wold Herman O. A., 1960. Recursive versus nonrecursive systems: an attempt at synthesis, *Econometrica*, vol. 28(2), p. 417–427.

Suppes Patrick, 1970. *A probabilistic theory of causality*. Amsterdam, North-Holland, 130p.

Swain Marshall W., 1989. Theories of causation, *in*: Josianne Duchêne, Guillaume Wunsch, and Éric Vilquin (eds.), *L'explication en sciences sociales: la recherche des causes en démographie*, p. 197–214. Brussels, CIACO éditeur and ARTEL, 478 p. (Chaire Quetelet 1987.)

Taylor Richard, 1966. *Action and purpose*. Englewood Cliffs, NJ, Prentice-Hall, 269 p.

Wilson Thomas P. and Bielby William T., 1983. Recursive models for categorical data, *Social Science Research*, vol. 12(1), p. 109–130.

Winship Christopher and Mare Robert D., 1983. Structural equations and path analysis for discrete data, *American Journal of Sociology*, vol. 88(1), p. 54–110.

Wright Sewall, 1921. Correlation and Causation, *Journal of Agricultural Research*, vol. 20(7), p. 557–585.

Wright Sewall, 1934. The method of path coefficients, *Annals of Mathematical Statistics*, vol. 5(3), p. 161–215.

Wright Sewall, 1968. *Evolution and the genetics of populations*. Chicago, University of Chigago Press, 3 vol., 470 + 512 + 614 p.

Wunsch Guillaume, 1986. Causal connections in causal modeling, *Genus*, vol. 42(3–4), p. 1–12.

Wunsch Guillaume, 1988. *Causal theory and causal modeling*. Leuven, Leuven University Press, 200 p.

Wunsch Guillaume, Duchêne Josianne, Thiltgès Évelyne, and Salhi Mohamed, 1996. Socio-economic differences in mortality. A life course approach, *European Journal of Population*, vol. 12(2), p. 167–185.

C H A P T E R

131

Multidimensional Exploratory Analysis

LAURENT TOULEMON

Institut national d'études démographiques (INED), Paris, France

Basic demographic data for a population are most often reported by age, sex, and year. Demographers may work with regional population counts by sex and age at different dates or events (e.g., births, deaths, marriages, entries or exits) categorized by sex, age, and year.

In addition to these fundamental components, demographic analysis can incorporate complementary information, such as rank among children of the same mother for births; causes of death; and reason, place of origin, or place of destination for migrations. The data can be extended to other areas for analysis dealing with more complex phenomena (e.g., transition to adulthood, which is marked by the occurrence of multiple events) or when the population is divided into groups with multiple characteristics (e.g., regions defined by a set of demographic and socioeconomic characteristics).

Regression models seek to *separate* the contribution of each *factor* to variation in the phenomenon being studied and can be seen as methods of standardization (see Chapter 132). Conversely, multidimensional exploratory methods serve to identify the *global structure* of the data and summarize it using indices constructed after the fact, with reference to optimality criteria based on the proportion of information explained. In this chapter, I present techniques for exploratory analysis of data and then describe several methods of multidimensional analysis and classification used in demography.

I. EXPLORATORY DATA ANALYSIS

Methods for *exploratory data analysis* are characterized by the refusal to impose an *a priori* model on the data. Their goal is to summarize as best as possible the structure of the information present in the data. Given a set of data, exploratory methods use different graphic procedures to summarize the variables, first individually and then in groups. They are also useful for identifying points that are outliers with respect to this structure.

Exploratory data analysis can focus on the distribution of a single variable, on conditional distributions, or on relationships between several variables. Once considered too simple for statisticians, this form of analysis has spread widely thanks to improved software, and it is now part of the basic statistical toolkit (Cleveland, 1993; 1994; Erickson and Nosanchuk, 1992; Hartwig and Dearing, 1979; Tukey, 1962). The practice of exploratory data analysis can also profit from the application of insights from statistical graphics (Tufte, 1990, 2000; Bertin, 1967). Eugen Horber (2004) has provided a well-annotated bibliography.

1. Distribution of a Single Variable

The box-and-whiskers graph, or *box plot*, is an efficient way to represent the distribution of a variable (Tukey, 1988; Chambers *et al*., 1983). Take, for example, a variable X defined by n elements and a set of values

$\{x_i\}_{i=1,n}$. This set of possible values can be described by a *five-number summary*. The median M divides the set into two equal parts; one-half of the elements has values lower than the median, and the other half has higher values. Around the median, the first and third quartiles Q1 and Q3 separate elements arranged by order of value into four quarters of equal size. Outside of the interquartile range [Q1; Q3], limiting values are defined to show the maximum dispersion of the variable. First, the limits within which the values should *normally* be found are calculated, with the interquartile range serving as a reference to determine the *normal* range of variation; $m = Q1 - 1.5\ (Q3 - Q1)$ and $m' = Q3 + 1.5\ (Q3 - Q1)$ are, respectively, the lower limit and the upper limit that can be assigned to this distribution. The observed extreme values m_1 and m_3 are, respectively, the smallest value larger than m and the largest value smaller than m'. Most often, m_1 and m_3 are the extreme values taken by X. If this is not the case, values outside the interval $[m_1; m_3]$ are considered to be outliers and are explicitly indicated as such. The distribution of *X* is represented by a *box* from Q1 to Q3, including the median M and *whiskers* linking the quartiles to the limiting values. Values outside the limits are shown as single points to complete the graphic (Fig. 131–1A). Sometimes, extreme values whose distance from the quartiles is more than three times the interquartile range are indicated using different points (NIST/SEMATECH, 2004). In the fictional example provided, the distribution of the variable *X* has a longer tail in the direction of the higher values; the median is toward the bottom of the box, the top whisker is longer than the bottom whisker, and there is an outlier in the upper half of the box.

Extreme values should first be verified, because the assumption is that they result from an error in measurement, coding, or data entry. They can then be the subject of a specific study justified by their exceptional position, or they may be suppressed when studying the distribution of X. The treatment and possible suppression of these outlying points is the first step in the examination of data. However, the general form of the distribution depends on the scale used to observe X, and scale may also influence the diagnosis of extreme values. If a transformation of X is observed instead of X itself, the form of the distribution is changed. For example, Figure 131–1B shows the logarithmic transformation ln (X) instead of X. An extreme value is now observed in the direction of the minimum values.

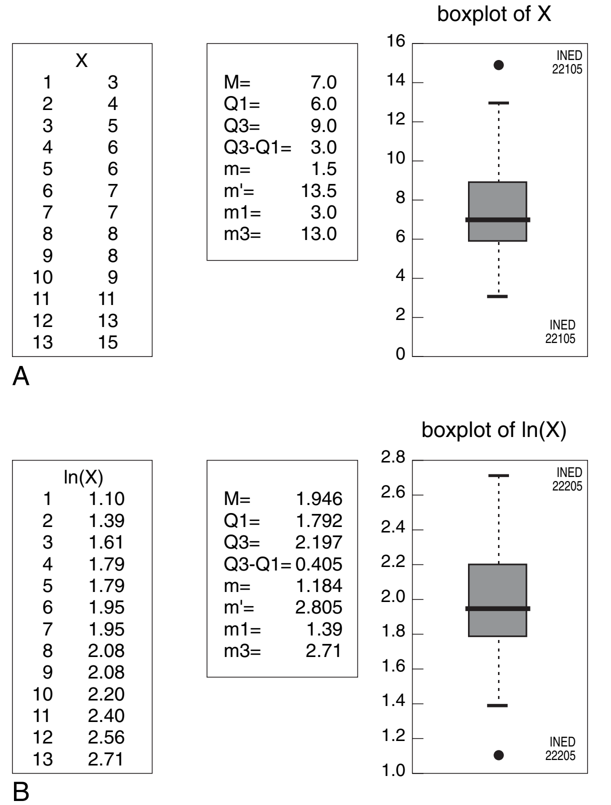


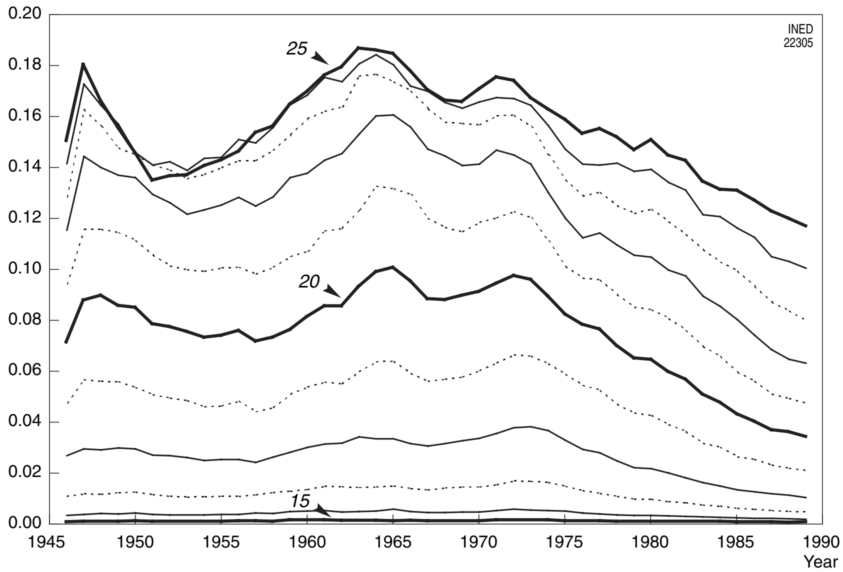
FIGURE 131–1 **A:** Example of a box-and-whiskers plot. **B:** Example of a box-and-whiskers plot after transformation of X to ln (X).

2. Conditional Distributions

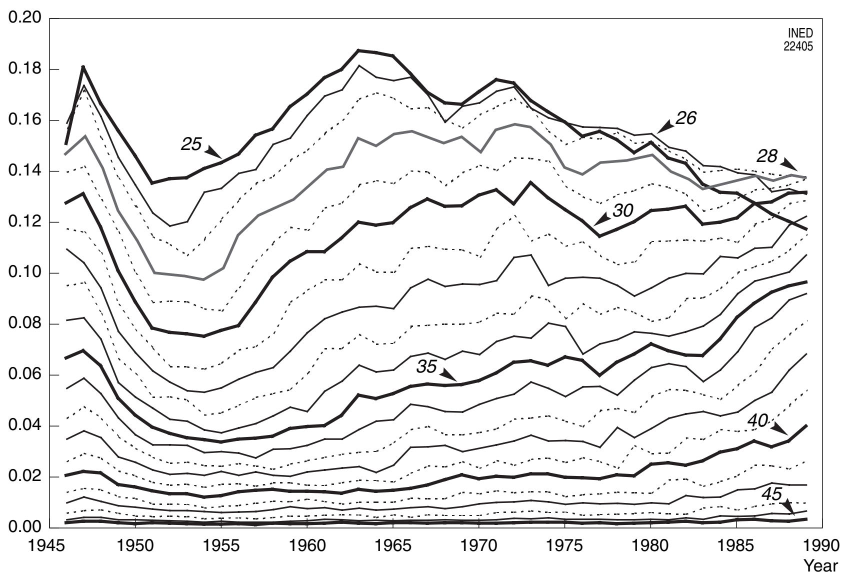
Conditional distribution is particularly useful for comparing different distributions. Let us start with a very simple example using demographic data by age and by year, such as the probability of first birth for French women between the ages of 15 and 49 years, as observed from 1946 to 1989 (Ni Bhrolcháin and Toulemon, 1996; Rallu and Toulemon, 1993).

The series of age-specific birth probabilities, presented in two graphs for clarity (Fig. 131–2), can be summarized in a series of box plots graphed by age (Fig. 131–3) or by year (Fig. 131–4). Figure 131–3, which presents a series of box plots by age, shows that probabilities before age 25 are more variable than those at older ages during the period under study. The distributions are not symmetric at young ages or at older ages. The distribution of probabilities at young ages is more spread out in the direction of the low values, whereas at older ages, the reverse is true. In both cases, the extreme values are found at the end of the period, showing that birth probabilities in the 1980s are moving away from the distributions in earlier years, falling at young ages and rising at older ages.

The same data are presented from a different angle in Figure 131–4, which shows years along the x-axis and, for each year, the distribution of birth probabili-



A



B

FIGURE 131–2 **A:** First birth probabilities by age from 1946 to 1989 in France for women between the ages of 15 and 25 years. **B:** First birth probabilities by age from 1946 to 1989 in France for women between the ages of 25 and 45 years. (From Ni Bhrolcháin and Toulemon, 1996).

ties for ages 18 to 37 years. The compression of interquartile intervals Q1 to Q3 is clearly visible in this graph. The probabilities are less and less differentiated by age; the decrease in probabilities at young ages stretches the distribution toward the lower end, whereas at the other end of the distribution, the maximum value decreases. Starting in 1973, the extreme values decrease; on one end, the minimum probability (at age 18) falls, and on the other end, the ages where fertility is highest spread out around age 29 years.

The exploration could be made more precise by separating the data in Figures 131–3 and 131–4 into subsets. For example, Figure 131–3 could be separated into two figures showing the distribution of birth probabilities at each age for the periods 1946 to 1972 and 1973 to 1989, whereas Figure 131–4 can be divided by separating groups younger than 28 years, for whom fertility has been declining since 1973, from older groups, for whom fertility is stable or rising.

After having described the distribution of birth probabilities according to each of the variables age and

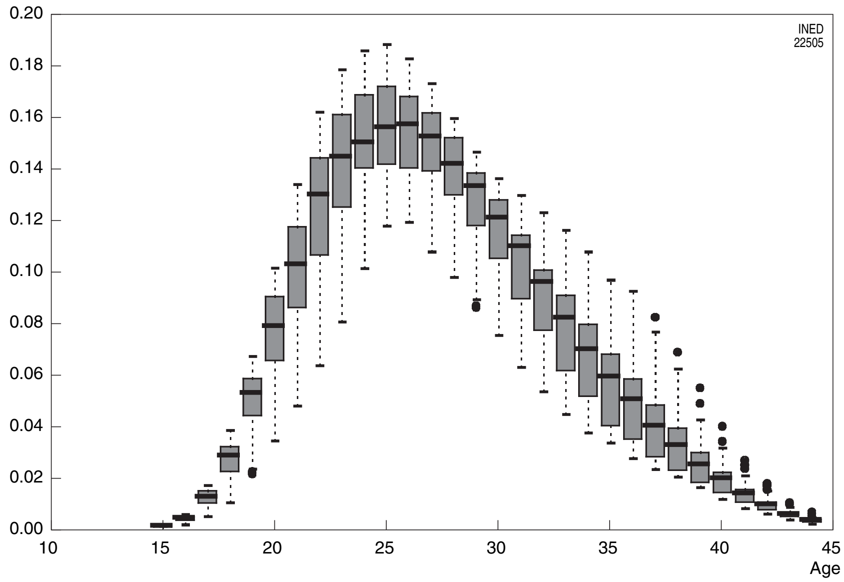


FIGURE 131–3 Box plots of first birth probabilities by age from 1946 to 1989 in France.

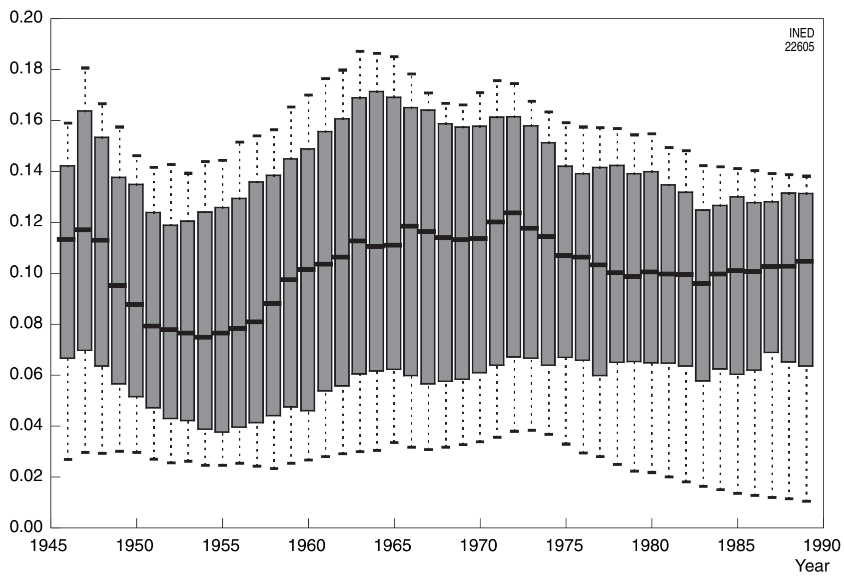


FIGURE 131–4 Box plot of first birth probabilities by year in France for women between the ages of 18 and 37 years.

year, it is possible to show variations in fertility by age and year simultaneously (see Chapter 7). Year and age can be used as geographic coordinates to map a landscape where altitude represents the variable of interest (here, the age-specific probabilities of first birth). Several possible representations are included in standard spreadsheet software, including the level curve (i.e., contour plots). In these figures, popularized among demographers by James Vaupel, Bradley Gambill, and Anatoli Yashin (1985), years are placed on the x-axis and ages on the y-axis, as in Lexis diagrams, and each age-year cell is shaded based on the value of the variable under study. This shading makes level curves visible, and the cartography makes it possible to observe all of the values taken by the variable of interest at a glance. Figure 131–5, for example, which presents the same data as in the previous figures, shows that the age when the probability of a childless women giving birth to a first child drops has grown steadily since the late 1950s. The increase in fertility in older ages preceded the decrease in fertility at young ages by 20 years.

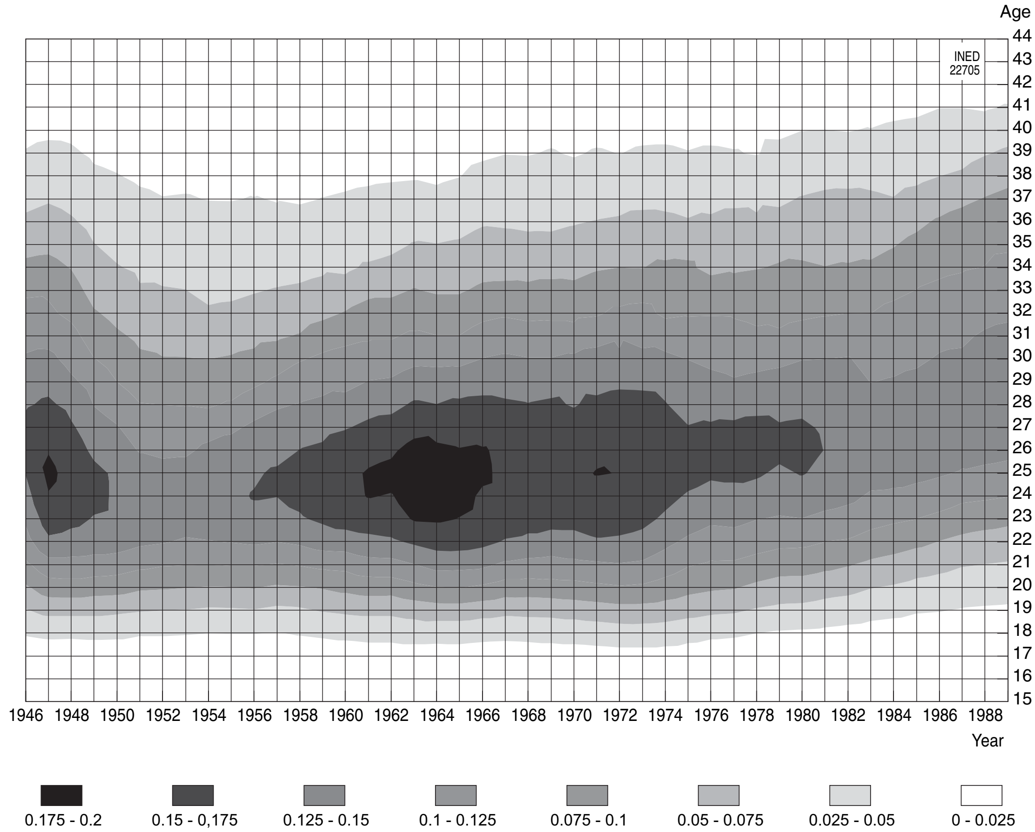


FIGURE 131–5 Contour plot of first birth probabilities among women between the ages of 15 and 45 years in France between 1946 and 1989.

3. Associations between Variables

The relationship between two variables can also be represented graphically by designating one variable on the x-axis and the other on the y-axis (Mosteller and Tukey, 1977). In graphs that cross two series of variables (i.e., scatterplots), each element is represented by a point; if the points are years, they can be connected. This is illustrated in Figure 131–6, which plots first birth probabilities in France for ages 25 to 30 against probabilities at age 24 for the years 1946 to 1989. The strong correlation between fertility at different ages reflects simultaneous variation at all ages from one year to another, a period effect. Figure 131–6 also shows a change in this relationship between 1947 and 1964, when the probabilities are strongly correlated among each other (see Fig. 131–6, black points), and the more recent period, where the significant decline in probabilities at age 24 is not accompanied by an equivalent decline at older ages (see Fig. 131–6, white points).

This type of figure can be generalized as a matrix of graphs (i.e., scatterplot matrix) presenting, for *n* variables described simultaneously, a table with *n* rows and *n* columns and with each cell (row *i*, column *j*) occupied by a graph comparing variables *i* and *j*.

4. Exploratory Analysis to Validate Models

The results of exploratory data analysis are rarely published as such. These analyses are useful as ways to become familiar with the data, but they do not provide easily quantifiable results. Nevertheless, exploratory data analysis is not solely limited to this first contact with the data. It can also be used as a tool to validate model-generated results. For example, in a study on the determinants of fertility decline in developing countries (Mauldin *et al.*, 1978), different regression models were developed as part of an attempt to separate or quantify the proportion of the decline due to economic development (i.e., social setting) and to family planning programs (i.e., program effort). The various regression models were compared and validated using exploratory techniques applied to the residuals of different models rather than to raw data. Figure 131–7 shows how exploratory analysis can be

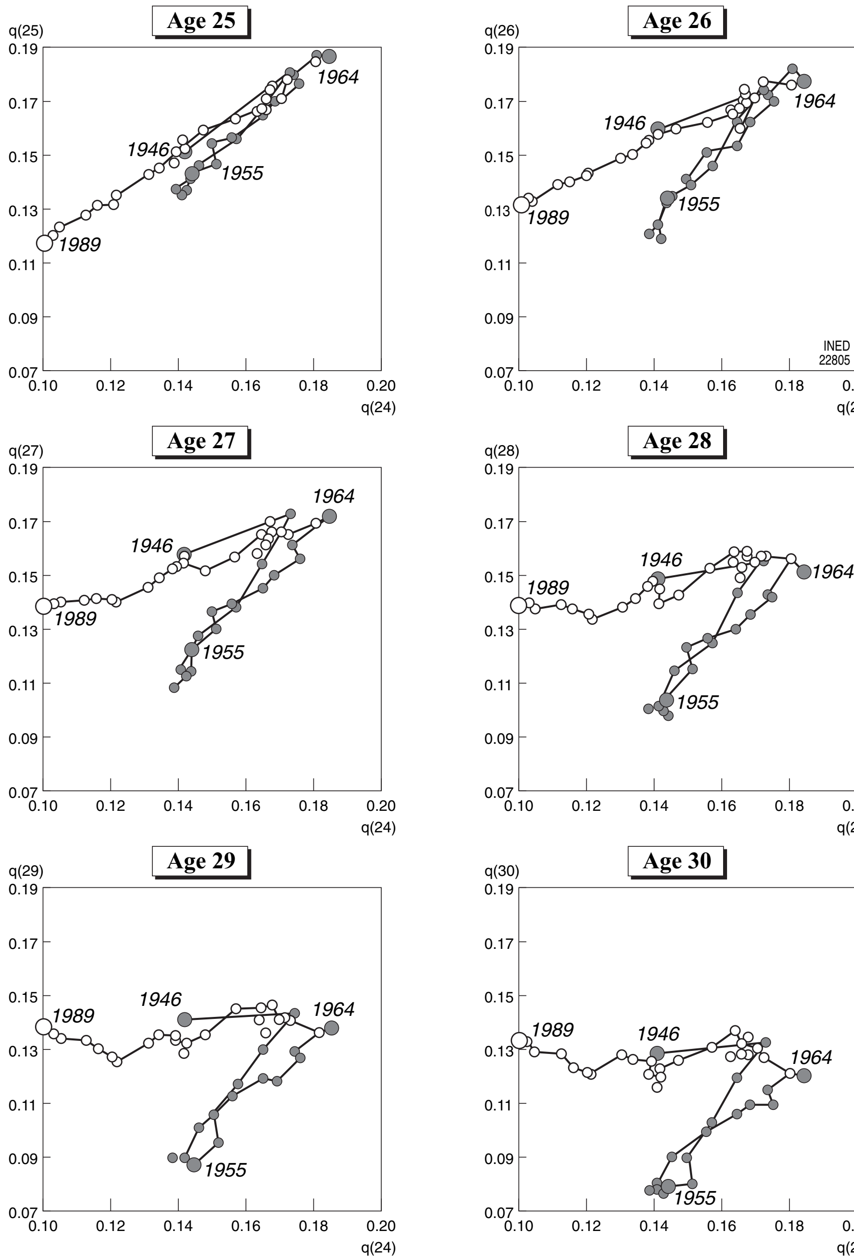


FIGURE 131–6 Scatterplots of first birth probabilities at ages 25 to 30 against probabilities at age 24, by year, France, 1946–1989. Interpretation: In each graph, first birth probabilities at a given age (age 25–30 years, respectively, in the six graphs) are plotted against the probability at age 24. Each point represents 1 year; points for the years 1946–1964 are blue, and points for more recent years (1965–1989) are white.

used to validate models. The magnitude of the residuals left by each model can be compared using box plots. The table referred to in the sixth box-and-whiskers plot in Figure 131–7 presents residuals from a model that includes an interaction between intensity of family planning program effort and socioeconomic context.

Exploratory analysis of the residuals confirms the regression results. Intensity of program effort is the primary factor differentiating country-specific fertility rates; residuals are smaller for model 3 than for model 2. There is also an interaction between family planning effort and socioeconomic context; differences in fertility due to the intensity of family planning programs are larger when the socioeconomic context is favorable. The two factors are correlated; family planning programs are more common and more active in countries where the socioeconomic situation is favorable. It is possible to estimate an indirect effect of socioeco-

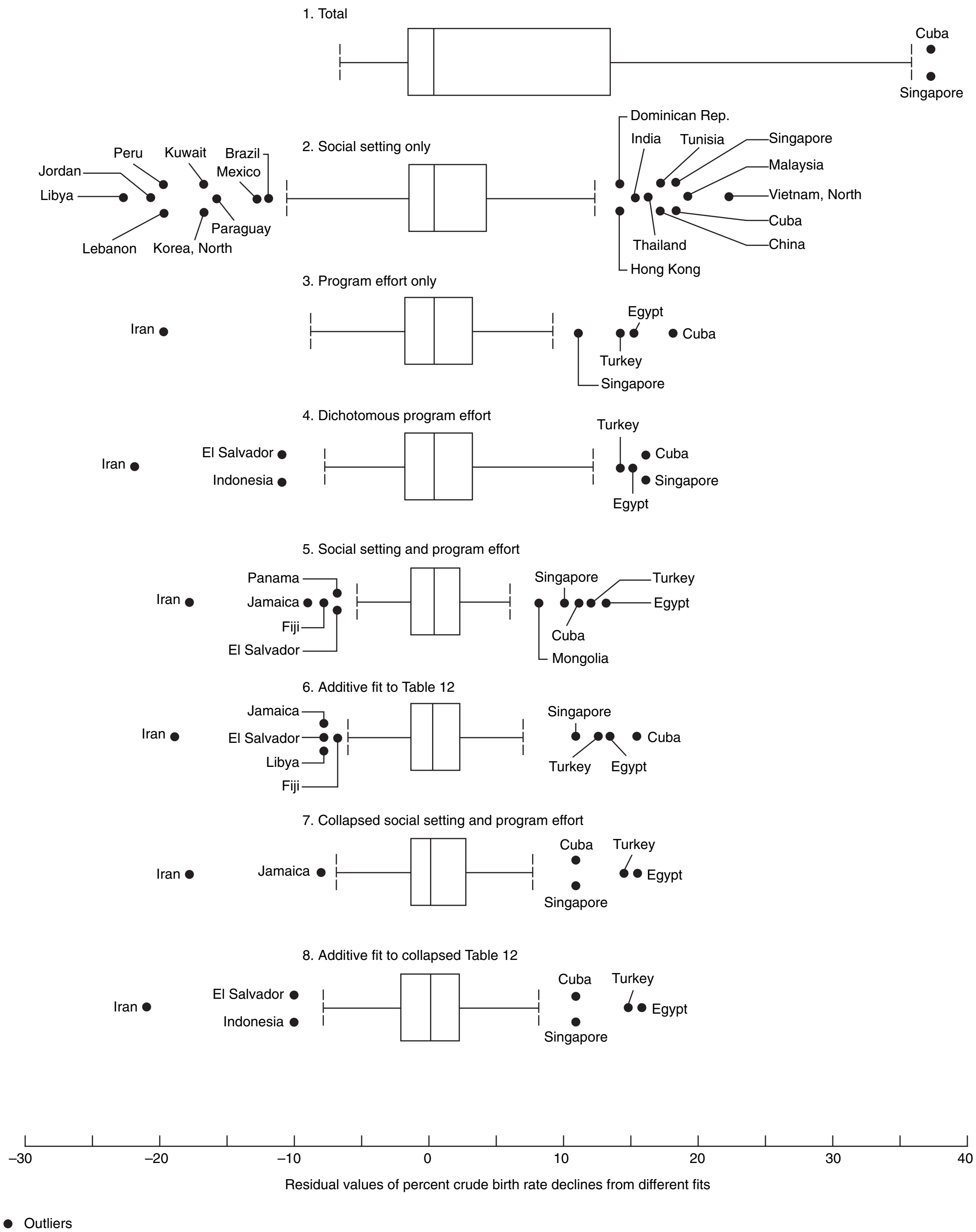


FIGURE 131–7 Residuals of different models explaining fertility decline between 1965 and 1975 in 94 countries as a function of socioeconomic conditions and family planning program effort. (From Mauldin *et al.*, 1978. Courtesy of Blackwell Publishing.)

nomic setting, because socioeconomic characteristics partially explain the presence of a family planning program.

The principal contribution of the exploratory analysis is to highlight the extreme values of the residuals for each model, showing which countries do not conform to the common model defined by each regression. There was almost no decline in fertility in Iran between 1965 and 1975, despite the presence of a family planning program introduced in 1967, whereas

Egypt, Turkey, Cuba, and Singapore experienced larger declines than would be predicted by the models.

II. METHODS FOR MULTIDIMENSIONAL ANALYSIS

Methods of *data analysis* deal with the common structure of different variables and aim to summarize the associations and oppositions among variables in the best way possible. For instance, the scatterplot could be generalized to more than two series by considering all the values taken by a set of variables (e.g., rates, probabilities) as points in a multidimensional space, a space with as many dimensions as there are variables. Given a set of observations of these variables, this type of plot produces a cloud of data that the analyst seeks to represent in the best way possible, losing as little information as possible and using the smallest possible number of indices. For example, in 1959, Sully Ledermann posed the question of the number of indices necessary to describe a mortality table and proposed "a solution independent of all *a priori* assumptions as to the structure of the phenomenon being studied" (Ledermann and Breas, 1959).

These methods of data analysis are all based on the effort to summarize the global structure of the data with a limited number of indices by reducing the number of dimensions of the information. We start by illustrating these principles using a concrete example and then present specific methods of analysis.

1. Principal Components Analysis

Following Sully Ledermann and Jean Breas, we consider the set of 157 mortality tables that served as the base for the construction of the model life tables published in 1955 by the United Nations (United Nations, 1955). These life tables can be described as a set of multidimensional data: two series of probabilities of death, one for men and one for women, defined for different countries (of which many are European) at different periods (groups of years from the early 20th century). For each sex, country, and group of years, we have a series of death probabilities for the age groups 0, 1 to 4, 5 to 9, . . . , 75 to 79, 80 to 84, or the series ${}_1q_0$, ${}_4q_1$, ${}_5q_5$, ${}_5q_{10}$, . . . , ${}_5q_{80}$.

The data are treated as 157 individuals (the life tables or country-years) characterized by their values on 38 variables (18 probabilities of death and life expectancy at birth for each sex) and represented by 157 points in a 38-dimensional space. The probabilities by sex and age are strongly correlated with each other. In countries (or periods) where mortality is high, it is high at all ages and for both sexes. This empirical observation is represented geometrically by the lengthened form of the cloud of data. Sully Ledermann and Jean Breas first presented 157 pairs of two probabilities (${}_5q_{30}$ and ${}_5q_{40}$, for men) as points on a two-dimensional graph (Fig. 131–8).

The cloud of points obtained by this method is two-dimensional. It can be represented on a new set of two axes, γ_1 and γ_2, chosen such that γ_1 corresponds to the major axis of the ellipse formed by the 157 points, with the axis representing the most inertia, and γ_2 is perpendicular to γ_1. The same principle—looking for the axes that contain the most inertia, that is, the axes along which the cloud of data points is most spread out—is then applied to the complete set of data series. Each data series (18 age-specific death probabilities and a life expectancy for each sex) is first transformed; the probabilities are replaced by their logarithm and the life expectancy by the logarithm of 100 years minus the life expectancy [$\ln(100 - e_0)$]. The variables are next centered and scaled[1] and each series is represented as a point in a 38-dimensional space. The 157 series form a *cloud* of 157 points in this space. If the mortality probabilities were all independent from each other, the points would be spread out in the shape of a sphere; because they are correlated with each other, the cloud is shaped like an extended ellipse.

Just as the pairs ${}_5q_{30}$ and ${}_5q_{40}$ were represented along the axes γ_1 and γ_2, the set of 38 transformed probabilities is analyzed along a certain number of axes. The first axis is chosen in the direction that maximizes inertia; the second axis is chosen according to the same rule, but with the constraint of being orthogonal to the first, and so on.

Principal components analysis is based on the correlation matrix for the 38 scaled and centered variables (Lebart *et al*., 2000).[2] The eigenvalues of this matrix are calculated along with the axes that correspond to them. The coordinate of each life table on each axis is a linear combination of the values taken by the variables in the table.

The coordinate on the first axis best summarizes the position of each table with respect to the others in the sense that, of all the combinations of variables,

[1] Variables are centered by subtracting their mean; the mean of the centered variable is 0. Variables are scaled by dividing by their standard error; the variance of the scaled variable is equal to 1.

[2] This work, the most recent in a long series of publications by the same authors, sometimes collaborating with Alain Morineau and Jean-Pierre Fénelon, offers an instructive overview of these methods and a complete bibliography covering French-language and English-language sources. Bry (1995, 1996) also offers brief graphical and mathematical presentations of these methods.

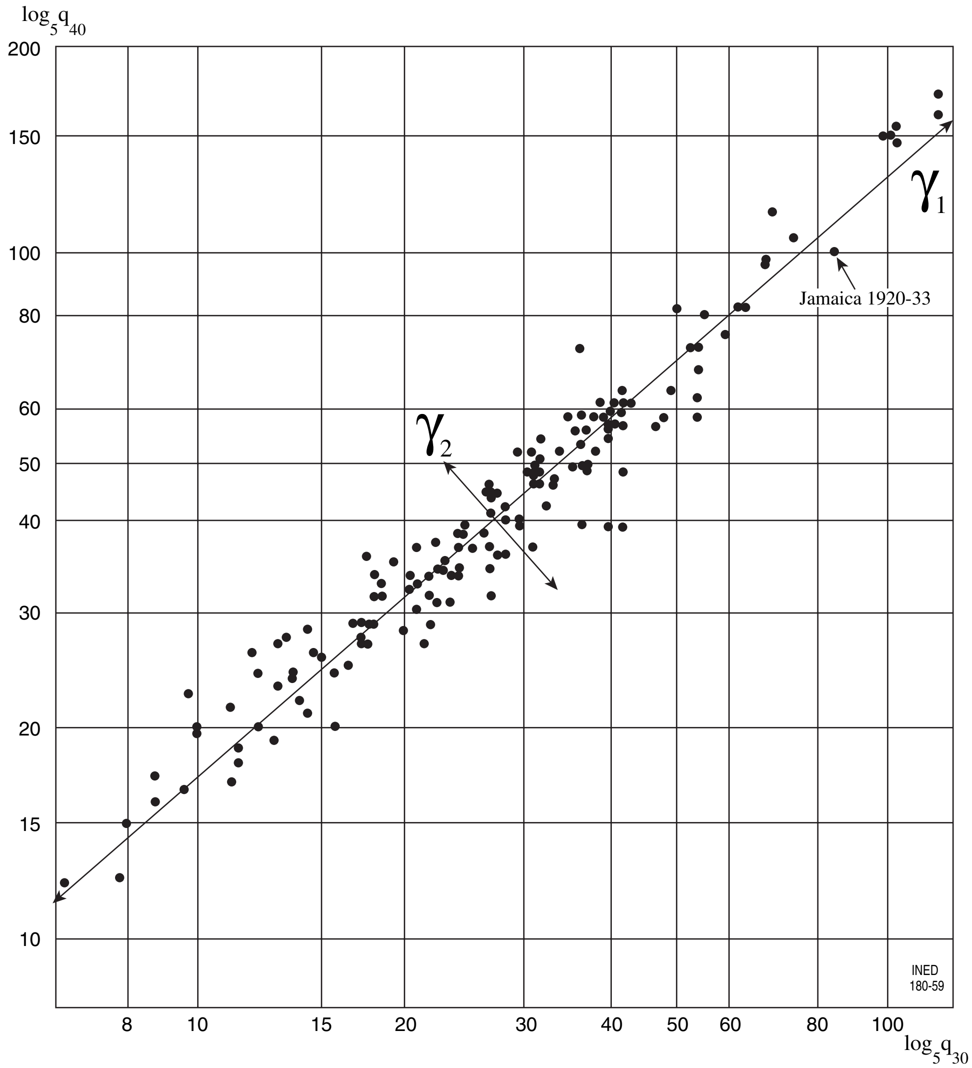


FIGURE 131–8 The probability of death for men between the ages of 40 and 45 years is plotted (logarithmic scale) against the probability of death between the ages of 30 and 35 years, using the United Nations life tables. (From Ledermann and Breas, 1959. Courtesy of *Editions de l'INED*, 1959.)

the coordinate on the first axis represents the largest proportion of the variance within the set of life tables. The first axis is therefore the best summary representation of the cloud. Similarly, the plane formed by the first two axes is the best two-dimensional representation, with each table represented by two coordinates on this plane.

To present summary indices that offer a concrete interpretation, the authors suggest restricting the analysis to the first three axes and then rotating these axes to give them a meaning more directly related to the mortality rates. The first three axes represent, respectively, 81.5%, 7.4%, and 3.7% of the total inertia of the cloud. The authors replace them with three new axes produced by two successive rotations. The first rotation, in the plane of the first two axes, transforms axes (1,2) into (1′,2′); the second rotation, in the plane (2′,3), transforms these axes into (2″,3″), as indicated in Figure 131–9. The first of these new axes corresponds to the general level of mortality and is very strongly correlated with the mortality of men between the ages of 10 and 34 years and women between ages of 5 and 44 years. It contributes 77% of the inertia. The second of the new axes (9.5%) represents adult mortality after age 30, and the third axis (6.5%) represents mortality at extreme ages, younger than 5 and older than 70 years (see Fig. 131–9). By construction, the cumulative inertia in the first three axes is conserved (92.6%), but the new axes correspond more closely to entries in the life tables. Sully Ledermann (1969) continued along this path in proposing single and double entry tables based on this principle.

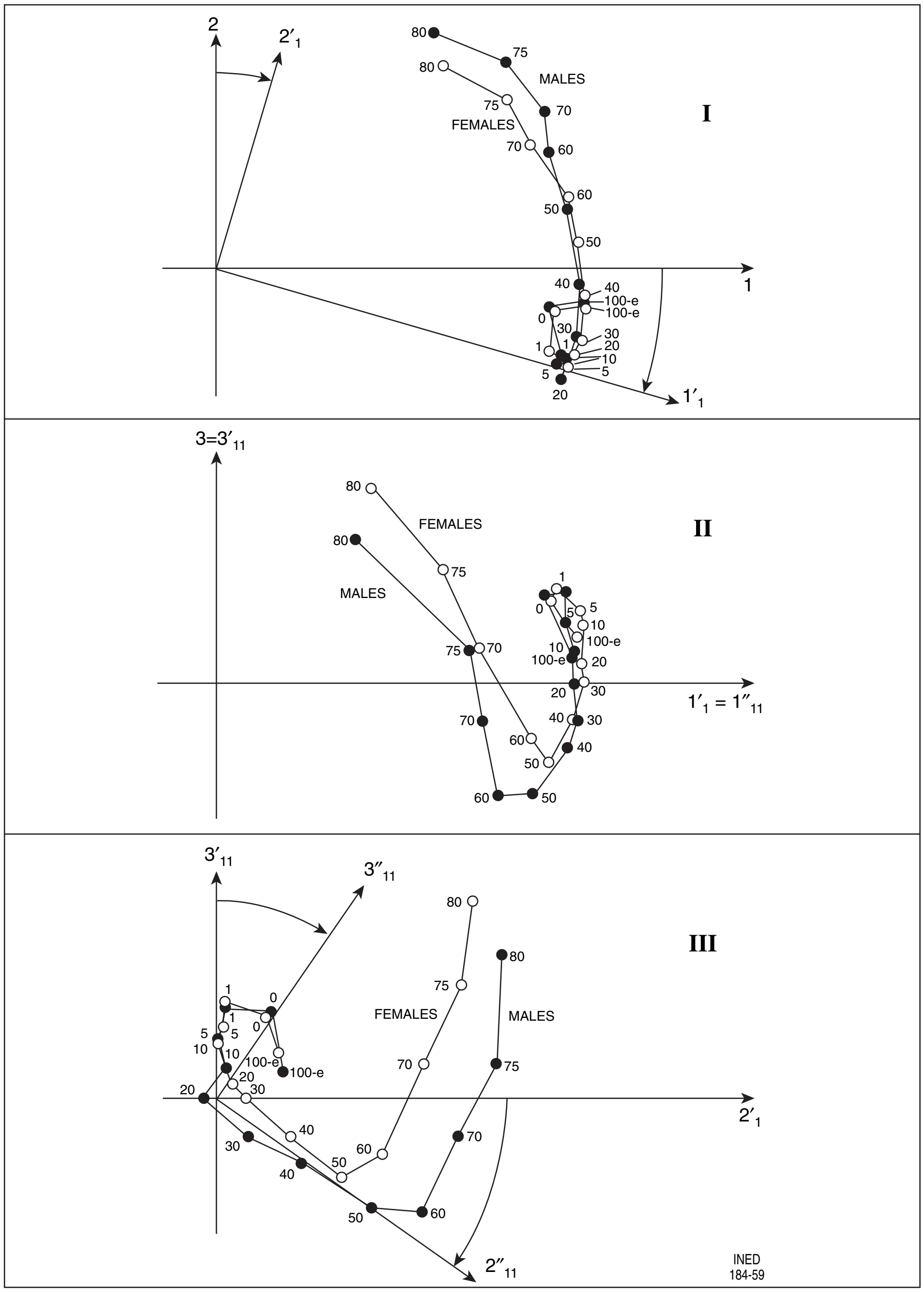


FIGURE 131–9 Rotation of the first three axes in principal component analysis. (From Ledermann and Breas, 1959. Courtesy of *Editions de l'INED*, 1959.)

2. Multiple Correspondence Analysis

Principal components analysis is based on the information contained in the correlations between different quantitative variables. When information is collected in the form of categorical variables, correspondence analysis is the method of choice (Lebart *et al.*, 2000). Given two variables X and Y, information is presented in a contingency table or cross tabulation, which divides the population according to the two variables simultaneously. Each row of the table *x*, corresponding to a value of the variable X, can be considered as an individual with a profile representing its distribution according to Y and a weight proportional to the number of individuals with the value *x*. We then compare each individual row *x* to the profile representing the average distribution of the variable Y. This comparison generates a distance between points x, which defines a cloud of individual rows in the space of values of the variable Y. The cloud of individual rows has defined axes of inertia, each of which is associated with an eigenvalue representing the proportion of inertia contained in the axis. Similarly, the individual columns y are distributed in a space defined by the possible values of X. The axes of inertia of these two clouds are associated with the same eigenvalues and can be linked by formulas known as transition formulas; both individual rows and individual columns can be represented in the same space.

Correspondence analysis can be generalized to include more than two variables. The method can be used to analyze vast sets of categorical variables, such as variables gathered in surveys. For a given variable, each value becomes an individual that takes its coordinates in the space defined by the set of other variables. As in principal components analysis, results are summarized using a limited number of dimensions, and the cloud of individual values is observed along the successive axes or in the multidimensional space defined by the first axes of inertia. Multiple correspondence analysis has been established and used in very diverse contexts, particularly in optimal scaling research based on categorical responses. Michel Tennehaus and Forrest Young (1985) describe some of the different labels for this method.

3. Active Variables and Illustrative Variables

The choice of variables used to summarize is arbitrary to some degree. For example, Sully Ledermann and Jean Breas (1959) reinforced the coherence of the set of mortality probabilities by adding life expectancy at birth—itself a summary of the life table—to the series of 18 probabilities for each sex. Alternatively, they could have not included life expectancy in the calculation of the axes of inertia, but instead used life expectancy at birth as an *illustration* of the axes by seeing where this indicator was located in the space defined by the three principal axes.

When analyzing data, it is always possible to supplement the active variables used to construct the axes of inertia with illustrative variables that do not contribute to the inertia of the cloud but illustrate the results by characterizing either diverse groups of individuals (for categorical variables) or specific directions in the variable space (for quantitative variables).

4. Canonical Analysis

Canonical analysis is a very general method, of which multiple correspondence analysis, linear regression, and discriminant analysis can be considered particular cases (Lebart *et al*., 2000).

Given a group of variables X and another group of variables Y defined for the same individuals, the goal of canonical analysis is to look for the linear combinations x and y that maximize the correlation between x and y. In an analysis of cause-specific male mortality in the age groups of 25 to 54 years and 55 to 79 years in different Italian provinces (Caselli and Egidi, 1981) (see Chapter 55), two sets of data are examined for each of 54 Italian provinces. The provinces are described by mortality from 16 groups of causes of death and by 38 variables related to socioeconomic conditions in the province (i.e., external variables). In the first step of the analysis, two principal components analyses are used to summarize the structure of each set of variables. A principal components analysis, including all of the variables, weights the external variables heavily. There are twice as many external variables as mortality variables, so this analysis provides very little additional information. Principal components analysis of the external variables makes it possible to suppress 11 redundant variables, reducing the number of external variables to 27. Canonical analysis shows four axes representing dimensions of correlation between provincial cause-specific mortality among men younger than age 55 and the socioeconomic characteristics of provinces. The first axis associates degenerative diseases, diseases of the digestive system, and accidents to variables measuring economic development, confirming the excess mortality of the north of Italy already revealed in the preceding analyses. The second axis associates diseases of the respiratory system and infectious diseases with indicators of poor economic conditions and indicators of isolation; a high proportion of bachelors and widowers and areas with mountains or inland hills. The correlation between cancers of the respiratory system and pollution and, at older ages, the consumption of tobacco appears more clearly using canonical analysis than by comparison of the two principal components analyses, although the dimensions demonstrated are not *a priori* optimal for describing the variation across provinces in cause-specific mortality or in socioeconomic conditions.

Canonical analysis is difficult to apply and to use, but it subsumes different methods of analysis as particular cases. If a group of variables Y is reduced to a single variable y, canonical analysis is equivalent to a linear regression of y on X. If X and Y are categorical variables, the analysis becomes correspondence analysis.

III. METHODS OF CLASSIFICATION

1. Ascendant Hierarchical Classification

Methods of data analysis can be used to summarize the structure of a set of data. Methods of classification are their complements. They are used to construct homogenous groups without specifying the criteria that define this homogeneity. The optimal classifications are those that minimize intragroup variance (i.e., variance among the members of a single group) and maximize intergroup variance (i.e., variance across members of different groups). The data set is represented as a multidimensional cloud, constructed from the full set or after having reduced the number of dimensions using a preliminary analysis. Groups are then constructed by minimizing the distance between members of the same group and maximizing the distance between members of a different group. The most frequently used methods are hierarchic classification methods. The principle of these methods is simple. First, the two closest individuals are grouped together, and the position of that group is calculated. The next two closest individuals or groups are then classed together. There are different types of classification methods, depending on how the position of a group, the distance between an individual and a group, and the distance between two groups are defined.

Each classification can be associated with the proportion of inertia that it absorbs, and the total inertia of the cloud can then be decomposed into the sum of the intergroup inertia and the intragroup inertia. Intragroup inertia is smaller when the members of the group are close, that is, when the grouping is appropriate. The association between group and inertia makes it possible to assign an index of inertia to each group. A hierarchic tree can then be drawn by placing each group along an x-axis according to the reduction of inertia associated with the grouping. Starting from the first grouping of two individuals and ending with the last grouping that gathers two groups into one single group, it is possible to see at what point the groupings become too costly in terms of the loss of information and to choose an optimal number of classes.

Figure 131–10, borrowed from France Meslé and Jacques Vallin (2002), summarizes classifications of four groups of 28 European life tables for men and for women in 1965 and 1995. For both sexes, heterogeneity across countries is best summarized by a dichotomy in 1995, whereas the trees for 1965 show more varied levels of groupings. For male mortality, a clear division between the western and eastern parts of the continent develops over the course of the period. By 1995, there are two very different groups of life tables; these groups almost exactly replicate the old political separation of the continent, with the exception of Portugal, which is grouped with the eastern countries because of its high adult mortality. For women, the countries of Central and Eastern Europe are grouped with Western Europe in 1995, with the exception of Bulgaria and Romania, which are grouped with the countries of the former USSR. These classifications do not provide explicit information on the details of the life tables that produce them, but they do show the uniqueness of the countries of the former USSR and the sex-specific divergence of the countries of the old Soviet bloc from other countries.

2. Discriminant Analysis and Segmentation

Multiple discriminant analysis can be a useful tool for data taking the form of a set of variables X and a variable y defining fixed classes of individuals. These methods are used to determine the linear combination of X variables that best separates the different classes. They are particularly useful if one wants to assign a class to individuals for whom the X variables are known but the y variable is unknown (Lebart *et al*., 2000). If y takes only two values, discriminant analysis is equivalent to a linear regression of y on X.

A competing approach to this problem is offered by segmentation methods based on decision trees, in which a series of choices is used to segment the population into subgroups assigned to one or the other class, rather than applying a linear combination of X variables. There are three primary advantages to these methods. First, the decision rules are easily readable and interpretable. Second, any kind of X variable can be used, without prior coding. Third, the method is robust to outliers and accepts missing data. These methods are especially efficient for prediction. They are not based on any model built a priori, and they follow specific strategies of inference. For instance, the replication of the same method on subsets of individuals leads to a “decision forest,” which leads to a very robust set of results; the predictive value of these methods is much better than for other strategies (Breiman, 2001). These methods are now used in demographic research. In a recent demographic application, Francesco Billari, Johannes Fürnkranz, and Alexia Prskawetz (2000) studied differences in the transition to adulthood in Austria and in Italy. These

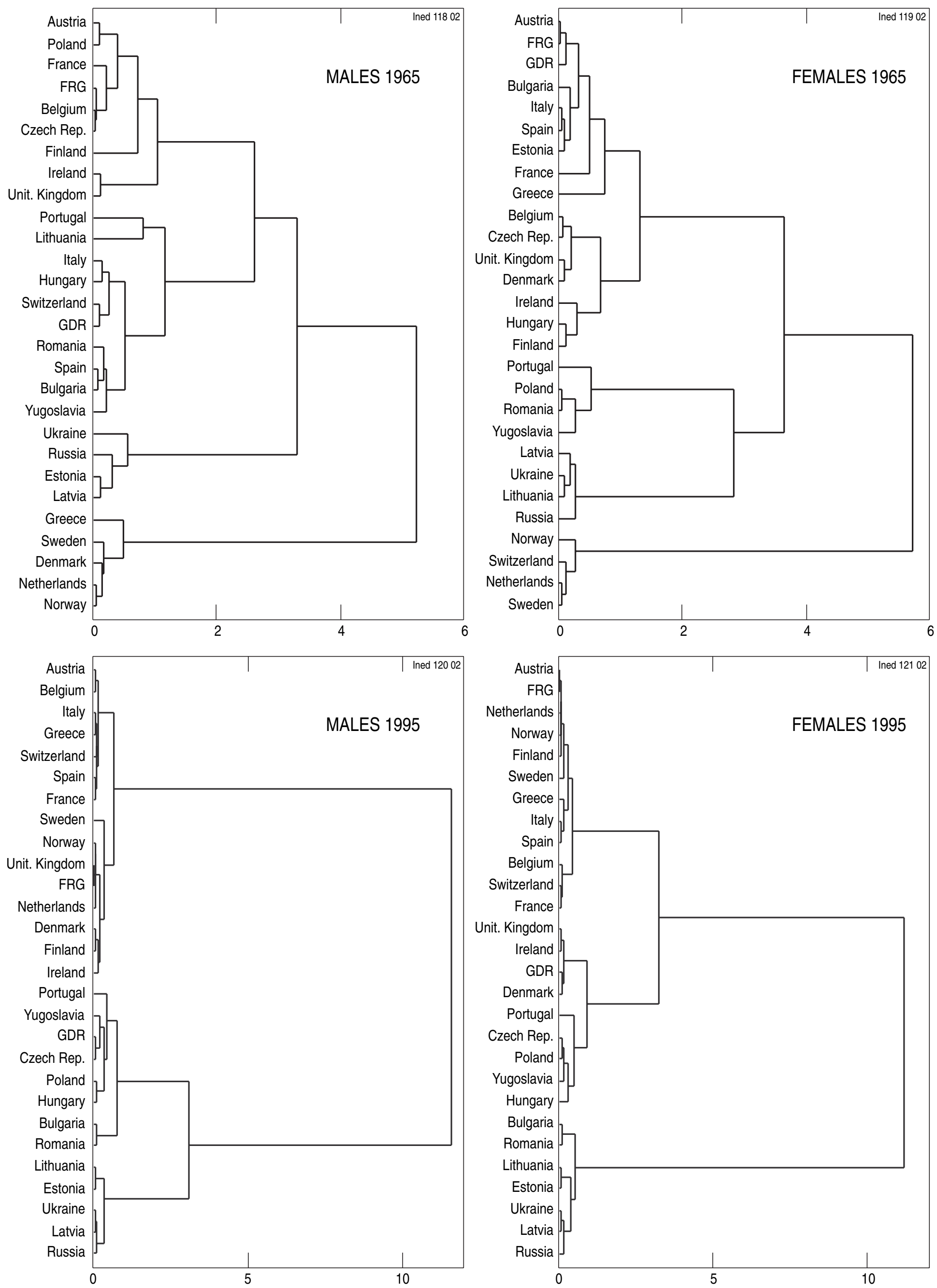


FIGURE 131–10 Tree diagrams resulting from the hierarchical analysis of age-specific mortality probabilities according to sex for 28 European countries in 1965 and 1995. (From Meslé and Vallin, 2002. Courtesy of *Editions de l'INED*, 2002.)

researchers constructed three sets of variables based on the dates of different transitions to adulthood collected by Family and Fertility Surveys: leaving school, first employment, leaving the parental home, first union, first marriage, and birth of the first child. For each of these transitions, they define dichotomous occurrence variables (i.e., quantum: whether the event took place), continuous calendar variables (i.e., timing: the age at which the event took place[3]), and placement with respect to the other transitions in four categories (i.e., sequencing, or placement relative to each of the other transitions: before, at the same time, after, neither took place). These variables are used along with sex, age

[3] For individuals who have not experienced the event, the age at the time of the survey is used.

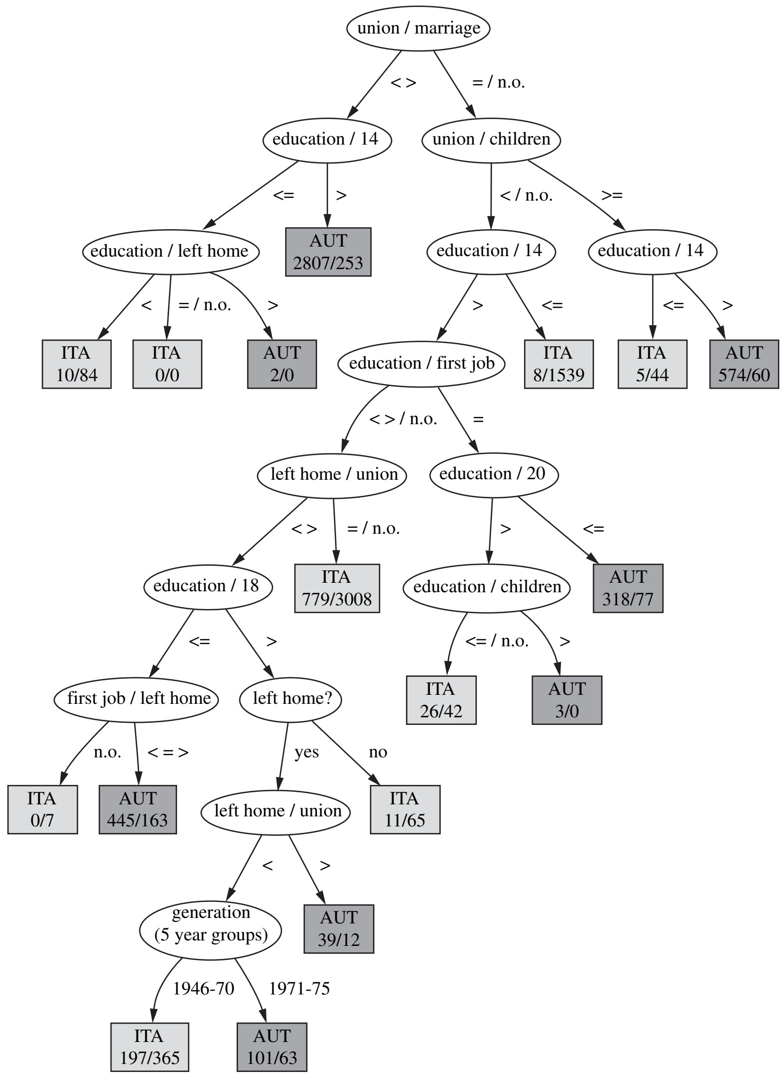


FIGURE 131–11 Decision tree applied to the transition to adulthood in Austria (AUT) and in Italy (ITA). Numbers at the end of each branch correspond to the number of Austrians and Italians, respectively, affected by the rule corresponding to the branch. (From Billari *et al.*, 2000. Courtesy of Max Planck Institute for Demographic Research, 2000.)

group, and cohort to best separate the respondents from Austria and Italy, using a sequence of tests to build a decision tree.

This decision tree is presented in Figure 131–11. Each ellipse presents a comparison between the dates of two events or of an event and an age. Depending on whether the first event takes place before (<), after (>), or at the same time (=) the second event or whether no event already took place (n.o.), the decision is made to put the individual on one branch or the other of the decision tree. The first test (which comes first because it is the most efficient and therefore the better predictor) is based on the comparison between marriage and first union; if the dates of these two events are different, the individual is assigned to the left part (most probably leading to Austria). If, the end of studies took place after the age of 14, the decision is made that the individual is supposed to be Austrian. Two tests are sufficient to isolate a group of 2807 Austrians and only 253 Italians. If marriage and first union were simultaneous or did not occur (i.e., right branch of the tree), the next test is based on a comparison between union and the birth of the first child. The principal result of this method in that example is the demonstration of the importance of sequencing of transitions to adulthood in distinguishing the two countries: Austrians are differentiated from Italians by having different dates of first union and first marriage, children born before the first union, and simultaneous first employment and completion of schooling.

Compared with classic demographic methods, which are attached to the timing of one or another event, this type of method can simultaneously capture the ensemble of transitions to adulthood. This analysis shows that it is the existence or absence of intermediate steps that best defines the differences between young adults in these two countries. Italians seem to be characterized by the simultaneity of familial passages; marriage is the first union, and there are no births before that union. Conversely, among the Austrians, steps toward economic independence appear to take precedence: first job (often simultaneous with the end of schooling) and departure from the parental home.

CONCLUSIONS

Methods of exploratory data analysis have three principal goals. First, by refusing to pose *a priori* hypotheses and by using criteria based on all available information, they make it possible to quickly identify the global structure of information. Second, they reveal outlying points (which could result from errors in measurement, coding, or data entry); because they are iterative in nature, they are a way to prepare for more detailed analyses.

Thanks to the progress of statistical software and spreadsheets, it is possible to become familiar with a complicated data set relatively quickly. Methods of exploratory data analysis provide varied tools for synthetic description that are complementary to methods of demographic analysis.

Acknowledgment

This chapter was translated from the French version by Sarah R. Hayford.

References

Bertin Jacques, 1967. Sémiologie graphique. Paris and La Haye, Mouton, Gauthier-Villars and EHESS, 431 p. (Les Réimpressions, 1999. Course presented online at http://www.sciences-po.fr/cartographie/cartographie_html/5_page5theorie/graphique_bertin2001.)

Billari Franceso, Fürnkranz Johannes and Prskawetz Alexia, 2000. Timing, sequencing and quantum of life course events: a machine learning approach. Rostock. Rostock, Max Planck Institute for Demographic Research (MPIDR), 27 p. (Working Paper no. WP-2000-010.)

Breiman Leo, 2001. Statistical modelling: the two cultures, *Statistical Science*, vol. 16(3), p. 199–215.

Bry Xavier, 1995. *Analyses factorielles simples*. Paris, Economica, 112 p.

Bry Xavier, 1996. *Analyses factorielles multiples*. Paris, Economica, 112 p.

Caselli Graziella and Egidi Viviana, 1981. L'analyse des données multidimensionnelles dans l'étude des relations entre mortalité et variables socio-économiques d'environnement et de comportement individuel, *Genus*, vol. 37(3–4), p. 57–91.

Chambers John, Cleveland William, Kleiner Beat and Tukey Paul., 1983. Graphical methods for data analysis. Boston, Duxbury Press, 395 p.

Cleveland William S., 1993. Visualizing data. Murray Hill, NJ, AT&T Bell Laboratories and Hobart Press, 360 p.

Cleveland William S., 1994 (1985). The elements of graphing data. Monterey, CA, Wadsworth, 297 p.

Erickson Bonnie and Nosanchuk Terry, 1992 (1977). Understanding data: an introduction to exploratory and confirmatory data analysis for students in the social sciences. Milton Keynes, Open University Press, 388 p.

Hartwig Frederick and Dearing Brian, 1979. Exploratory data analysis. Beverly Hills, CA, Sage University Papers, 83 p. (Quantitative Applications in the Social Sciences, no. 16.)

Horber Eugen, 2004. Bibliographie succincte, partiellement commentée, sur l'analyse exploratoire des données. Geneva, Mouvement International pour le Développement de la Recherche en Analyze Graphique et Exploratoire (MIRAGE), 5 p. (http://www.unige.ch/ses/sococ/mirage/biblio.pdf.)

Lebart Ludovic, Morineau Alain and Piron Marie, 2000. Statistique exploratoire multidimensionnelle. Paris, Dunod, 439 p.

Ledermann Sully, 1969. Nouvelles tables-types de mortalité. Paris, INED and Presses Universitaires de France, 260 p. (Travaux et documents, cahier no. 53.)

Ledermann Sully and Breas Jean, 1959. Les dimensions de la mortalité, *Population*, vol. 14(4), p. 637–682.

Mauldin W. Parker, Berelson Bernard and Sykes Zenas, 1978. Conditions of fertility decline in developing countries, 1965–75. *Studies in Family Planning*, vol. 9(5), p. 89–147.

Meslé France and Vallin Jacques, 2002. Mortality in Europe: the divergence between East and West, *Population-E*, vol. 57(1), p. 157–198.

Mosteller Frederick and Tukey John, 1977. Data analysis and regression: a second course in statistics. Addison-Wesley, Pearson Education POD, 588 p.

Ní BhrolchÁIN Maíre and Toulemon Laurent, 1996. Exploratory analysis of demographic data using graphical methods. Paris, INED, 49 p. (Dossiers et recherches de l'INED, no. 56.)

NIST/SEMATECH, 2004. e-Handbook of statistical methods. Boulder, CO, National Institute of Standards and Technology, 789 p. (http://www.itl.nist.gov/div898/handbook, consulted in April 2004.)

Rallu Jean-Louis and Toulemon Laurent, 1993. Les mesures de la fécondité transversale. I. Constructions des différents indices. II. Application à la France de 1946 à 1989. *Population*, vol. 48(1, 2), p. 7–26, 369–404. (An abridged version in English was published in 1994: *Population: An English Selection*, vol. 6, p. 59–94.)

Tenenhaus Michel and Young Forrest W., 1985. An analysis and synthesis of multiple correspondance analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data, *Psychometrika*, vol. 50(1), p. 91–119.

Tufte Edward, 1990. Envisioning information. Cheshire, CT, Graphics Press, 126 p.

Tufte Edward, 2000 (1986). The visual display of quantitative information. Cheshire, CT, Graphics Press, 197 p.

Tukey John, 1962. The future of data analysis, *Annals of Mathematical Statistics*, vol. 33(1), p. 1–67.

Tukey John, 1977. Exploratory data analysis. London, Addison-Wesley, 499 p.

United Nations, 1955. Age and sex patterns of mortality. Model life-tables for underdeveloped countries. New York, United Nations/Nations Unies, 22, 38 p. (Population Studies, no. 22.)

Vaupel James W., Gambill Bradley A. and Yashin Anatoli I., 1985. Contour maps of population surfaces. Laxenburg, Austria, IIASA, 84 p. (Working Paper no. WP-85-47.)

Vaupel James W., Wang Zhenglian, Andreev Kiril F. and Yashin Anatoli I., 1997. Population data at a glance: shaded contour maps of demographic surfaces over age and time. Odense, Odense University Press, 98 p. (Odense Monographs on Population Aging, no. 4; http://www.demogr.mpg.de/Papers/Books/Monograph4/start.htm.)

C H A P T E R

132

Regression Models

LAURENT TOULEMON

Institut national d'études démographiques (INED), Paris, France

Regression techniques have grown in popularity and sophistication, and they have earned a rightful place in every demographer's toolbox (Hoem, 1991). However, despite attractive flexibility and power, they, like all methods of demographic analysis, demand care in use and application.

The chapter begins with a description of features common to all analytic models, followed by an account of regression specifics. I conclude with an overview of different types of regression.

I. MODEL CONSTRUCTION

Even basic parameter estimation, such as the calculation of crude death rates, depends on an underlying model (Keilman, 1993). Although frequently overlooked in practice, a careful understanding of the underlying model is essential for proper interpretation of results.

1. The General Model

We begin with a simple example. Suppose we observe the ratio $m = D/P$, in which D is the number of deaths in a given year, and P is the average annual population. Then, m is a rate, expressed in events per person per year. The simplest (and most naïve) interpretation of m depends on two assumptions:

1. *Homogeneity*: the rate m applies to all individuals in the population and does not vary among individuals with sex, or age, or other characteristics.
2. *Independence*: the risk of death of any individual is proportional to the duration of exposure (e.g., once having moved into the area, incoming migrants face the same risk per unit time as original inhabitants).

Homogeneity is unsupported by the evidence so any interpretation of crude death rates, whether in comparing two populations against each other or making comparisons within one population across time, will be limited by the extent that the homogeneity assumption is violated.

Independence between mortality and external perturbations such as migration is often easier to accept because the perturbations are usually relatively small compared with the size of overall mortality.

In terms of statistical inference, exploratory or predictive methods can be used to describe the data as accurately as possible or to forecast with the least possible error future or unknown results (Breiman, 2001) (see Chapter 131). To link these data to the real world, it is necessary to come back to the modalities governing the construction of the data and the transition from the real world to the data (which are constructed and not given) and then from the data to the real world (Cox, 2001).

The lesson is that results can be interpreted only within the framework of a general model. We will build that model in three steps.

Step 1. Construct a general model of the real world with measures or indicators that capture the behavior of interest (e.g., mortality, fertility) and the relationships between the measures.

Step 2. Within the framework of the model, calculate results. The formalization of the model in step 1 makes it possible to estimate model parameters, to perform tests of statistical significance, and to assess goodness-of-fit.

Step 3. Link back to the real world, using the model parameters obtained in step 2 to describe real world behavior. The more complex the model, the more difficult it may be to make that link, but it is still critical for simpler models.

For the previous example, we construct the death rate m, intending to measure the general level of mortality (step 1). We calculate the rate (step 2) and then examine the results and build conclusions on mortality (step 3). If we see a secular decrease in the value of our model parameter, m, we conclude that the general mortality level in our population also decreased over time.

In practice, our model is only a rough representation of reality. For example, although populations are never homogeneous, we still calculate crude death rates but warn the reader about their limitations. The more we want our models to reflect reality, the more sophisticated our models become; for example, we may construct more homogeneous groupings (such as by calculating age- and sex-specific death rates) and formalize the relationships between groupings. For this, regression models are the tools of choice.

Before describing regression models, let us examine two assumptions formalized in regression but also common in the estimation of demographic rates, even if in an implicit manner, and contrast their interpretations.

For assumption, the rates we observe are the consequence of forces that each individual may experience in a probabilistic way. We begin by assuming the opposite, for example, by supposing that the population can be disaggregated into two distinct groups: those who are fated to die within the year (probability of death = 1) and those who will survive (probability of death = 0). In this case, the calculation of the death rate is exactly the proportion of persons who will die, and it suffices simply to enumerate these two groups. The measured rate represents the proportion fated to die, not the overall death rate of the population. The comparison between two populations will then focus on the proportion fated to die in both populations, not on the overall mortality. This assumption is often unrecognized in demography, so much so that a formalization of it has been described as a "new paradigm" (Courgeau and Lelièvre, 1996). We will return to this point later.

For assumption 2, the rates we observe are the result of an underlying stochastic process. The time path of mortality across years is interpreted as being decomposable into an underlying secular trend and random fluctuations around it. In the long run, we seek to identify the random component (e.g., year-to-year variations in influenza) from the secular trend.

We often consider survey results as if they were coming from one sample among the universe of all possible samples. These super-population models let us describe annual variations as a realization of individual variations (as in assumption 1) and macro variations, some of which are themselves stochastic. This is the basis for testing whether a particular annual variation in rates is exceptional or not (Suchindran *et al.*, 1979; Udry *et al.*, 1979).

2. Regression Models

Regression models are characterized by a *dependent* (sometimes called *endogenous* or *response*) variable, usually denoted y, and one or more *independent* (sometimes called *exogenous* or *factor*) variables, X, related by the equation $y = f(X)$. In general, we seek to estimate f based on the observed data. Most often, we postulate that the function f applies at the individual level (i.e., for an individual i), and we estimate the parameters of a relationship $y(i) = f[X(i)]$.

Once X, y, and the general form of f have been constructed (step 1), regression consists of choosing a particular and specific form for f that satisfies some optimality constraint, usually resulting from trade-offs between simplicity of form and fit to the data.

a. Linear and Least Squares Regression

The simplest form of regression is linear regression: X is composed of one or more variables $x_1 \ldots x_p$, and we write the following:

$$y(i) = \sum_{j=1}^{p} a_j\, x_j(i) + \varepsilon(i) \quad (1)$$

We are usually not concerned with the particular values taken by each X variable, but rather the manner in which y varies across individuals i, with each individual being characterized by the values of the variables X. Assumptions about ε allow us to make estimates of the parameters a and therefore of the value of the function f that minimizes the values of ε. In least squares regression, we calculate the values of the parameters a that minimize the sums of squares (thus the term *least squares*) of the residuals, $\varepsilon(i)$.

The Gauss-Markov theorem states that if certain conditions on ε apply, least squares regressions have several highly desirable properties. The conditions are that ε must have a zero mean, and that its variance, σ^2, is the same for all observations and independent of the values taken by X; in that case, the estimate of a based on least squares criterion is unbiased and of minimum variance among the unbiased estimates.

The Laplace-Gauss theorem further states that if ε follows a normal (or gaussian) distribution, the distribution of the parameters a can be estimated and statistical tests of significance can be performed.

b. Log-linear Models and Likelihood

In demography, we often use log-linear models, which are based on a different formulation of f (Hoem, 1987). When y is a rate, the log-linear model presumes a linear relationship between the logarithm of y and the variables X:

$$\ln(y(i)) = \sum_{j=1}^{p} a_j \, x_j(i) \tag{2}$$

For any individual, we cannot observe the true underlying event rate, $y(i)$, but we see the outcome of the event. In this case, the least squares method does not apply, and we use the maximum likelihood criterion to find the values of the parameters a that maximize the probability of seeing the $y(i)$ that we do (see Chapter 8). We create a likelihood function that conditions the observed data on the parameters a and then find the values of a that maximize the likelihood.

In practice, it is often easier to work with the logarithm of the likelihood function than the likelihood function itself. Because the logarithm is a monotonic function, the values of a that maximize the likelihood will also maximize the log likelihood. If L is the likelihood and A is the universe of individuals at risk for the event, the log likelihood over all individuals is given by the following equation.

$$\ln(L) = c + \sum_{i \in A} y(i) + \sum_{i \notin A} (1 - y(i)) \tag{3}$$

We estimate the most likely value of a (i.e., the one that maximizes L), and then we get an estimate of the function f.

c. Comparisons of Models

After the general framework has been defined (step 1), we can use it to analyze and compare models of increasing complexity. More sophisticated models can fit the data better—but any improvement in fit (e.g., by decreased mean squared error or by increased likelihood) must be checked against simpler models to determine whether the improvement occurred by chance.

II. WHY REGRESSION?

The regression approach has two compelling advantages. First, explicit specification of an equation such as $y = f(X)$ allows for the identification of the separable contributions of each independent variable x to the variability in y. Second, regressions make it easier to analyze alternative model specifications.

1. Correlation between Independent Variables

Regression treats each of the independent variables in an identical way. In practice, we often distinguish between one of the X variables as the variable of interest (e.g., x_1) for its effect on y, and one or more other X variables as confounders (e.g., x_2), whose effect we wish to control.

In that case, we could compare a model including only x_1 [$y = f(x_1)$] with one including both X variables (i.e., $y = f[x_1, x_2]$).

The first model lets us see how y varies with x_1 in a crude or uncontrolled way, whereas the second lets us see how y varies when x_1 changes and x_2 is held constant. This is a type of standardization. We examine the effect of x_1 on y, with this effect assumed to be constant whatever the level of x_2 may be. These adjusted variations differ from crude adjustments if two conditions hold: x_2 has an effect on y once adjusted for x_1, and x_1 and x_2 are not statistically independent (i.e., the distribution of x_1 differs over groups selected by their value of x_2). From this perspective, regression is a form of standardization midway between direct and indirect standardization (see Chapter 15). It offers two advantages. First, it is possible to standardize across many confounding variables. Second, the model lets us test the hypothesis of independence. This lets us think of regression as an enhanced form of indirect standardization (Hoem, 1991).

2. Interaction Effects

Sometimes, the way that y varies with x_1 can differ depending on the value of another variable x_2. In that situation, there is an interaction between x_1 and x_2 in their effect on y. The interaction between the x_i terms can be weak in the sense that the effect of x_1 on y can vary with x_2 but without changing the sign of the x_1 effect, or it can be strong if the sign of the x_1 effect changes (see Chapter 15). When the interaction is weak, sometimes it is possible to re-express the x_i terms in a way to minimize the interaction, but when the interaction is strong, in a strict sense, it may not be possible to speak of "the" effect of x_1 on y, because the effect is sometimes positive and sometimes negative, and it cannot be summarized by a single value.

Regression allows us to compare the fit of models with and without interactions. They provide one (but not the only) criterion for evaluating the importance of including interactions in a model.

3. Choosing a Model

Even without considering interactions, we can combine p potential independent variables into 2^p possible models. Systematic comparison and evaluation of all possible models is usually impossible. Some statistical software applications include procedures that can automatically select independent variables that meet certain statistical properties, such as stepwise selection. Forward stepwise selection first chooses the single independent variable whose inclusion improves model fit the most; subsequent independent variables are added to improve overall fit, building on the previously selected variables. Backward stepwise selection proceeds in the opposite direction by sequentially dropping the variable that adds the least to overall fit. Some procedures use a mixed strategy of both forward and backward steps, and they may allow the specific inclusion of certain key independent variables.

As in any standardization, the choice of the model can have a critical influence on any result. The choice of an *optimal* model depends on two disparate criteria: one related to step 2 and the other related to steps 1 and 3.

a. Statistical Criteria: Step 2

In a strict sense, it is only possible to compare two models that are nested: one model with a number of independent variables, X_1, and the second comprising all of the X_1 variables plus additional variables X_2. The null hypothesis is that the additional X_2 variables do not add any explanatory power. In practice, even when the null hypothesis is true, additional explanatory variables can improve model fit through randomness alone. For a test of the null hypothesis, we choose a critical value α, often 5%, which represents the chance we are willing to take that, through randomness alone, we could see an improvement in model fit even though the X_2 had no real contribution, leading us to conclude erroneously that X_2 had a real effect on y. After a critical value has been chosen, we can calculate the level of improvement in model fit needed to meet the threshold. If including the X_2 produces a larger improvement, we say that the effect of X_2 on y is significant. More precisely, this means that if the observed improvement in fit would have been observed with probability less than α under the null hypothesis, we reject the null hypothesis and accept the alternative that the X_2 do affect y.

Why is 5% the critical value? In general, the choice is a compromise. We want a high enough critical value so that we will know if the X_2 variables contribute to the model, but we also want a low enough value to avoid being fooled by random chance that they contribute when they do not. To ensure that we are not fooled, we would choose a low numerical critical value that is difficult to meet, such as 1 in 1000. In that case, if the X_2 variables did not truly contribute to the model we would run only a 1 in 1000 risk of being fooled into thinking that they did; the test of the hypothesis would reveal only highly significant relationships.

However, with such a low critical value, not being able to reject the null hypothesis does not mean that the null hypothesis is true (i.e., that the X_2 terms have no role in the model). In choosing a threshold that is so difficult to meet, we run the risk of overlooking important effects. The power of a test is its ability to ferret out true relationships and to reject the null hypothesis when it is false. In most cases, the power of the test cannot be estimated. However, it increases with the level α; with increasing the probability to reject erroneously the null hypothesis if it is true (i.e., no real effect of X_2 on y), we also increase the probability to correctly reveal that the null hypothesis is false (i.e., the variables X_2 have an effect on y).

Raising the level α increases, by definition, the risk of erroneously believing a relationship to be significant. This is called type I error, and in general, the scientific community is very concerned about this type of mistake. Peer reviewers of articles submitted to journals usually set by default a maximum critical value of 5% (i.e., when there is no effect, they are willing to tolerate a risk of 5% that you would erroneously conclude that there was, which means that one erroneous result will be given for each 19 cases without a conclusive result).

Type II error, the risk of erroneously overlooking a significant result when it truly exists, is considered a less damaging error. The inability to reject the null hypothesis is viewed as leaving the question inconclusive and in doubt, rather than proof of no relationship. It is the researcher's responsibility to reduce the risk of type II error as much as possible by ensuring that his or her research is properly performed, with sufficient sample sizes and correct methods.

b. Balancing Theory and Practice

The power of a test depends on effect size and sample size. Given a particular set of data, there are two ways to improve the power of a test to identify significant variables: one good; the other, bad. The proper way is to clarify as precisely as possible the alternative hypothesis. For example, if the analyst suspects that females experience lower death rates than males, but the test has insufficient power to spot this effect, he or she can construct the null hypothesis "female mortality is the same or higher than male mortality." Rejecting this null hypothesis leads to the conclusion that "female mortality is less than male mortality." We then perform the hypothesis test with a 10% critical value, and we reject the null hypothesis only if there is a difference in the supposed direction: females, rather than males, experience lower mortality. Under the null hypothesis, the critical value is one-half of 10%, or 5%. The trick consists of trading off the ability to spot higher female mortality for more power in detecting lower female mortality. If the *a priori* assumption about the relationship between female and male mortality is wrong, we are worse off; we are pointing more precisely but in the wrong place. We need as much knowledge as possible about the subject to produce the best testable hypotheses.

The wrong way to improve test power is to specify the null and alternative hypotheses *after* having examined the data. One variant of this approach is to perform many analyses and present only those with significant results. For example, with a 5% critical value and 20 different tests, even when the null hypothesis is true, we will erroneously find, on average, one positive result, and there is more than an even chance of finding at least one (falsely) positive result: 64% if the tests are independent and less if they are linked. If we present only significant results at a nominal 5% critical level, we will overstate the findings because the real critical value for the multiple comparison problem is much higher than 5%.

This is a very common type of error. In France, a physician reported that the number of births varied with the phases of the moon after having performed 30 separate tests (Toulemon, 1986). For another example, numerous researchers have reported a link between breast cancer and the contraceptive pill, including one meta-analysis (CGHFBC, 1996) that reviewed the studies available at the time. One-half of the studies concluded that no significant link was found, and the other half concluded just the opposite. However, among the studies that found a link, the groups at risk were never exactly the same; some studies found a higher risk for women who had never borne children, some found it for female smokers, and others found it for women who had started using the pill at an early age. Even if the null hypothesis of no statistical relationship were true, if each study had performed 20 tests to publish one positive result, it would be entirely consistent to see one-half of the studies (rather than 5%) finding an erroneous link between breast cancer and use of the pill. Moreover, the studies finding a link are published more often and faster than those that reported no link. Although there is still doubt about an overall link, one study appears to support the conclusion that the link reported by these studies are consistent with random chance and that there may not be a significant link between breast cancer and use of the pill (Lê, 2002; Marchbanks *et al*., 2002).

c. Nonstatistical Criteria: Steps 1 and 3

Model specification does not rely solely on statistical criteria, nor is the choice of the model specific to regression problems. Three examples are discussed in this section.

In example 1, interactions are ignored. In calculating the total fertility rate (TFR), each age gets an equal weight, and in essence, it is simply the sum of the age-specific rates. The TFR, reported in units of children per woman, is based on the artifice of an artificial (or synthetic) cohort (see Chapter 14). To analyze changes in fertility over time, calculation of the TFR represents a mean rate, or a regression in which fertility is a function of age and time. However, in most developed countries since the 1980s, age-specific fertility rates have decreased at the youngest ages and increased at older ages. There is a strong interaction between age and year (see Chapter 15), and it is therefore not possible to speak strictly about a rise or fall in fertility. It is only by the convention of an artificial cohort that we can examine the change in the TFR over time. We can see evidence of the interaction by examining how the mean age of childbearing changes over time.

In example 2, significant explanatory variables are omitted or overlooked. We know that the age of marriage strongly influences the risk of marital dissolution. To measure its change over time, we could describe a model in which the risk of divorce depends on the age at marriage and the year of marriage. One implicit model is that it is the age at marriage that specifically affects the risk of divorce. An increase in the mean age at marriage is accompanied by a decrease in the proportion of marriages at very young ages and a decrease in the overall divorce rate. An alternative model is that the age at marriage is an indicator of marriage at young ages, but the proportion of early marriages is fixed although the mean age at marriage rises. In the latter model, the significance of age at marriage changes with the year of marriage. A marriage at age

25 in the year 1980 occurs at the mean age at marriage, whereas a marriage at age 25 in the year 2000 is among the youngest quintile. Accordingly, a regression that examines changes in divorce over time may omit age of marriage because the same age of marriage at two different times has a different meaning; comparable ages are incomparable. Instead, we could introduce a variable that measures precocity of marriage, constructed to be statistically independent from the year of marriage and which splits each year's marriages into earlier or later groups (Hoem, 1996). When the meaning of a value of an explanatory variable changes according to the value of the variable of interest, it is often better to omit this explanatory variable or to construct a new indicator variable that is statistically independent of the variable of interest.

In example 3, several alternative strategies are possible. In studying marital dissolution, should we standardize for prenuptial cohabitation, the number of children, or premarital births (Toulemon, 1995)? The answer is not straightforward, and each model focuses on a specific question for which it was constructed (step 1). None of these models may be considered as the best one, giving the only answer. The usefulness and validity of the conclusions (step 3) depend more on the general model (step 1) than on the specific estimation (step 2).

4. Regressions and Demographic Analysis

a. Sex, Age, and Period

Demographers focus on three particular variables: sex, age, and period. Each of these must be carefully handled.

Male-female differences rarely can be reduced to a simple difference in levels, and in general, it is preferable to estimate one regression for each sex, which is equivalent to taking into account all possible interactions between sex and other independent variables.

In the previous section, we saw that age occasionally is better interpreted in a relative rather than absolute sense. In general, when studying nuptiality or migration, a good model systematically tries to eliminate age effects so that when all relevant variables are taken into account, the remaining effect of age on demographic behavior disappears. This happens, for example, in migration studies in which age effects and residence duration effects diminish when family status (e.g., marital status, number of children), housing status (e.g., owner, renter), and profession are taken into account. Age in and of itself is not a variable that directly affects mobility, but life cycle events (e.g., marriage, divorce, children leaving home, buying or selling a house) that are associated with age (Courgeau, 1985) are the variables affecting migration. When studying mortality, we could equally imagine that specific biologic indicators of aging or of health could substitute for age in determining risk of death. Nonetheless, we are not yet there, and age continues to be a key variable in describing demographic behavior. Calculating rates by age can be easy and informative, but we must keep in mind that systematic use of age as an explanatory variable can be abused.

After age and sex, period also holds a special place. When an individual's characteristics change, there is, by definition, a statistical dependence of that variable with time. To measure the effect on behavior of that variable, we can use event histories, in which time plays a peculiar role. The unit of analysis is no longer the individual but his or her complete biography, described as the succession of durations at which we observe the occurrence of events (see Chapter 23).

b. Regression Complements Demographic Analysis

A great advantage of regression is that it allows us to be more specific and precise in our analysis and then use the results in a conventional demographic context. For example, age- and parity-specific fertility rates calculated by regression methods show that, between 1975 and 1989 in France, the period fertility rate fell for women younger than 28 years of age having their first birth and rose above that age, whereas the rates for women having a second birth rose over all age groups. However, these changes in the period rates can be described in terms of quantum and tempo. The mean age at birth of first child rose, and parity progression ratios from zero to one child and from one to two children remained stable. Having their first child at a later age, women have less time to have a second child. The period fertility instant rate rose, but the length of duration at risk decreased, and the total fertility rate was constant (Rallu and Toulemon, 1993). This insight would not have been available if we were limited only to regression analysis.

III. TYPES OF REGRESSION

All forms of regression are based on the same general principles and seek a form that minimizes error (when we can directly measure the dependent variable) or maximizes the likelihood (when we can only observe the realization of an event). The choice between the two often depends on the characteristics of the dependent and independent variables.

1. Dependent Variables

Dependent variables can take several forms: quantitative variables, proportions, or rates. Characteristics of each of these types determine the type of regression used (Leridon and Toulemon, 1997).

Linear regression is designed for models characterized by quantitative dependent variables that can be measured along an additive scale, such as the number of children or the amount of income. We then examine how a change in one of the independent variables, Δx, translates into a change in the dependent variable, Δy, through some constant ratio a (i.e., $\Delta y = a\ \Delta x$). For several X variables, we often write an equation similar to the following:

$$y = \sum_j a_j\ x_j \tag{4}$$

Additive dependent variables can always be transformed if needed to make interpretation easier. A common and useful transformation is the logarithmic transform, which makes proportional comparisons easy because it transforms differences into ratios. Then, a change in one of the independent variables, Δx, is related to a change in the dependent variable, Δy, by $\exp(a\ \Delta x)$, and the change is proportional to the value of y: $\Delta y = y\ a\ \Delta x$. For several independent variables, we often write an equation of the form:

$$\ln(y) = \sum_j a_j\ x_j \tag{5}$$

or

$$y = \exp\left(\sum_j a_j\ x_j\right) \tag{5′}$$

and

$$\frac{\partial y}{\partial x_i} = a_i \exp\left(\sum_j a_j\ x_j\right) = a_i\ y \tag{5″}$$

These rates are most often interpreted on a multiplicative scale and are useful for studying relative risks or relative growth rates, such as in studies of income.

Probabilities are well modeled with a logit, or log-odds, transform. This transformation corresponds to the hypothesis that Δx is related to Δy in a way that is proportional both to y and to $(1 - y)$: $\Delta y = y(1 - y)\ a\ \Delta x$.

If y is the probability that an event will occur, then $y/(1 - y)$ is the odds of the event, and we can model the logit in the now familiar regression context:

$$\text{logit}(y) = \ln\left(\frac{y}{1-y}\right) = \sum_j a_j\ x_j \tag{6}$$

or

$$y = \frac{1}{1+\exp\left(-\sum_j a_j x_j\right)} \tag{6′}$$

and

$$\frac{\partial y}{\partial x_i} = \frac{a_i \exp\left(-\sum_j a_j\ x_j\right)}{\left(1+\exp\left(-\sum_j a_j x_j\right)\right)^2} = a_j\ y(1-y) \tag{6″}$$

2. Independent Variables

Explanatory variables can be quantitative or categorical; in the latter case, we often transform variables that take p categorical values into p dichotomous dummy variables; we create indicator variables x_p where $x_p = 1$ if $x = p$ and $x_p = 0$ if $x \neq p$.

3. More Complex Models

Regression methods have grown in number and sophistication as models have become more complex. Among the refinements most often used in econometrics are seemingly unrelated regression (SUR) systems of equations, which exploit correlations between different dependent variables measured over the same individuals; models with endogeneity, which allow for multiway dependencies among the independent variables; multilevel models, which allow for the unified handling of individual and group-level variables (see Chapters 24 and 133); and log-linear models, which build on structural effects and groups of variables without making the distinction between dependent and independent variables (see Chapter 92). The description of these models is beyond the scope of this chapter.

CONCLUSIONS

Regression methods have become indispensable in demography. They have allowed for the formalization of previously implicit hypotheses in the estimation of demographic rates (Courgeau and Lelièvre, 1996), and they allow us to generalize from the usual standardizations.

They also encourage the investigation and evaluation of many alternative model specifications by comparing their statistical significance. On a spectrum from crude rate models in which each independent variable's effect on the dependent variable is measured in isolation, "all other things kept unequal" (Héran, 1996), to full models in which the effect of each variable is measured, "all other things being equal," regressions allow us to measure and to test different intermediate formulations.

Regressions, like the other methods of analysis, can be understood only as being the middle step of a three-step process. Upstream from the regression, step 1 consists of defining the general model that links the data collected and the relationships between data to the real world. Downstream, conclusions are drawn from the results of the regression by linking back from the model to the real world.

Regression and classic demographic analysis complement each other. Some insights can be derived only by building on the structure of classic demographic behavioral relationships, but at the same time, regression analyses have quickly become essential in summarizing the diversity of behavior by using independent variables other than age and sex.

Acknowledgment

This chapter was translated from the French version by Robert Chung.

References

Breiman Leo, 2001. Statistical modeling: the two cultures, *Statistical Science*, vol. 16(3), p. 199–215.

CGHFBC (Collaborative Group on Hormonal Factors in Breast Cancer), 1996. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53297 women with breast cancer and 100239 women without breast cancer from 54 epidemiological studies, *The Lancet*, vol. 347(9017), p. 1713–1727.

Courgeau Daniel, 1985. Interaction between spatial mobility, family and career life-cycle: a French survey, *European Sociological Review*, vol. 1(2), p. 139–162.

Courgeau Daniel and Lelièvre Éva, 1996. Changement de paradigme en démographie, *Population*, vol. 51(3), p. 645–654.

Cox David, 2001. Comment [on Breiman 2001], *Statistical Science*, vol. 16(3), p. 216–218.

Héran François, 1996. L'école, les jeunes et les parents. Approche à partir de l'enquête Éducation, *Économie et Statistique*, vol. 293, p. 5–17.

Hoem Britta, 1996. The social meaning of the age at second birth for third-birth fertility: a methodological note on the need to sometimes respecify an intermediate variable, *in*: Population Research Institute (ed.). *Yearbook of population research in Finland*, vol. 33, p. 333–339. Helsinki, Population Research Institute (PRI), 350 p.

Hoem Jan M., 1987. Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review, *International Statistical Review*, vol. 55(2), p. 119–152.

Hoem Jan M., 1991. La standardisation indirecte améliorée et son application à la divortialité en Suède (1971–1989), *Population*, vol. 46(6), p. 1551–1568.

Keilman Nico, 1993. Du fondement et de l'usage des taux, *in*: Les mesures de la fécondité transversale. Réflexions autour d'un article, *Population*, vol. 48(2), p. 423–427.

Lê Monique, 2002. Analyze commentée: contraception orale et risque de cancer du sein, *Revue d'épidémiologie et de santé publique*, vol. 50(5), p. 501–503.

Leridon Henri and Toulemon Laurent, 1997. *Démographie. Approche statistique et dynamique des populations.* Paris, Economica, 440 p. (Économie et statistiques avancées, chapitre 14.)

Marchbanks Polly A., McDonald Jill A., Wilson Hoyt G., Folger Suzanne G., Mandel Michele G., Daling Janet R., Bernstein Leslie, Malone Kathleen E., Ursin Giske, Strom Brian L., Norman Sandra A., and Weiss Linda K., 2002. Oral contraceptives and the risk of breast cancer, *New England Journal of Medicine*, vol. 346(26), p. 2025–2032.

Rallu Jean-Louis and Toulemon Laurent, 1993. Les mesures de la fécondité transversale. III. Réponses aux commentaires, *Population*, vol. 48(2), p. 434–442.

Rallu Jean-Louis and Toulemon Laurent, 1994. Period fertility measures. The authors' reply, *Population: An English Selection*, vol. 6, p. 123–130.

Suchindran Chirayath M., Clague Alice S., and Ridley Jeanne, 1979. Estimation of stochastic variation in vital rates: a simulation approach, *Population Studies*, vol. 33(3), p. 549–564.

Toulemon Laurent, 1986. Nouvelles données sur les variations du nombre des naissances selon les rythmes lunaires et circadiens, *Population*, vol. 41(4–5), p. 847–853.

Toulemon Laurent, 1995. La standardisation des quotients instantanés, *Revue européenne de démographie*, vol. 11(4), p. 333–342.

Udry J. Richard, Teddlie Charles, and Suchindran Chirayath M., 1979. The random variation in vital rates based on total enumeration of events, *Population Studies*, vol. 33(2), p. 353–364.

C H A P T E R

133

Bridging the Gap between Micro-demography and Macro-demography

FRANCESCO C. BILLARI

Istituto di Metodi Quantitativi, Università Bocconi, and Innocenzo Gasparini Institute for Economic Research, Milan, Italy

I. MACRO 1–MICRO 1: WHY DO DEMOGRAPHERS NEED TO BRIDGE THE GAP?

Demography can be defined as the study of human populations (Caselli *et al*., General Introduction). Populations are by definition aggregates of individuals, and they constitute entities at the aggregate or *macro* level of analysis. In the past, demography could be said to be a *macro-demography*, because individual lives entered in formal analysis when contributing to numbers of events, persons, years, and survivors—also the statistical analyses used to be located at the macro-level. During the last decades of the 20th century, a *micro-demography* emerged with a specific emphasis on the unfolding of individual-level demographic trajectories and on the consequences of individual heterogeneity for the study of population dynamics. The gap between macro-demography and micro-demography has since then undoubtedly widened, and this is rightly seen as a problem. This chapter documents the emergence of this gap and the answers to the challenges posed by macro-micro-micro links in other disciplines. It also focuses on recent proposals to bridge this gap and avoid a conflict between macro-demography and micro-demography.

Some hints about the reasons micro-demography has emerged can be found by having a look at disciplines that are strictly linked to demography and that have attempted micro-founding the study of specific types of behavior using some type of "methodologic individualism" approach. We look at three disciplines in particular: ecology, sociology, and economics. In *ecology*, what is defined as *individual-based modeling* (IBM) for the study of animal and plant populations has emerged starting in the mid-1970s as a research program that has lead to significant contributions. In a review, Grimm (1999) discussed 50 IBMs of animal populations published mostly during the 1990s. According to Grimm and Railsback (2005), individual-based models in ecology fulfill four criteria. First, they explicitly consider individual-level development. Second, they represent explicitly the dynamics of the resources an individual has access to. Third, individuals are treated as discrete entities, and models are built using the mathematics of discrete events rather than rates. Fourth, they consider variation between individuals of the same age. Individual-based models in ecology are aimed at producing "patterns" that can be compared with patterns observed in reality (Grimm and Railsback, 2005).

In *sociology*, the approach proposed by James Coleman (1990) proposes to found social theory ultimately on the micro-level decisions of individuals. Coleman suggests a three-part schema for explaining macro-level phenomena, consisting of three types of relations: (1) the macro-to-micro transition, which is how the macro-level situation affects individuals; (2) purposive action of individuals, which is how individual choices are affected by micro-level factors; and (3) the micro-to-macro transition, which is, how macro-level phenomena emerge from micro-level

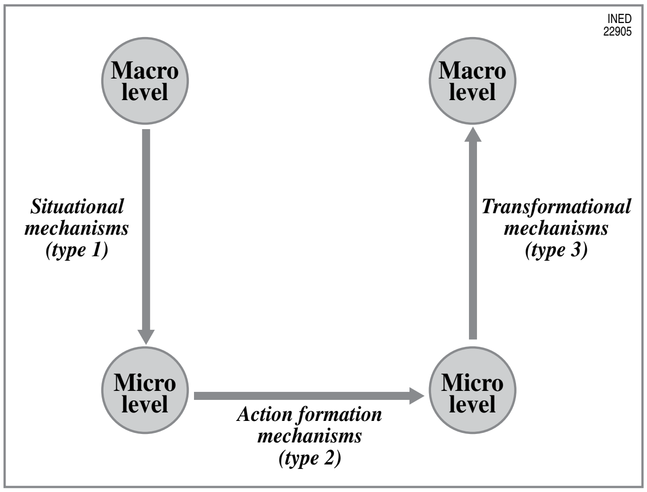


FIGURE 133–1 Coleman's conceptual framework in terms of social mechanisms. (Adapted from Hedström and Swedberg, 1999.)

action and interaction. Figure 133–1 represents Coleman's conceptual framework. It is embedded in the notion of "social mechanism" as the key concept to explain behavior in the social sciences, a concept proposed by Hedström and Swedberg (1999), who see the three types of relationships as *situational mechanisms*, representing the case in which "the individual actor is exposed to a specific social situation, and this situation will affect him or her in a particular way"; *action formation mechanisms*, representing "a specific combination of individual desires, beliefs, and action opportunities (that) generate a specific action"; and *transformational mechanisms*, specifying "how these individual actions are transformed into some kind of collective outcome, be it intended or unintended." The framework is very similar to the one presented recently by Daniel Courgeau (2003) in a review on the macro-micro link.

The micro level is now the natural point of departure in *economics* and when pointing to the macro level as the important outcome. As observed by van den Bergh and Gowdy (2003, p. 65) "During the last quarter century, the microfoundations approach to macroeconomic theory has become dominant." Mainstream economics, also known as "neoclassic" economics traditionally considers a "representative agent" that maximizes a potentially complex utility function subject to potentially complex budget constraints. This and other hypotheses lead to mathematically tractable models of macro-level outcomes. The new home economics approach that exactly applies the toolkit of neoclassic economics to demographic choices has been a key to the success of the work of Gary Becker (1980). This approach has reached a level of maturity that can be attested to by looking at recent textbooks on the subject (Ermisch, 2003) and at the importance of the literature on population economics (Zimmermann and Vogler, 2003). That we ought to start from the micro is also clearly stated by an economist who is particularly interested in population matters, Jere Behrman, who states, "For both good conditional predictions and good policy formation regarding most dimensions of population change and economic development, a perspective firmly grounded in understanding the micro determinants—at the level of individuals, households, farms, firms, and public sector providers of goods and services—of population changes and of the interactions between population and development is essential" (Behrman, 2001, p. 317–410). The attention on the policy relevance of research on population (including policy implications of results) is undoubtedly the main characteristic that comes to the surface when looking at research on population economics. Micro-based theories of behavior are used to cast "conditional prediction" of reactions to a given policy, with these reactions affecting macro-level outcomes. Within economics, several scholars have objected to the neoclassic paradigm from various perspectives (Behrman, 2001, offers objections to critiques concerning population-development relationships). Of particular interest for the reasoning of this paper are the critiques on mainstream economics that concern the assumption that agents are homogeneous and the lack of explicit interaction between agents (Kirman, 1992). Kirman's point is that even if individuals are all utility maximizers (an idea that has also been criticized by several scholars), the assumption that the behavior of a group of heterogeneous and interacting agents can be mimicked by that of a single representative individual whose choices coincide with the aggregate choices of the group is unjustified and leads to misleading and often wrong conclusions. To overcome this micro-macro "aggregation" problem, which is the transformational mechanism in Coleman's scheme, some economists have proposed to build models that resemble that of IBM in ecology. Models in agent-based computational economics explicitly allow the interaction between heterogeneous agents (Tesfatsion, 2001).

Let us now consider the approach of "micro-demography." The success of "micro-demography" is strictly linked to the emergence of the *life course approach* as an interdisciplinary program of study in the social and behavioral sciences (Dykstra and van Wissen, 1999) and to the development of appropriate statistical tools and data collection efforts (see Chapter 23). The life course approach has been under development since the mid-1970s (Mayer, 2000; Mayer and

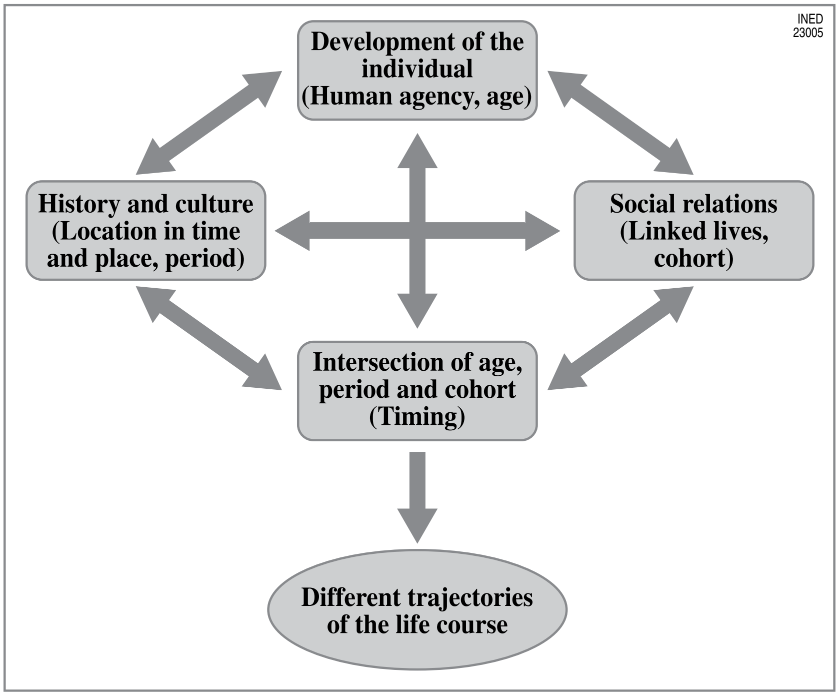


FIGURE 133–2 The conceptual framework of the life-course approach. (Adapted from Giele and Elder, 1998.)

Tuma, 1990). In their end-of-millennium review, Giele and Elder (1998) identify four chief elements as fundamentally shaping individual biographies; the first three are individual development (i.e., human agency), history and culture (i.e., location in time and place), and social relations (i.e., linked lives). In terms of the classification of social mechanisms by Hedström and Swedberg, location in time and place and linked lives are potential factors in situational mechanisms, whereas human agency is a potential factor in action formation mechanism. The intersection of such elements constitutes the fourth chief element: the timing of lives. Figure 133–2 shows the framework adopted by the life course approach. A set of interconnected *trajectories* lies at the heart of the life course analysis; trajectories are themselves shaped by *events*. The key elements outlined by Giele and Elder are naturally linked to notions used to define the coordinates of populations in demography. The human agency concept lies behind the use of *age* as a privileged time axis. Location in time and, to an extent, the idea of linked lives suggest *cohorts* or groups of cohorts as basic descriptive units for comparison. History and culture emphasize the importance of *period* and *location* in space. Because events lie at the very heart of the tradition of demography, it is not a surprise that the life course approach has been particularly influential in demographic research from a substantive point of view (Dykstra and van Wissen, 1999) and from a methodologic point of view, with a field of *life course analysis* emerging as an important part of research in population (Billari, 2003). Different from the conceptual frameworks developed in rational choice sociology or in economics, the life course approach does not explicitly provide a model for the micro-to-macro link (i.e., a scheme including a transformational mechanism). For this reason, the diffusion of the life course approach has so far probably contributed to widen the gap between scholars focusing on macro-level issues and those focusing on micro-level issues.

An important discussion of the tensions between what we may call macro-demography and micro-demography is made by Ronald Lee (2001). Lee says, "Demography is changing in profound ways, in its methods and topics of research, and in its training. There is less aggregate level (macro) analysis and more individual level (micro) analysis. There is less emphasis on process and dynamics, and more emphasis on individual decisions about demographic behavior. There is less formal demography, and more data analysis. There is much less funding of aggregate demography and formal demography, and much more funding of micro level empirical studies. Formal demography is not the same thing as aggregate or macro demography. There is also micro level formal demography, and much formal demography integrates micro and macro approaches. But for the most part, when I refer to formal demography, I will have in mind either macro level or integrative formal demography" (Lee, 2001, p. 1). Lee connects the rise of micro-demography to some factors, of which some are important for our argument. First is the increasing availability of survey data, contrary to the situation in the past, when data were mostly arising from aggregate-level population statistics. Second is the importance of the links with economics, particularly the development of the New Home Economics approach centered on the work of Gary Becker. Third, the links with other disciplines have pushed demographers to try to communicate as much as possible with colleagues from neighboring disciplines, and as we have seen before, micro-foundations have become important in neighboring disciplines. Lee concludes that demography is abandoning its core, mostly macro and mostly formal, and that this is a danger, because "key issues are macro demographic" (Lee, 2001, p. 3).

In the remainder of this chapter, I argue that there should be no "disciplinary soccer game"[1] between macro-demography and macro-demography. Micro-demography could and should be directed to the study and explanation of the key macro demographic issues mentioned by Ron Lee and should not limit itself to explaining how life courses unfold. I first review the way the macro-micro transition is dealt with (an issue that is discussed in depth in other chapters of this book). The next section discusses in particular the links between formal demography and statistical demography as related to the gap between micro-demography and macro-demography. I then present some challenges that are not solved if an explicit model of the macro-micro-macro link is unavailable. The next section of this chapter discusses the importance of looking at both levels when analyzing empirically demographic behavior. Choosing to adopt a Coleman-type of framework to bridge the gap and in parallel to approaches such as individual-based modeling in ecology, I propose simulation as a way to perform the micro-macro transition. A later section of this chapter discusses the potential role of computer simulation, particularly agent-based computational demography to bridge the gap between micro and macro, and the test ideas on transformational mechanisms. In the last section, I discuss research directions.

[1] Here I borrow and change a quotation on "interdisciplinary soccer games" from Lesthaeghe (1998), who starts from a paper by Hammel with "Economics 1: Culture 0" as the first words in the title.

II. FORMAL DEMOGRAPHY AND STATISTICAL DEMOGRAPHY: THE PARADIGM SHIFT

Formal demography has traditionally developed the advanced means of dealing with macro-level data that are presented in the first volume of this book, and these methods are proudly exhibited as flagship achievements of demography by Lee. Some specific examples are important. The approach developed for the analysis of demographic rates was specifically aimed at the analysis of homogeneous groups in a population (Henry, 1972), although Louis Henry took heterogeneity into account in his analysis of the components of fertility (see Chapter 34). Mathematical models of population dynamics that have been used in demography and in ecology have treated traditionally populations in an aggregate fashion and used differential calculus to derive results that are beautiful and useful at the same time (see Chapter 20). Age, period, and cohort models have been developed to analyze macro-level data with the aim of decomposing the effect of those three main factors (see Chapter 18).

In the first volume (fifth part) of this treatise we find a discussion of new pathways for formal demography. These pathways have been important in shaping the shift from macro-demography to micro-demography. First of all, the importance of recognizing that demographic "agents" are heterogeneous has brought important attention to the possible fallacies we incur in as we use mean rates as measures of the micro-level hazard of experiencing a demographic event. The importance of heterogeneity beyond traditional variables such as age, sex, cohort, and period was recognized by actuaries as early as the 1930s (Perks, 1932). The work of Vaupel and his colleagues (see Chapter 21) has shown that ignoring changes in the composition of a cohort leads to wrong inference about the dependence on age of the force of mortality. Since then, the importance of micro-level variation in frailty (also known as unobserved heterogeneity) has become primary in the interpretation of changes in demographic rates; this has been a key issue since the late 1970s for the statistical analysis of micro-demographic data. A similar movement toward including the possibility that individuals are heterogeneous has existed in ecology and economics.

The success of micro-demography is even more linked to what we could define as the new statistical approach to demographic data: "event history analysis." Although event history analysis clearly shares common origins with formal demography and with life tables (e.g., the title of the widely cited paper by

David Cox [1972] "Regression models and life-tables"), its development in the 1970s has been very much independent of progress in formal demography, with some exceptions. Early links between event history analysis and formal demography from the probabilistic and statistical point of view have been present in the work of Jan Hoem (Hoem, 1991; Borgan and Hoem, 1988; Hoem, 1976, 1987). The general idea of Hoem's approach was to show that macro-level quantities actually used in formal demography could be embedded in a probabilistic setting and be naturally linked to regression models. These regression models could be linked to indirect standardization techniques, focusing on counts, or to micro-level event history models. In their 1989 manual, *Event History Analysis in Demography*, Daniel Courgeau and Éva Lelièvre linked event history analysis to formal demography, explicitly saying that the move from longitudinal analysis in formal demography to event history analysis marked a paradigm shift in demographic research (Courgeau and Lelièvre, 1996). Since then, event history analysis had an important impact on demographic research, an impact that is likely to have triggered reactions such as the one documented earlier by Ron Lee, given a possibly excessive focus on the micro side (see Chapter 23).

The diffusion of event history analysis as the main technique for micro-demography has taken advantage of the theoretical basis of the life course approach, including the explicit reference to age, period, and cohort as key dimensions of that approach (see Fig. 133–2). The link was evident especially among scholars on the border between sociology and demography. Mayer and Huinink (1990), for instance, discussed the use of event history analysis to estimate age, period, and cohort effects—seeing event history analysis as a "farewell to Lexis," implying that in the era of micro-level data, age-period-cohort analysis would have overcome its classic problems deriving from the use of macro-level data (see Chapter 18). Event history analysis could be a way to fit micro-level age-period-cohort models when having access to detailed micro-level data, as shown in the application by Billari and Rosina (1998) to mortality in pre-transitional settings.

Fortunately, although the focus of studies using event history analysis has mainly been on the micro-level, on what in terms of Figure 133–1 we would call "action-formation mechanisms," some later studies have included macro-level determinants, seeing the timing of events as affected by "situational mechanisms" in a multilevel settings. Earlier papers have included macro-level covariates in micro-level event history models. Blossfeld and Huinink (1991) included a measure of the business cycle in an event history model of the timing of first marriage and of first birth. Later, event history models were generalized for the inclusion of unobserved heterogeneity at the aggregate level, creating fully multilevel event history models (see Chapter 24). These models have been systematized in their discrete-time version (Barber *et al.*, 2000) and in their continuous-time version, with a generalization to multilevel and multiprocess event history models (Lillard and Panis, 2004).

The development of multilevel event history analysis has influenced data collection. More advanced demographic data collection efforts that are based on micro-level data (e.g., getting information by individual- or household-level interviews) are more and more oriented toward the collection of macro-level data, including time-varying macro-level data. Axinn and colleagues (1997), with their Neighborhood History Calendar, design an instrument that is aimed at collecting neighborhood- or community-level histories to be used as inputs in multilevel event history models. This instrument could provide useful retrospective reports of a community history. An example of the outcome of such instrument is show in Table 133–1. The need to complement micro-level histories with macro-level time-varying data has been influencing recent efforts to collect internationally comparable demographic data. The Generations and Gender Program launched by the UN/ECE in 2000 (UNECE/UNFPA, 2000) includes the collection of retrospective and prospective individual-level survey data, complemented by the development of a longitudinal contextual-level database, allowing for a fully multilevel approach to the analysis of demographic event histories (Spielauer, 2004).

From the micro-level statistical modeling point of view, the study of situational mechanisms, including those due to unobserved factors, is allowed by statistical modeling techniques and increasingly by the col-

TABLE 133–1 Example of the Outcome of Aggregate-Level Data Collection using a Neighborhood History Calendar: Number of Neighborhoods with a School within a 15-Minute Walk

Year	Number	Percent
1964	42	32
1974	86	65
1984	93	70
1994	113	86
1996	119	90

From Axinn *et al.*, 1997, p. 371.

lection of relevant paired micro- and macro-level data. Part of the micro-macro opposition within demography is overcome at least on this side. What event history models, including their multilevel generalization, could not and cannot do, by definition, is to bridge the final gap, specifying how "transformational mechanisms" lead from micro-level to macro-level outcomes. We shall return to this point later.

III. ON EMPIRICAL ANALYSES: MACRO-LEVEL CORRELATIONS AND TRUE MICRO-LEVEL EFFECTS

The risk of incurring in an *ecologic fallacy* (i.e., the inference that a macro-level relationship reflects a micro-level relationship) has been one of the motivations to introduce multilevel analysis in the social sciences and to move away from a pure macro-demographic approach (see Chapter 23). However, critics of micro-demography have rightly underlined the risk of incurring in an *atomistic fallacy* (i.e., the lack of attention to the context in which individual demographic action occurs) (Courgeau, 2003). At the beginning of this chapter, I argued that the life course approach provides a natural framework to study the impact on demographic choices of the *macro-level* context, the *meso-level* context constituted by relevant individual interactions among members of a group (Lelièvre *et al*., 1997), and *micro-level* determinants. The next section is devoted to bridging the gap from the micro to the macro level (and to the aggregation issue represented in the rightmost *arrow* in Fig. 133–1), and this section describes the issue from an empirical point of view, focusing on examples related to fertility.[2]

1. Are Micro-level Data Panaceas for Causal Analysis in Demography?

At the individual level, most studies report that cohabitation has the effect of lowering fertility. De Rose and Racioppi (2001), in analyzing data from Fertility and Family Surveys (FFS) with a multilevel statistical model, show that expected fertility in European countries is lower for cohabiting couples compared with married couples.[3] The causal relationships between partnership status and fertility are, however, not simple to isolate. The first issue that comes to mind is "macro": Some of the countries with the highest proportions of cohabiting couples and earlier ages at first union formation also have the highest levels of fertility in Europe. We know that macro-level relationships are subject to ecologic fallacy. Nevertheless, having access to micro-level event histories is no panacea for the study of causal relationships, because causal relationships may be estimated in a biased way in standard event history models. For instance, using United States data, Brien and colleagues (1999) show that the timing of partnership formation and of nonmarital conception depend on common unobserved factors; in other words, people may get married or start cohabiting *because* they plan to have a child. This type of unobserved heterogeneity, in part "shared" by two demographic processes—and often used in analyses of mortality to assess the role of common factors that affect the survival of two related individuals (see Chapter 21)—makes it difficult to evaluate the outcome of event history models using a causal interpretation. A solution is to use multiprocess event history models, explicitly allowing for the presence of correlated unobserved factors affecting more than one demographic choice at a time (Lillard and Panis, 2004). Using those models, the causal impact of being in a union has been confirmed by studies that take into account the presence of unobserved heterogeneity, and a comparative study between West Germany and Sweden has shown that for the hazard of first birth the distinction between marriage and cohabitation is not significant in Sweden (Baizán *et al*., 2004). Another point that seems clear at the micro level is the relationship between the number of unions and fertility; most scholars see separation or divorce as having an intuitive "clearly negative effect on fertility" (Pinnelli *et al*., 2002, p. 79). This intuitive idea may also be challenged, especially in reference to very low fertility settings, to hypothesize that in some specific situations the dissolution of a union may have a positive impact on fertility—union dissolution may trigger fertility. We can consider a simple, paradoxical example. In a case such as the one observed in lowest low-fertility countries, childlessness is relatively rare, and so is the situation of living as a single forever, but the progression to higher parities is also at particularly low levels. This implies that almost all couples have one child and that not very many progress to a second child. If the rule is one child per couple, the only way to reach replacement is to have individuals experience two-couple relationships! Children may be union-specific capital, symbols of the partners' commitment to their relationship (Griffith *et al*., 1985). Single-country analyses have shown that the first "shared" birth of a couple has a major commitment value (Vikat *et al*., 1999), although

[2] Part of the material presented in this section has been adapted from Billari (2005).

[3] Pinnelli and collegues (2002) provide a more detailed discussion on this issue.

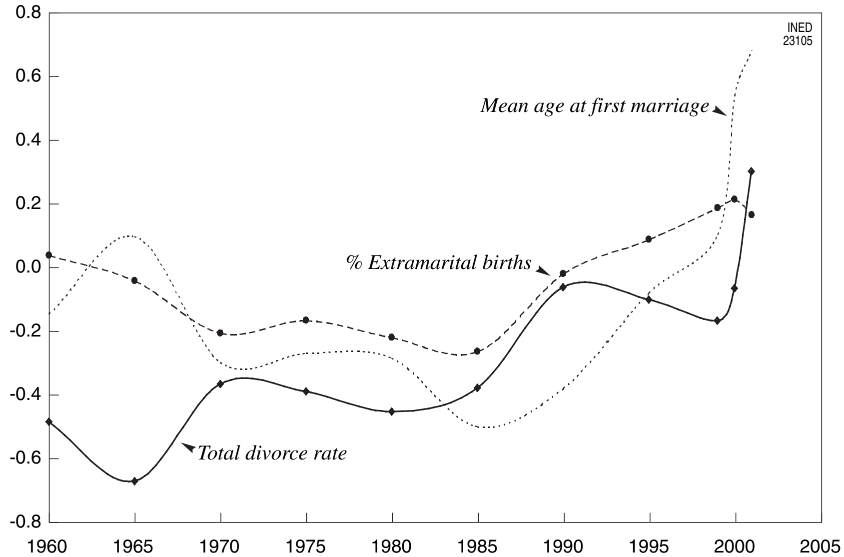


FIGURE 133–3 Changing correlations between indicators concerning union formation and fertility in the countries of the Council of Europe: cross-country correlation coefficients between the total fertility rate and total divorce rate (**A**), percentage of extramarital births (**B**), and mean age at first marriage (**C**). (From Billari, 2005.)

the effect tends to disappear with higher parities (Buber and Fürnkranz-Prskawetz, 2000). Evidence for the commitment value of a first shared birth has also been detected in the FFS analysis by Thomson and colleagues (2002).

Micro-level data alone are no panacea for assessing causality in studies concerning demographic choices. From the purely micro point of view, there is no warranty that the outcome of an event history model gives the causal account of the process. Event history modeling does not rule out the presence of unobserved heterogeneity, and this implies that the outcome of models based on analyses of micro-level data are not to be taken as true causal effects, although discussions in the econometric and epidemiology literature provide a series of tools to assess such effects (Billari, 2006).

2. Macro-level Correlations and Micro-level True Effects

Macro-level correlations between indicators are not necessarily linked to causal relationships at the micro level. The risk of incurring in the so-called ecological fallacy is one of the main justifications to move from macro-demography to micro-demography. We have, however, seen that using a micro-demographic approach does not necessarily lead to grasping causal relationships. What can macro-level relationships tell us? We argue that they help us to start considering the links between macro and micro, that is, the "situational mechanisms" of Figure 133–1. If "micro matters," as Behrman (2001) rightly claimed, we should be aware that macro also matters. Simple correlations can tell us about changes at a general level and give hints on situational mechanism and on the interaction between situational and action formation mechanisms (i.e., macro-micro interactions). The latter can be grasped in a better way using comparative research and/or multilevel modeling.

In the literature, a great deal of attention has been given to the fact that the cross-country correlation between total fertility rates and the share of women participating in the labor market has changed from negative to positive for Organization for Economic Cooperation and Development (OECD) countries. Billari and Kohler (2004) argue that this is one of the several cross-country correlations that have changed in correspondence with the emergence of lowest low fertility in Europe during the 1990s. The attention given to the study of the relationship between total fertility and labor force participation has been much more meticulous than any dedicated to the changing relationships between fertility and other related behaviors.[4] Figure 133–3 shows analyses for countries of the Council of Europe, including the correlation of total fertility rates and mean age at first marriage, the per-

[4] An exception is Dalla Zuanna (2001), who documents the change in the correlation between total fertility rates and the share of nonmarital births in 16 Western European countries from 1981 to 1996.

centage of nonmarital births, and total divorce rates. For all three indicators, the correlation with total fertility rates changes to become positive during the 1990s. Two lessons can be drawn. First, there is an additional fallacy with respect to the ecologic fallacy when looking at macro-level cross-country relationships; it is what we may call a situational fallacy. Cross-country correlations refer only to specific periods, and they may change and even shift in direction as the relevant institutional context of decisions change. Second, changing correlations may help us in discussing the potential role of situational mechanisms; even if at a micro level marriage is conducive to higher fertility, a macro-level focus on marriage may not necessarily be conducive to higher fertility.

Comparative research and multilevel statistical models are useful to understand that there is no such a thing as a "true" causal effect, independent on the context, even at the micro-level. This seems a relatively obvious assertion, but in our opinion another key problem of micro-demographic approaches is to ignore that conclusions based on micro-level analyses are to be declared conditional on the context they are embedded. Cohabitation and marriage have different impacts on fertility (e.g., in the United States, Spain, or Sweden), and the same is true for the impact of divorce on subsequent lives. Referring again to the framework of Figure 133–1, the action-formation mechanism may assume different configurations according to the situation they are inserted in. This can be summarized in the notion of macro-micro interaction effects; some micro-level effects can be different in different macro-level contexts. Some examples of such interactions in terms of multilevel statistical modeling are presented by Courgeau and Lelièvre (see Chapter 23). This is shown in an example about educational enrollment and the timing of first birth in young women's lives.

Most literature on the impact of pregnancy on students focuses on analyzing the United States (and rarely the United Kingdom) and stresses that becoming a mother has the effect of pushing women out of education. In terms of micro-level event history modeling, this translates to saying that when women become pregnant (ignoring cases of abortion), their hazard of dropping out of education increases substantially (Billari and Philipov, 2004). However, leaving education is seen as a major prerequisite to become a mother (Blossfeld and Huinink, 1991). Billari and Philipov (2004) conduct a comparative analysis using multi-process event history models to evaluate the (causal) impact of leaving education on the hazard of first birth and the (causal) impact of the conception leading to the first birth on the hazard of leaving education. Some results in terms of log-hazard related to the change in the relevant time-varying covariate are reported in Table 133–2. Table 133–2 offers two basic messages. First, it is true that finishing education triggers the transition to motherhood, because being involved in education has low compatibility with the role of mother. Nevertheless, the impact is different in different contexts, with the lowest levels in countries with a high welfare state support (Nordic countries) and a high support from parents and relatives (Italy and Spain). Looking at the impact of becoming a mother on educational career, the picture shows that the micro-level effect can be exactly opposite: Becoming a mother

TABLE 133–2 Estimate from a Series of Multiprocess Event History Models of the Impact of Concluding Education on the Hazard of First Birth and of the Impact of Conceiving the First Birth on Concluding Education[a]

	Finland	Norway	Sweden	Austria	Belgium	France	Germany	Switzerland	Greece	Italy	Spain
Impact of concluding education on first conception hazard	0.75	1.02	0.90	1.17	1.75	1.50	1.52	1.02	1.26	0.79	0.68
Impact of first conception on the hazard of concluding education	−0.60	−0.60	−0.83	0.22	0.80	0.43	0.27	−0.71	0.10[b]	−0.60	−0.56

[a] Age at conception leading to first birth is estimated as age at first birth minus 9 months and using the Fertility and Family Surveys data for 11 European Countries.

[b] The effect is not statistically significant.

From Billari and Philipov, 2004.

slows down the exit from education in Nordic countries (where welfare support may help mothers to continue their education at a slower pace) and in Italy and Spain (where the support is likely to come from parents and relatives). Becoming a mother *accelerates* the end of education in other Western European countries. The sign of the effects and the nature of action-formation mechanism may vary with the context and comparative research contributes to illuminate this weakness of pure micro-demographic analyses.

IV. FROM MICRO TO MACRO: THE ROLE OF AGENT-BASED COMPUTATIONAL DEMOGRAPHY

Using sophisticated multilevel models and comparative analyses, we can study macro-micro relationships, particularly the situational and action-formation mechanisms of Coleman's framework depicted in Figure 133–1. The missing part of the micro-macro gap concerns transformational mechanisms. How do we get from the findings and the ideas of micro-demography back to the macro level? This is perhaps the key issue raised by authors who think that demography should give answers about population processes at the macro level.

I previously mentioned the debates about aggregation problems that often pervade neighboring disciplines, particularly economics. Without the strong paradigm of the "representative agent" that underlies mainstream economics, demography has to solve aggregation problems taking into account that demographic choices are made by heterogeneous and interacting individuals and that demographic choices sometimes are made by more than one individual (e.g., couple, household). For these reasons and for the natural links to current micro-demography, computer simulation provides a way to transform the micro into macro without having to impose assumptions on the micro level one would not like to impose (e.g., homogeneity, lack of interaction). This is the way some scholars in neighboring disciplines have chosen when introducing individual-based modeling in ecology. The literature on agent-based modeling in social sciences has flourished. This has particularly been the case in economics (e.g., agent-based computational economics), political science (Johnson, 1999), and to a lesser extent, sociology (Macy and Willer, 2002; Halpin, 1999).

Agent-based computational demography (ABCD)[5] is a set of tools that models population processes, including their macro-level dynamics from the bottom up, starting from assumptions at the micro level (Billari and Prskawetz, 2003). ABCD includes microsimulation that has been used to derive macro-level outcomes from empirical models of micro-level demographic processes (i.e., event history models) and formal models of demographic behavior that describe micro-level decisions and for which we are also interested in macro-level outcomes.

Demography has for a long time been using simulation techniques, and microsimulation has become one of the principal techniques in this discipline; it is a widely discussed and applied instrument in the study of family and kinship networks and family life cycle (Tomassini and Wolf, 2000; Hammel and Wachter, 1996; Ruggles, 1993; Wachter, 1987; Hammel *et al*., 1979). Microsimulation has also been widely used in the study of human reproduction and fecundability (Nakazawa and Ohtsuka, 1997; Riley and Sheps, 1966), migratory movements (Courgeau, 1995), or whole populations (Land, 1986), and its role has been discussed in the general context of longitudinal data analysis (Wolf, 2001). Evert van Imhoff and Wendy Post (1998) provide a general overview of the topic. Microsimulation has been used to study and predict the evolution of population using a model for individuals.

What does ABCD add to demographic microsimulation in helping to bridge the gap between micro-demography and micro-demography? The emphasis of demographic microsimulation has been on the macro-level impact of a certain set of parameters estimated at the micro-level from actual empirical data. There has been no particular emphasis on the theoretical side. Agent-based models do not necessarily include only parameters estimated from actual empirical data, but it may include parameters that are relevant for a specific theoretical meaning. Microsimulation is to the event history analysis what macrosimulation (i.e., population projection based on aggregate-level quantities as in the cohort-component model) is to traditional, macro-level, formal demography. ABCD is the micro-based functional equivalent of mathematical demography. As outlined in Axelrod (1997, p. 4), agent-based computational modeling may be compared with the principles of induction and deduction. "Whereas the purpose of induction is to find patterns in data and that of deduction is to find consequences of assumptions, the purpose of agent-based modeling is to aid intuition." As with deduction, agent-based modeling starts with assumptions. However, unlike deduction, it does not prove theorems. The simulated data of agent-based models can be analyzed inductively, however the data are not from the real world as in the case of induction.

[5] This section drawn heavily from Billari and colleagues (2003).

There are reasons why ABCD may help bridging the macro-micro gap in demography (Billari *et al*., 2003, offers a full discussion). First, it is relatively easy to include feedback mechanisms and to integrate micro-based demographic behavioral theories (and results from individual-level statistical models of demographic behavior such as event history models) with aggregate-level demographic outcomes. This ability to include feedback is possibly the most important advantage of ABCD models. In such models, space and networks can be formalized as additional entities through which the agents will interact. Second, compared with mathematical modeling, it is relatively easy to introduce heterogeneous agents that are not fully rational—hence, the paradigm of the representative, fully rational agent that has and often still penetrates many economic and sociologic applications can easily be relaxed in agent-based modeling. Third, when building agent-based computational models, it is indispensable to adopt simple formulations of theoretical statements. Although agent-based modeling employs simulation, it does not aim to provide an accurate representation of a particular empirical application. Instead, the goal of agent-based modeling should be to enrich our understanding of fundamental processes that may appear in a variety of applications. This requires adhering to the KISS principle, which stands for the slogan "keep it simple, stupid" (Axelrod, 1997). Fourth, using agent-based approaches, it is possible to construct models for which explicit analytic solutions do not exist, such as social interaction and generally nonlinear models. Agent-based models are used to understand the functioning of the model and the precision of theories need not be limited to mathematical tractability. Simplifying assumptions can then be relaxed in the framework of an agent-based computational model. As Axtell (2000) observed, even when models can be solved analytically or numerically, agent-based modeling techniques may be applied because their output is mostly visual and therefore easier to communicate to people outside academia. In general, we can see formal modeling of population dynamics using differential equations and agent-based computational models as two ends of a continuum along the macro-micro dimension (Rahmandad and Sterman, 2004). It is possible to conceive "artificial societies" that need not necessarily resemble present societies; such artificial societies can be seen as computational laboratories that may allow us to reproduce past macro-events from the bottom up.

An example of macro-demographic issue that can be addressed using ABCD is the dynamics of marriage. Billari (2000) has proposed to derive what has been long known within demography as the shape of the hazard rate of marriage (i.e., a macro-level empirical regularity that has been sometimes modeled using "demographic laws") by simulating a marriage market at the micro level. Todd and colleagues (2005) present a series of such models. Figure 133–4 shows the shape of the hazard rate of marriage derived from a model in which individuals marry after a period of "learning" during adolescence (which has a normal distri-

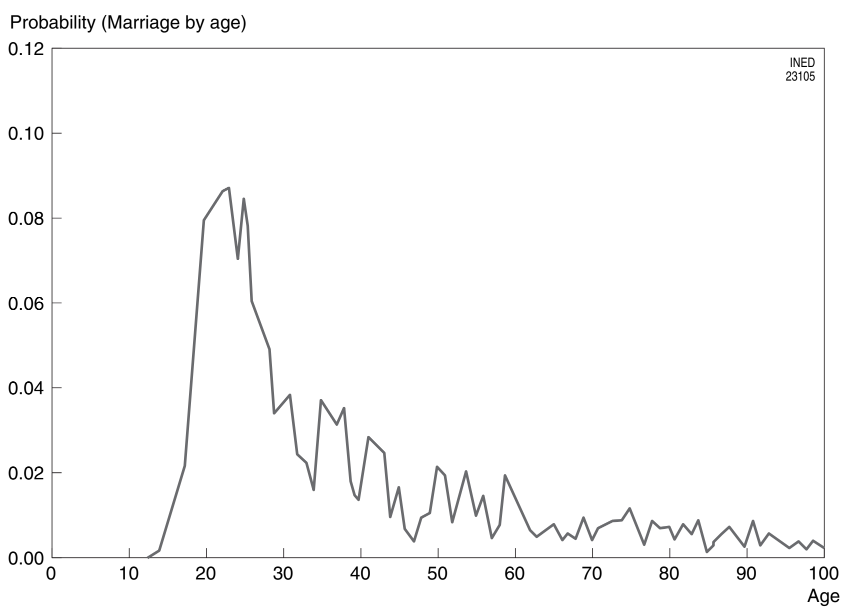


FIGURE 133–4 Hazard rate for marriage from an agent-based simulation of marriage. (From Todd *et al*., 2005.)

bution) as long as they find an acceptable partner who accepts to marry them. The shape of the curve is qualitatively similar to the ones observed in reality.

DISCUSSION

In this chapter, Coleman's framework for the analysis of social dynamics was presented as a potentially unifying scheme to bridge the gap between micro-demography and macro-demography. First, the study of the influence of macro-level factors on micro-level demographic decisions, as well as the interaction between macro- and micro-level factors, has been pushed by recent developments. The life course approach has provided a theoretical framework for demographic decision-making; multilevel statistical models possibly applied to event histories have been developed as key analytic tools; and data collection efforts with true multilevel dynamics have started and are currently planned. With these developments, micro-level analyses that do not take into account macro-level factors can be less justified by the difficulty of connecting the two levels. Second, I have argued that much less progress has been made in demography on the problem of transformational mechanism (i.e., the aggregation of micro-level results to obtain macro-level outcomes). If we are not willing to accept the idea of representative agents that has helped economists to micro-found macro-economics in a formal way, ABCDs (including microsimulation) provide an important toolkit to build a bridge from micro-level findings to macro-level dynamics. Future models may also be developed in a more statistical fashion by estimating the simulation parameters more directly from actual data.

Most of the ideas outlined for bridging the macro-micro gap, including multilevel statistical models, event history analysis, and the role of heterogeneity in the population, have become embedded in the mainstream of demographic thought. To build the final bridge, we need to accept that computer programming, specifically programming aimed at building simulation models, must become part of the mainstream technical tools in demography. As in the case of other approaches, learning and interacting with disciplines that are close to demography constitute an important advantage. Demographers ought not create a gap between micro and macro, and the existing gap needs to be bridged.

Acknowledgments

I would like to thank Paolo De Sandre for useful suggestions, Ron Lee for the permission to borrow ideas and quote from his 2001 paper, and Volker Grimm for sending me a prepublication version of his book with Steven Railsback. Many people have contributed to shaping my ideas on issues dealt with in this chapter. I am especially grateful to Tom Burch, Giuseppe Micheli, Alexia Prskawetz, and Fausta Ongaro.

References

Axelrod Robert M., 1997. The complexity of cooperation. Agent-based models of competition and collaboration. Princeton, Princeton University Press, 232 p.

Axinn William G., Barber Jennifer S., and Ghimire Dirgha J., 1997. The neighborhood history calendar: A data collection method designed for dynamic multilevel modeling, *Sociological Methodology*, vol. 27, p. 355–392.

Axtell Robert, 2000. *Why agents? On the varied motivations for agent computing in the social sciences.* Washington, The Brookings Institution, Center on Social and Economic Dynamics, 22 p. (CSED Working Paper, no. 17; available online at: http://www.brook.edu/dybdocroot/es/dynamics/papers/agents/agents.htm).

Barber Jennifer S., Murphy Susan A., Axinn William G., and Maples Jerry, 2000. Discrete-time multilevel hazard analysis, *Sociological Methodology*, vol. 30, p. 201–235.

Becker Gary S., 1981. *A treatise on the family*. Cambridge, MA, Harvard University Press, 288 p. (Also available: 2nd ed., 1991, 424 p.)

Behrman Jere R., 2001. Why micro matters, *in*: Nancy Birdsall, Allen C. Kelley, and Steven W. Sinding (eds.), *Population matters. demographic change, economic growth and poverty in the developing world*, p. 371–410. Oxford, UK, Oxford University Press, 440 p.

Billari Francesco C., 2000. Searching for mates using "fast and frugal" heuristics: a demographic perspective, *in*: Gérard Ballot and Gérard Weisbuch (eds.), *Applications of simulation to social sciences*, p. 53–65. Oxford, Hermes Science Publications, 461 p.

Billari Francesco C., 2003. Life course analysis, *in*: Paul Demeny and Geoffrey McNicoll (eds.), *Encyclopedia of population*, revised ed., p. 588–590. New York, Macmillan., 1040 p.

Billari Francesco C., 2005. Partnership, childbearing and parenting: trends of the 1990s, *in*: UNECE and UNFPA (eds.), *The new demographic regime. Population challenges and policy responses*, p. 63–94. Geneva and New York, United Nations, 301 p.

Billari Francesco C., 2006. Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood, *in*: Paolo Ghisletta, Jean-Marie Le Goff, René Lévy, Dario Spini, and Eric Widmer (eds.), *Towards an interdisciplinary perspective on the life course*. New York, Elsevier Science (in press).

Billari Francesco C., Ongaro Fausta, and Prskawetz Alexia, 2003. Introduction: agent-based computational demography, *in*: Francesco C. Billari and Alexia Prskawetz (eds.), *Agent-based computational demography. Using simulation to improve our understanding of demographic behavior*, p. 1–17. Heidelberg, Physica/Springer, 210 p.

Billari Francesco C. and Philipov Dimiter, 2004. *Education and the transition to motherhood: a comparative analysis of Western Europe*. Vienna, Vienna Institute of Demography of the Austrian Academy of Sciences, 40 p. (European Demographic Research Papers, no. 3.)

Billari Francesco C. and Prskawetz Alexia (eds.), 2003. *Agent-based computational demography. Using simulation to improve our understanding of demographic behavior*. Heidelberg, Physica/Springer, 210 p.

Billari Francesco C. and Rosina Alessandro, 1998. Does cohort matter in pre-transitional mortality? Analysis of adult mortality using an event history approach: the case of Chioggia in the 17th century, *Genus*, vol. 54(3–4), p. 327–347.

Blossfeld Hans-Peter and Huinink Johannes, 1991. Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation, *American Journal of Sociology*, vol. 97(1), p. 143–168.

Borgan Ørnulf and Hoem Jan M., 1988. Demographic reproduction rates and the estimation of an expected total count per person in an open population, *Journal of the American Statistical Association*, vol. 83(403), p. 886–891.

Brien Michael J., Lillard Lee A., and Waite Linda J., 1999. Interrelated family-building behaviors: cohabitation, marriage, and non-marital conception, *Demography*, vol. 36(4), p. 535–551.

Coleman James S., 1990. *Foundations of social theory*. Cambridge, MA, Belknap Press, 993 p.

Courgeau Daniel, 1995. Migration theories and behavioral models, *International Journal of Population Geography*, vol(1), p. 19–27.

Courgeau Daniel, 2003. General introduction, *in*: Daniel Courgeau (ed.), *Methodology and epistemology of multilevel analysis. Approaches from different social sciences*, p. 1–23. Boston, Kluwer Academic Publishers, 235 p.

Courgeau Daniel and Lelièvre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p. (Published in English: *Event history analysis in demography,* Oxford, Clarendon Press, 1992, 226 p.)

Courgeau Daniel and Lelièvre Éva, 1996. Changement de paradigme en démographie, *Population*, vol. 51(3), p. 645–654.

Cox David Roxbee, 1972. Regression models and life tables (with discussion), *Journal of the Royal Statistical Society*, no. B34, p. 187–220.

Dalla Zuanna Gianpiero, 2001. The banquet of Aeolus: a familistic interpretation of Italy's lowest low fertility, *Demographic Research*, vol. 4(5), p. 131–162.

De Rose Alessandra and Racioppi Filomena, 2001. Explaining voluntary low fertility in Europe: a multilevel approach, *Genus*, vol. 57(1), p. 13–32.

Dykstra Pearl A. and van Wissen Leo J. G., 1999. Introduction: the life course approach as an interdisciplinary framework for population studies, *in*: Leo J. G. van Wissen and Pearl A. Dykstra (eds.), *Population issues. An interdisciplinary focus*, p. 1–22. New York, Kluwer Academic and Plenum Publishers, 287 p.

Giele Janet Z. and Elder Glen H. Jr. (eds.), 1998. *Methods of life course research. Qualitative and quantitative approaches*. Thousand Oaks, Sage, 344 p.

Grimm Volker, 1999. Ten years of individual-based modeling in ecology. What have we learned, and what could we learn in the future? *Ecological Modeling*, vol. 115(2–3), p. 129–148.

Grimm Volker and Railsback Steven F., 2005. *Individual-based modeling and ecology*. Princeton, NJ, Princeton University Press, 412 p.

Halpin Brendan, 1999. Simulation in society, *American Behavioral Scientist*, vol. 42(10), p. 1488–1508.

Hammel Eugene A., 1995. Economics 1: Culture 0. Fertility change and differences in the northwest Balkans 1700–1900, *in*: Susan Greenhalgh (ed.), *Situating fertility: anthropology and demographic inquiry*, p. 225–258. Cambridge, UK, Cambridge University Press, 304 p.

Hammel Eugene A., McDaniel Chad K., and Wachter Kenneth W., 1979. Demographic consequences of incest tabus: a microsimulation analysis, *Science*, vol. 205(4410), p. 972–977.

Hammel Eugene A. and Wachter Kenneth W., 1996. Evaluating the Slavonian Census of 1698. Part II. A microsimulation test and extension of the evidence, *European Journal of Population*, vol. 12(4), p. 295–326.

Hedström Peter and Swedberg Richard, 1999. *Social mechanisms. An analytical approach to social theory*. Cambridge, UK, Cambridge University Press, 340 p.

Henry Louis, 1972. *Démographie: analyse et modèles*. Paris, Larousse, 341 p. (Re-ed. 1984, INED, 314 p.).

Hoem Jan M., 1976. The statistical theory of demographic rates: a review of current developments (with discussion), *Scandinavian Journal of Statistics*, vol. 3(4), p. 169–185.

Hoem Jan M., 1987. Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review, *International Statistical Review*, vol. 55(2), p. 119–152.

Hoem Jan M., 1991. La standardisation indirecte améliorée et son application à la divortialité en Suède (1971–1989), *Population*, vol. 46(6), p. 1551–1568.

Johnson Paul E., 1999. Simulation modeling in political science, *American Behavioral Scientist*, vol. 42(10), p. 1509–1530.

Kirman Alan P., 1992. Whom or what does the representative individual represent?, *Journal of Economic Perspectives*, vol. 6(2), p. 117–136.

Land Kenneth C., 1986. Method for national population forecasts: a review, *Journal of the American Statistical Association*, vol. 81(347), p. 888–901.

Lee Ronald D., 2001. *The decline of formal and aggregate analysis: demography abandons its core*. Berkeley, CA, University of California, Center on the Economics and Demography of Aging, 6 p. (Paper presented at the Annual Meeting of the Population Association of America. Online available at: http://www.demog.berkeley.edu/~rlee.)

Lelièvre Èva, Bonvalet Catherine, and Bry Xavier, 1997. Analyse biographique des groupes, les avancées d'une recherche en cours, *Population*, vol. 52(4), p. 803–830. (Special issue: Nouvelles approches méthodologiques en Démographie, edited by Daniel Courgeau. Published in English: Event history analysis of groups. The findings of an on-going research project, *Population: English Selection*, 1998, col 10(1), p.11–37).

Lesthaeghe Ron J., 1998. On theory development and applications to the study of family formation, *Population and Development Review*, vol. 24(1), p. 1–14.

Lillard Lee A. and Panis Constantijn W. A., 2004. *Multilevel and multiprocess modeling, aML version 2. User's guide and reference manual*. Los Angeles, CA, EconWare, 446 p.

Macy Michael W. and Willer Robert, 2002. From factors to actors: computational sociology and agent-based modeling, *Annual Review of Sociology*, vol. 28, p. 143–166.

Mayer Karl U. and Huinink Johannes, 1990. Age, period and cohort in the study of the life course: A comparison of classical APC analysis with event history analysis or Farewell to Lexis? *in*: David Magnusson and Lars R. Bergman (eds.), *Data quality in longitudinal research*, p. 211–232. Cambridge, UK, Cambridge University Press, 285 p.

Mayer Karl U. and Tuma Nancy B. (eds.), 1990. *Event history analysis in life course research*. Madison, University of Wisconsin Press, 297 p.

Nakazawa Minato and Ohtsuka Ryutaro, 1997. Analysis of completed parity using microsimulation modeling, *Mathematical Population Studies*, vol. 6(3), p. 173–186.

Perks Wilfred, 1932. On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries*, no. 63, p. 12–57.

Pinnelli Antonella, De Rose Alessandra, Di Giulio Paola, and Rosina Alessandro, 2002. Interrelationships between partnership and fertility behavior, *in*: Miroslav Macura and Gijs Beets (eds.), *Dynamics of fertility and partnership in Europe. Insights and* *lessons from comparative research*, vol. 1, p. 77–98. New York, United Nations, 159 p.

Rahmandad Hazhir and Sterman John, 2004. *Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models*. Cambridge, MA, Massachussets Institute of Technology, 31 p. (MIT Sloan Working Paper 4512–04.)

Ridley Jeanne C. and Sheps Mindel C., 1966. An analytic simulation model of human reproduction with demographic and biological components, *Population Studies*, vol. 19(3), p. 297–310.

Ruggles Steven, 1993. Confessions of a microsimulator, *Historical methods*, vol. 26(4), p. 161–169.

Spielauer Martin, 2004. *The generations and gender contextual database. Concepts and content*. Rostock, Max Planck Institute for Demographic Research, 47 p. (MPIDR Working Paper WP 2004–026.)

Tesfatsion Leigh, 2001. Introduction to the special issue on agent-based computational economics, *Journal of Economic Dynamics and Control*, vol. 25(3–4), p. 281–293.

Todd Peter M., Billari Francesco C., and Simăo Jorge, 2005. Aggregate age-at-marriage patterns from individual mate search heuristics, *Demography* (in press).

Thomson Elizabeth, Hoem Jan M., Vikat Andres, Prskawetz Alexia, Buber Isabella, Toulemon Laurent, Henz Ursula, Godecker Amy L., and Kantorova Vladimira, 2002. Childbearing in stepfamilies: how parity matters?, *in*: Erik Klijzing and Martine Corijn (eds.), *Dynamics of fertility and partnership in Europe. Insights and lessons from comparative research*, vol. 2, p. 87–99. New York, United Nations, 238 p.

Tomassini Cecilia and Wolf Douglas, 2000. Shrinking kin networks in Italy due to sustained low fertility, *European Journal of Population*, vol. 16(4), p. 353–372.

UNECE and UNPFA, 2000. *Generations and gender programme. Exploiting future research and data collection options*. New York and Geneva, UNECE and UNPFA, 166 p.

van den Bergh Jeroen C. J. M. and Gowdy John M., 2003. The microfoundations of macroeconomics: an evolutionary perspective, *Cambridge Journal of Economics*, vol. 27(1), p. 65–84.

van Imhoff Evert and Post Wendy J., 1998. Microsimulation methods for population projection, *Population: An English Selection*, vol. 10(1), p. 971–938. (Special issue: New Methodological Approaches in the Social Sciences.)

Wolf Douglas, 2001. The role of microsimulation in longitudinal data analysis, *Canadian Studies in Population*, vol. 28(2), p. 313–339.

Zimmermann Klaus F. and Vogler Michael (eds.), 2003. *Family, household and work*. Berlin, Springer, 427 p.

CHAPTER

134

Fertility, Child Mortality, and Nuptiality Indicators at the Individual Level

DOMINIQUE TABUTIN

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Over the past 30 years, notably due to the increase in surveys, demographic research has moved increasingly toward the analysis of individual data, with the hope of improving the explanation of observed phenomena. We have therefore moved from the long-preferred macro, aggregated level to the micro level[1] by attempting to extract the best from the various statistical tools available in the social sciences, including classification and regression (see Chapters 131 and 132). In a survey or a census, individuals are characterized by age, sex, marital status, level of education, occupation, and sometimes by their ethnic group or religion, which make up the variables known as explanatory or independent. How can dependent variables (i.e., those requiring explanation), such as a woman's or a man's fertility, a woman's or a couple's history regarding child mortality, or the intensity of marital mobility, be simply defined?

Can composite indicators be created at an individual level? Can indicators be created that summarize a life or an experience and avoid the perverse effects of confounding variables, which in demography are frequently the individual's age, duration of marriage, or the number of children already born? These indicators would enable the direct comparison of individual fertility of women (or men) and their integration into individual statistical analyses or (if needed) the creation of groups, according to the levels of excess fertility or subfertility (in comparison to the average level) or of excess child mortality or child undermortality.

There are sophisticated methods for analyzing individual data based on the occurrence of events during a given period. For example, has there been a birth (or not) over the last 5 years? Has a child died (or not)? Has contraception been used (or not)? Others are based on the number of events observed during a recent given period (the last 5 years, for example). However, this is not the main subject here, and I will not expand on the vast and complex issue of the statistical tools required for these analyses of individual determinants.

The aim of this chapter is to specify to what extent it is possible at a given time to summarize the fertility history of a woman (or a man), the mortality history of her children, or the history of her marital life through simple individual indicators that are representative of her fertility, her children's mortality, or of her nuptiality. Such indicators can be valuable tools for explanatory analysis in processing abundant data,

[1] With a tendency today to try and associate the various levels (i.e., multilevel approaches).

such as those provided by surveys such as the World Fertility Survey (WFS)[2] and the Demographic and Health Surveys (DHS).[3]

I. DATA AND GENERAL PROBLEMS

The WFS surveys[4] of the 1970s, the DHS surveys from 1985, and in the Arab countries, the Pan Arab Project for Children (PAPCHILD) surveys of the 1990s, and the Pan Arab Project for Family Health (PAPFAM) surveys from 2001, as well as most censuses carried out in developing countries over the past 20 years, have included simple questions on each woman's total number of live births and on the total number of these children who died. In most cases, the fertility surveys also provide the complete maternal history and the survival status of each child,[5] sometimes even including the marital history (i.e., marriages and successive remarriages with the dates of these events and the causes of separation).

Although the total number of events experienced since a given age (e.g., 15 years) or an original event (e.g., marriage) up to the time of the census or survey are individual measurements that are directly identifiable at the level of each woman, they are not measurements of fertility, nuptiality, or child mortality because they depend considerably on the different characteristics of each woman, such as her current age, age at marriage, or the duration of her marriage, to mention only the three factors that most affect the level of attained fertility.

Let us consider a few examples. Within the same context, one woman declares seven live births (of which three died); another, five (with two deaths); another, three (with one death). Without knowing the woman's age or the duration of her marriage, we cannot comment on her fertility level, and we cannot come to any precise conclusions on the mortality level of her children. The ratio of deceased children to the total number of children born provides a fair idea of their survival rate, but we do not know the time scale. Of these three women, we cannot judge which is the most fertile or which has had the most negative experience with regard to her children's mortality.

The idea is to create indicators of a woman's parity or of the mortality of one woman's children independently of variables other than fertility and mortality, which modify the level throughout each woman's life.

Two approaches are possible for creating these indicators:

1. *Adjusting the number of declared events to the duration of exposure to risk*, such as adjusting the parity by dividing a woman's number of declared children by the duration of her exposure to the risk of fertility (of which several definitions are possible)
2. *Calculating an indicator of relative level for each individual*, defined as the ratio of the number of declared events to a theoretical (or expected) number, taking into account notably her age and the duration of her marriage. This process of indirect standardization makes it possible to place each individual (or group of individuals) in excess fertility or subfertility, in excess child mortality or child submortality, in excess mobility or submobility, and so on, compared with a standard or average reference level.

These indicators allow much analytic freedom, because they can be used in aggregated approaches (e.g., comparison between subpopulations, social groups), as well as in analyses of individual level determinants (e.g., regressions). In the latter case, they make it possible to go beyond traditional demographic analysis (aggregated) and sometimes to synthesize various elements of information, to better specify risk factors or explanatory factors relevant to the individual. Having said this, they are not without their limitations or intrinsic problems, which we will be able to discuss after having defined the proposed indicators.

II. INDIVIDUAL FERTILITY INDICATORS

The most traditional cumulated fertility indicator and the most immediate at the individual level, which resumes a woman's (or a man's) fertility history, is *parity* (i.e., the number of live-born children declared by a woman at a given time). However, a woman's parity at a given time does not only depend on her fertility. To control for effects relating to the woman's age, her age at marriage, or the duration of her marital life, the *corrected parity*, taking into account the

[2] For the World Fertility Survey, the first large international program of fertility surveys, carried out in the 1970s and 1980s.

[3] For the Demographic and Health Surveys, the second large international program of surveys, which succeeded WFS during the mid-1980s. It is currently in its fourth phase.

[4] Chapter 121 is devoted to information systems in demography. Almost 300 national surveys have been conducted since 1970 in 120 developing countries.

[5] By calculating the number of successive births and deaths, we obtain the completed fertility and the number of deceased children per woman at the time of the survey.

duration of the risk of fertility can be calculated, or a *standardized indicator* (e.g., *duration ratio* [DRAT]) can be created that compares a woman's fertility to a standard level, independent of age or the duration of marriage.

1. From Observed Parity to Adjusted Parity

It is possible to directly analyze the individual parities observed, and this is even common practice, but it requires taking precautions that deprive us of a considerable part of the information available. The idea of calculating adjusted parities resulted from this problem.

a. Observed Parities

To study the relationship between fertility and socioeconomic variables, many authors use *parity* (i.e., the cumulated number of live births that a woman has had from the beginning of her reproductive life up to the time of the survey). Most frequently, the analysis only refers to married women. It is often available and simple (i.e., there is no need to date the events, there is no reference period, and therefore there are no problems in localizing the event in time).[6] In statistical analyses on observed parity, it is necessary to circumvent the effects of age or marriage duration. There are several possibilities:

1. *Using the parity of each woman as a dependent variable*, by introducing age and marriage duration among the explanatory variables; in this case, however, a considerable part of the explanation focuses on these two variables and, in particular, we must make the hypothesis that the relationship between fertility and the explanatory variables do not vary according to women's age or cohort.
2. *Limiting the analysis to a certain age group* (i.e., cohort) of women; in this case, several choices are available:
   a. *Working on only the oldest women*, those who have more or less completed their reproductive life (those between the ages of 40 and 49 or 45 and 49 years), but controlling for the marriage duration; the behavior of these women is primarily anchored in the past and the explanatory analysis will provide little information on the reasons behind current practices if fertility has changed recently. Furthermore, the characteristics of the women observed at the time of the survey can differ from what it was at the time in which they made their important decisions with regard to fertility, and this temporal shift between variables to be explained and explanatory variables can result in significant bias and errors in interpretation.
   b. *Centering the analysis on an intermediary age group*, in the prime of fertility and with sufficiently high numbers.[7]
   c. *Stratifying the sample of mothers into age groups* and undertaking separate analyses on those in the age groups of 25 to 34, 35 to 44, and 45 and older by introducing marriage duration as a control variable. This makes it possible to expose the explanatory factors specific to these three cohort groups, to different histories and experiences, but it also leads to working on smaller numbers and to the loss of statistical precision.

Each of these approaches has the advantage of simplicity, but they also have disadvantages, particularly that of limiting the analysis to small groups that require, in the best cases, at least much larger samples.

b. Adjusted Parities

To retain simplicity and to avoid this limitation at the same time, we attempted to control for age and marriage duration by dividing each woman's parity by the duration of her exposure to the risk of fertility. We create a *standardized indicator*, the equivalent of a rate at an individual level.

If P_i is the parity of a woman i and n_i the number of years of exposure to risk:

$$p_i = \frac{P_i}{n_i}$$

For a woman with four children in 8 years, $p_i = 0.50$; for another with four children in 12 years, $p_i = 0.33$. This indicator can be used in individual and aggregated statistical analyses (e.g., social groups, cohort

[6] This type of information is not without risks of omission (e.g., children dying at a young age) or confusion (e.g., between stillbirths and live births). These risks are generally greater among older women and those with greater completed fertility and within a higher mortality context. The women's age must also be correctly reported. It is common, especially in Africa, for young women to be declared as older and for older women to be declared as younger, which results in underestimates and overestimates of fertility.

[7] This is what Bruno Schoumaker (2001) did in a study of individual and contextual fertility determinants (i.e., multilevel analysis) in rural Morocco by taking the number of live-born children of women between the ages of 25 and 39 as a dependant variable and by introducing the age groups of 30 to 34 years and 35 to 39 years, as well as marriage duration, among the 20 explanatory variables. Access to the media, the contextual mortality level, age, and particularly marriage duration showed a significant effect on fertility.

groups), because it makes it possible to place women or groups into categories of excess fertility or subfertility according to an average level.

By considering only married women, *n* can be the period since the first marriage, or better still, if we also want to check periods of divorce or widowhood, the total number of years spent in union.[8] In societies in which a considerable proportion of births occur before marriage, the beginning of a woman's period of exposure could be fixed as being the year of her first birth minus 1 year.[9]

This indicator of individual performances can be calculated for women who are still married at the time of the survey or for all ever married women. It is advisable to exclude young women (15 to 19 years) and to limit the calculation to women who have been married for 4 or 5 years (or who had been married for at least 4 or 5 years). Although the construction of this indicator is simple in principle, it requires correct information concerning age at first marriage (or the time interval since the marriage) or the duration of each union, which is far from always being the case.

These adjusted parities (by duration of exposure to risk) have not often been used in statistical analyses of fertility determinants.[10] Nevertheless, we can mention some studies carried out using data from the WFS of the 1970s and 1980s (Pullum, 1987; Little, 1977).

In the same way, we can also work with *births occurring over a given period or within a given age range*. An already old example is provided by a study of the socioeconomic determinants of legitimate fertility carried out in Sri Lanka by Roderick Little and Soma Perera (1981), from the WFS of 1975. These authors analyzed the fertility of three large female cohorts defined by their marriage duration[11]: less than 10 years of marriage, 10 to 19 years, and 20 years and more. In each of these groups, the women had different durations of exposure. Let us take, for example, those with between 10 and 19 years of marriage; the duration of exposure is complete for the first 10 years of marriage but not for the following years (some women are at 11 years, and others are at 19 years of marriage). To control for these durations of exposure and be able to compare individual fertility, the authors created, for each woman, an indicator of the estimated number of births over 10 years; in other words, the fertility constituted between 10 and 20 years of marriage. For the cohorts with 10 to 20 years of marriage, if $N(10–19)$ is the parity of a woman who has had 10 to 20 years of marriage and $M(10–19)$ is the number of months lived by this woman since her 10th wedding anniversary, the estimated number of births over 10 years would be equal to

$$120\frac{N(10-19)}{M(10-19)}$$

and so on for the other cohorts. This fertility indicator can be used in individual analyses (e.g., regression) and in aggregate analyses (by marriage cohort or socioeconomic group).

c. A Variant on Male Fertility

Unions of various types, high marital mobility, high polygamy, and sometimes fertility outside of marriage are characteristics of numerous societies in sub-Saharan Africa. However, there are very few surveys of male fertility that trace the reproductive and marital lives of men, union by union, woman by woman, child by child. Nevertheless, this was the case of a survey carried out in South Benin in 1989 on 1650 married men aged 20 years or more (Donadjé and Tabutin, 1994). The completed fertility of married men was 11.4 children, with polygamy affecting about one-half of them. However, in such a context, how can the individual fertility of men be measured?

We could have created a DRAT indicator that controlled for age and marriage duration (but did not because we had not thought of it). To take into account the significant variability in the ages of the respondents and their experiences, we should verify the duration of each union. For each man, we calculated *the number of children per man-year* by dividing the total number of live-born children of each man (i.e., his parity) by the total number of years spent in marriage with each of his wives.[12] To eliminate very young husbands, cases of sterility, and quick divorces, this indicator was calculated using men who had spent at least 5 years in union and had at least one child.

This indicator made it possible to confirm high male fertility (the overall average was 0.30, or one child every 3 years) and to study the social or spatial dis-

[8] If we have the required data and it is of good enough quality to calculate it (i.e., duration of different unions or starting and ending dates of each union).

[9] Taking into account the nine months of pregnancy.

[10] Without really knowing why, the number of births occurring over the 5 years preceding the survey are often preferred, which is a better period indicator.

[11] The authors rightly preferred the duration of marriage to age because of the rapid increase in age at marriage in Sri Lanka and the decline in fertility that began in the 1970s.

[12] More precisely, from marriage up to the age of 49 for women, who at the time of the survey were older than 50 years or who left their union after the age of 50 years.

parities and the individual determinants, independent of union duration.

2. Duration Ratio

Let us consider the second approach, based on comparing each woman's parity with a reference model. We will then extend the procedure to men. With the dual aim of avoiding the segmentation of the samples due to the necessity of verifying both the age and marriage duration and to improve the explanation, Bryan Boulier and Marc Rosenzweig, in 1978, proposed an individual standardized measurement of cumulated fertility. It is an indicator of the *relative fertility level*, defined as the ratio of an observed fertility to a theoretical fertility. For each married woman, the number of children that she has had is compared with the theoretical number of children that she could have had in a natural fertility context considering her age and union duration. The DRAT of a woman aged *a* is therefore equal to:

$$\mathrm{DRAT(a)} = \frac{C(a)}{\int_m^a f(a)da}$$

where $C(a)$ is the number of children of the woman aged a, $f(a)$ the natural fertility rate by age and m the woman's age at marriage.

This is an indirect standardization, a kind of replica at the individual level of the marital fertility indicator that Ansley Coale (1973) created at the aggregate level.[13] DRATs are fertility indicators that make it possible to compare women (or groups of women) who have been exposed to the risk of reproduction at different durations. It is advisable to calculate them for women who have been married for at least 4 or 5 years.

DRAT controls the effect of age on fertility only to a certain extent as we make the hypothesis that theoretical fertility only depends on age, thereby excluding the role of age at marriage on fertility, frequently demonstrated in contemporary or historical realities. This model also assumes that observed and theoretical fertilities have the same tempo. This can lead to distortions. The choice of fertility reference is therefore important. The series of natural fertility rates calculated by Ansley Coale and colleagues (1975)[14] generally are used. However, as Bryan Boulier and Marc Rosenzweig have already stressed, other reference models more appropriate to the population studied can be chosen. Rather than a theoretical model or one based on old European data, there is no reason not to create one's own fertility reference based on the realities of the country or region studied. This is what Aloys Ilinigumugabo (1989) did in his multivariate study of fertility determinants in Rwanda, for example. The two standards (his and that of Ansley Coale and colleagues) lead to the same relationships, but the Rwandan standard doubles the explained variance. Another solution is to take the average fertility by age of the country studied as a reference; the subfertile women (or groups) would have a DRAT below 1, the overfertile would be above 1.

The DRAT reduces heteroscedasticity. This problem, which is quite common in regressions of the number of children according to age, is due to the fact that the variance of the dependent variable is related to an explanatory variable and that it is not constant. This is the case with the number of children whose variance increases significantly with the age of the women.

Without going as far as to supplant the observed parities, DRAT has been used in a considerable number of analyses of fertility determinants in developing countries. For example, Ron Lesthaeghe and colleagues (1985) took it as a dependent variable in their study of individual and contextual effects of education on the intermediate fertility variables in Kenya. Bruno Schoumaker and Dominique Tabutin (1999) used it in the same way in multilevel analyses of fertility in rural Morocco.

a. Is it Possible to Calculate a Male Duration Ratio?

In the same way, in a polygamic context or one of high marital mobility, it is possible to compare the total number of children that a man has had with his different wives (successive or simultaneous) to the theoretical number that he would have had if each of his wives had followed a natural (or theoretical) fertility reference plan during the period spent with this man.

If $C_1, C_2, C_3 \ldots, C_n$ are the number of live-born children that a man has had with his wives $1, 2, 3, \ldots, n$, if $m_1, m_2, m_3, \ldots m_n$ are the ages of entry into union of these women and $a_1, a_2, a_3 \ldots a_n$ their ages (at the

[13] Defined as the ratio of the real number of births observed in a population to the number of births that would have been observed if the women of this population had had the same fertility as the Hutterites, who had one of the highest marital natural fertility rates in history (i.e., total fertility rate of 12.4 children per married woman).

[14] A standard series calculated from 13 populations, mainly European, with natural fertility leading to a total legitimate fertility rate of 11.8 extrapolated for annual age groups to calculate the DRAT.

time of the survey if they are still married or at the end of their union if they are no longer married), the male DRAT is equal to

$$\frac{C_1 + C_2 + C_3 + \ldots + C_n}{\int_{m_1}^{a_1} f(a)da + \int_{m_2}^{a_2} f(a)da + \int_{m_3}^{a_3} f(a)da + \ldots} = \frac{\sum_1^n C_n}{\sum_1^n \int_{m_n}^{a_n} f(a)da}$$

This model adopts the same hypotheses as the female model. We should also assume, which is a considerable assumption, that a man's behavior (e.g., his sexuality) is independent of his age (i.e., he is young at the first union, significantly older at the *n*th) and the age of the different wives in the case of polygamy (i.e., no extreme discriminatory behavior). The calculation should no doubt be limited to men age 30 to 55 years. When both spouses are still together in their first union, the man's DRAT is equivalent to that of the woman's.

3. Individual Measurement of Recent Fertility

If we truly wish to understand and explain the current situation, it is not sufficient to work with cumulated fertility indicators (e.g., those that we have just discussed) that are based on a stock of children constituted over many years (Farooq, 1985). In the study of changes in fertility, we should attempt to measure recent behavior. In most cases, the available explanatory variables (e.g., education, place of residence, occupation, standard of living) relate to the survey date.

It is therefore necessary to use *the number of births reported by the (married or ever married) women over a recent period*. We could take the births of the last 12 months, information that is often available in censuses but that is sometimes of poor quality (e.g., misperception of the reference period, nondeclaration of the birth of recently deceased children); in detailed retrospective surveys, the information is better, but the number of events is quite low and therefore subject to significant sampling errors. It is better to increase the reference period.

It is most common to use *the number of births that occurred in the 3 or 5 years preceding the survey* (or sometimes the interval since the first marriage when that is less than 5 years), a kind of compromise between the wish to study current changes and the need to minimize reporting or sampling errors. Over a period of 5 years, it is even possible to study the intervals between births for women who have had more than one child in the period.

The duration of exposure to marriage during this period should be controlled for: from 5 years for a large number of women, less for others if they were married during the interval or have divorced. The ratio of the number of births in interval N to the marriage duration d leads to the following indicator:

$$f = \frac{N}{d}$$

At the individual level, it is equivalent to the marital fertility rate. This type of indicator is used as a dependent variable in many studies of individual fertility determinants in developing countries.

III. INDICATORS OF CHILD MORTALITY PER WOMAN

A few words on the principles and data are required before a presentation of the possible indicators.

1. Principles and Data

The study of mortality in demography remained for a long time very descriptive and was based on traditional tools (e.g., rates) (see Chapters 8, 11, and 14) or indirect approaches (e.g., Brass' methods) (see Chapter 127). We were satisfied with some differential approaches aggregated in terms of education level, social group, and zone of residence of the mother or the parents. It was and still is useful but completely insufficient to explain mortality, notably in terms of individual risk or populations at risk. Consequently, such as was the case for fertility but with a slight delay, during the 1980s, individual approaches were developed for child mortality that consisted of measuring for each woman an indicator of excess or submortality of her children compared with an average or a standard. Like any new methodology, this approach has its advantages and its limitations. Explanatory studies using various methods of multivariate analyses of the survival of children born in a recent period also multiplied, but we will not examine them here.

The creation of indicators per woman was started and developed by James Trussell and Samuel Preston at the beginning of the 1980s, then taken up again during the 1990s, notably by Amadou Noumbissi. Aimed at going beyond traditional analyses of the determinants of child mortality, it fulfills the dual aim of:

1. Improving differential analysis and research of the determinants of mortality at young ages by integrating individual heterogeneity into the study

of differences between groups, because at a given age, the children of the same group are not all exposed to identical risks of death[15]

2. Using in an optimal manner the mass of individual information on mothers and their children provided by the large samples of surveys such as WFS, DHS, and others.[16]

The principle in the construction of these child mortality indicators[17] (CMI) at the level of each mother is simple: we attempt to measure the relative experience of a mother (i) with regard to infant and child mortality by dividing the number of deaths that she experienced (Di) by the estimated number that she would have had (E_i) if mortality was at a standard level (theoretical or empirical):

$$CMI(i) = \frac{D_i}{E_i}$$

We postulate that the proportion of a woman's (or a group of women's) children who die give a measurement of child mortality if it is adjusted to take into account marriage duration or a woman's age.[18] This idea is already old[19] and is the basis for all indirect estimation methods (at the aggregated level) of child mortality from proportions of deceased children. The different variants of the approach come from the method adopted for estimating theoretical deaths.

2. Trussell and Preston's Indicators

James Trussell and Samuel Preston (1982, 1984) developed three versions of their models that differ in the way that they estimate theoretical deaths and in the data required (Amadou Noumbissi, 1996, p. 91–114). These go from the simple numbers of each woman's live-born and deceased children (resulting from the Brass questions) to the complete description of the fertility tempo by the birth dates of each child. Let us now consider the simplest and most common method based on the total number of live-born children and deceased children, simple information that is available in all surveys and in numerous censuses in developing countries, and the age and marriage duration of each woman.

The indicator for a woman i aged j (or with marriage duration j) is equal to the number of her children who have died D_i divided by an expected number of deceased children E_i:

$$CMI_{ij} = \frac{D_i}{E_i}$$

For a given woman, the number of deaths observed is the product of her births N_i by the real proportion of deceased children PD_j while the number of expected deaths is the product of her births by the theoretical proportion adopted PD'_j:

$$CMI_{ij} = \frac{D_i}{E_i} = \frac{N_i \cdot PD_j}{N_i \cdot PD'_j} = \frac{PD_j}{PD'_j}$$

The estimate of the expected proportion is based on the Brass method, which makes it possible to estimate the probability of death among children (from birth to age a) as being the product of the proportion $d(j)$ of the deaths of children of mothers of a given age j by a multiplying coefficient $k(j)$, reflecting the varying precocity of fertility (Brass, 1975):

$$q(a) = d(j) \cdot k_{(j)} \quad \text{or} \quad d(j) = \frac{q(a)}{k(j)}$$

These coefficients $k(j)$, which are a function of the fertility pattern, should be calculated from declared parities according to age (e.g., 15 to 19, 20 to 24, 25 to 29 years) or according to marriage duration (e.g., 0 to 4, 5 to 9, 10 to 14 years). Rather than using the original method of William Brass (which only uses age), we will use the variant by James Trussell (United Nations, 1984, p. 76–86; Trussell and Hill, 1980) based on the model life tables of Coale and Demeny.[20]

To estimate the expected proportion of deceased children PD', we select the probabilities of dying of a standard model life table, $q_s(a)$, and convert them into proportions of deceased children:

$$PD'_j = \frac{q_s(a)}{k(j)}$$

[15] It challenges the hypothesis of homogeneity within the group, which is typical of aggregated analysis based on averages meant to represent each person's risk. All children are born with a personal health capital that changes with varying rapidity over the course of their lives according to the external threats that they face, their health and nutrition, and other factors.

[16] These samples are frequently about 5000 women and make it possible to observe 15,000 to 20,000 live births.

[17] An excellent presentation, critique, and illustration is made in the book of Amadou Noumbissi (1996) on the methods of analyzing child mortality.

[18] A necessary adjustment insofar as the older the woman or the longer her marriage duration, the older her children and the longer they are exposed to the risk of mortality and the higher her proportion of deceased children.

[19] Developed by William Brass in 1964, they were then largely improved by him and by many other analysts during the 1970s and 1980s (see Chapter 127).

[20] In our opinion, nothing prevents the use of another version of the method, such as that of Alberto Palloni and Larry Heligman (1985), based on the United Nations' model life tables for developing countries.

The chosen standard model life table is the one that best represents the actual situation. James Trussell and Kenneth Hill (1980) give a choice between the North, South, East, and West families of Coale and Demeny's tables. Alberto Palloni and Larry Heligman (1985) propose a choice between five models of the United Nations' tables.[21] Attempts should be made to ensure that the number of real deaths is more or less equal to the number of expected deaths for the whole population.

The child mortality indicator of a woman *i* aged *j* is calculated as follows:

$$CMI_{ij} = \frac{D_i}{N_i \cdot q_s(a)/k(j)}$$

This is a standardized mortality indicator as it takes into account the age or marriage duration of the mothers. Like DRAT, it can be used in individual or aggregated analyses (in which it represents the average of a group). It is relatively simple to implement.

However, this indicator makes it possible to measure the relative experiences of mothers with regard to mortality, but not mortality levels themselves. It does not refer to any period or to a precise age group of children; the older the women, the older the average period and the more the age groups of mortality concerned are extended (e.g., infants, children, adolescents, sometimes adults). The choice of a standard level poses a problem that is inherent to all processes of indirect standardization, which is discussed in the section on the indicator proposed by Amadou Noumbissi.

The Trussell-Preston indicator was used in a large United Nations study (1985) on the socioeconomic differences of infant and child mortality in 15 developing countries (i.e., seven from Africa, five from Asia, and three from Latin America) based on WFS data or census data that included Brass' questions. This study uses the CMI as a variable to be explained, and according to country, from 10 to 17 explanatory variables (of which some were contextual). The dozen privileged variables[22] of mother or father's education, the zone and region of residence, ethnic group, religion, income, family structure were systematically made the subject of univariate, bivariate, and multivariate analyses, combining aggregated and individual analyses. This methodologically pioneering study demonstrated the benefits of using an indicator such as the Trussell-Preston CMI to identify the factors of child mortality.

3. Noumbissi's Indicator

At the beginning of the 1990s, Amadou Noumbissi took up the generalized problem of measuring child mortality per woman and taking individual heterogeneity into account and attempted to improve the Trussell-Preston method. Preserving the approach and the general hypotheses of the method, he proposed simplifying the calculation of the indicator by dividing the number of deaths derived from a standard model life table by the number of deaths observed in the whole of the population. For a woman *i* aged *j* (or with a marriage duration of *j*):

$$CMI_{ij} = \frac{D_{ij}}{E_{ij}} = \frac{D_{ij}}{N_{ij} \cdot PD_{tj}} = \frac{PD_{ij}}{PD_{tj}}$$

where N_{ij} represents the births of woman *i* aged *j*, PD_{ij} the proportion of her deceased children and PD_{tj} the proportion of children deceased for all women of age *j*.

The formula remains the same, but the reference to observed average mortality among women of the same age group (or marriage duration) avoids the problems resulting from using a standard model life table. Amadou Noumbissi (1996, p. 98–102) has demonstrated that the three types of indicators proposed by James Trussell and Samuel Preston (of which the simplest was presented above) were not sensitive to the choice of model life table, with the various possible standards leading more or less to the same results as the reference to actual average mortality. Logically, it is recommended to use Noumbissi's method, which is the simplest, and to apply it to women between the ages of 20 and 45 or 25 and 40 years, where it is most effective.[23]

Amadou Noumbissi (1996) applied these various methods to the data from two fertility surveys carried out in Cameroon in 1978 and 1991, combining univariate and multivariate analyses and multiplying the comparisons between different approaches. He concludes that socioeconomic and cultural factors have more weight than the bio-demographic factors of each child, stressing the importance of the individual heterogeneity of the mothers' characteristics, regardless of social group.

[21] "Latin America," "Chile," "South Asia," "Far East," and "General" (i.e., the average of four others).

[22] Each was the subject of a chapter in the United Nations' publication.

[23] It is better to exclude the youngest and oldest women because of the excess mortality of children specific to these extreme ages. There will be a slight loss in precision due to the decrease in the sample size, but there will be an improvement in quality.

IV. POSSIBILITIES REGARDING NUPTIALITY

In surveys in developing countries, nuptiality is generally treated more succinctly than fertility and child mortality. However, the WFS of the 1970s and 1980s focused some attention on it, revealing various elements of the nuptiality of each woman through eight questions in the questionnaire on the history of her current marriage and four questions on each previous marriage (if any). This information makes it possible to recreate the precise marital history of each woman (e.g., duration of the unions, causes of breakdowns). However, in the DHS, there is little information (e.g., date of first marriage, number of co-spouses) on the history of nuptiality.[24]

The stability of marriages, the number of marriages, and the total duration of marital life are crucial determinants of a woman's parity at a given age. All else being equal, the more stable a woman's marital life, the higher her fertility. The degree of marital mobility can also be considered as a psychosocial identity characteristic of personal life. Let us consider some examples of nuptiality indicators that can be created when past data on each union is available.

1. Marital Stability Indicators for Women

A marital stability indicator can be computed by comparing the total duration that a woman has spent in union to the total duration she would have spent in union if she had not experienced any disruption (e.g., divorce, widowhood). The numerator is obtained by summing the durations of all the unions the woman has experienced since her first marriage, and the denominator is equal to the difference between her current age and her age at first marriage:

$$MSI = \frac{D_1 + D_2 + D_3 + \dots}{x_a - x_m} = \frac{\sum_{m=1}^{n} D_m}{x_a - x_m}$$

where D_m is the duration of each union, from the first to the *n*th, x_a is the woman's current age, and x_m is her age at first marriage.

For a woman whose first union has not yet ended, the indicator is equal to 1; as soon as there is a union breakdown, it falls below 1, and the difference is greater when the duration of the union is shorter. It is advisable to limit the calculation to women whose first marriage goes back at least 5 years.

Inversely, a *marital instability indicator* can be defined by adding all of the periods spent outside marriage to the number of years since the first marriage:

$$MII = \frac{\sum_{m=1}^{n} H_m}{x_a - x_m}$$

where H_m is the series of periods of widowhood or divorce, from the first to the *n*th, and x_a and x_m are the current ages and the ages at first marriage. This indicator is the difference of the previous one at 1.

We can calculate a *relative individual marital stability indicator* by relating the indicator of a woman j of a given age *a* to the average indicator of women of the same age *a* (i.e., average of various individual indicators):

$$RIMSI_{aj} = \frac{MSI_{aj}}{MSI_{aT}}$$

where MSI_{aT} is the average indicator for all women aged *a*.

A woman has a lower marital mobility if her indicator is above 1 and a higher mobility if her indicator is below 1. The values of the indicators can also be separated into a few large groups (e.g., <0.70, 0.70 to 0.85, 0.85 to 0.95, 0.95 to 1.05, 1.05 to 1.15, >1.15), and a corresponding category can be attributed to each woman. Each one is therefore grouped into a type of relative mobility (e.g., very high, high, medium, low).

2. Male Indicators for Polygamous Societies

In monogamous societies, the same indicators that we presented for women can be used for men. However, the problem is different in polygamous societies. In these matrimonial systems that frequently combine cohabitation (more or less stable), monogamous marriage, and polygamous marriage, the marital life cycle of a man can be extremely diversified, from monogyny to bigyny, trigyny, or higher orders, with reversals after a marriage breakdown, all with periods between unions and extremely variable causes of separation and conditions according to his social status and age.

The study on male fertility in Benin (Donadjé, 1992; Donadjé and Tabutin, 1994) aimed at measuring the influence of different male marital models on the fertility of Beninese men. From the precise fertility and marital histories of each man (e.g., dates, durations, causes of breakdowns, fertility in each union) *four*

[24] In our opinion, this is an error on the part of the designers of these surveys, which has been condemned many times but not corrected in the recent phases of this program of surveys.

marital profiles combining stability and type of marriage were defined[25]:

Stable monogamy: a man having only experienced one union in his life, at least up to the time of the survey
Unstable monogamy: a man having had several spouses, but always successively
Stable polygamy: a man having taken, after his first marriage, a second and then possibly a third or a fourth wife, but without ever reverting to monogamy
Unstable polygamy: a man having gone from monogamy to polygamy, then reverting to monogamy before possibly taking another wife.

Simple individual indicators (which require appropriate data of good quality) of the same type as those established for women were created for each of these four categories of marital profiles. They made it possible to analyze the linkages between the nuptiality of Beninese men and their completed fertility after the age of 55 or their number of children per woman-year, controlling for the total duration of marriages. The monogamous men appear to be slightly more fertile (up to the age of 50) than polygamous men (0.35 children per woman-year versus 0.29),[26] whereas union stability does not have a significant effect on the fertility of the monogamous or the polygamous. The role of the intensity and rapidity of the remarriage of widowers or divorced men is very important.

V. LIMITATIONS

It is unnecessary to further emphasize the advantages of the individual indicators that make it possible, at low cost, to go beyond traditional aggregated analysis by controlling for the main confounding variables from the start, which are age or marriage duration. They make it possible to better specify the risk factors and other individual explanatory factors by taking into account the heterogeneity of the categories generally used in aggregated analyses. However, it is advisable to remember the principal limitations, which mainly come from the nature of the information used:

[25] After various other attempts at more detailed typologies that came face to face with a problem of small numbers while distinguishing groups of cohorts (or age groups) to avoid the effects of age.

[26] This is no longer the case after the age of 50, at ages when the wives of monogamous men are no longer fertile, whereas polygamous men—notably in unstable polygamy—continue to have children with their younger wives.

Selection risk. By definition, most of these indicators, like all the other demographic indicators created using data from retrospective surveys, are based on declarations made by the surviving women or men at the time of the survey. All retrospective observations (see Chapter 121) include a selection risk that is greater with higher mortality or mobility.

Gaps between explained variables and explanatory variables. All information on fertility and child mortality is by definition based in the past, but a past of varying remoteness according to the type of questions and the ages of the individuals. There is most frequently a chronologic gap between the variable to be explained, which is measured by these indicators, and the explanatory variables, which in most cases are prevalent characteristics at the time of the survey. This problem cannot be solved without complete social and economic biographies (see Chapter 24).

Quality of the information. The quality of the results depends on the quality of the data. However, we know that all retrospective questionnaires are faced with particular problems. For fertility, child mortality, and nuptiality we must rely on the memory of individuals to gather a mass of crucial information on ages, dates, and durations. Major errors come from omissions (e.g., of births and especially deaths of children), confusion (e.g., between still births and live births), or false declarations (e.g., of dates or ages). Individual indicators are highly likely to be more sensitive to these problems than aggregated measurements, which can benefit from compensation effects. This type of approach is only possible if, after rigorous verification, the data quality proves sufficiently high.

Random fluctuations. The low number of births or deaths of children observed for each woman is the cause of the considerable random variability of individual indicators. The effects can be reduced, but not avoided, by eliminating women with exceptional characteristics with regard to the phenomenon studied (e.g., very young, short duration of a stable marriage, not very fertile). In particular, this type of indicator is certainly not recommended in contexts of low fertility or low mortality.

Acknowledgments

I am grateful to Bruno Schoumaker for his comments on this text. This chapter was translated from the French version by Accenta Ltd.

References

Boulier Bryan L. and Rosenzweig Mark R., 1978. Age, biological factors and socioeconomic determinants of fertility: a new measure of cumulative fertility for use in the empirical analysis of family size, *Demography*, vol. 15(4), p. 487–497.

Brass William, 1975. *Methods for estimating fertility and mortality from limited and defective data based on seminars held 16–26 September 1971 at the Centro Latinoamericano de Demografía (CELADE), Costa Rica*. Chapel Hill (North Carolina), Laboratory for Population Statistics, University of North Carolina, 160 p. (Occasional Publications of the Laboratory for Population Statistics.)

Coale Ansley J., 1973. The demographic transition, in: IUSSP (eds.), *International Population Conference, Liège 1973. Proceedings / Congrès international de la population, Liège 1973. Actes*, vol. 1, p. 53–73. Liège, IUSSP, 494 p.

Coale Ansley J., Hill Allan G., and Trussell James, 1975. A new method of estimating standard fertility measures from incomplete data, *Population Index*, vol. 41(1), p. 203–213.

Donadjé Florentin, 1992. *Nuptialité et fécondité des hommes au Sud-Bénin: pour une approche des stratégies de reproduction au Bénin*. Louvain-la-Neuve, Université catholique de Louvain, Institut de démographie and Académia, 222 p.

Donadjé Florentin and Tabutin Dominique, 1994. Male nuptiality and fertility in Southern Benin, *in*: Thérèse Locoh and Véronique Hertrich (eds.), *The onset of fertility transition in sub-Saharan Africa*, p. 135–162. Liège, Ordina Editions and IUSSP, 308 p.

Farooq Ghazi M., 1985. The definition of fertility: measurement effects, *in*: Ghazi M. Farooq and George B. Simmons (eds.), *Fertility in developing countries: an economic perspective on research and policy issues*, p. 125–148. Basingstoke, Hampshire, UK, MacMillan, 533 p.

Ilinigumugabo Aloys, 1989. *L'espacement des naissances au Rwanda: niveaux, causes et conséquences*. Louvain-la-Neuve, CIACO Editeur and Institut de Démographie, 243 p.

Lesthaeghe Ron J., Vanderhoeft Camille, Becker Stan and Kibet Moses, 1985. Individual and contextual effects of education on proximate fertility determinants and on life-time fertility in Kenya, *in*: John B. Casterline (ed.), *The collection and analysis of community data. WFS seminar on collection and analysis of data on community and institutional factors, 20–23 June 1983*, p. 31–63. Vooburg and London, International Statistical Institute and World Fertility Survey, 286 p.

Little Roderick J. A., 1977. *Regression models for differentials in fertility*. Voorburg, ISI, 55 p. (WFS Technical Paper, 635.)

Little Roderick J. A. and Perera Soma, 1981. *Illustrative analysis: socio-economic differentials in cumulative fertility in Sri Lanka. A marriage cohort approach*. Voorburg, ISI, 77 p. (WFS, Scientific Reports, no. 12.)

Nations unies/United Nations, 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York, United Nations, Population Division, 324 p. (études démographiques, no. 81, ST/ESA/Ser.A/81.)

Noumbissi Amadou, 1996. *Méthodologies d'analyse de la mortalité des enfants. Applications au Cameroun*. Paris and Louvain-la-Neuve, L'Harmattan and Academia, 304 p.

Palloni Alberto and Heligman Larry, 1985. Re-estimation of structural parameters to obtain estimates of mortality in developing countries, *Population Bulletin of the United Nations*, vol. 18, p. 10–33.

Pullum Thomas W., 1987. Analytical methodology, *in*: John Cleland and Christopher Scott (eds.), *The world fertility survey. An assessment*, p. 644–676. Oxford, UK, Oxford University Press, 1058 p.

Schoumaker Bruno, 2001. Déterminants de la fécondité et contexte local au Maroc rural: une application des modèles multi-niveaux, *in*: Francis Gendreau (ed.), *Les transitions démographiques dans les pays du Sud*, p. 129–143. Paris, AUPELF-UREF/Estem, 505 p.

Schoumaker Bruno and Tabutin Dominique, 1999. Analyse multi-niveaux des déterminants de la fécondité. Problématique, modèles et applications au Maroc rural, *in*: UEPA-NSU (eds.), *La population africaine au 21e siècle*, vol. 1, p. 299–332. Dakar, UEPA, 630 p. (Proceedings of the Third African Population Conference, Durban.)

Trussell T. James and Hill Kenneth, 1980. Fertility and mortality estimation from the Panama retrospective demographic survey, 1976, *Population Studies*, vol. 34(3), p. 551–563.

Trussell T. James and Preston Samuel H., 1982. Estimating the covariates of childhood mortality from retrospective reports of mothers, *Health Policy and Education*, vol. 3(1).

Trussell T. James and Preston Samuel H., 1984. Estimating the covariates of childhood mortality from retrospective reports of mothers, *in*: Jacques Vallin, John Pollard, and Larry Heligman (eds.), *Methodologies for collection and analysis of mortality data*, p. 331–364. Liège, IUSSP, Ordina Éditions, 530 p.

United Nations, 1983. *Manual X. Indirect techniques for demographic estimation*. New York, United Nations, 304 p. (Demographic Studies, no. 81.)

United Nations, 1985. *Socio-economic differentials in child mortality in developing countires*. New York, United Nations, 319 p. (ST/ESA/SER.A/97.)

C H A P T E R

135

Cohort and Period Measures of Fertility Based on Microdata

ALBERTO PALLONI AND PAOLO DE SANDRE

Center for Demography and Ecology, University of Wisconsin, United States

Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova, Italy

Conventional demographic analysis of fertility is based on measures that require information on the final outcomes of the reproductive process—births or pregnancy termination—that are usually classified by age and less frequently classified by parity (see Chapters 12 and 14). In the case of total births, these measures are classic occurrence-exposure rates for repeatable events (i.e., fertility rates and age-specific fertility rates), whereas in the case of birth by parity, we normally employ occurrence-exposure for non-repeatable events (see Chapter 8). With the information data normally available to demographers, fertility analysis is almost always carried out using age patterns of fertility and summary measures of fertility, such as the total fertility rate (TFR) (see Chapter 14). Information of births by parity is used less frequently, and seldom are we in a position to evaluate patterns of fertility by age and parity, as well as synthetic measures derived from them, such as parity progression ratios (PPRs).

There are three problems associated with the use of these measures. The first is that we rarely interpret synthetic period measures such as TFR with the caution we should employ given that these measures are sensitive to the parity composition of women at any given period and, therefore, to the reproductive histories of women belonging to different cohorts. The second is that when we do focus on parity-specific measures, we seldom employ proper tools to establish a relation between them and the more commonly used age-specific rates. These relations are complex because they incorporate information on the reproductive history of women belonging to different age groups and or cohorts. We rarely are able to identify patterns of fertility by social groups within a population, and consequently, we are unable to offer even approximate causal interpretations.

All three problems can be addressed with microdata retrieved from fertility or more general-purpose household surveys that include birth histories as well as a number of characteristics for the population being studied. There are a number of ways to take advantages of the richness of information available in these studies to better understand fertility dynamic (De Sandre, 2000), to properly interpret conventional synthetic measures of fertility and, more generally, to produce an integrated set of indicators with well defined specification regarding their advantages and shortcomings.

In fact, it has been the advent of new, more detailed information on fertility processes, that prompted demographers to question the validity of conventional measures of fertility (Keilman, 1994a, 1994b, 1999; Rallu and Toulemon, 1993, 1994; Calot, 1994; Desplanques, 1994; Festy, 1994; Ni Bhrolchain, 1994; Sardon, 1994) (see Chapter 17). They argue forcefully in support of the use of modified procedures reflecting better the details and nuances of birth interval dynamics, which are only weakly if at all captured by conventional measures, such TFR (Rallu and Toulemon, 1993, 1994; Ni Bhrolchain, 1992; Feeney and Yu, 1987; Palloni, 1984; Feeney, 1983). Rethinking of fertility measures occurs on the heels, though independently, of critiques to the theoretical framework that assigns primacy of cohort over period approaches to the analysis of fertility (Ni Bhrolchain 1987, 1992; Ryder, 1964, 1965; Hajnal, 1959). The revisionist approach questions the importance of cohort effects on fertility, plays down the relevance of cohort-based approaches, and emphasizes instead the need to fine-tune and adjust period measures rather than de-emphasizing their importance on the grounds that they do not convey properly potential cohort effects. Newly proposed measuring tools can be categorized in two classes. The first are synthetic life table-based procedures that adjust for compositional effects by using full information on birth interval dynamics and are appropriate as a foundation for making broader inferences regarding the course of the fertility process. The second are extensions of life table-based procedures designed to adjust period-based measures of fertility levels (quantum) for possible changes attributable to the timing of births, not to changes in the underlying intensity of fertility.

In this chapter, we formulate initial outlines of a generalized framework for a new class of fertility indicators that include those referred to previously as a subclass and establish the relations between several subclasses. The chapter stops short of evaluating the performance of the various classes of indicators. In the second section of the chapter, we show that conventional measures of fertility are contaminated by compositional characteristics and derive synthetic measures to attenuate some of the distortions they may cause. The next section shows a general application of synthetic procedures to identify fertility profiles by subpopulations, and in the last section of this chapter, we review the effects of changes in the tempo of fertility on its quantum.

I. GENERAL FRAMEWORK

Paradoxically, the controversy about proper measurement of fertility in general and the TFR in particular (Bongaarts and Feeney, 1998; Keilman, 1994a, 1994b; 1999; Rallu and Toulemon, 1993; 1994) is not replicated with measures of mortality. Empirical research suggests that cohort effects are more likely to be present in and relevant for mortality trends than they are for fertility. It should be in this area where researchers could find plenty of room for controversy and where the selection of measures should be accompanied by a plethora of careful justification. This is far from being the case. In mortality analysis we rely exclusively on the use of life expectancy (e_o), a pure synthetic indicator. This practice is generalized despite the fact that time trends in life expectancy, for example, conceal and confound changes that affect all cohorts simultaneously (i.e., period effects) and those associated with particular birth cohorts (i.e., cohort effects) as much or more than measures of total fertility do in the case of fertility analysis. However, there is where similarities between mortality and fertility indicators end for there is an important difference that leaps into the open. Whereas we can assert that a synthetic measure of mortality, such as e_o, is not contaminated by compositional factors, it is not possible to make the same claim about TFR. This does not refer to the fact that both e_o and TFR are functions of the population composition according to factors that directly affect fertility or mortality levels (e.g., education, occupation), but to a less obvious feature. The e_o term characterizes the experience of a fictitious cohort that experiences an age pattern of mortality risks observed at time t, $[\mu(x,t)]$. TFR characterizes the experience of a fictitious cohort that experiences a sequence of observed fertility rates $[F(x,t)]$. Although the sequence $[\mu(x,t)]$ constitute the primary object of interest in mortality analysis, the sequence $[F(x,t)]$ is a result not just on fertility risks associated with women of different parities but also of the composition of women by age and parity. The latter introduces the same kind of compositional distortions that age distributions do when comparing crude mortality rates. The value of e_o is given by the following equation:

$$e_0(t) = \int_0^{\omega} \exp\left(-\int_0^{x} \mu(y,t)dy\right) dx \qquad (1)$$

where $\mu(y,t)$ is the age-specific mortality risk at exact age y in year t (experienced at time t by those born $t - y$ years before). Although one may quarrel about the plausibility of the assumption that the entire sequence of observed values $\mu(y,t)$ can ever be experienced by any of the cohorts contributing to it (e.g., due to possibly strong cohort effects), there is no question that $e_o(t)$ depends on these rates and nothing else. This is not the case of TFR.

1. Total Fertility Rate as a Function of Age-Parity–Specific Fertility Rates

We begin with the simplest of formulations according to which TFR is a weighted average of age-specific fertility rates.

$$TFR(t) = \sum_j f_j(x,t) w_{j-1}(x,t) \qquad (2)$$

where $f_j(x,t)$ is the conditional risk of a jth birth for women aged x and of parity $j - 1$ at time t, and $w_{j-1}(x,t)$ is the fraction of all women aged x at time t who are of parity j – 1. Thus, $w_{j-1}(x,t)$ is the parity distribution of women age x at time t. The quantity $f_j(x,t)$ is an occurrence-exposure rate so that $f_j(x,t)$dt, where dt is a small interval of time, is approximately equal to the conditional probability of having a jth birth among women aged x and of parity $j - 1$. The distribution $w_j(x,t)$ is not always the outcome of the set of observed values of $f_j(x,t)$ because it may depend on rates other than those we are interested in. To see this, notice that the expression for $w_j(x,t)$ is

$$w_j(x,t) = \int_0^x f_{j-1}(x-y,t-y)\exp\left(-\int_{x-y}^{x} f_j(v,t-v)dv\right)dy \quad (3)$$

where $f_{j-1}(x - y, t - y)$ is potentially different from $f_{j-1}(x - y, t)$ for all j, x and y.

Comparisons of TFR across groups or over time reflect disparities in the values of $f_j(x,t)$, the quantities of main interest to the analyst, and the differences in the distributions of women by achieved parity, a distribution that may not reflect current fertility risks at all.[1] When parity-specific fertility rates are constant, TFR will reflect correctly the prevailing fertility levels by parity. However, when fertility experiences short run shocks or even longer term fluctuations, the parity distributions at various point in time will differ as they will reflect parity-specific fertility rates experienced during the past. As a result TFR will not always be an accurate measure of the period levels of age-parity-specific fertility rates. Arguably, it is $f_j(t,x)$ that should be the main object of fertility studies, just as $\mu(x,t)$ is the main object of interest in mortality analysis. If so, TFR is no less contaminated than other demographic indicators that happen to be joint functions of rates of interest and of nuisance distributions.

The expression for TFR suggested previously also makes salient the fact that the analogy between it and crude mortality rates is of limited utility only. In fact, whereas the distribution that distorts crude mortality rates (the age distribution) is a result of past fertility and mortality rates $\mu(x)$, the distribution distorting TFR is a function solely of the past values of the age-parity–specific rates themselves.[2]

The solutions available to neutralize the aforementioned shortcoming of TFR are the same as those we resort to in the case of mortality. The first is to recalculate TFR using direct or indirect standardization (see Chapter 15). Indirect standardization could be applied when only the $w_j(x,t)$ terms but not when the $f_j(x,t)$ are known. Indirect standardization requires that we search for a standard set of age-parity–specific fertility rates and apply them to the observed and a standard parity distribution by age. Direct standardization requires knowledge of both $w_j(x,t)$ and $f_j(x,t)$, as well as a standard set of $w_j(x,t)$.

The second solution (Rallu and Toulemon, 1993, 1994; Palloni, 1984) is to create a fictitious cohort and apply to it the observed rates $f_j(x,t)$, much as we do in the construction of life tables for mortality or nuptiality. Systematic application of $f_j(x,t)$ to an arbitrary cohort of women results in the parity distribution intrinsic to the $f_j(x,t)$. As a consequence, all derived measures, including TFR, will be a function solely of these rates, not of the joint influence of $f_j(x,t)$ and of some potentially confounding observed parity distribution. As in the case of mortality, the second solution is preferable to the first and deserves closer scrutiny.

2. Description of the Life Table Procedure

The formulation proposed follows closely solutions developed by Rallu and Toulemon in a recent paper (1994) and much earlier on by Palloni (1984). Let us assume that we condition on age of entrance into union and we overlook altogether births that take place outside a union. These assumptions are not limiting because the procedure can be adapted easily to include out-of-union pregnancies. The fraction of a union-cohort who entered union at age y and who are of parity j at age x is given by the following equation:

$$W_j(x/y) = \int_0^{x-y} f_{j-1}(x-z)\exp\left(-\int_{x-z}^{v} f_j(v)dv\right)dz \quad (4)$$

We use W instead of w to emphasize the fact that this is the distribution by parity that results from the application of the observed age-parity–specific rates. From [4] one obtains the conditional (on age of entrance into union) intrinsic set of age-specific fertility rates and TFR (Palloni, 1984). To calculate the unconditional rates we integrate out the density of ages of entrance into unions. If TFR(y) is the expected TFR among those who enter a union at age y, then TFR $= \int_y$ [TFR(y) $dH(y)$], where $H(y)$ is the distribution function of age of entrance into unions. The intrinsic values of the age-specific fertility rates and of all other derived quantities (e.g., TFR, PPRs, means ages at birth for parity) are all life table quantities and therefore possess all the virtues and shortcomings of measures derived from a life table in the case of mortality (e.g.,

[1] The fact that $f_j(x,t)$ is also a function of parity-specific rates by duration and the duration distribution by parity is evaluated later in the chapter.

[2] This statement is accurate only when fertility and mortality are unrelated.

life expectancy, mean ages at death). In particular, they are synthetic measures and apply to a fictitious cohort. As in the case of mortality, the life table–based quantities for fertility reflect changes in quantity and changes in tempo of fertility. This topic is discussed later in the chapter.

3. Total Fertility Rate as a Function of Age-Duration-Parity-Specific Fertility Rates

The quantities $f_j(x,t)$ are themselves a function of the age-duration-parity–specific rates, $f_j(x,d,t)$, and the composition of the age group by parity and duration.

$$f_j(x,t) = \int_0^x f_j(x,d,t) v_{j-1}(x,d,t) dd \tag{5}$$

where $v_{j-1}(x,d,t)$ is the fraction of women aged x and achieved parity $j-1$ who have been d years in that parity. Equivalently, these are the women who experienced parity $j-1$ at age $x-d$ and have not progressed to the next parity since. As a consequence, an age-specific fertility rate and the corresponding TFR are affected by composition by duration and by parity; the fertility rate at age x is given by the following equation:

$$f(x,t) = \sum_j \int_0^x f_j(x,d,t) v_{j-1}(x,d,t)\, dd$$

and

$$TFR(t) = \int_0 \left\{ \sum_j \int_0^x f_j(x,d,t) v_{j-1}(x,d,t)\, dd \right\} \tag{6}$$

To remove the compositional effects of duration composition from the fertility measures we can resort to the standardization procedures (direct and indirect) or to the life table.

4. A Generalization of the Life Table Procedure

The procedure described earlier (Sect. 2) can be generalized in a straightforward manner to deal with composition by duration (Rallu and Toulemon, 1994; Palloni, 1984). Assume we start with an arbitrary cohort of women at some age y (exact age of union). We then recursively apply the rates $f_j(x,d)$ to generate the fraction of women at age x and duration d of parity j:

$$W_j(x,d) = \int_z \int_y f_{j-1}(x-z, r) \exp\left(-\int_{x-z} f_j(v, v-(x-z)) dv \right) \tag{7}$$

This expression looks intractable but its application amounts to nothing more complicated than building several life tables by duration for each age. If durations with nonzero risks are, say, not longer than 5 years, then for each age we will have 5 life tables. It is easy to calculate $f_j(x)$ from these life tables and, from these, all other pertinent quantities, including $W_j(x)$, $F(x)$, TFR, and PPRs.

Rallu and Toulemon (1994) apply the life table procedures described previously to the case of France and compare their performance to that of conventional fertility measures. They conclude that they are superior in terms of the accuracy with which they capture changes in fertility. In particular, they verify that the life table-based measures lead to lower values of TFR when the parity distributions are "too favorable" and the values $w_j(x,t)$ are positively correlated with the values $f_j(x,t)$ and lead to higher values of TFR when the quantities $w_j(x,t)$ and $f_j(x,t)$ are negatively correlated.

In an earlier application, Palloni (1984) used the same life table procedure to quantify the effects of intermediate variables on fertility in developing countries. As this represents an extension of the life table procedure, we will treat it separately.

II. EXTENSIONS AND APPLICATION OF THE LIFE TABLE PROCEDURE

Although the quantities $f_j(x,t)$ or, better yet, $f_j(x,d,t)$, are in some sense the elementary particles of fertility measurement, an analyst may be interested in evaluating the effects that certain conditions have on TFR or on associated measures. Equivalently, the researcher may be interested in creating scenarios by calculating what TFR, or other fertility measures, would be if the population distribution of conditions that directly impact fertility were changed in a predefined manner. A clear example of this is the evaluation of the role played by the so-called intermediate variables (Bongaarts, 1978). As shown elsewhere (Palloni, 1984) this can be achieved with an extension of the life table procedure suggested earlier.

1. Summary of the Procedure

Consider a simple case where the researcher is interested only in quantifying the effects of oral contraception, C. The observed population is consequently divided into those who practice contraception ($C = 1$) and those who do not ($C = 0$). Similarly the basic fertility rates, $f_j(x,d,t)$ attain two possible values, $f_j(x,d,t; C = 1)$ and $f_j(x,d,t; C = 0)$. Assume further that through the use of hazard models on appropriate birth interval data we establish that the ratio of the age-duration-parity–specific rates for individuals practicing contraception to the age-duration-parity–specific rates for those not practicing contraception is $\exp(\beta_j) = \zeta_j$ for parity j, with $\beta_j < 0$. Deployment of the life table procedure described before on two hypothetical cohorts, one practicing and the other not practicing contraception, will result in two sets of fertility rates $F(x; C = 1)$ and $F(x; C = 0)$ with corresponding TFR($C = 1$) and TFR($C = 0$). Similarly, the procedure can produce parity-specific measures, such as mean age and PPR for parity j. The overall effects of C on TFR can be gauged using the relative difference

$$\Delta_T = \frac{[TFR(C=1) - TFR(C=0)]}{TFR(C=0)}$$

The observed TFR is a weighted average of TFR($C = 1$) and TFR($C = 0$), and the weights are given by the observed distribution of the population according to the trait C. The total effect, Δ_T, can be decomposed further into contributing parts. In particular, it is possible to estimate the magnitude of the effects of C that operate by reducing the fertility rate at lower parities, say at parities $j < 2$. This can be achieved by first calculating the TFR that results after imposing contraception, ($C = 1$), for the first- and second-order births, say, TFR′:

$$TFR' = [P'_o + P'_o * P'_1] + \Sigma_{k=2,w}(P'_o * P'_1)\Pi_{[j=2,k]}P_j$$

where P_j is the observed PPR of order j and P_j' is the PPR of order j estimated after taking into account the effects of contraception. Finally, an estimate of the effect we seek can be calculated as the ratio $\Delta_p = [TFR' - TFR]/TFR$, and the contribution of contraception through the first two birth orders is the fraction Δ_p/Δ_T.[3]

The decomposition technique can be extended further. For example, we could investigate the effects of contraception on the age pattern of fertility. This is easily done computing age-specific fertility rates from life tables for synthetic cohorts that practice and do not practice contraception and then contrasting the two.

2. Estimation of Effects

The key quantity in the foregoing procedure is ζ_j. To estimate this parameter we need information on the birth interval dynamics of a population where some members practice contraception and others do not. A typical source are birth histories and recollections of associated behavior, including abortion, contraception, breastfeeding, and participation in unions. To retrieve an estimate of β we can choose between alternative models. In what follows, we describe the use of multistate hazard models, although other models could yield similar estimates. The idea is to represent parities 1, 2, 3 . . . as states that women attain at various points in their childbearing history. Transitions can only occur among successive states and only in one direction. Whether women experience a transition from j to $j + 1$ and, if they do, the timing (duration) of such transition depend on a number of factors, including the age at which she attained state j, behaviors adopted during the period of time during which she waits in state j, external conditions that take place during her sojourn in j, and some of her more or less permanent characteristics, such as family upbringing, education, age at first union, and even perhaps some community characteristics. The most general model for the conditional risk of passing from parity j to parity $j + 1$ at duration d for the ith woman has the following form:

$$f_{i(j)}(d; x, \mathbf{Z(t)}) = \mu_o(d) * \exp(\alpha * X + \zeta\,\mathbf{Z(t)}) * \exp(\varepsilon_i)$$

with probability $\Gamma_{i(j)}(\mathbf{W})$ and

$$f_{i(j)}(d; x, \mathbf{Z(t)}) = 0 \qquad (8)$$

with probability ($1 - \Gamma_{i(j)}(\mathbf{W})$, where d is duration in state (parity) j for woman i, x is her current age, $\mathbf{Z(t)}$ is a vector of covariates some of which are fixed (time invariant) and others are time dependent, such as contraception and breastfeeding behavior, ζ is a vector of effects, α is the effect of age, and ε_i is the effect of a fixed (time invariant) individual unmeasured condition characteristic of woman i. Although some women may experience a transition from state j to state $j + 1$, others may never do so. An estimate of this probability in the population is an estimate of the PPR of order j and we have denoted it as Γ_j ($\mathbf{W}$). This quantity can be function of individual or contextual covariates included in the vector $\mathbf{W}$. The population based PPR is the average of the individual probabilities of experiencing the transition or the average of Γ_j ($\mathbf{W}$). The parameters of this model can be estimated by maximum likelihood using software such as CTM (Yi *et al.*, 1987).

3. Application of Estimates

After the parameters of the model are estimated, we can deploy the life table procedure described previously and calculate the "predicted" parity distributions and occurrence exposure rates for any desired population profile. A population profile is simply a

[3] Expressions for the standard errors of each of these estimates can be obtained by straightforward application of the delta method. An analyst can test whether measured changes or contributions are of any statistical significance.

subset of the population with given values of the vectors of characteristics Z and W. For example, if Z and W each include only one fixed dichotomous covariate, we will have four possible population profiles, a two by two table. Application of the life table procedures described earlier using predicted occurrence-exposure rates for each of the four profiles will result in four different values for the quantities of interest, one for each cell of the two by two table, including TFR, mean ages at parity j and PPRs. The overall population values associated with these quantities are simply their weighted averages, where the weights are given by the proportional distribution of the population in the two by two table.[4]

In theory, at least, we can generate fertility estimates for any desired population profile we wish to study. It is certainly possible to assess the relevance of counterfactual scenarios involving, for example, no practice of contraception, significant delays in the age of entry into first union, absence of breastfeeding at higher parities, elimination of abortion and so forth.

4. Shortcomings

At first glance the suggested procedure may appear exceedingly complicated due to the large number of calculations involved. However, with the speed and capacity of modern desktops, estimation of life tables for various subgroups of the population cannot take more than a few minutes of actual computer time. The real difficulty is rooted in the decision making process in the hands of the investigator to determine the relevant subpopulations and scenarios for which calculation must be undertaken. These should be chosen so that they correspond to theoretically relevant constructs and are consistent with substantive questions and conjectures.

A second shortcoming is that we cannot ever be certain that a particular scenario is internally consistent. Imagine for example that the investigator poses a scenario where contraception is practiced for timing purposes in the first two parities but not thereafter. There is nothing intrinsically difficult in the life table calculations for such a scenario. The main difficulty is that calculations rely on two unverifiable assumptions. First, that the dynamic of birth intervals other than the first two, will be the same as those we observe in the current population even in the counterfactual scenario. Second, that characteristics with some influence on the birthing process will remain unchanged even after changing the behavior in the first two birth intervals. The first assumption pertains to interrelations between birth intervals. It is tantamount to believing that the estimates from the data capture thoroughly all relevant interrelations. The second assumption has to do with direction of causality, an issue neglected in the model formulated previously. It is tantamount to assuming that relevant individual characteristics are exogenous to the birthing process of higher order.

These assumptions may not be entirely realistic, but a similar level of unreality pervades the application of the life table procedure itself. The implicit assumption we make in constructing the life table is that the observed $f_j(d,x,t)$ can prevail in an hypothetical population even though we have no way of verifying this. It should be remembered that these rates are cohort-specific rates and could potentially be a function of the past history of some cohorts. Except for the trivial case when fertility is constant, cohorts will have different histories and we will have no way to state unequivocally that the observed values $f_j(d,x,t)$ and $f_j(d',x + r,t)$, for example, can actually be experienced by members of a single cohort.

III. ESTIMATING THE TEMPO AND QUANTUM DIMENSIONS OF PARITY-SPECIFIC FERTILITY

1. An Informal Introduction to the Problem

Previous sections of the chapter reviewed problems with conventional measures of fertility caused by the fact that these measures, as other crude demographic indicators, are the joint outcome of both rates and distributional characteristics. As in the case of mortality, the life table approach is an attractive solution for it attenuates the impact of distributional differences and permits to infer the contribution of well-identified determinants that are unequally distributed in the population.

There is, however, another issue that requires some attention. This is that in many situations it will be important to parse the effects of changes in the timing of events and of their intensity. This is a classic issue in fertility analysis identified as the problem of disentangling changes in *tempo and quantum* of fertility (1999; Keilman, 1994a, 1994b; Bongaarts and Feeney, 1998; Ryder, 1964). Although less popular or less commonly studied, a completely analogous problem exists in mortality analysis as well. In the case of fertility, the

[4] Throughout, we assumed that fertility risks are conditional on age of entrance into union. To calculate overall measures of fertility, such as TFR, we must integrate the density of waiting time to first entry into union. For example, TFR = $\int$TFR(y) dH(y), where H(y) is the distribution function of first age at union.

issue can be described as follows: Even when the overall fertility level experienced by women in various groups in a population may be identical, the timing of their births could be different. For example, although the total number of children born to women belonging to different cohorts may be the same–the intensity of fertility is constant–the youngest cohorts may be having children earlier (anticipating childbearing) relative to the older cohorts. If this phenomenon takes place over a period of time, conventional measures of levels of fertility or those associated with the life table approach, will reveal changes even though all women, regardless of cohorts, will eventually have the same number of children. The condition for this to happen is that there be a relation between fertility rates to ensure that if women of a particular cohort anticipate childbearing at some point during their reproductive lives, there will be an offsetting reduction later on.

Characteristics of a particular period, for example, may impel anticipation (postponement) of births without, however, altering the lifetime value of fertility. The change, however, will increase (reduce) measures of fertility for that particular period and could incorrectly lead to the impression that lifetime fertility levels are changing, when what is taking place is simply a change in the timing of births. Confusion between changes in tempo and quantum has been normally associated with conventional measures of fertility such as TFR, but they also apply to measures derived from life table procedures.

A similar phenomenon can occur in mortality. Suppose for example, that some cohorts experience unusually high mortality at young ages and that, as a consequence, these cohorts will benefit from unusually mild mortality rates when they reach older ages. Under some regular conditions constraining the relation between the force of mortality at various ages, it could happen that, even though the overall life expectancy by cohorts remains unchanged, the life expectancy for several periods of time will experience changes. However, because death is inevitable, the mortality quantum is always equal to 1, whereas the fertility quantum by birth order is usually less than 1 and can vary over time. Only tempo effects influence mortality; in the case of fertility, both the tempo and the quantum may change.

2. Formalization of Tempo and Quantum Changes in Overall Fertility Rates

Assume that age patterns of fertility differ only in their first moment so that we can speak unambiguously of a younger or older fertility pattern. A regime of fertility anticipation will occur when the population of interest shifts from an older to a younger fertility pattern without changes in the level of overall fertility. To formalize this idea we first write a age-specific fertility rate for time t as the product of a factor controlling the tempo (age pattern) and another controlling the quantum (TFR) or

$$f(y,t-d) = Q\,(t-d)\lambda(y,\,t-d)$$

where $f(y,\,t-d)$ is the age-specific fertility rate at age x (y) for women born in year $(t-d)$, $\lambda(y,\,t-d)$ is an age pattern of fertility for the same cohort, $Q(t-d)$ is the cohort TFR, and $\int \lambda(y,t-d)dy = 1$. Changes in the tempo of fertility imply only that while all cohorts experience $Q(t-d) = Q$ (for all d), the age pattern changes so that the cohort aged z at time t experiences $f(y,\,t-z) = Q\lambda'(y,\,t-z)$, subject to the restriction $\int \lambda'(y,t-d)dy = 1$. We now approximate the age pattern of fertility appropriate for a cohort born d years ago as

$$\lambda(x,t-d) \approx \lambda(x) + d\left.\frac{\partial\lambda(y,t-d)}{\partial d}\right|_{y=x} \tag{9}$$

This is a representation closely related to one suggested by Keilman (1994a,b; 1999). Two restrictions must be noted. First, the function $\lambda(x)$ can be thought of as a standard fertility schedule and must be a suitable representation of fertility in the population being studied. Second, because we require that $\int \lambda(y,t-d)dx = 1$ for any all d, it must be the case that $\int (\partial\lambda(y,t-d)/\partial d)dx = 0$ for all d. It can now be shown that the TFR at time t, $Q_{per}(t)$ is given by

$$\begin{aligned} Q_{per}(t) &= \int_x f(x,t-x) \\ &= Q\left(\int_x \lambda(x)dx + \int_x x\frac{\partial\lambda(x,t-x)}{\partial x}dx\right) \end{aligned} \tag{10}$$

or equivalently by

$$Q_{per}(t) = Q(1+\mu') \tag{11}$$

where μ′ is the annual change in the mean of the cohorts' fertility schedule, such as $\mu' = \partial M_{coh}(t-d)/\partial d$, and $M_{coh}(t-d)$ is the mean age of the childbearing schedule for the cohort born d years before t. The latter quantity can be approximated from the observed annual change in the mean age of the fertility schedules over a short period before t, such as between t and $t-h$:

$$\mu' \approx (1/h)(M_{per}(t) - M_{per}(t-h))$$

where $M_{per}(t)$ and $M_{per}(t-h)$ are the period mean ages of the fertility schedules at times t and $t-h$.

For all purposes, the function defined as the partial derivative of the cohort fertility schedule is nothing but an age pattern of fertility changes, such as $\phi(x)$. These are well known in demographic analysis dealing with the nature of fertility behavior and can be inferred by assuming regular and known patterns of fertility. However, not all of them satisfy the restriction. For example, if we assume that fertility behavior can be represented well by a Coale-Trussell function in which the only change is in the level of birth control, the function $\phi(x)$ has a very simple form that is proportional to the time derivative of the birth control parameter. In this case, however, the sum of all age-specific changes does not add up to 0. This is because when only the birth control parameter changes, there will be also changes in the cohorts' completed fertility unless offsetting changes in other parameters of the fertility pattern take place.

3. Formalization of Tempo and Quantum Changes in Parity-Specific Fertility Rates

Let $f_j(y, t-x)$ be the occurrence-exposure rate of order jth at age y for a cohort born $t-x$ years ago so that the corresponding cohort-specific PPR is $R_{j(coh)}(t-x) = 1 - \exp\left(-\int f_j(y, t-x)dy\right)$. Cohort-specific occurrence-exposure rates can also be represented as the product of a standard pattern of occurrence exposure rates and a level parameter, $f_j(y, t-x) = \lambda_i(y, t-x)C_j(t-x)\mathrm{f_j}$. To adhere to the framework developed before, assume also that $C_j(t-x) = C_j$; that is, cohorts only experience changes in the age pattern of occurrence-exposure rates, not in their overall levels. The cohort of age z at time t experiences the rates $f_j(y, t-z) = C_j\lambda_i'(y, t-x)$.

Because the mean age of the jth-order rates is only a function of the age pattern, not of the level parameter, a reasonable way to represent changes in the tempo of the jth birth is through changes in the mean age of occurrence-exposure rates. In particular, we assume that

$$\lambda_j(x, t-d) \approx \lambda_j(x) + d\left.\frac{\partial\lambda_j(y, t-d)}{\partial d}\right|_{y=x} \tag{12}$$

subject to the constraint $\int(\partial\lambda_j(y, t-d)/\partial d)dy = 0$. In what follows we let $\phi_j(y) = \partial\lambda_j(y, t-d)/\partial d$, an age pattern of changes in the parity-specific occurrence-exposure rates. Under these conditions, we can again derive an expression for the period PPR as a function of the cohort PPR and a measure of the rate of changes in the age patterns of the occurrence-exposure rates. In particular, the PPR of order j at time t, $\mathrm{R}_{j(\mathrm{per})}(t)$, will be equal to

$$R_{j(\mathrm{per})}(t) = 1 - (1 - R_j)^{(1+\eta j)}$$

where $\mathrm{R_j}$ is the parity progression rate for the standard occurrence exposure rates of order j, and η_j is the annual change in the mean age of the jth-order cohort-specific occurrence exposure rates [defined as the product $(\int\lambda_j(x)dx)^{-1}$ $(\int x\phi_j(x)dx)$]. This expression shows how changes in cohort measures (changes in the timing of births) translate into period measures of the quantum of fertility. It closely resembles the one inferred before to link period measures of fertility and changes in the timing of overall fertility.

From this expression we can derive a relation between cohort-specific changes in the timing of births of any order and period TFRs. Using the life table approach proposed before we calculate the period TFR, $Q_{per}(t)$, as

$$Q_{per}(t) = \Sigma_{\{k=o,w\}}[\Pi_{\{j=o,k\}}p_j]$$

where $\mathrm{p_j}$ is the synthetic (life table) parity progression rate of order j. We now use the expression relating the synthetic PPRs of order, say R, to their cohort equivalent and show how the timing in births of order R influences period TFR.

As an illustration, assume that the timing of first birth for all cohorts experiences a change following the pattern assumed before. It must then be the case that $Q_{per}(t)$, the period TFR at time t, is equal to the cohort TFR times the ratio of the cohort PPR of order 0 to the period 1:

$$Q_{per}(t) = Q(1 - (1 - R_0)^{(1-\eta_0)})/R_0 \tag{13}$$

where η_o is the annual rate of change in the mean age of the first birth occurrence-exposure rates and R_o is the parity progression rate associated with the standard first-order occurrence-exposure rates. The most important implication from this expression is that to adjust the period TFR for changes in the timing of first birth it is not sufficient to know what η_o is. We must also account for the (scaling) effects of the standard first-order cohort PPR R_o.

4. Three Generalizations

The foregoing relations are just a small sample of a much larger set of relations linking changes in quantum and tempo of fertility of various orders. First, it is possible to investigate the effects of changes in the intensity (quantum) of fertility on measures of fertility tempo (mean and variance of fertility schedule) using overall and parity-specific fertility rates. This is the inverse problem treated before, albeit a less interesting one; when intensity changes take place without affecting the timing of births. Second, the assumptions made at the outset about gradual changes across cohorts though convenient and tractable are not very realistic because they rely on the idea that gradual fertility across cohorts are tenable. As indicated at the outset there is no compelling reason to believe that any consequential fertility change is purely cohort based and that, on the contrary, there is evidence suggesting that period effects are of equal or even more import. Third, empirical illustrations based on historical cases and on simulation help us understand the accuracy of the relations suggested and to appreciate the set of conditions, some of which are more relevant than others.

These three generalizations are taken up elsewhere (Palloni, 1984). There we develop a more extensive set of linkages under more realistic assumptions about the nature of fertility changes, including cases with period changes, such as anticipation and postponement of childbearing, that affect all cohorts simultaneously. Empirical evaluation of relations through simulated data reveals that there are simple ways of inferring the nature of these changes from a very small set of observables.

CONCLUSIONS

This chapter identifies a seldom mentioned problem associated with conventional fertility measures and proposes a general framework for attenuating its confounding effects. The gist of the problem is that conventional measures of fertility may be contaminated by the parity composition of the female population. To circumvent the problem we propose, as have done before us Rallu and Toulemon (1993; 1994) and Palloni (1984), measures based on a life table approach. This solution can be made as complex and fine-tuned as available data allow. The life table procedure can shed light on the nature of fertility changes and leads to precise decomposition of changes by parity and by determinants promoting those changes.

Life table-based and conventional measures of fertility can lead to misleading interpretations that confound changes in the timing of fertility across cohorts and changes in the intensity of fertility rates. Following Keilman (1994a, 1994b, 1999), we adopt a simple model of fertility changes and show that there are simple relations between period-based life table measures of fertility and their cohort counterparts.

The methods described in this chapter require very good information on the evolution of birth intervals. These data can be obtained from a longitudinal prospective survey following women over time or from retrospective data on fertility histories. The latter are readily available but are affected by recall lapses concerning the dating of events. Longitudinal prospective surveys lead to better information on the number and dating of events, but they are very costly to carry out and are therefore rare. From a practical viewpoint, the methods developed here do not always yield reliable results. In some cases, for example, due to inadequate data, the biases may be higher than in the case of traditional period measures. One could then choose the simplest method, as the procedures proposed in this chapter require intensive computation and the use of software that is not widespread. The choice depends on the case in point.

Our review is limited in that we only focus on effects of changes in tempo on measures of quantum and not the reverse, rely on very simplistic representation of a regime of cohort-based fertility change and, finally, we provide no numerical illustration of the robustness of relations and estimates proposed.

References

Bongaarts John, 1978. A framework for analyzing the proximate determinants of fertility, *Population and Development Review*, vol. 4(1), p. 105–132.

Bongaarts John and Feeney Griffith, 1998. On the quantum and tempo of fertility, *Population and Development Review*, vol. 24(2), p. 271–291.

Calot Gérard, 1994. Synthetic measures based on rates or on probabilities, *Population, An English Selection*, vol. 6, p. 95–107.

De Sandre Paolo, 2000. Patterns of fertility in Italy and factors of its decline, *Genus*, vol. 56(1–2), p. 19–54.

Desplanques Guy, 1994. Principal results of INSEE's Family Survey, *Population, An English Selection*, vol. 6, p. 109–111.

Feeney Griffith, 1983. Population dynamics based on birth intervals and parity progression, *Population Studies*, vol. 37(1), p. 75–89.

Feeney Griffith and Yu Jingyuan, 1987. Period parity progression measures of fertility in China, *Population Studies*, vol. 41(1), p. 77–102.

Festy Patrick, 1994. Does adding more variables necessarily improve analyses? *Population, An English Selection*, vol. 6, p. 120–122.

Hajnal John, 1959. The study of fertility and reproduction: a survey of 30 years, *in*: Milbank Memorial Fund (eds.), *Thirty years of research in human fertility: retrospect and prospects*, p. 11–37. New York, Milbank Memorial Fund, 158 p. (Papers presented at the Annual Conference of the Milbank Memorial Fund, New York October 22–23, 1958.)

Keilman Nico, 1994. "Period fertility indices": a comment, *Population, An English Selection*, vol. 6, p. 111–115.

Keilman Nico, 1994. Translation formulae for non-repeatable events, *Population Studies*, vol. 48(2), p. 341–358.

Ní Bhrolcháin Maíre, 1987. Period parity progression ratios and birth intervals in England and Wales, 1941–1971, *Population Studies*, vol. 41(1), p. 103–125.

Ní Bhrolcháin Maíre, 1992. A critique of the cohort approach to fertility, *Population and Development Review*, vol. 18(4), p. 599–630.

Ní Bhrolcháin Maíre, 1994. Past history, synthetic indicators and period fertility, *Population, An English Selection*, vol. 6, p. 116–120.

PALLONI Alberto, 1984. Assessing the effects of intermediate variables on birth interval-specific measures of fertility, *Population Index*, vol. 50(4), p. 623–657.

RALLU Jean-Louis and TOULEMON Laurent, 1993. La mesure de la fécondité transversale. I. Construction des différents indices, *Population*, vol. 48(1), p. 7–26.

RALLU Jean-Louis and TOULEMON Laurent, 1994. Period fertility measures. The construction of different indices and their application to France, 1946–89, *Population, An English Selection*, vol. 6, p. 59–93.

RYDER Norman B., 1963. The translation model of demographic change, *in*: Milbank Memorial Fund (eds.), *Emerging techniques in population research*, p. 65–81. New York, Milbank Memorial Fund, 306 p.

RYDER Norman B., 1964. The process of demographic translation, *Demography*, vol. 1(1), p. 74–82.

SARDON Jean-Paul, 1994. The virtues of the synthetic index, *Population, An English Selection*, vol. 6, p. 107–109.

YI Kei Mu, HONORÉ Bo, WALKER James, and YATES George, 1987. *CTM: A program for the estimation and testing of continuous-time multi-state multi-spell models*. Chicago, University of Chicago, ERN/NORC (Economic Research Center/National Opinion Research Center), 40 p. (Preliminary Draft.)

C H A P T E R

136

The Contribution of Qualitative Methods to Demography

MICHEL BOZON

Institut national d'études démographiques (INED), Paris, France

The association of qualitative and quantitative methods in the study of population has not always been considered a delicate match. In British social and cultural anthropology and French ethnology up to the mid-20th century, it was customary to start every population case study by its outlining structure and, consequently, by a demographic description (Kertzer and Fricke, 1997, p. 3–4). Marcel Mauss, in his famous *Manual of Ethnography*, written in the 1930s, described the framework of an ethnographic case study: "A social morphology study should be completed by a proper statistical survey in demography. We should obtain trend statistics on each house, family, clan, tribe (when the society has many tribes). We can measure fertility, births by sex, morbidity, and mortality for which we should carefully distinguish between mortality by accident or violent causes and natural mortality" (Mauss, 1947, p. 20). Thereafter, anthropologists' interest in demography declined as the demographers developed and fine-tuned its statistical techniques.

Demography developed its autonomy then as it does in the current period, through the availability and the production of large quantities of numerical data from civil registration, censuses or registers, and single or multiround demographic surveys. Demographers invented increasingly sophisticated statistical methods that were at first meant to make sense of the figures.[1] At the same time, a group of demographers have always expressed an interest in methods belonging to anthropology or qualitative sociology, or shown a tendency to engage in dialogue with representatives of these disciplines (Courgeau, 1999; Basu and Aaby, 1998; Caldwell *et al*., 1988; Gérard and Loriaux, 1988; Roussel and Bourguignon, 1976, 1979).

However, not all demographers have an open mind about nonquantitative approaches, and interest in cooperation with demographers is not viewed with benevolence by all social science researchers, especially by the most qualitatively oriented among them (Kertzer and Fricke, 1997). According to some authors, there is a radical paradigmatic heterogeneity between research based on quantification (to which demography belongs) and that based on qualitative methods. Maryse Marpsat (1999, p. 13) sums up this opposition: "Quantitative methods are supposed to be attached to a strictly positivist or empirical vision inspired by natural sciences, which aim at testing hypotheses by showing the correlation between variables, while qualitative methods, by the very importance accorded to the meaning that actors themselves give to their own actions, are supposed to belong to other intellectual traditions (often cited is Weber's comprehensive sociology, phenomenology to which we can attach ethnomethodology and symbolic interactionism)." With this perspective, the opposition between the science of facts and that of interpretations becomes indomitable.

Other researchers believe that there are two different methods, but that they can be combined and articulated within complex research projects and

[1] Allan Hill (1997, p. 224) refers to such a phenomenon in demography as "the corrosive effect of too many numbers."

multidisciplinary teams according to the issues for analysis. I place myself unambiguously in the second stream, in the belief that to approach demographic issues and advance the discipline, it is possible and desirable to associate the two types of methods and to develop a comprehensive approach. Consequently, I would think that demography should be less defined by its sources or methods of analysis than by its object, which is the analysis of the reproduction of populations, seen from a broad perspective (Townsend, 1997; Greenhalgh, 1995).

The contribution of qualitative methods will first be illustrated here by examples of the diversity of the types of demographic research in which it can be found. The most frequently used qualitative techniques will then be discussed. The objectives assigned to these methods and the various roles that they play will be enumerated in detail. The effect of comprehensive approaches on demography itself will be envisaged. In certain cases, they only modify the presentation of results. In other cases, they have theoretical effects, invoking new quantitative methods, new interpretations, and even new issues.

It is hardly surprising that a large part of the examples are borrowed from research conducted in developing countries, where statistics are often poor and where it is often necessary to look for complementary data (Quesnel, 1988; Pison, 1980). On the other hand, we shall deliberately put aside historical demography, even though research associated with oral history (Leboutte, 1991) or historical anthropology (Segalen, 1985; Flandrin, 1976) can be considered as qualitative research in historical demography.

I. SAMPLES OF COMPREHENSIVE DEMOGRAPHY

In demographic research recourse to nonquantitative methods is often presented as an effort to widen the explanatory base of a phenomenon. This approach can be qualified as "comprehensive demography" (Bozon, 1992a). Research in this category does not ignore quantitative methods, it can be said that it attempts to introduce a continuous to-and-fro critique, a productive scientific tension between the determination of the area of validity of a phenomenon and the comprehension of its meaning.

1. Age Difference between Spouses and Social Age

In a study of the male-female age difference among (heterosexual) couples (Bozon, 1991), I used data collected from a survey on couple formation in France (Bozon and Héran, 1989). In this study, the questionnaire administered to 3000 respondents, who were aged below 45 and living as a couple, was part of a larger project, including 30 preliminary semi-structured interviews whose systematic analysis helped to design the questionnaire. After the main survey, a second series of 75 interviews were conducted among selected respondents of the questionnaire. The questionnaire itself included 40 open questions out of a total of 250. The phenomenon of spousal age difference is not considered a universal determinant that would be relatively invariant in marriage and that consequently would not require any explanation. The questionnaire survey showed that young women, more than men of the same age adhere to the model of couples in which the man is older. The analysis of the semi-structured interviews highlighted that women's idealization of men considered to be *mature* is accompanied by a clear rejection of young people of their own age group, considered to be immature and too young in a pejorative sense. "Available men are perceived, classified, and judged on the basis of their social age, which depends less on physical appearance than on social status and autonomy" (Bozon, 1991, p. 138). A man's social maturity at the time of the first meeting is linked by women to his experience, his professional status, knowledge, and residential autonomy. A return to quantitative data shows that this idealization of qualities and social assets associated with a man's maturity is higher among women who themselves have few educational or professional assets and are more dependent on the social status of a man.

2. Irregularities and Anomalies Are Theoretically Interesting

Demography may deliberately focus on irregularities. For example, in a study of an African population with very high fertility (i.e., rural Gambia), Caroline Bledsoe, Fatoumatta Banja, and Allan Hill (1998) started off with an ironical finding: women who had experienced reproductive mishaps (abortions, stillbirths, death of a child at an early age) tended to adopt modern Western contraception more frequently than others. This unexpected behavior according to Western thinking about high fertility patterns is interpreted by the authors as an effect of a local representation of the reproductive life of women. The study consisted of an initial survey to identify women of childbearing age in 40 rural villages (a total of 2,980 women), followed by a series of open interviews and field observation as well as a monthly survey over 15 months on a subsample of 270 women who had had a pregnancy in the last 3 years. This survey had both quantitative and open-ended questions. The anthropologic survey highlighted the fact that a woman's childbearing life is perceived as a *body resource expenditure*. The reproductive potential of a women is limited by her total endowment in fetuses (a number fixed by God, intangible for a woman, but variable from one woman to another) and this diminishes progressively, not according to age but according to the ordeals suffered by the body: irreversible loss of "muscles" associated with pregnancies and childbirth (including stillbirths), as well as loss of strength, greater during obstetrical trauma, but which can be partially restored by rest between pregnancies. Fertility is considered as a "potential to be expended" rather than a "time-bound capacity" (Bledsoe *et al*., 1998).

There is no idea of an individual biologic time, which would flow in a continuous manner, or of menopause as an upper age boundary of reproduction. Individuals refer to a reproductive age that is a discontinuous reality illustrated by different names borne by the woman at different phases of the expending of her potential.[2] In this system, it is not surprising that the most significant contraceptive initiatives take place after considerable aggression against the body, which, despite its harshness, does not entitle the woman to the sexual rest of postpartum abstinence. In certain cases, contraception is also a strategy that enables women to save their reproductive endowment while waiting for a possible remarriage. The study of one of the theory's anomalies reveals a whole hidden world of representations of the body and of fertility, showing that recourse to a practice considered as modern (Western contraception) may be perfectly adapted to a traditional rationality.

3. Demography without Numbers

Nancy Sheper-Hughes' research on infant mortality in a *favela* of the State of Pernambuco in the Nordeste region of Brazil may seem further from the traditional perspectives of demography (Scheper-Hughes, 1997). She qualifies it herself as "demography without numbers." It starts traditionally. The author studies the death records of children younger than 5 years at the civil registry office of a medium size town. One-third of the effective deaths are not reported (in most cases, the birth was unrecorded too). Among the deaths recorded, three-fourths do not indicate cause of death; the declared causes are characteristic: "heart stopped, respiration stopped," "hunger," "dehydration," "weakness," "fall," and so on. The low official presence of these demographic events encouraged the author to attempt to reconstitute the social and cultural reality of the death of babies and children through participant observation in the *favela* and an oral survey among all those who are involved with these deaths (e.g., mothers, neighbors, fathers, brothers, sisters, healers, priests, gravediggers). In this extremely precarious environment of sugar-cane cutters, mothers showed a certain fatalistic attitude toward their babies. This is displayed by a *laissez-faire* attitude and calm resignation in the face of destiny. It is current belief that certain children from the very beginning have no wish to live nor to fight, and consequently, it is not necessary to make great efforts to keep alive those who are destined to become "angels." An implicit selection takes place in the care given to children. Mothers' love takes root only very progressively when there is certainty that the child is no longer a temporary visitor to earth. This remarkable field observation helps to highlight the subjective internalization process of living conditions and social constructions of maternal attitudes on which history and historical demography, with notably research by Philippe Ariès on the emergence of feelings on childhood in the classic period (Ariès, 1962), had already cleared a path.

Research on the homeless carried out at the Institut National d'Études Démographiques (INED) by Maryse Marpsat and Jean-Marie Firdion (2000) is a good example of the systematic and reasoned injection of qualitative methods in a mainly statistical data collection. The principal survey is based on a questionnaire administered on a representative sample of users of shelters and food distribution centers during a winter month in Paris. The questionnaires were preceded by street surveys at night and preliminary interviews among homeless persons, followed by an interview survey among managers of the shelters, homeless persons as well as more ethnographic observations among persons who obtain their livelihood from begging. Exploratory interviews made it possible to identify "the recurrence of the trajectories of single men who had worked in itinerant professions (e.g., long distance drivers, sailors, the military)" (Marpsat, 1999), which led to the inclusion in the questionnaire of a specific question on this issue. In open-ended questions on housing, the respondents gave different meanings to this term according to their experiences and social situations.[3] It became evident that ques-

[2] A young woman without a child or a with a single child (*janka* in Wolof) is different from a young spouse or a woman at the beginning of her reproductive life (*jongoma*), a young married woman, or a woman in the middle of her reproductive career (*jegg*). (Hill, 1997, p. 238).

[3] For single men living in a shelter, a hotel room is already considered to be a home, whereas a woman with a child or children will consider home only as rented accommodation without time limitation (Marpsat, 1999).

tioning without further detail on respondents' last home tended to confirm the implicit definition of each individual. The interviews were conducted after the questionnaire so as to specify and qualify certain results: Although many men reported in the questionnaire that they had lost their homes after a family break-up, the biographical interviews showed that this was a respondent's interpretation within a more complex process. In the end, many difficulties encountered while administering the questionnaires in the shelters encouraged the researchers to embark on qualitative research on the shelters. This research in the field brought to light the existence of a hierarchic system of shelters in homology with a hierarchy of users (Soulié, 2000), which was confirmed by the results of the questionnaire.

4. Micro-demography, Community Studies, and Local Continuous Demographic Observation

In the use of qualitative methods in demography, a particular mention must be made of micro-demography as illustrated by the Australian demographer Jack Caldwell, in his own work in Africa and India and through those of the Committee on Anthropological Demography,[4] set up by the International Union for the Scientific Study of Population (IUSSP) (Caldwell *et al.*, 1988). These micro-demographic studies are community studies that, in a geographic area or a relatively limited group, associate the methods of demographic or epidemiologic surveys with various anthropologic techniques, often within the context of small, pluridisciplinary teams. *Pluridisciplinarity* refers to the collaboration of demographers and representatives of other disciplines rather than the conversion of demographers' to qualitative research. Micro-demographic studies require a generally prolonged stay by demographers in the field, providing the opportunity to set up real local demographic surveillance systems and conduct a number of in-depth investigations. Micro-demography is different from anthropology in its strong interest in change and innovation, observed at the community level that is also that of many population programs or projects.

The initial aim of this collaboration with other disciplines was to enrich demographic data collection and the interpretation of results. Gradually, these community studies have freed themselves of their initial objectives, taking on themes that were traditionally only considered in analyses as simple determinants of demographic behavior (particularly of fertility). In a study published in 1996 in *Population Studies* (Niraula and Morgan, 1996) comparing two Nepalese villages, the authors chose to look at women's personal autonomy through examination of its ties with marriage formation and with postmatrimonial contacts between women and their families of origin. The two villages (one in the hills of Katmandu and the other near the Indian border) are very different in terms of their social organization and general degree of female autonomy. The comparative double case study showed that individual differences between the biographic characteristics of the women in each environment were secondary in explaining the overall variation between localities in terms of specific contextual differences between the two villages. Even though a woman's personal autonomy is measurable at an individual level, it is above all the result of social organization and collective control and these must be observed in a holistic approach rather than through individual surveys.

The use of qualitative methods in demography occurs generally within complex and composite research strategies. A thorough examination of the major techniques used shows that each of them is adapted to particular issues, corresponds to a type of fieldwork, and enables specific types of analyses to be carried out.

[4] Created and initially chaired by Jack Caldwell.

II. QUALITATIVE OBSERVATION TECHNIQUES

It is possible to rank the various qualitative observation techniques used in demographic studies according to their degree of dependence on the statistical surveys with which they are associated.

1. Open-ended Questions and Nonstandardized Items in Demographic Questionnaires

An apparently simple practice to import into a demographic questionnaire is the use of the *open question* for which, contrary to closed pre-coded questions or questions that record quantities (number of children, date of entry into union), the content, form, or length of response is not fixed in advance. Recourse to open-ended questions is particularly relevant to investigate the representations associated with a term (e.g., the concept of environment, the word sexuality, the concept of bad health or that of a real home), or the justifications spontaneously given for a behavior or an attitude (for example, why one would refuse to marry a younger spouse). Semi-open questions in which a non pre-defined "other" category, is added to a series of items can also be included. In the procedures of questionnaire enrichment, we can also mention the systematic recording of observations on the interview itself: respondent's reaction to the questions (e.g., discomfort, hesitation, surprise), additional comments, possible presence, or reaction of a third party that may affect the answer. It is easy to introduce nonstandardized elements into the survey questionnaire, even if this requires further training of interviewers. It is more difficult to analyze them adequately. The remarks and comments on the interview could be simply used to remove ambiguities or resolve problems of consistency during data capture (by correcting the interviewer's pre-codification). In the same way, the statistical analysis of open-ended questions can take a simplified form, which through coding transforms them into closed questions with only a few modalities. On the other hand, it could be more thorough, with the objective of capturing the full extent of native expressions and formulations by designing a detailed nomenclature organized according to theoretically relevant principles[5] or through an exhaustive capture of all responses for onward computer-aided textual analysis.[6] This second option is more applicable to enumerative types of responses, whereas the first one is better for a complex narrative or argumentative structure. The exact wording of certain responses to open questions may be retained as examples in the research report.

2. A Flexible Procedure: The Interview Series

A research procedure often indirectly associated with quantitative demographic surveys is the *semi-structured interview series*, conducted before, during or at the end of a questionnaire survey. These interviews target a population and an issue expressed in an interview guide. The target population may be divided into subpopulations that are expected to provide specific information. The choice of the subpopulations depends on the hypotheses and available knowledge of the structural dimensions of the phenomenon. In an interview survey on family recomposition in France, it is necessary on the one hand, to interview persons from the lower classes and from other social classes and, on the other hand, to try to include all the actors concerned: children, parents with child custody, step-parents, parents without custody. Rather than representative samples, interview samples are like laboratory experiments of experience. The interview guide, designed according to initial hypotheses, is similar to a closed questionnaire but "in practice, it structures the interrogation but does not direct the speech" (Blanchet and Gotman, 1992, p. 64). It sets the initial instructions (e.g., "Tell me about your marital life since your marriage"), as well as intermediate ones that indicate the theme or themes. It includes a detailed thematic guide, which the interviewer is expected to know without having to specify and which corresponds to all the points that should be discussed freely by the respondent or after prompting by the interviewer. This procedure does not imply standardizing the way the interview takes place.

What most clearly differentiates an interview from a questionnaire is the great flexibility in its use. When an exploratory interview series is used to design a future questionnaire, the interview guide may be modified between the beginning and the end of the series. When a series of interviews is conducted after administering the questionnaire, possibly on a subgroup of the survey sample, the interview guide may also be adapted for each individual respondent by taking account of information already provided in the quantitative questionnaire. The advantage of an interview series over a questionnaire survey is to allow the introduction of a real to-and-fro reflection within empirical data collection. Even the number of interviews necessary is not decided beforehand, because it obeys the principle of saturation (Blanchet and Gotman, 1992); in principle, the qualitative survey ends at the moment, unknown beforehand, when further interviews no longer provide new information or processes.[7] To be analyzed, semi-structured interviews must be recorded and then transcribed. In its simplest form, this could be a thematic analysis, which gathers observations according to items or themes of the guide and highlights the range of situations and terms used by the actors. This type of analysis is generally well suited for the design of a questionnaire. More elaborate analyses can highlight the structure of a system of representations, for instance through "the analysis of relations of opposition" (Blanchet and Gotman, 1992), typical chains of events, correspondence between events experienced and perceptions, for example, the suffering of women who are unable to have a male child in Korea (Das Gupta, 1998).

[5] On the coding instructions of the circumstances of the "first encounter" (Bozon and Héran, 1987, p. 970–984).

[6] On the analysis of the question on representations of the environment, see Guérin-Pace (1997) and Collomb and Guérin-Pace (1998).

[7] On the contrary, in a quantitative survey, the number of respondents is calculated *a priori* to enable good statistical inference on the entire population or on a subpopulation.

3. Observing Interactions among Agents

Participant observation, which is very frequently used in anthropology, is less so in demography. Yet it is through this method, rather than by explicit reports in questionnaire surveys or semi-structured interviews, that we can highlight real and concrete inequalities in the treatment of children (quantity of food, health care) as observed in Africa between children of the current marriage and those of a previous marriage (Bledsoe, 1995), or in a certain number of Asian countries between boys and girls (Das Gupta, 1998), or in the Brazilian example cited above between the children who are considered to possess the energy to survive and the others (Scheper-Hughes, 1997). Direct observation is the best way to evaluate family planning programs or other health programs (e.g., diarrhea control programs, immunization programs, AIDS prevention or treatment programs) or institutions like dispensaries, health centers, or shelters for homeless persons. In these situations, direct observation makes it possible to define the object of inquiry as the interaction between the characters, especially interaction between those who are culturally or socially very far apart and whose representations of the body and of illness are very different (e.g., a nongovernmental organization doctor from a developed country and an Indian woman in Bolivia). The observer must objectively view all the elements of the situation and consequently take into account when analyzing what was observed or heard, the effects of the location, and the effects of his or her presence and his or her status with regard to the observed persons. Observation can in particular contribute to a better perception of the strategies and interests of those who are often considered to be none other than mere users or targets of programs or actions. The qualitative survey on shelters for homeless persons (Soulié, 2000) showed that centers select certain profiles of persons and that homeless persons themselves, depending on their situation, experience and aspirations, navigate through the system by choosing or avoiding certain shelters.

4. Focus Group: A Questionable Technique

Some demographers use focus group discussions especially in the exploratory phase of research. The use of this type of interview, organized by the researcher, may not be considered a rigorous research technique. It is difficult to systematically record the words spoken and the very complex nature of the interaction between participants, as the researcher is at the same time the group host, participant and observer. Moreover, the elements of these meetings are always cited in a vague manner. One can consider it as a technique for making initial contact or mobilization, an action or training technique, rather than a proper research tool. Demographers are fascinated by this method, which provides rapid results for a low investment (Knodel, 1998), but the data produced are rarely considered scientific, as they cannot really be transposed and used beyond the context that produced them. The method is misleading because it produces results, but these are simply reactions to a created situation: local representations are always shown as relatively homogenous because the effects of spokesmen, silences, and nonresponse are not considered and the expressed representations are not matched with individual practices. The artefact is close to the one produced in political science, when we assume the existence of a "public opinion" on all issues, whereas in most cases, it is simply a reaction to a situation, mainly induced by the practice of the opinion poll itself (Meynaud and Duclos, 1996; Bourdieu, 1980). Because elements of individual biographies are not collected, it is impossible for example, to identify the tactical or strategic usage of certain traditional or innovative representations in the justification discourse directed at the observer, and even less so, the existence of circumstantial contradictions between the official discourse and practice, which are very important for analysis. An argument often used to recommend the use of focus groups is the difficulty or impossibility of undertaking individual interviews. It is true that in many cases, creating the circumstances and the appropriate location for a face-to-face interview involves a relatively long process. However, a focus group discussion may not necessarily replace that which interviews and a good participant observation can offer.

5. Case Studies in Demography

A considerable amount of research in *comprehensive demography* takes the form of case studies. In itself, a case study is neither a technique nor a method. It is a type of approach and a scientific product, in which the objective is to highlight the social processes existing in an area or a human group. It is the opposite of a national demographic survey, aimed at putting in variables the behaviors and attitudes of a vast sample of the population to establish correlation between these variables. Unlike the case studies of anthropologists, those of demographers generally include quantitative surveys that provide background data or a local census for the rest of the investigation. They may be completed by multiple-round surveys. The choice of which group to study is generally not left to chance. The fact that a health program covers an area or that a nongovernmental organization settles there can weigh on the choice of that community. The fact that anthropologic research has previously been carried out in a given area may encourage demographers to revisit the area. The major methodologic asset of a case study, which requires a prolonged stay in the study area, is that it offers the possibility of varying points of view on the subject, by a combination of different sources and different approaches: "It is not the accumulation of data that is determinant, but rather the critical interplay that can be established between the various sources" (Bozon, 1984, p. 15). Maryse Marpsat cites the metaphor of triangulation used by American authors (Jick, 1983, quoted by Marpsat, 1999) to describe the combination of methods that clarify an object, like the adjustments that come from multiple viewpoints. A combination of methods can be used in strictly quantitative demography, such as the comparison between data from a population register and data from a retrospective biographical survey (Poulain *et al*., 1991). However, in a case study, the articulation of methods is the fundamental principle of a long procedure that functions in a to-and-fro manner and by a continuous adaptation of hypotheses. A quantitative survey on users of a health program or on a population may be associated with an anthropologic observation *in situ*, which helps to discover new elements. The use of a multiplicity of perspectives enables the description of processes and complex interactions, which do not reduce behavior to variable mechanics.

Recourse to qualitative methods is often considered by demographers as a simple preliminary step or a subsidiary guarantee of quality in data collection, but which does not modify the essence of things. Sometimes, but this is rare, a focus on the renewal of methods and on dialogue with anthropology and qualitative sociology contribute to the development of the demographic discipline toward other research issues, other concepts, and methods.

III. QUALITATIVE METHODS AND THE QUALITY OF DEMOGRAPHIC DATA COLLECTION

Demographers are permanently concerned with improving the quality and reliability of their data, and tracking errors and their effects on observations makes them sensitive to reflexivity and interactivity in the data collection exercise. The use of qualitative approaches in the initial phases of planning a survey has become a common way of enriching the initial stock of hypotheses and suggesting, then testing, different acceptable modes of questioning respondents. The idea of qualitative checks after data collection is equally becoming accepted.

The very first hypotheses of a study are necessarily derived from the literature already published or from more general theoretical premises. They must be made to undergo a first empirical test to give them substance. The improvement of research hypotheses on couple formation mentioned earlier (Bozon and Héran, 1989) is largely the result of a series of preliminary interviews: the choice of a spouse was redefined as the outcome of a *judgment* based on very structured systems of classification and evaluation of potential partners. These hidden classification systems become apparent in the terms employed by respondents to qualify their partners (e.g., serious, intelligent, simple, nonconformist, mature), which become significant indicators. More commonly, the vocabulary that appears spontaneously during semi-structured interviews contains native terms and formulations that can be used again in the wording of questions and response items.

A preliminary qualitative survey does not only contribute to the improvement of hypotheses, it helps in the design of practical procedures for the quantitative survey. Traditionally, only women were questioned on many demographic themes (e.g., fertility behavior, nuptiality, family), but qualitative surveys showed that questioning men, especially the spouses of female respondents, made it possible to reach a more realistic view of spousal negotiation and interaction.[8] Consequently, a strategy that involved the two spouses, at least on a part of the sample, was adopted in large scale surveys, like the Demographic and Health Surveys from the 1990s.

A demographic questionnaire is essentially a process where human behavior is defined in variables (and indicators), which seem natural when the behavior in question is already partially or highly objectified and formatted within a reality recognized by social statistics (Héran, 1984). Consequently, it is not difficult to formulate questions on the number of living children, on legal marital status, on housing that gives the status of home-owner or tenant, nor on contract employment. The design of variables is much more difficult and must be based on an initial qualitative approach when research is on poorly formalized, stigmatized, or hidden realities, or when the theme appears rarely in the literature: in the questionnaire on the homeless, designed by Maryse Marpsat and Jean-Marie Firdion (2000) after the traditional questions on current occu-

[8] Hertrich (1997) provides an example of a case study based on dual data collection from women and their husbands.

pation and last occupation, different types of probes were added concerning "work carried out without a pay-slip," "periodic, irregular work carried out for someone," "selling street newspapers," "selling salvaged or self-produced goods." On themes like informal work, the various types of informal housing, sexual violence (treated in depth in the Colombia Demographic and Health Survey, 1995), it is not possible to propose adequate response categories *ex nihilo*; an initial extensive exploration is always necessary.

Generally, an initial exploratory study, even indirect (*armchair anthropology*) of local social processes, including their ritual and symbolic aspects, makes it possible to avoid ethnocentrism or excessive legalism in the questionnaire design. Étienne van de Walle and Dominique Meekers, in their article *Marriage Drinks and Kola Nuts* (1994), showed that it is not possible for demographers in many developing countries to measure the date of marriage (even defined very loosely) in a simple way without taking into account the diversity of the traditional components of entry into union. After a review of the anthropologic literature on the subject, they analyzed the function of the very polysemous ritual of presentation of drinks and kola nuts, which was asked about in the Côte d'Ivoire Fertility Survey (1980–1981). In developed countries, the growing informality of transition to adulthood is greatly challenging the imagination of those who still wish to date the entry into union or the residential independence of young persons (Villeneuve-Gokalp, 1998).

The decision to include open-ended questions in a questionnaire, far from representing an incapacity to close questions (i.e., to limit response items beforehand), must be regarded as a deliberate wish to collect a certain number of spontaneous formulations, which are expressed through the social categories used by respondents themselves and not only through those of the analyst. In an INED biographic survey on transition to adulthood carried out among 25 to 34 year olds (Bozon and Villeneuve-Gokalp, 1995), the retrospective questions on family, marital, occupational and residential trajectories of respondents were followed by the open-ended question: "Tell me if an event in your youth has had a major impact on you (no matter whether you have already mentioned it or not). What event was this and at what age?" The answers showed that a number of events subjectively considered by respondents as striking are not even mentioned in the biographical reading grid of researchers: for example, the death of close relatives or friends, health problems or even historical events. This open-ended question could be linked with other closed questions in the questionnaire, such as, for example, the experience of depressive phases in a person's life (Archambault, 1998).

Apart from the now common use of qualitative approaches in exploratory research, complementary qualitative surveys are now being conducted after the administration of questionnaires. These complementary operations have various aims. Certain operations are essentially to verify and explore the contradictions in responses or even refusals. Unlike a traditional control survey, which essentially tries to measure the consistency between the initial questionnaire and the control questionnaire, a qualitative control operation does not consider problematic responses as errors but as an indicator of the respondent's perception of the survey or the fieldwork style. In the United States, a qualitative post-enumeration survey showed a high level of incorrect responses associated with respondent indifference to the task of filling the questionnaire and a feeling of social disaffiliation and distance from the State (Rehner-Iversen *et al.*, 1999).

It is also possible to organize complementary interviews to clarify the meanings of certain responses (contribution to questionnaire data capture) and evoke explanatory hypotheses on results (contribution to the construction of interpretative hypotheses). In INED's survey on couple formation, the post questionnaire interviews had this dual function. Many respondents had reported that in the process of building their relationship with their future spouse, they had gone through a phase of part time cohabitation (defined in the questionnaire as "a period when you were not yet permanently living together under one roof, but were already spending some days and some nights together in one another's home during the week"). The interviews showed that this response mixed up realities that were so different that it could not be interpreted as a distinct and real phase (Bozon, 1992b).

At the end of a quantitative operation, anthropologists and sociologists sometimes undertake a relatively independent qualitative survey focusing on a subpopulation or a group of individuals sharing a characteristic or relative to a problem that the questionnaire unveiled. In the 1970s, for example, Louis Roussel, in his work on the family, introduced independent qualitative surveys after traditional questionnaire surveys (Roussel and Bourguignon, 1976, 1979). The study on reproductive mishaps and Western contraception in Gambia cited earlier (Bledsoe *et al*., 1998) derives from a result of a questionnaire that would not adjust to standard data interpretation, thereby leading to a parallel anthropologic research that ended with another very different research framework.

The role of hypotheses in quantitative research is often presented unrealistically as a set of questions or rather of temporary responses that the *ad hoc* apparatus of a numerical survey would confirm or invalidate. Anthropologists and qualitative researchers see themselves rather as constantly redesigning and refining their hypotheses, exposed to the challenge of new observations. In reality, even in statistical procedure, the development of hypotheses that, at a given moment become interpretative hypotheses (or interpretations), is a very gradual process: "The progressive construction of hypotheses found in most qualitative surveys, corresponds to that which occurs in successive phases of a survey and statistical analysis, with the difference that it is impossible to return to the field at the analytical phase, unless by a new investigation, a qualitative approach or reiteration of statistical surveys" (Marpsat, 1999, p. 14). Integrating a qualitative approach in an essentially statistical strategy contributes empirically to the continuous reflection on the survey while it is taking place, then to self-analysis. The continuous production of hypotheses can be supported by more field trips, rather than by the demographer's imagination or presuppositions.

Greater flexibility and interactivity in demographic data collection contributes to making the demographic discipline "thicker" (Fricke, 1997). The interrogations brought about by nonstatistical approaches can renew certain fundamental concepts of the discipline and guide it toward new issues. Sometimes, demography invents new methods by systematizing initially qualitative approaches.

IV. QUALITATIVE APPROACHES AND THE RENEWAL OF DEMOGRAPHY

When demographers give an account of the changes in their discipline, for example at the anniversary-issues of their great journals (*Demography* in 1993, *Population* in 1995, *Population Studies* in 1996), they indicate the evolution in methods as well as in the issues (questions or debates), but rarely mention theoretical innovations, even if authors like Ron Lesthaeghe insist heavily on this dimension (Lesthaeghe, 1992, 1998; Lesthaeghe *et al*., 1994). It is particularly rare for recourse to qualitative methods and dialogue with disciplines that use them to be mentioned as factors in the evolution of the discipline. Yet, even though their contribution is not willingly recognized,[9] there are reasons to believe that the persistent existence of other methods for describing social phenomena explains a part of the development of the demographic discipline.

1. Using Variables or Describing Processes

Traditionally, demography and quantitative sociology focused on phenomena that could be defined in terms of quantifiable variables. After transforming the subject into variables by schematization[10] and data collection, analysts try through various statistical procedures to link the variables to be explained with the explanatory variables. The objective is to parsimoniously use a minimum of explanatory variables (also called determinants) to describe the phenomenon to be explained. Inversely, qualitative approaches focus on the *processes*, that is, the relations established between the attitudes, social behavior and objective structures. The process and its given descriptions may be relatively complex as no initial transformation of the behavior into simple variables has been made. The analyst nevertheless makes a parsimonious hypothesis as he assumes that the processes to be described, although complex, are few. In this type of approach, the opposition between dependent and independent variables or between phenomena to be explained and determinants fades, enabling us to move our focus more easily toward issues that theory did not initially consider as central or that appear difficult to transform into variables.

Take the example of nuptiality in African countries (Bledsoe and Pison, 1994). For a long time, it had been considered by demographers as a simple determinant of fertility that did not deserve any particular interest and second, as a difficult phenomenon to describe. When asking about date of marriage in the National Fertility Surveys of the 1980s, one was not too sure what was being measured. Through anthropology and its methods, a certain number of demographers learned how to observe and describe marriage as a process of union formation with a duration and with significant stages that were not all identical. The observation of these processes then took on the function of analyzer of social organization, especially of inter-generational relations (e.g., what is the influence of elders in the lineage?) and gender relations (Locoh, 1995).

2. Cultural Context of Individual Behaviors

Another characteristic of the approaches inspired by anthropology or comprehensive sociology is the

[9] This difficulty among demographers to recognize the contribution of qualitative approaches to their discipline is perhaps related to the obsessive desire by a large number of them to belong to the "hard sciences."

[10] Schematization is necessary and legitimate according to the research hypothesis; according to Gaston Bachelard, quoted by Hubert Gérard, there is a "right to overlook what is negligible" (Bachelard, 1938).

importance given to the analysis of social and cultural structures in which people function. They are different from certain analytical models traditionally employed in demography, for example in the analysis of fertility determinants, which place social actors in a strictly individual framework, with a possible sprinkling of general cultural traits that are meant to characterize them because of their membership of aggregate ethnic, national, or geographic groups. The impact of culture or social organization on demographic behavior remained for a long time *terra incognita* to population studies. To consider the effect of social and cultural organization other than as mere residuals unexplained by individual variables, some demographers, again influenced by anthropology, tried to integrate the analysis of individual behavior and social organization (Kertzer, 1995).

Take the example of the prolonged *post-partum* abstinence observed in African countries. The prevalence of this behavior was for a long time considered to be a determinant of fertility, something specific to each cultural group and not arousing any particular interest. The change in perspective that led to considering abstinence as an issue requiring explanation led to two developments both influenced by anthropologic approaches. The macro-level analyses of Lesthaeghe and colleagues (1994) looked at the relationship between the characteristics of the social and economic organization of various ethnic groups, systematized from earlier anthropologic studies, and the prevalence rate of *post-partum* abstinence of each group. Two social organization characteristics seem particularly related to low abstinence levels: matrilineal (versus patrilineal) lineages, and low productive value of women in the economy (which is the case of pastoral societies). Research on Gambia by Caroline Bledsoe and Alan Hill (Bledsoe *et al.*, 1998; Hill, 1997) made it possible to observe internal variations within a society. The authors first linked the practice of abstinence to a local representation of reproductive life, which constitutes the mental framework in which choices are made. Because the rule of abstinence is traditional does not mean that it cannot be strongly manipulated. First, abstinence is better expressed as a negotiated reduction of frequency rather than a total suppression of sexual relations (Hill, 1997, p. 233). Second, average birth intervals have been observed to regularly increase with parity: young women rather tend to reduce abstinence periods while women who already have many children increase them. On the whole, independently of factors of change associated by demography with progress in education, there are social and economic organizations that favor or discourage the tradition of postpartum abstinence, which appears to be a more strategical and negotiated practice for women than is generally assumed.

3. Comprehensive Demography and Emergence of Gender Analysis

Another example of the evolution of demography toward issues and perspectives that were initially foreign is the increased interest among demographers in gender analysis (Riley, 1997a, 1997b; Greenhalgh, 1995; Federici *et al.*, 1993). This new development is largely due to research and methods of less quantitative disciplines. Demography has evidently never ignored the sex of individual members of populations and has learnt very early to calculate life expectancy or infant mortality rates by sex as well as to establish the existence of inequalities in the status of men and women in levels of education, age at marriage and levels of occupational activity (Véron, 1997). The central role played traditionally in demography by the analysis of fertility has even promoted converged interest in a certain number of specifically feminine indicators: it has become the tradition to link female education to age at marriage, number of children, frequency of contraceptive use, and even level of infant mortality (see Chapter 93). In the classic versions of the demographic transition theory, which is basically an application to population studies of a sociologic theory of modernization, increase in women's education is considered as one of the factors that directly influences the decline of fertility and infant mortality. This focus on female education may distract from the fact that neither the rise in levels of education nor the drop in fertility necessarily lead to an increase in equality between the sexes. All societies have their own social and cultural construction of the relationship between men and women, called the gender system, which determines male and female roles, their respective resources and the power attributed to them. Although the result is not achieved in the same way in each society, men always have more power than women in the public as well as the private spheres.

Gender-sensitive demography is interested in the persistent factors that produce discrimination against girls and women, those that promote decisional autonomy in women, as well as modes of negotiation and constraint between men and women. Qualitative methods are invaluable for highlighting and linking these elements. They were first examined at macrosocial levels where anthropologic data, gender systems, and demographic processes could be analyzed. Fertility decline in several major Asian countries (India, China, Korea) occurred at the same time as an increase in the proportion of missing daughters, mainly by selective induced abortion. In these patrilineal societies, sons, as opposed to daughters, live with their parents after marriage and contribute to their upkeep. Technologic modernization in this case is reinforcing the expression of the traditional son preference. If demography is particularly suited to providing a quantitative diagnostic of the phenomenon, it needs to associate with more anthropologic approaches to describe the set of processes and dramatic consequences on the condition of women by contextualizing behavior beyond the individual level. Another possible application of a gender approach in demography relates to the description of the consequences of the AIDS epidemic on women in Africa. Essentially qualitative studies have shown on the one hand that women were rarely in a position to refuse unprotected sex from HIV-positive husbands and on the other, that their fertility was increasing rather than declining when they were HIV positive (Desgrées du Loû, 1998). Demographic or epidemiologic studies integrating an anthropologic approach would be needed to inscribe the overall logic of these behaviors in their cultural setting, without omitting the gender dimension and constraint and even the violence.

If demography is the analysis of the reproduction of populations, one can hardly be surprised *a priori* that one of the major issues of interest is the quantitative analysis of fertility. Contact with disciplines such as anthropology and sociology and more qualitative approaches have led to a critique of the demographic concept of reproduction as being too biologic and individualistic, as a reproduction of persons in which children are always assigned to a precise woman (Townsend, 1997). In the wider focus of other disciplines, reproduction must be understood as "the social reproduction of structures of relationship and social positions" (Kertzer and Fricke, 1997, p. 21). In this perspective, demography needs to also look at the processes surrounding social parenthood, like adoption, family recomposition, child fostering, and even guardianship, as well as the mechanisms behind the reproduction of gender relations. Contact with other approaches will immediately widen its scope and possibilities for interpretation.

4. Quantitative Systematization of Qualitative Approaches

A part of the qualitative methods used by demographers can finally be considered as a systematization of preexisting qualitative approaches. Family reconstitution in historical demography is based on the technique of genealogy. Biographic analysis prolongs the life history approaches of sociologists and anthropologists. The development of multilevel analysis results from concern with integrating the micro, meso and macro levels within the same analysis, and this is traditional in anthropology. The sampling methods used in statistical surveys on the homeless imitate the field procedures used by anthropologists (night surveys, surveys of shelters). The use of textual analysis in demography is a systematization of content analysis. The invention of new methods by the standardization of empirical field methods is particularly common in countries with poor statistics. Although quantitative methods produce schematization and homogenization of data corresponding to variable transformation, they often depend on a logic of interrogation and analysis of reality that is initially much more complex. In this sense, there is no barrier between qualitative and quantitative methods and dialogue with the disciplines that do not use quantitative methods remains essential to the development of demography.

Such dialogue does not imply conversion to another discipline, but the continuous recognition by demographers of results, issues, and tools suggested by qualitative methods. However, it does not seem desirable to imagine purely qualitative demography, which would exist as a subdiscipline next to mainstream demography. For example, even if a qualitative approach reveals in various cultures that age is generally perceived in discontinuous categories (what may be called *social age*), it is still necessary to link these native categories to a continuous measurable age. Comprehension is brought about by the tension between meaning and measurement, rather than a simple focus on meaning.

CONCLUSIONS

Even if demography possesses an incomparable statistical toolbox, refusing to use different approaches would be a short-term perspective that might block its development as a discipline. A wide and unrestricted analysis of the reproduction dynamics of populations requires *comprehensive demography*, both reflexive and interpretative. A tradition of critique of sources and rigor in data collection helps the demographer to accept the contribution of qualitative methods in the preliminary phases of the construction of research; the requirement of reflexivity should lead to more frequent invitations for qualitative confirmations of obtained numerical results.

Believing itself to be a discipline of facts, demography is often reluctant to the idea of interpretation. However, the demand for interpretation is merely a requirement for contextualization of behavior, to which qualitative approaches can be of assistance. This expansion of focus can take place in three directions: the association between the observed behavior and the meaning conferred on it by the actors, the analysis of interpersonal interactions that produce the behavior, the situation of the observed phenomena in a social and cultural setting that should be described beyond what is perceived by the respondents. This triple contextualization is necessary to reveal, for example, the constraints that weigh on the demographic behavior of women and their possibilities for taking initiatives within a given gender system. Interpretative demography is likely to appropriate new issues and invent other methods to treat them.

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

ARCHAMBAULT Paul, 1998. États dépressifs et suicidaires pendant la jeunesse: résultats d'une enquête sociodémographique chez les 25–34 ans, *Population*, vol. 53(3), p. 477–515.

ARIES Philippe, 1962. *Centuries of childhood*. New York, Vintage Books, 447 p.

BACHELARD Georges, 1999. *La formation de l'esprit scientifique*. Paris, Vrin, 256 p. (1st ed., 1938.)

BASU Alaka Mawade and AABY Peter (eds.), 1998. *The methods and uses of anthropological demography.* Oxford, UK, Oxford University Press, 329 p. (International Studies in Demography.)

BLANCHET Alain and GOTMAN Alain, 1992. *L'enquête et ses méthodes: l'entretien.* Paris, Nathan, 128 p.

BLEDSOE Caroline, 1995. Marginal members: children of previous unions in Mende households in Sierra Leone, *in*: Susan Greenhalgh (ed.), *Situating fertility. Anthropology and demographic inquiry*, p. 130–153. Cambridge, Cambridge University Press, 304 p.

BLEDSOE Caroline, BANJA Fatoumatta, and HILL Alan, 1998. Reproductive mishaps and Western contraception: an African challenge to fertility theory, *Population and Development Review*, vol. 24(1), p. 15–57.

BLEDSOE Caroline and PISON Gilles (eds.), 1994. *Nuptiality in sub-Saharan Africa, contemporary anthropological and demographic perspectives.* Oxford, Clarendon Press, 326 p. (International Studies in Demography.)

BLEDSOE Caroline and HILL Alan, 1998. Social Norms, Natural Fertility and the Resumption of Postpartum Contact in the Gambia, *in*: Alaka Mawade Basu and Peter Aaby (eds.), *The methods and uses of anthropological demography*, p. 268–297. Oxford, Oxford University Press, 329 p. (International Studies in Demography.)

BOURDIEU Pierre, 1980. L'opinion publique n'existe pas, *in*: Pierre Bourdieu (ed.), *Questions de sociologie*, p. 222–235. Paris, Éditions de Minuit, 277 p.

BOZON Michel, 1984. *Vie quotidienne et rapports sociaux dans une petite ville de province. La mise en scène des différences.* Lyon, Presses Universitaires de Lyon, 300 p.

BOZON Michel, 1991. Women and the age gap between spouses. An accepted domination? *Population, An English Selection*, no. 3, p. 113–148.

BOZON Michel, 1992a. Par-delà le quantitatif et le qualitatif. Pour une analyze des mécanismes de différenciation, *in*: AIDELF (eds.), *Démographie et Différences*, p. 559–566. Paris, PUF, 662 p. (Colloque International de Montréal, Colloques 4.)

BOZON Michel, 1992b. La diversification des modes d'entrée dans la vie en couple, *in*: AIDELF (eds.), *Démographie et Différences*, p. 567–572. Paris, PUF, 662 p. (Colloque International de Montréal, Colloques 4.)

BOZON Michel and HÉRAN François, 1987. La découverte du conjoint. I. Évolution et morphologie des scènes de rencontre, *Population*, vol. 42(6), p. 943–986.

BOZON Michel and HÉRAN François, 1989. Finding a spouse. A survey of how French couples meet, *Population*, *An English Selection*, no. 1, p. 91–122.

BOZON Michel and VILLENEUVE-GOKALP Catherine, 1995. L'art et la manière de quitter ses parents, *Population et Société*, no. 297, p. 1–4.

CALDWELL John C., HILL Alan G. and HULL Valerie J. (eds.), 1988. *Micro-approaches to demographic research*. London, Kegan Paul, 500 p.

COURGEAU Daniel, 1999. Preface, *in*: GRAB (Groupe de réflexion sur l'approche biographique) (eds.), *Biographies d'enquêtes. Bilan de 14 collectes biographiques*, p. ix–xvi. Paris, INED, IRD, Réseau Socio-Economie de l'habitat, PUF, 336 p. (Méthodes et Savoirs.)

DAS GUPTA Monica, 1998. *"Missing girls" in China, South Korea and India: causes and policy implications*. Boston, MA, Harvard Center For Population, 19 p.

Demography, 1993. Thirty years of *Demography, Demography*, vol. 30(4).

DESGRÉES DU LOÛ Annabel, 1998. Santé de la reproduction et sida en Afrique subsaharienne: enjeux et défis, *Population*, vol. 53(4), p. 701–730.

FEDERICI Nora, OPPENHEIM MASON Karen and SOGNER Solvi (eds.), 1993. *Women's position and demographic change*. Oxford, Clarendon Press, 368 p.

FLANDRIN Jean-Louis, 1976. *Familles. Parenté, maison, sexualité dans l'ancienne société*. Paris, Hachette, 288 p.

FRICKE Tom, 1997. Culture theory and demographic process: toward a thicker demography, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 248–278. Chicago, University of Chicago Press, 294 p.

GÉRARD Hubert, 1988. Au-delà du quantitatif. Discours d'ouverture, *in*: Hubert Gérard and Michel Loriaux (eds.), *Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie*, p. 15–52. Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quetelet 1985.)

GÉRARD Hubert and LORIAUX Michel (eds.), 1988. *Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie*. Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quetelet 1985.)

GREENHALGH Susan, 1995a. Anthropology theorizes reproduction: integrating practice, political economic, and feminist perspectives, *in*: Susan Greenhalgh (ed.), *Situating fertility: anthropology and demographic inquiry*, p. 3–28. Cambridge, Cambridge University Press, 304 p.

GREENHALGH Susan (eds.), 1995b. *Situating fertility: anthropology and demographic inquiry*. Cambridge, Cambridge University Press, 304 p.

GUERIN-PACE France, 1997. La statistique textuelle, un outil exploratoire en sciences sociales, *Population*, vol. 52(4), p. 865–888.

GUÉRIN-PACE France and COLLOMB Philippe, 1997. *Les Français et l'environnement. L'enquête "Populations-Espaces de vie-*

Environnements". Paris, INED and Presses universitaires de France, 255 p. (Travaux et documents, cahier no. 141.)

Héran François, 1984. L'assise statistique de la sociologie, *Economie et statistique*, no. 168, p. 23–36.

Hertrich Véronique, 1997. Les réponses des hommes valent-elles celles des femmes? Une double collecte sur les questions génésiques et matrimoniales dans une population du Mali, *Population*, vol. 52(1), p. 45–62.

Hill Allan G., 1997. Truth lies in the eye of the beholder: the nature of evidence in demography and anthropology, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 223–247. Chicago, University of Chicago Press, 294 p.

Jick Todd D., 1993. Mixing qualitative and quantitative methods: triangulation in action, *in*: John van Maanen (eds.), *Qualitative Methodology*, p. 135–149. Beverly Hills, CA, Sage, 272 p.

Kaufmann Jean-Claude, 1992. *La trame conjugale. Analyze du couple par son linge*. Paris, Nathan, 216 p.

Kertzer David, 1995. Political-economic and cultural explanations of demographic behavior, *in*: Susan Greenhalgh (eds.), *Situating fertility: anthropology and demographic inquiry*, p. 29–52. Cambridge, Cambridge University Press, 304 p.

Kertzer David and Fricke Tom (eds.), 1997. *Anthropological demography. Toward a new synthesis*. Chicago, University of Chicago Press, 294 p.

Kertzer David and Fricke Tom, 1997. Toward an anthropological demography, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 1–35. Chicago, University of Chicago Press, 294 p.

Knodel John, 1998. Using Qualitative Data for Understanding Old-Age Security and Fertility, *in*: Alaka Mawade Basu and Peter Aaby (eds.), *The methods and uses of anthropological demography*, p. 57–80. Oxford, Oxford University Press, 329 p. (International Studies in Demography.)

Leboutte René, 1991. Motivations des acteurs de la transition démographique. De l'analyse quantitative à l'enquête orale dans la région liégeoise (fin XIX[e] siècle-XX[e] siècle), *in*: SBD (eds.), *Historiens et Populations*, p. 281–300. Louvain-la-Neuve, Academia, Société belge de démographie (SBD), 836 p. (*Liber amicorum* Étienne Hélin.)

Lesthaeghe Ron J., 1992. Beyond economic reductionism: the transformation of the reproductive regimes in France and Belgium in the 18th and 19th Centuries, *in*: Calvin Goldscheider (ed.), *Fertility transitions, family structure and population policy*, p. 1–44. Boulder, CO, Westview Press, 283 p.

Lesthaeghe Ron J., 1998. On theory development and applications to the study of family formation, *Population and Development Review*, vol. 24(1), p. 1–14.

Lesthaeghe Ron, Kaufmann Georgia, Meekers Dominique and Surkyn Johan, 1994. Post-partum abstinence, polygyny and age at marriage: a macro-level analysis of sub-Saharan societies, *in*: Caroline Bledsoe and Gilles Pison (eds.), *Nuptiality in sub-Saharan Africa, contemporary anthropological and demographic perspectives*, p. 25–54. Oxford, Clarendon Press, 326 p. (International Studies in Demography, IUSSP.)

Locoh Thérèse, 1995. Divorce et remariage des femmes en Afrique de l'Ouest. Le cas du Togo, *Population*, vol. 50(1), p. 61–93.

Marpsat Maryse, 1999. *Les apports réciproques des méthodes quantitatives et qualitatives: le cas particulier des enquêtes sur les personnes sans domicile*. Paris, INED, 24 p. (Documents de travail no. 79.)

Marpsat Maryse and Firdion Jean-Marie (eds.), 2000. *La rue et le foyer. Une recherche sur les sans-domicile et les mal-logés dans les années 1990*. Paris, INED et PUF, 416 p. (Travaux et documents, cahier no. 144.)

Mauss Marcel, 1947. *Manuel d'ethnographie*. Paris, Payot, 210 p.

Meynaud Hélène and Duclos Denis, 1996. *Les sondages d'opinion*. Paris, La Découverte, 128 p.

Niraula Bhanu B. and Morgan S. Philip, 1996. Marriage formation, post-marital contact with natal kin and autonomy of women: evidence from two Nepali settings, *Population Studies*, vol. 50(1), p. 35–50.

Pison Gilles, 1980. Calculer l'âge sans le demander: méthode d'estimation de l'âge et structure par âge des Peuls Bandé, *Population*, vol. 35(4–5), p. 861–892.

Population, 1995. Cinquante années de *Population, Population*, vol. 50(6).

Population Studies, 1996. The first fifty years, *Population Studies*, vol. 50(3).

Poulain Michel, Riandey Benoît and Firdion Jean-Marie, 1991. Enquête biographique et registre belge de population: une confrontation des données, *Population*, vol. 46(1), p. 65–88. (Published in English: Data from a life history survey and the Belgian population register: a comparison, *Population*. *An English Selection*, 1992, vol. 4, p. 77–96.)

Quesnel André, 1988. Objectifs et formes de l'approche qualitative dans les études démographiques menées dans les pays à statistiques incomplètes, *in*: Hubert Gerard and Michel Loriaux (eds.). *Au delà du quantitatif. Espoirs et limites de l'analyse qualitative en demographie*, p. 321–345. Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quételet. 1985).

Rehner Iversen Roberta, Furstenberg Frank F., Jr., and Belzer Alisa A., 1999. How much do we count? Interpretation and error-making in the Decennial Census, *Demography*, vol. 36(1), p. 121–134.

Riley Nancy, 1997. Gender, power and population change, *Population Bulletin*, vol.52(1), p. 1–48.

Riley Nancy, 1997. Similarities and differences: anthropological and demographic perspectives on gender, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 115–138. Chicago, University of Chicago Press, 294 p.

Roussel Louis and Bourguignon Odile, 1976. *La famille après le mariage des enfants. Etude sur les relations entre générations*. Paris, INED and PUF, 268 p. (Travaux et documents, cahier no. 78.)

Roussel Louis and Bourguignon Odile, 1979. *Générations nouvelles et mariage traditionnel. I. Enquête auprès de jeunes de 18–30 ans*. Paris, INED and PUF, 292 p. (Travaux et documents, cahier no. 86.)

Scheper-Hughes Nancy, 1997. Demography without numbers, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 201–222. Chicago, University of Chicago Press, 294 p.

Segalen Martine, 1985. *Quinze générations de Bas-Bretons. Mariage, parentèle et société dans le pays Bigouden sud: 1720–1980*. Paris, PUF, 874 p.

Segalen Martine, 1997. Comment se marier en 1995? Nouveaux rituels et choix sociaux, *in*: Gérard Bouchard and Martine Segalen (eds.), *Une langue, deux cultures. Rites et symboles en France et au Québec*, p. 149–166. Paris, La Découverte et Université Laval, 350 p.

Singly (de) François, 1987. *Fortune et infortune de la femme mariée. Sociologie de la vie conjugale*. Paris, Presses universitaires de France, 229 p.

Soulié Charles, 2000. Le dualisme du réseau d'hébergement pour les personnes sans abri à Paris, *in*: Maryse Marpsat and Jean-Marie Firdion (eds.), *La rue et le foyer. Une recherche sur les sans-domicile et les mal-logés dans les années 1990*, p. 211–256. Paris, INED and PUF, 413 p. (Travaux et documents, Cahier no. 144.)

Townsend Nicholas, 1997. Reproduction in anthropology and demography, *in*: David Kertzer and Tom Fricke (eds.), *Anthropological demography. Toward a new synthesis*, p. 96–114. Chicago, University of Chicago Press, 294 p.

van de Walle Étienne and Meekers Dominique, 1994. Marriage drinks and kola nuts, *in*: Caroline Bledsoe and Gilles Pison (eds.), *Nuptiality in sub-Saharan Africa, contemporary anthropological and demographic perspectives*, p. 57–73. Oxford, Clarendon Press, 326 p. (International Studies in Demography, IUSSP.)

Véron Jacques, 1997. *Le monde des femmes. Inégalité des sexes, inégalité des sociétés*. Paris, Seuil, 205 p.

Villeneuve-Gokalp Catherine, 1997. Le départ de chez les parents: définitions d'un processus complexe, *Économie et statistique*, no. 304–305, p. 49–162.

C H A P T E R

137

Methods of Social Network Analysis

ALEXIS FERRAND AND AINHOA DE FEDERICO DE LA RÚA

Institut de Sociologies et d'anthropologie, Universite des Sciences et Technologies de Lille, Villeneuve d'Ascq, France

Institut Federatif de Recherche sur les Economies et les Societes Industrielles (IRESI), Universite des Sciences et Technologies de Lille 1, Lille, France

The concept of networks is found "in almost all fields"[1] of science, including the social sciences. It has also become part of our everyday language, with a variety of meanings. In the social sciences, the word began to acquire a specific definition and to be used in a particular way more than 50 years ago. The examples of two pioneers will show how the concept has helped to advance knowledge in psychosociology and social anthropology.

In the early 1930s, Jacob Moreno in the United States suggested speaking of the "social atom," referring not to the individual as such, but to the individual plus the interpersonal relations of attraction and rejection centered on him. This introduced an intermediate object in psychosociology, between the individual, locus of emotion and personality, and the small group that is the basic form of social order. Moreno was studying a re-education institution and wondered why a series of girls ran away, following each other. He discovered they were linked in *a network*, and he identified this concept as applying to relations that cross boundaries of groups. He also gave it a specific theoretical status. Moreno defined it as a *structured* phenomenon; however, informal such relational networks may be, they are regulated, organized and structured by principles that need to be brought to light. He also attributed a function to such networks. They provide channels for communicating information, playing a part in the circulation of rumors and thereby helping to form public opinion, which has a social regulation function. By 1934, networks had gained an acknowledged place in psychosociology as functional, supra-individual entities, detached from and transcending social organizations, informal but not formless.

Twenty years later, English social anthropology—which, like psychosociology, practices direct observation of small social interaction systems—also discovered the importance of interpersonal relationship systems. John Barnes was studying a Norwegian parish, aiming to investigate its class system and how power operated. He distinguished three fields: an economic activity (e.g., fishing), a territorial field (e.g., neighborhoods, agricultural cooperation), and a third field that united the other two. "The third social field has no units or boundaries; it has no coordinating organization. It is made up of the ties of friendship and acquaintance which everyone growing up in Bremnes society partly inherits and largely builds up for himself. . . . Most of the ties are, however, between persons who accord approximately equal status to one another, and it is these ties which, I think, may be said to constitute the class system of Bremnes" (Barnes, 1954, p. 237). This enabled Barnes to answer his questions. In the first place, actors assess their social position within the network, where they are mainly in contact with fairly similar people; even if they know there are wealth differences, they can think of themselves as equal. Second, if each resident is linked to many others in the network, and in a variety of ways, these links would be endangered if sharp disagreements or conflicts were allowed to develop. Here, the

[1] "Jeux de réseaux," *Cahiers S. T. S*, no. 9–10, Paris, Center National de la Recherche Scientifique.

network's functional effect is to impose conflict avoidance on the community's power structure.

These two examples reflect the range of approaches that have adopted and refined the network concept. The first starts from relations between individuals in small groups and arrives at the notion of social regulation by networks through which public opinion is formed. The other starts from institutionalized groups to reconstitute the matrix of a society's institutions. Both recognize networks as a structured and structuring reality that fulfills specific functions. Although the two approaches are profoundly different and regardless of the major developments that have since taken place in the discipline,[2] both remain characteristic of the vast intellectual field now known as network analysis.

I. OBSERVED NETWORKS AND RELATIONAL SYSTEMS

1. Two Methods

Like these pioneers, we will start from empirical considerations to define the subject of this chapter; we will say that a social network can be identified if one can define *a set of actors*[3] and a *particular type of relation*, and *can observe relations of this type* (their presence or absence) between the actors. If we can examine the relations each person has with all the others, we can speak of a *total network* or *complete network*. In this way surveys can be used to observe relations between members of an association, elected officials in a local community, or co-workers in the same department, where the networks involved are not too large. One can also work on very large networks using relations that are systematically recorded (e.g., in administrative data, directories, registers).

If the number of actors is not defined, the network is an *open* one. Examples are friendship networks, kinship networks, neighbor networks, and sexual networks. By definition, open networks cannot be observed in their entirety. *Atomistic* surveys are used, with each respondent being asked to describe the partners with whom he or she has the type of relationship observed. The set formed by the direct relations between the respondent (called *ego*) and other people (called *alters*) is a *relational star*. If *ego* describes the links between his or her *alters*, *we speak of personal network*, *personal community* or *ego network*.

[2] For the historical development of social network analysis, see Scott (1991) or Freeman (2004).

[3] The term *actor* is used here in a very broad sense and may refer to an individual, a group, or an organization.

These definitions can be related to more conventional objects of study. *Kinship systems*, for example, are networks made up of three basic types of links: filiation, consanguinity, and marriage. The psychosociology of small groups also looks at relational structures within groups and some of their effects (e.g., communication, influence). It is often based on laboratory observation, which can be used to create small complete networks but does not take into account the fact that actors may belong to more than one relational system at a time. Network analysis is more precise than many sociability studies because it records the distinctive properties of the relationship with each partner and their overall sociometric organization. That is, rather than mentioning e.g. what kind of things he usually does with neighbors in general, the respondent will precisely describe what he does with each specific neighbor.

Complete networks and personal networks are different kinds of objects, constructed using different observation methods. However, they take similar theoretical approaches to their models of actor rationality and relational constraints, they use some similar methods for prompting respondents to elicit their alters, and, taking personal networks as small complete networks, they use some of the same sociometric indicators (e.g., network density, connectedness). The two types of networks are located at different levels: all personal networks are subsidiary parts of broader social networks, whether bounded or open, just as a nuclear family is a subsidiary part of a kinship network.

When observing a complete network, the analysis (a) concerns a *case*, (b) highlights a pattern of relations that actors may not be aware of, (c) can be used to systematically observe subgroups, the links between them and the rare positions occupied by central actors or intermediaries. When observing a personal networks, the analysis (a) identifies regularities, (b) concerns the network perceived by the actor, and (c) often provides for a more systematic analysis of an actor's diverse group memberships and the variety of roles he plays in different relationships.

2. Approaches Rather Than a Unified Theory

John Barnes argued that there is no such thing as a theory of social networks and that perhaps there never would be. Ten years later, Ronald Burt wrote, "Network analysis is not a single corpus of knowledge cumulating with each passing year. . . . A loose federation of approaches, progressing on many fronts as a result of the efforts of many persons, is currently referenced as network analysis" (1980, p. 79). A few years later, Michel Forsé and Simon Langlois identified at least some epistemologic attitudes common to practitioners of network analysis, which "bring together researchers around the world who, although they do not have exactly the same theoretical perspectives, do have a common conception of sociology as a positive and cumulative science" (1997, p. 28).[4]

The diversity does not belie a considerable underlying body of shared ideas. First of all, there is the idea that a network may display *regularities*, a certain *pattern* of relations. The term *structure* (in the weak sense of that term) refers to the main empirical features, measured by indices drawn from matrices of the links between actors. These indices include density, connectedness, segmentation into cliques or classes of structural equivalence, distances and centrality. How are such regularities to be explained? At the theoretical level it may be assumed that the creation of a relation or the impossibility of creating one depends on the way in which existing relations are structured. In that case the network is regarded as a *relational system*. This is a more theoretical notion than a network, because to define a system, hypotheses have to be formulated about the interdependence of the relations. Here the term *structure* refers to the principles and rules that *produce* this particular system (Ferrand, 1997). *Structure* is taken in the strong sense, referring to a generative model (e.g., the incest taboo is one of the structural rules of kinship, but other rules are needed to define a particular kinship system).

Depending on the analysis, the network as a collective form is treated as an independent variable explaining actors' strategies or collective effects or as a dependent variable explained by the motives and principles that lead actors to form a relation or to end it.

There are four main orientations in network analysis:

1. Explaining actors' behavior on the basis of the relations in which they are involved and the positions they occupy in a network. Their network positions are regarded as explanatory principles at least as powerful as the personal attributes that define them regardless of context, such as age, income and level of education (Wellman and Berkowitz 1988, p. 30). This approach considers three effects. One is a cognitive and normative effect: the network can affect the actor's *orientations*, opinions, standards and beliefs (Ferrand and Mounier, 1993, 1996). One is an instrumental effect: relations offer possibilities for mobilizing resources of all kinds, hence *means of action* (Lin, 1995, 2001). The third is a controlling effect: relationships impose *constraints* and limit the range of possible actions.

2. Explaining collective effects from the characteristics of the network concerned. The type of structure conditions the dissemination of information, social control and cohesion, differentiation of powers, and conflict (Lazega, 1992, 2001; Stokman and van den Boos, 1992; Galaskiewicz, 1985).

3. Explaining the network itself: why it exists, the preconditions for its existence, its particular properties and the conditions that determine whether it changes and how. Considerable progress is now being made in defining statistical models to assess changes in a complete network at different moments in time, and in building models to simulate such changes (Snijders, 2002; de Federico de la Rúa, 2004a).

4. Explaining the characteristics of personal networks (and therefore partly enlightening the formation of the complete network) by the influence of cultural models (Héran, 1988; Fischer, 1982a, 1982b), actors' strategies or the influence of structural constraints. Taking this approach, a dyadic relation between two actors may be understood in terms of its *embeddedness* within the actors' other relations. This reveals a feedback loop between actions and existing relations: actions may mean forming relationships and so changing the network, which affects actions and so on (Burt 1982). The *actors*, who are the network *nodes*, are distinguished by their reciprocal relations, but they are not determined by them. They have a margin of freedom that varies depending on the organization of the network and the position they occupy in it (Degenne and Forsé, 1994). The actor's position in the structure is never definitively settled, and it is a central issue for actions.

Almost any phenomenon or entity can reasonably be defined as a network of relations. There are few social realities that do not involve at least minimally organized relations between actors, and pointing out the existence of a network is stating the obvious.[5] If an observer thinks he is making an original contribution by showing that some phenomenon involves a network it is because he assumes (usually implicitly) that only a relational system with particular charac-

[4] "Regroupe à travers le monde les chercheurs du domaine qui, sans partager exactement les mêmes perspectives théoriques, ont en commun une conception de la sociologie comme science positive et cumulative."

[5] The reverse exercise is more challenging. Only in very special situations can it be shown that there are no interrelations or feedback between actors.

teristics could form a network. What he is actually doing is to show what *the particular structural features* of that network are—which is quite a different thing. One can only make scientifically interesting propositions by *comparing* the structural properties and effects of different types of networks.

II. PROPERTIES OF RELATIONSHIPS AND NETWORKS

Before defining the particular features of a network, one must define the type of relations that form the network's links. In the following sections, we consider the properties most often examined in personal and complete networks.

1. Characterizing Relations

How can the *type of relation*, the basic unit for identifying a network, be defined? Some have formal definitions, such as marriage, shareholdings between firms, hierarchy of command, and membership of an association. Others are less precise. For example, what is friendship? There is a paradox; relationships between people vary widely, but our everyday vocabulary for describing them is limited. We have words to define *categories* of kin, close or distant, and a few other general categories like acquaintance, colleague, confidant, old friend, true friend, neighbor, tennis partner or team member. These words define typical roles that more or less depend on the contexts in which people meet. They do not define the content of the relationship.

a. Role, Context, and Content of a Relation

These limitations arise in the procedures for observing the simplest social networks, in which the respondent is prompted with everyday words to define their relationship with each of their contacts: "neighbor," "colleague," "family member," "member of the same club," "friend," and so on. The interpersonal tie is characterized by the name of a *typical role* or the situation where it takes place. It may also be characterized by *content*, i.e. whatever exchange of resources it makes possible. Content is often complex and a long, meticulous ethnographic observation is required to reveal all its dimensions. However, network analysis is as much concerned with the way relations are organized as it is with their content, gaining in scope what it loses in detail. As a rule only a few *typical contents* are considered, such as informational links (e.g., exchanging information about job opportunities), emotional support (e.g., talking about a personal problem), practical help (e.g., stepping in to settle a problem with a public service, help with odd jobs about the home) and social links (e.g., going out together, or being regularly in conflict).

The naming of relationships raises two classic sociologic questions. One concerns *regulating relationships*: to what extent do role norms (or context[6]) impose a standard content on a relation, and how free are the actors to negotiate their own agreements (interactionist perspective)? The other concerns the *social variability of role content*. For example, Claude Fischer (1982b) has shown that there is no consensus among Americans as to what "friend" means; different groups use different rules, and a person may use the term imprecisely or in a very individual sense. François Héran (1987, p. 45) has shown that in France the definition of "neighbor" varies according to housing environment. Linton Freeman and Danching Ruan (1997) have shown that the link between relationship role and content is fairly similar among European countries, but differs sharply between Europe and China.

b. Formal Characteristics of Relations

Multiplexity

Most studies cover more than one type of link. In that case, a relation between two actors may involve one type of link only (a "simple" relation) or several (usually called a "multiplex" relation [Knoke and Kuklinski, 1982, p. 15.]). A multiplex relationship between two people has more than one role, or more than one type of content. This factor tells us whether different spheres of social life overlap, or are strictly separate. Stereotypically, communities where each person is constantly meeting the same others but in different roles (e.g., as kin, neighbor, colleague) contrast with life in the metropolis, where work, home life and leisure activities involve separate relational systems (e.g., a colleague is unlikely to be a neighbor, friends are unlikely to live near each other).

Frequency of Interaction

It is easy to observe whether contacts are frequent or rare. What does frequency tell us, sociologically? Does repeated contact indicate an intense relationship? "High interaction rates do not necessarily imply . . . high 'intensity' of links . . . , for which it is difficult to devise measures" (Price, 1981, p. 291). In a few cases, interaction frequency is not significant. In all other

[6] The notions of "context" and "role name" are often used interchangeably in the literature.

cases, it should be regarded as a measure specific to the type of link examined. For example, "frequency exaggerates the strength of ties to coworkers and neighbors" (Marsden, 1990).

Duration

Not all relationships are lasting, and sometimes the researcher can find out how long a relationship has existed. What can be said about duration? There are relationships that do not change over a 10-year interval, others that change fast. According to some approaches, duration is the condition for exchanges to increase and trust to develop (Blau, 1964). Others stress the "irretrievable investments"[7] that accumulate in a relationship (Rusbult, 1980; Sprecher, 1988); others allude to a natural history of relations (Bozon, 1998; Ferrand, 1989).

Symmetry and Reciprocity

Many social ties are directionally biased, implying that the two partners are differently positioned in expressing an attitude (choosing someone as a friend, or being chosen) or in circulating resources (giving advice or receiving it). *Symmetry* refers to a two-way link (A chooses B as a friend and B chooses A) and *reciprocity* means that resources circulate both ways (A gives to B and B gives to A). Symmetry and reciprocity are crucial questions for psychosociology (i.e., sociometry), anthropology, and sociology (i.e., social exchange theory).

Particular exchanges apart, the partners' subjective perception of the relation may not be the same. Edward Laumann (1973) asked 985 Detroit men to name "three men who are your closest friends and whom you see most often." When he cross-checked with the partners named, he found that fewer than half named the first respondent as one of their three closest friends. At this basic level of identifying best friends, one-half of all relations were asymmetric.

Homophily

When two actors are linked by a relationship, is it obligatory, encouraged, tolerated, or forbidden that they belong to the same category? All relations listed can be placed on a "choice matrix" showing the actors' attributes in rows and columns. The classic example of this type of analysis is the *choice of spouse*; in this case, homophily becomes homogamy. An *individualistic* interpretation suggests that subjective freedom in choosing a spouse involves interiorized, reciprocal selection, whereas a *structural* interpretation suggests that subjective freedom in choosing a spouse applies only within the margin of freedom allowed by the actor's sociability network.

The homophilic choice hypothesis receives much attention from analysts. It is an attractive model, because it can appeal to macro sociologic postulates: Social stratification imposes strict limits on chances of contact between different categories. However, some social interactions are possible precisely by virtue of a difference, or a gap, between the partners. Examples are helping relations that require a particular resource one partner has, or depend on a genealogic gap.[8]

Strength

Some observers have tried to rank social links and distinguish those that bind more strongly than others. To do this the researcher must construct an indicator from the available information. A variety of attempts have been made, often with doubtful results. Mark Granovetter's now famous distinction between strong and weak ties (Granovetter, 1973) enabled him to identify ties with a specifically sociometric function, acting as *bridges* between cliques that are not otherwise linked. However, *strength* has little intrinsic consistency.

Sense of Affective Closeness

The interviewer can ask the respondent for his or her subjective appreciation in terms of closeness, intimacy, or trust. From this information, the examiner can identify the "core network."

One can regard each relation as a *specific observation* and make statistical descriptions of the relative frequencies of their characteristics. This has shown that reciprocal relations are more common in some networks than in others. One can also use the psychosociology of relationships or certain role theories to *explain*, for example, the possibility of particular resource exchanges by the characteristics of the partners' roles in the relation (Petite, 2002; Freeman and Ruan 1997; Bonvalet *et al.*, 1993; Fischer, 1982a). These analyses are innovative in two ways. First, they concern interactions between two coordinated actors and produce somewhat unusual statements because the relationship itself is the subject of propositions, such as "a relationship is more easily established

[7] All the resources (e.g., time, energy, emotional cost, intimate confidences, money) that an actor devotes to maintaining a relationship and that cannot be taken back if the relationship is broken off.

[8] Ferrand and colleagues (1999) published a study of homophilic choices in different kinds of relations in France according to social class.

between two similar actors." Second, each actor may be a term in several relations observed, which raises the problem of co-dependency between observations; this suggests that multilevel statistical procedures should be used (Snijders and Bosker 1999; van Duijn *et al.*, 1999; Snijders *et al.*, 1995). However, we shall not further pursue the question of dyad analyses.

2. Composition of Personal Networks

In a structural approach, it is considered that relations may depend on each other. This is easier to analyze if the composition of the personal network is examined (i.e., the ways in which one individual's relationships are compatible, complementary, cumulative).

a. Network Size

The simplest property of a personal network is its size—the number of relations of a particular kind that an actor can accumulate. This property is socially determined: Numbers of links of different kinds cited routinely vary according to the age, sex, and status of the respondent. Moreover numbers of relations are not a mere matter of quantity, because they imply qualitative thresholds that, above two partners, involve interdependence. Zero relations implies a psychological inhibition, practical impossibility or a social prohibition; one means a mirror relationship and, for *ego*, dependency; two provide *ego* with an alternative, create competition, make a *tertium gaudens* possible and, with *ego*, form a triad—the smallest pertinent structure for sociology (Fischer, 1982a; Simmel, 1917 [1981]; Blau, 1964; Caplow, 1968 [1984];). With more than two relationships *ego* has resource allocation problems. When the number of relations named is small, that number in itself can be interpreted as a basic structural indicator rather than a quantitative measure. Wherever at least two people are named, this is a structural property that raises the question of competition and gives *ego* some choice: what happens in one relationship depends on what happens in the other.

Taken on its own, size is a very crude indicator. It basically indicates the respondent's social isolation or expansiveness, his or her *ability* to maintain active links with others and hence (if the relations examined involve exchange) his or her overall potential for mobilizing relational resources.

b. Composition in Terms of Individuals' Attributes

If the respondent counts the partners possessing this or that sociodemographic attribute (e.g., sex, age, qualification), one can see how many *alters* have this or that status or nationality, how many live far away, and so forth (de Federico de la Rúa, 2004b; Ferrand, 2002).

c. Composition in Terms of Relational Characteristics

Composition in terms of relational characteristics refers to the distribution of relations according to the characteristics mentioned in the last section. In examining roles, the interviewer asks how many kin or neighbors are in the network. This reveals, for example, that manual workers' personal networks include a higher proportion of family members than do those of senior executives. Examining content, how many people in the network provide emotional support, and how many are leisure companions or advisers? This may show that the respondent invests most in kinship network, friendship network, or exchanges among workmates. The proportion of multiplex relations—those fulfilling more than one role—indicates the strength of connection between different networks (e.g., kinship and exchanges between workmates).

3. Structural Characteristics of Networks

We now leave simple counting of relations to look at the sociometric organization of relations. A first step in structural sociology is to describe the links between relations and the way they form networks with identifiable characteristics. With more than ten actors, any graphical representation is unintelligible, misleading or defies any proper theoretical interpretation. The best solution is therefore to build numerical indicators from the matrix of relations between actors. Jacob Moreno (1934 [1970]) had begun to create such indicators. They were perfected and became far more numerous when graph theory was used as a formal language to suggest new measurements (Degenne, 1978; White *et al.*, 1976; Harary, 1969; Flament, 1963, 1965; Harary *et al.*, 1965). They were operationalized once data processing capacity made it possible to make iterative calculations on matrices. Highly sophisticated indicators were created, sometimes at the initiative of computer scientists and formalists rather than in response to substantive questions. In this chapter, we mention only a few to illustrate typical approaches used in introductory course books (Freeman, 2004; Mercklé, 2004; Molina, 2001; Lazega, 1998; Pizarro, 1998; Requena Santos, 1996; Rodriquez, 1995; Scott, 1992; Knocke and Kuklinski, 1982; Chiesi, 1980, 1981), and we suggest some reference works on methodology for readers wishing to specialize (Carrington *et al.*, 2005; Doreian *et al.*, 2005; de Nooy *et al.*, 2004; Wasserman and Faust, 1996; Degenne and Forsé, 1994; Freeman *et al.*, 1991; Marsden, 1990; Marsden and Lin, 1982).

a. Density: A Global but Poorly Informative Indicator

Network density is the ratio between the number of links between actors and the maximum possible number of links.[9] When many actors are interlinked, density is high, and it gives the image of a cohesive group, with all the associated theoretical connotations, such as strong social control and a sense of belonging. However, density is only a very rough measure of network structure.

b. Distance between Actors: Network Openness versus Disconnectedness

Density only counts *direct* links between two actors. However, common sense readily identifies a crucial feature of networks, which is that links are themselves interconnected, forming *paths* that can *indirectly* link two actors. Instead of a matrix showing the presence or absence of a direct link between two actors, one can calculate a matrix showing the *minimal distance* between two actors. This feature has attracted the curiosity of researchers who, like explorers trying to reach the edge of the world, wondered whether networks have boundaries or are open.

If two actors cannot contact each other (even by a large number of intermediaries), then the network is said to be *disconnected*. This is a normal property when all relations are exclusive (e.g., some kinds of commercial relations, monogamous love relations). If the network is open, one may ask how far any actor is from any other. Stanley Milgram's "small world" research (Milgram, 1967) suggests that two randomly chosen individuals can contact each other by a relatively limited number of intermediaries (in their 1969 survey, Travers and Milgram found an average of five intermediaries[10]). These findings have proven useful in connection with contagious disease transmission, but less useful for social science research. The substantive question is to *decide* whether, with the particular type of relationship examined, one can consider that two actors at distance n are still effectively connected or whether the link is a purely formal one.

c. Cliques: Subsets of Strongly Interconnected Actors

A *clique* is a subset of actors in a network, such that each clique member is linked to all others (density within the clique is 1). This mechanical criterion is rarely applicable, and has been replaced by less rigid requirements, e.g. that each be connected with others at least by indirect paths of a given maximum length ($n \le 2$ or $n \le 3$). The term used is *n-cliques*. This reticular clique notion is reminiscent of the sociologic concept of *group*. In practice, however, cliques may overlap or include smaller cliques, requiring a more complex concept of "group" that takes subgroups into account.

For any study of circulation processes (e.g., of information, goods, help), connectedness shows whether there are breaks in the overall structure,[11] and *n-cliques* identify subsets within which exchanges can take place, taking paths that are limited by the particular circulation constraints of the resource. These characteristics are far more differentiating than density.

d. Rare and Crucial Connections: Bridges and Break Points

A network may consist of two cohesive cliques, with only one or just a few links between them. Such links are called bridges. The existence of a bridge is a major structural characteristic, because without it the network becomes disconnected; it is unique (or rare) but crucial. Bridges counteract the paradoxical risk run by networks that are highly connected locally (in dense cliques) of being disconnected on a broader scale (if the cliques are not interconnected). For many purposes, networks made up of cliques with no bridges are not very operational. Mark Granovetter has suggested that an urban district made up of dense but poorly interconnected micro-neighborhoods would not have much capability to resist the threat of urban renewal (Granovetter, 1982).

[9] If P is the number of actors (network size) and L the number of actual links (here assumed to be symmetric links), density D is obtained as follows:

$$D = \frac{L}{(P \cdot (P-1))/2}$$

The denominator gives the number of squares in the total matrix, minus the diagonal of ego-centerd links: it indicates the number of possible choices. However, because reciprocal choices have been counted, only a half-matrix can be filled in, so the denominator is divided by 2. The numerator indicates the number of actual existing links.

[10] There has been new research into the "small world" idea (Newman, 2003; Barbási, 2002; Watts, 1999).

[11] Attention must be paid to *boundary effects* introduced by observation. If the respondent is asked to list no more than five friends, the resulting network may appear disconnected, but if 10 friends were listed, it might prove not to be.

e. Describing Positions in the Structure: Centrality Indicators

Centrality defines a position in the overall organization of relations in the network's structure. Three very different indicators are used to measure it. Let us call an actor *ego* and consider several propositions:

(a) *Ego* is central if directly linked to a large number of other actors: *degree centrality* counts the number of actors with whom *ego* has a direct link.

(b) *Ego* is central if linked to central actors. In practical relational terms, this formula is far more economical than the first one, but it ultimately records the same characteristic: the direct or indirect ability to contact numerous other actors.

(c) *Ego* is central, if able to prevent interactions between a large number of other actors because the paths that link them go through *Ego*; this is *betweenness centrality*.

f. Subsets of Actors Occupying Similar Positions: Classes of Structural Equivalence

Classes of structural equivalence are subsets of actors who maintain relationships with approximately the same other actors throughout the network. They need not be in contact with each other. Because the members of such a class have the same relational profile, that class is said to form a position in the structure. We can suggest a synthetic model of this structure by describing the typical relations between the few positions so identified; the very extensive matrix of relations among actors is replaced by a synthetic matrix of relations between a small number of typical positions. Emmanuel Lazega describes the informal structure of a law firm of 71 lawyers by grouping them into 11 different classes of actors who were, in a similar way, givers or receivers of professional advice (Lazega, 1998 p. 66; 2001 p. 103). For example, young trainees who were not linked to each other but who asked the same senior members for advice would be put in the same class of equivalence.

g. Modeling a Network: Counting Types of Triads

We are looking at the *structural* properties of networks that may be of any size. However, with increasing network size an exhaustive description of the components of a network soon becomes unintelligible and comparisons between networks impossible, as the number of different combinations of elements to be considered becomes astronomically large. Various ways are sought to synthesize the information. One way is to divide the network into triads: three actors and any relations between them, reciprocal or not.[12] This is the smallest subunit of a network that can be examined structurally (i.e., from the standpoint of how *one relation may be linked to other relations*. They present only a limited number of significant combinations (Fig. 137–1).[13] Any triad can be seen *from outside*, as in Figure 137–1 and *from inside* if one puts oneself in the place of one of the actors involved. A triad as seen from the standpoint of any one of its three members is called a *triplet*. The sociometric organization of a network is described by counting the frequency of each type of microstructure encountered (triad or triplet). This method gives a fairly accurate picture of structural properties of a network.

If the subject of the research is an open network (e.g., friendship or kinship networks, sexual networks) observed by a sample of independent personal networks, these microstructures will be counted using standard statistical procedures.[14] Conversely, if the research subject is a complete network in which one is observing all relations between the members of a defined social group, one must use procedures specifically developed to take account of the interdependence between the relationships observed.[15]

By observing these very elementary structures one can analyze basic propositions about the constitution of a social structure. For example, it is possible to examine the strength of the principle of *transitivity*, which is the extent to which the existence of relation *Rx* between P1 and P2, and between P2 and P3, implies the existence of the same type of relation between P1 and P3. Similarly, we can examine how relations *exclude* each other because they compete or because

[12] Any network, whatever its size, can be broken down into *x* triads, each formed by a different trio of actors.

[13] For a detailed presentation of structural analyses based on triads, see Burt (1980) or Holland and Leinhardt (1978, 1979).

[14] Preferably, multilevel analyses to take account of the fact that every relation observed depends on the ego it belongs to.

[15] Since the first U|MAN model (Uniform distribution, conditional on the dyad count: number of Mutual dyads, number of Asymmetric dyads, and number of Null dyads) suggested by Holland and Leinhardt (1976) and the *p1* model developed by Holland and Leinhardt (1981), proposals have gradually become more sophisticated. The model chosen must depend on the question asked; the *p2* model (van Duijn *et al*., 2004; Lazega and van Duijn, 1997; van Duijn, 1995) can explain the existence of a link between two actors based on the actors' relational characteristics (e.g., an analogous centrality) or on their attributes (e.g., sex, status, age). The *p** model (Snijders, 2002; Wasserman and Pattison, 1996; Frank and Strauss, 1986) makes it possible to count the types of triplets (properly speaking) in a network. The Siena model (Snijders, 2002) can be used to study the evolution of a network by observing typical changes in triads (e.g., closure of a network by transitivity, as when "my friends' friends become my friends") in combination with the effects of actors' attributes and relational preferences.

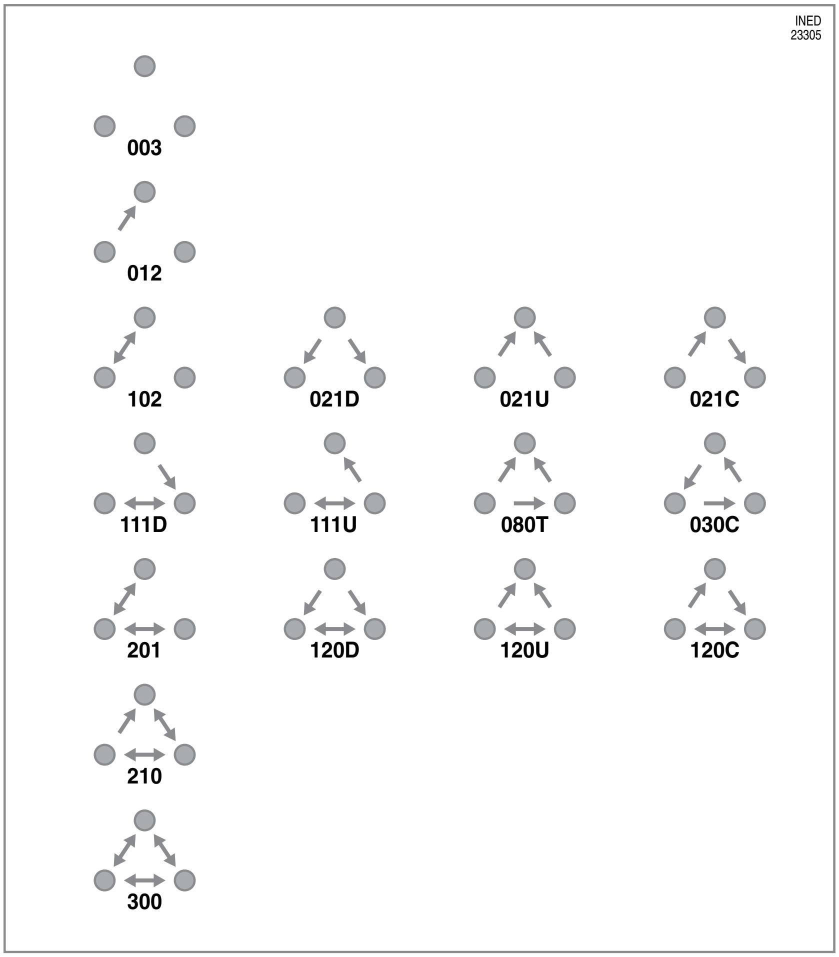


FIGURE 137–1 Isomorphic classes of triads according to Paul Holland and Samuel Leinhard (1976).

there is a rule forbidding them. With resource circulation and exchange, we can observe simple *reciprocity* (i.e., within a dyad) or the elementary form of generalized exchange in which A gives to B, who gives to C, who gives to A. These are major indicators of forms of solidarity and social control.

h. Complete Networks and Personal Networks

Although most of the characteristics we have just mentioned were defined in the context of complete network analysis, many can also be used to characterize personal networks. High density will mean that the actors are involved in networks where knowing each other can facilitate social control and make it easier to obtain help. However, density is a weak indicator. It is more discriminating to ask whether, apart from *ego*, the *alters* are all in direct or indirect contact with each other, or if the network is a disconnected one consisting of subsets isolated from each other. This may be the micro-social expression of a collective organization in which the domains of kinship, residence, and work are separate. Alternatively, it may be seen as resulting from *ego's* relational strategy: *alters* not connected with each other are more likely to be simple relations (*weak* links) giving access to *varied* sources of information (Granovetter, 1974). The investigator could stress the fact that, if the alters can only communicate through *ego*, *ego* occupies a privileged position as central intermediary. Ronald Burt (1992, 1995) has shown that a person can benefit from occupying this kind of *structural hole* position in a network.

This preliminary glimpse has identified the main network properties and relational characteristics that researchers have examined. The formal complexity of networks offers a vast domain to the methodologic imagination, and the specialized literature has produced an abundance of specific measurements. However, these basic indications are enough to give an idea of the many formally and qualitatively different types of possible measurements. Even these basics can be used to empirically validate many propositions suggested by the sociologic imagination Wright Mills extolled (Mills, 1959 [1970])—this being the sole source of actual knowledge.

III. SURVEY PROTOCOLS

Survey methods have been developed in which respondents are prompted to name and describe their relationships, enabling analysts to study networks. The problems posed by descriptions of relationships are the same whether the network to be analyzed is a complete one or a personal one, but naming relations is a more delicate matter when the network being explored is a personal one. Below, we mainly examine techniques that can be used in sample surveys for describing personal networks. These seem to be the most useful for addressing the questions about large populations that are the core of demographic analysis.

Every individual has some relationship with a large number of partners. In urban situations with a high degree of division of labor, social division, spatial specialization, and sharply segregated activities, relations take place in many different, often separate social contexts. An individual's complete relational universe is extensive and complex. Experiments have shown that the network a person accumulates over his or her lifetime may vary in size between 2000 and 5000 people (Pool and Kochen, 1978; Gurevich, 1961). The active relationships in an individual's personal network at a given time have been estimated at 290 on average, with a standard deviation of 250, people in the higher social strata having more contacts (McCarty *et al.*, 1997; Freeman and Thompson, 1989; Killworth and Bernard, 1978). Looking at an individual's close circle and most intimate contacts, figures of about 19 are given for close associates, using a wide range of criteria (Fischer, 1982a), with 3 to 5 taking specific criteria such as people with whom the respondent discusses important problems (this is roughly the number found in the General Social Survey's "confidant" module), and fewer than 3 confidants with whom emotional and sexual matters are discussed (Ferrand *et al.*, 1993). Only a *clinical* analysis could hope to explore an individual's complete personal network, the total sum of his or her relations.[16]

1. Selecting Relations: Name Generators

Most studies do not explore an individual's whole network but *parts* of it, *selected extracts* so to speak. The object of the analysis is an effect of the method—a *construction resulting from observation*. This is more radically true, and has more important consequences, than in other research fields. Because of the constructed nature of the findings, special care is needed when comparing different surveys, and conclusions and generalizations always need to be controlled by a careful examination of the observation procedures.

If the research concerns a fairly small social group (e.g., a company, a school class), the first methodologic preoccupation will be how to define the relations to be observed. If the research concerns open networks (e.g., friends, kin, neighbors) in large populations (i.e., societies), which is the approach on which we focus in the following sections, then observations must be selected in two successive steps. The first step is the procedure for choosing the survey respondents—the classic problem of designing a sample. The second step, which is specific to social network analysis, is the procedure by which each respondent is asked to select relations of a particular kind from their system of mutual acquaintances. In this way an ad hoc experimental sample of relations is extracted.

The problem here is how to get respondents to select what is wanted. It is a crucial question, because any distortion will not just result in variations in responses to a question, but will introduce what can be thought of as a *sampling bias*: it will change the lists of objects observed (an individual's relations). Respondents will select different types of relations depending on the instructions given. The questions asked to obtain a list of names determine the types of people regarded as network members, and so determine the operational definition of the network that is analyzed afterward (Bass and Stein, 1997; Campbell and Lee, 1991; McCallister and Fischer, 1978). The first methodologic preoccupation is *how to elicit the subset of relations that is relevant for the purposes of the study*. This is always the most important step for constructing, defining, and delimiting the networks studied.

Various types of questions can be asked to prompt a respondent to list their relations. They are called *name generators*, as the respondent replies by naming people according to a selection criterion (i.e., a particular type of relationship they have with them). The name generators are the most important questions in a social network survey, because the amount and quality of information obtained largely depends on them.

The process must select from among all the respondent's relations the ones that form the particular network the survey is intended to examine. The criteria applied vary; procedures may be based on *content* of the relation or a typical interaction (e.g., Who do you

[16] Jeremy Boissevain (1973) describes the "networks of two people" with 1300 and 840 relations, respectively.

discuss work problems with?, Who do you borrow money from?, Who do you ask to baby-sit?'); on a *role* (e.g., Who are your family members? Your friends? Neighbors?); or on *subjective assessments* about relations (e.g., Who are your closest or most intimate friends?). As well as the criterion for type of relation, an analyst can use *temporal* criteria (e.g., limiting the choice to people the respondent has frequented in the past 6 months) or *numerical* criteria (e.g., limiting the number of people mentioned to three, five, or more). The type of name generator and the limits imposed will depend on the purpose of the survey. Some surveys combine several name generators, as in the practical example set out below to illustrate in detail the types of questions used. After this example, we review a few classic or otherwise noteworthy surveys to show the particular data gathering techniques they have used.

2. Survey Example: The Social Survey on the Networks of the Dutch between 1999 and 2000

The Social Survey on the Networks of the Dutch (SSND) examined social network content and particularly its role in neighborhood life and in the workplace. The research sought to answer theoretical and methodologic questions about social capital, defined as the resources present in personal social networks. The aim was to measure the general distribution of social capital in a population and the efficacy and specificity of that social capital according to aims and contexts (Van der Gaag and Snijders, 2004; Flap, 1999, 2002; Lin, 2001). The survey was jointly conducted by the universities of Utrecht, Groningen, and Amsterdam. It used a representative sample of 1004 people from the Dutch population.

The emphasis was on access to the resources found in a social network rather than their use.[17] The fact that a resource exists in the network is not enough for it to be considered usable social capital. Hypotheses have to be formulated as to its availability, i.e. how *alters* may allow *ego to* use their resources. The authors' hypothesis was that resources are potentially available through a relation depending on the investment *ego* has made in that relationship in the past. Because the survey used a general population sample, the authors had to take care in choosing the resources referred to in their questions; respondents from different social milieus had to see them as relevant, useful resources.

The SSND survey uses three tools for measuring social capital, which the authors call the *name generator/interpreter* (McCallister and Fischer, 1978), the *position generator* (Lin and Dumin, 1986), and the *resource generator* (Snijders, 1999). The first two are adapted from the classic literature on social networks and social capital, and the third was used for the first time in this survey.

a. Name Generator or Interpreter

This is the oldest type of tool and is used in many surveys (Fischer, 1982a). Each *ego* is asked to list the people with whom he or she has exchanged (given or received) a particular resource. Each question about the exchange of a given type of resource *generates names*, identified by first name or initials, which are set down in a list. This list is then used to ask *ego* to characterize their relationship with each person using *name interpreter* questions. The list is then set out as a table to facilitate systematic coding.

Collecting name-generated data is a heavy burden for both interviewer and interviewee, because the number of data that are to be collected can become enormous, especially when many name-interpreting questions are included and larger social networks are encountered. However, it is especially recommended when detailed analyses of social network contents in specific populations are pursued.

Boxes 137–1 and 137–2 below present the questions used as name generators and name interpreters respectively. Tables 137–1 and 137–2 illustrate the grids produced for interpreting results.

b. Position Generator

This method, which follows Nan Lin's theory of social capital (Lin, 1982, 2001), measures relational access to what are regarded as high-prestige occupations and therefore, in a stratified and hierarchical view of society, regarded as social capital (Lin, 1982, 2001). The availability of each contact is measured by the strength of the link (e.g., family, friend, acquaintance) connecting *ego* to a person in each occupation. Box 137–3 presents the question asked and the data collecting grid.

This questionnaire is quick and easy to administer. However, it gives only an indirect measure of social capital and does not provide a basis for characterizing the relations because no reference is made to specific people, only to occupations and types of roles. For different populations in different countries, the choice of occupations needs to be adjusted.

[17] The actual use of social capital can be studied, but it would require a large number of additional questions. For the reasons why this particular choice was made, see Van der Gaag and Snijders (2002).

BOX 137–1 Name Generators: Questions to Elicit Names from the Respondent

1. *Contact-person for current/last job*

Now we get to a question that is important to discover your personal relations. I would like to know who helped you getting this job. May I have the first name and the first letter of the family name of the person who helped you get your current/last job? We will come back to this person later.

2a. *Asking advice*

If you have a problem at work, whom do you go to for advice?

2b. *Giving advice*

How is it the other way around? Are there also people who come to you for advice regarding problems that they have at their work?

3. *Sour social capital*

At work people not only cooperate, but also bother each other. How is that in your situation? Has any of your colleagues lately disturbed you in doing your job?

4. *Direct colleagues/cooperation*

Who are the two colleagues with whom you work most often?

5. *Boss*

May I have the first name and the first letter of the family name of your boss?

6. *Help to get house*

I would like to know the first name and the first letter of the family name of the person that helped you get this house, or from which you directly bought the house.

7. *Help with small jobs in and around the house*

If you are busy with a small job at home, and you need someone that gives you a hand—for instance, if you need furniture moved or a ladder held, whom do you ask for help?

8. *Keys*

Is there someone from outside your household that keeps a spare key to your house?

9. *Direct neighbors*

Who are your direct neighbors? We mean the people that live closest to you; for instance people live directly right, left, above or under your home? I would like to have two names of the people living directly next to you. For clarity, can you also give me the numbers of their address?

Neighbors are from house number ______ and ______

10. *Visiting others*

Many people sometimes visit others in their leisure time. Who do you go to for a visit?

11. *Core network*

Life is usually not only about going out and enjoying company. Everybody needs someone to talk about important matters from time to time. With whom did you discuss important personal matters during the last 6 months?

12. *Open network question*

Let's go through the list of names we have made together. Is there anybody else who is important to you and whose name is not yet on the list? If yes, I would like to add this person to the list. What are the activities you usually share with this person?

TABLE 137–1 Example of Register Used to Record Responses Describing Partners and Relations

Person no.	List of names elicited in name generators 1 to 12	Sex	Age	Role1	Role2	Role3	Religion	Other
1								
2								
3								
4								
5								
6								
7								
8								
9								
. . .								

BOX 137–2 Name Interpreters: Questions for Describing Relationships

The following questions all refer to network members identified with name generator questions 1 through 12 in Box 137–1. I would like to go through the list of people we have made once again.

1. Which of these persons is a woman?
2. How old are these persons?
3. How are you connected to these persons? (max. 3 role relations)
(1) partner, (2) parent, (3) child, (4) parent in law, (5) brother/sister, (6) another family member, (7) friend, (8) boss, (9) direct colleague, (10) another colleague, (11) someone who is working for you, (12) someone from the neighborhood, (13) direct neighbor, (14) a fellow club/association member, (15) acquaintance
4. Do you know the religion of these persons?
(1) Roman Catholic, (2) Netherlands protestant, (3) reformed, (4) other, (5) none, (6) I don't know
5. Which of these persons is married, or lives together with a partner?
6. Which of these persons has children that still live at home?
7. How often do you usually have contact with these persons? (six items for frequency)
8. How long have you known these persons?
9. Where, and at which opportunity did you first meet these persons?
10. Where and at which opportunity do you usually meet each other nowadays? (The list that follows is used for question 9 and 10.)
(1) during my education, (2) at a club or association, (3) at work, (4) at kin's homes, (5) at friends' homes, (6) at my place, (7) at their place, (8) in the neighborhood, (9) at a place to go out, (10) in church, (11) during holidays, (12) at a party, (13) elsewhere
11. Which of these persons lives within a radius of 5 kilometers from your house?
12. On a scale from 1 to 5, can you indicate how much you like these persons? Mark persons you like very much with a 5 and persons you don't like with a 1.
Interviewer: record corresponding marks in column Liking on list A.
(1) don't like, (2) don't especially like, (3) quite like, (4) like, (5) like very much
13. Can you also indicate how much you trust these persons? Mark persons you trust very much with a 5 and persons you don't trust with a 1.
14. Do you think that 5 years from now you will still have a relationship with these persons? With whom do you think you will not have a relationship 5 years from now?
15. How intensive is the relationship with these persons? Mark persons with whom the relationships is very intensive with a 5, and persons with whom the relationship is weak with 1
16. What kind of education did these persons complete?
(1) basic lower education, (2) high school, (3) higher vocational, (4) college or university
17. Which of these persons has a paid job at the moment?
18. What is the occupation of these persons? If they do not have work at the moment, I would like to know their last occupation.
19. Finally, I would like to now whether some people you know also know each other. How well do the following persons know each other? Do they know each other, and if yes, do they get along well or do they avoid each other?
1 = persons avoid each other, 2 = persons don't know each other, 3 = persons hardly know each other, 4 = persons know each other well, 5 = persons know each other well and get along well

c. Resource Generator

A third tool was invented for this survey (Snijders, 1999) (Box 137–4). It combines the positive aspects of the position generator (i.e., economy and internal validity) and the name generator (i.e., detailed information about resources). The questionnaire used for asking the respondent about their access to resources is structured like the position generator: a set list of resources, each one supposedly representing a subset of social capital covering several spheres of life. The potential availability of each of these resources is recorded by the strength of the tie between *ego* and *alter* (e.g., family, friend, acquaintance).

This questionnaire is quick to conduct and gives an easy representation of social capital. It can be directly processed statistically and quickly interpreted. Because social interaction and network formation are culturally determined, the content of the resource generator must be adapted to the population concerned.

TABLE 137–2 Sample Matrix Used to Record Responses Describing Relations between Partners

How well do persons no. 1 to no. 6			. . . know persons no. 2 to no. 6				
NG	Name[a]	No.	No.1 and no. 2	No. . . . and no. 3	No. . . . and no. 4	No. . . . and no. 5	No. . . . and no. 6
2a		1					
2b		2					
3		3					
7		4					
8		5					
11		6					

NG, name generators.

[a] In the Name column, the respondent puts the first name mentioned in reply to each of the name generators listed in column NG. In the boxes of the grid, the person indicates whether partner 1 knows partner 2, and so on.

BOX 137–3 Questions for the Position Generator

Before asking you more questions about your work and your daily activities, I would like to know what are the occupations of the persons you meet and have contact with. I have here a list of different occupations that people can have. Does anyone in your family have one of those occupations? Anyone among your friends? Among your acquaintances? By acquaintance, I do not mean the salespersons you come across in the shop, but somebody that you have a small talk with or would have a small talk with if you meet him/her on the street and that you know by his/her name.

Job/function	Family	Friend	Acquaintance	No.
1 Physician	1	2	3	0
2 Cook	1	2	3	0
3 Engineer	1	2	3	0
4 High ranking public servant	1	2	3	0
5 Construction worker	1	2	3	0
6 Director of a company	1	2	3	0
7 Etc., up to 30 occupations				

3. Particular Features of Some Exemplary Surveys

There have been a number of sample surveys designed to describe personal networks from different angles. They concern a range of issues, such as how information is mobilized in the search for work (Requena, 1991; Granovetter, 1973), the influence of personal network structure on promotion in an organization (Burt, 1992), or political attitudes and sense of belonging with regard to the European Union (de Federico de la Rúa, 2003). They may also look at macrosocial factors that modify networks, such as. the effects of political regime change in East Germany (Völker, 1995) or the impact of urbanization (Fischer, 1982a). We present a few surveys to highlight their original methods or because their impact is well known.

a. East York

Barry Wellman's surveys of East York (Toronto, Canada) in 1968, 1978, and 1981 (Wellmann and Berkowitz, 1988) had three aims: to show that in an urban environment social integration does not simply mean integration in the home neighborhood; to explore the dynamics of interpersonal exchanges aimed at obtaining resources; and to back up a structural interpretation of actors' behavior and attitudes. The research is original for its successive use of in-depth interviews, face-to-face questionnaires and postal questionnaires and for the steps it took to observe the creation and dissolution of relations.

b. Detroit Area Study

Edward Laumann conducted a study in Detroit (Laumann 1973) to show the influence of the personal network on people's behavior and attitudes. It took a sample of 1013 white males between the ages of 21 and 64 years, born in the metropolitan Detroit area. The name generator is formulated as follows: "Now would you think of the three men who are your closest friends and whom you see most often. They can be relatives or nonrelatives, as you wish." Here the generator limits the selection to a single type of relation defined

BOX 137–4 Questions for the Resource Generator

Here I have a list of skills and resources. Does anyone in your family have this kind of skill or resource? Anyone among your friends? Among your acquaintances? By acquaintance, I do not mean the salespersons you come across in the shop, but somebody that you have a small talk with or would have a small talk with if you meet him/her on the street and that you know by his/her name.

	I. Do you know anyone who II. . . . and are you someone who . . .	No	Family	Friend	Acquaintance	Yourself
1	. . . can repair a car, bicycle, etc.	0	1	2	3	4
2	. . . owns a car	0	1	2	3	4
3	. . . can repair household appliances	0	1	2	3	4
4	. . . can speak and write in a foreign language	0	1	2	3	4
5	. . . can work with a computer	0	1	2	3	4
6	. . . can play music	0	1	2	3	4
	. . . etc., up to 20 skills or resources					

by a role label (close friends) and, within that, limits the answer to three names. For each friend named, the respondent is asked for sociodemographic information (age, occupation, educational level, how far away they live) and political preferences. *Ego* has to say how he perceives the mutual acquaintanceship links between his friends; this gives the basis for a simple analysis of the sociometric structure. These advantages counterbalance certain limits of the survey: close friendship is quite an ambiguous notion and is not representative of friendship in general, still less of the diverse types of relations that make up personal networks.

c. Northern California Community Study

Claude Fischer (1982a) set out to study the extent to which the size, composition, functional and cultural diversity of an urban area affected individuals' sociability and the emergence of subcultures. It was a contribution to the debate launched by the Chicago School pioneers. Its sample of territorial communities was stratified by degree of urbanization, and within that, random samples of housing units were taken, and 1050 English-speaking people were interviewed in 1977 and 1978. Claude Fischer wanted a relatively ample description of their personal networks, and used nine name generators based on precise exchanges: Who has helped you with a small job around the home? Who do you talk to about the way your job should be done? Who do you go out with socially or go to dinner with? Eight answers to each question were allowed. The questionnaire also allowed respondents to add "other people who are important to you." Claude Fischer then asked for information about each *alter*. The questionnaire then made a second selection, retaining only the first five people named in response to certain name generators. For this subnetwork, the respondent was asked for more ample information about the history of each relation and about linkages between the *alters*. With these data, quite refined sociometric indices could be used. This approach gives a fairly open description of the personal network, and a more thorough analysis of a small number of relations. It is a fruitful compromise between extensive enumeration and analysis of structural complexity.

d. Contacts between People

A 1982 survey on contacts between people conducted for the Institut National de la Statistique et des Études (INSEE), which is the French national statistics office, and the Institut National d'Études Démographiques (INED) by François Héran (1987, 1988), was designed to describe the sociability of the French from a representative population sample. A particular feature of the survey was that it described the practices of households taken as a whole (5900 households were used) and the sociability of one individual chosen at random within each household (4700 individuals). There were two levels of information. Another particularity was that, as in some *time budget* analyses, respondents were asked to note, for every quarter hour for 8 days, what they were doing and whether they had "talked with people not living in their household (strictly work-related discussions excluded). If several people were involved, note each one individually." The respondent is asked for the gender, first name, and age of each person they talked to. This "catch-all" name generator is the survey's strong point (describing exhaustively, on a representative basis, what people the respondent has relations with) but also its weakness, because it says nothing about the content of those relations.

e. General Social Survey

The General Social Survey (GSS),[18] conducted by the University of Chicago's National Opinion Research Center (NORC), is a periodic representative survey of the population of the United States (sample of about 1500 people) with core modules repeated every year and specific modules used only in particular years. The name generator used in the 1985 survey was straightforward: "From time to time, most people discuss important matters with other people. Looking back over the last 6 months—who are the people with whom you discussed matters important to you?" The specialists advising the NORC chose this question because it would be meaningful to any interviewee and would elicit a varied range of types of relations.

Respondents are then asked for information about the first five *alters* they mention: the characteristic roles of each relation (e.g., type of family relationship, friend, colleague, leisure companion) and which of them knows which others. Limiting names to five is not a problem because only 6% of respondents mention more than five people. Here the procedure examines the core network—a smaller and more specialized network than in the *contacts between the people* survey. However, the composition of this micro-network is still a good indicator of peoples' overall sociability. As Ronald Burt points out (Burt, 1990, p. 418), based on a precise analysis of the relations listed, "The GSS sociometric name generator is less a definition of relation content than it is a window through which one can view the respondent's interpersonal environment." One advantage of the procedure is that it is particularly quick and economical. Small relational modules on precise issues are also included. In 1987, for example, the GSS described three political discussion partners and the way in which *ego* perceived these partners' party support choices (Knoke, 1990). The module identified interpersonal consensus with regard to voting. In 1988, questions were asked about sexual behavior.

f. Lifestyles

Between 1988 and 1989, the INSEE and the Center National de la Recherche Scientifique (CNRS) conducted a joint survey on domestic production in French households (e.g., home cleaning and maintenance, feeding and caring for family members). The survey covered a representative sample of 6,807 households. One module took an original approach to examine exchanges of help in household tasks, because it examined a network of households rather than individuals, and recorded the reciprocity of the exchanges. The survey used 18 name generators for different types of help (e.g., baby-sitting, care functions, housework, loans of money) that had been exchanged, received, or given over the past year. For each household, respondents described the type of relation (e.g., kin, neighbor, friend) and a few sociodemographic attributes of the household head. This made it possible to study what types of relations allow people to exchange these types of help, and the way in which these particular resources circulate in our society (Degenne and Lebeaux, 1991).

g. Confidants for Emotional and Sexual Issues

The module on confidants for emotional and sexual issues was part of a larger survey funded by the French national AIDS research agency (Agence Nationale de Recherches sur le Sida [ANRS]), covering 20,000 individuals in France. Between 1991 and 1992, it was conducted among a subsample of 2343 people. The aim was to study the normative influence of a person's network of confidants on their sexual behavior. The central idea was that types of sexual activity are the result of transactions within the couple, involving normative references to, and models inspired or strengthened by, each partner's close relational environment (Ferrand and Mounier, 1993, 1998). The name generator used the question "apart from the person you live with, with how many people do you discuss your love affairs, sexual problems, sexually transmitted diseases or what goes on in your couple?" The number of people mentioned was noted, and the attributes of the first three: sex, age, occupation, duration of the relation, and frequency of interaction. The respondent was asked how they perceive the sexual behaviors of their confidants. The particularity of this survey is that it shows that people can be asked about relations involving very intimate subjects and taboos, and will give answers about quite private aspects of their confidants' lives.

h. The International Social Survey Programme's Social Networks and Social Support Surveys

The International Social Survey Programme (ISSP) coordinates major periodic surveys in several countries, making sure the data produced are comparable.

[18] For full, precise information about how the GSS works, see Marsden (1987) and Burt (1984).

The 1986 and 2001 surveys investigated social networks and social support systems. The 1986 survey used representative samples from seven countries.[19] It described frequency of encounter with close kin (parents, siblings, children), the size of the extended family network, and close friends. It identified the type of relation (spouse, parent, child, brother or sister, other kin, friend, neighbor, colleague, other relation) to whom the respondent turned first, and second, when in need of help in any of the six fields covered: help around the home and garden, illness, borrowing large sums of money, problems with spouse, feeling depressed, needing important advice. The 2001 survey was conducted in 34 countries[20] and concerned extended family, friends in different contexts (work, neighborhood, other), membership of organizations, information about job opportunities and questions about satisfaction, trust and political opinions. The comparative potential of the ISSP surveys is particularly interesting and has been exploited by Linton Freeman and Danching Ruan (1997) among others.

i. The Institut National d'Études Démographiques Survey on Close Associates and Kin

The 1990 Close Associates and Kin (*Proches et Parents*) survey conducted by Catherine Bonvalet on a representative sample of 1946 people from the adult French population (Bonvalet *et al.*, 1993) followed on from earlier research into family networks (Gokalp, 1978). It had three aims: (1) increase knowledge of the extended family (2) explore the affinity network (as a rule, the kinship group a person can describe is broader than the group with which they have close ties); (3) study the social habits of the family and friendship network. The method used was to record the resources respondents could mobilize through their network of family and friends: help and support, transmission of goods, exchanges of services and information. There were three parts to the questionnaire. The first concerned help received and given by *ego* over their lifetime in the fields of education, employment and housing. The second described the network of close associates using a name generator: "Now we are going to talk about the people you regard as being close to you. Can you list them, regardless of whether they are friends, members of you family or of your spouse's family?" The third part, which is highly original, consisted of drawing up a family record book over five generations (from grandparents to grandchildren and including uncles, aunts, cousins, nephews, nieces and their spouses). In this way the survey identified three types of network: extended family, those considered "close," and the self-help network. The three do not necessarily coincide.

CONCLUSIONS

To conclude, it must be said that not all fields have been covered. We have not discussed survey methods on complete networks or the use of archive data for reconstituting complete networks.[21] The questions raised by data analysis on complete or personal networks have not been examined in depth; nor have the particular statistical approaches used for relational matrices. We set out the general principles and then focused on survey methods for large populations that use common demographic research procedures among other methods.

Surveys on personal networks certainly have much in common methodologically with studies designed to give a detailed description of kinship groups. However, this does not mean that ultimately they are much like conventional surveys. In the first place, at the strictly empirical level, many social network studies have developed comprehensive analyses of different aspects of socialization. The same observation takes into account relations formed through work, in the neighborhood, or through membership of an organization, as well as kin. Such studies never claim to give an exhaustive description. They observe interdependencies between different types of relations. This leads us to a second, more theoretical distinction. Social network analysis methods can be used for research inspired by a wide range of theoretical paradigms, but it is fair to say that the structural approaches are based on the hypotheses that can best exploit the demonstrative potential of social network observation. They lead us to define an actor less in terms of their individual attributes than their position in a particular set of intersecting relational systems. There the actor will find resources, be subject to control and prioritize their preferences. Each actor is limited by the position they occupy in these systems' struc-

[19] Australia, 1250 individuals surveyed; Austria, 1027; United Kingdom, 1416; Hungary, 1747; Italy, 1027; United-States, 1470; West Germany, 2809.

[20] South Africa, 2563 individuals surveyed; Germany, 936 in the West and 433 in the East; Australia, 1352; Austria, 1011; Brazil, 2000; Canada, 1115; Chile, 1504; Cyprus, 1006; Denmark, 1293; Spain, 1214; Finland, 1439; France, 1398; Hungary, 1524; Northern Ireland, 1407; Israel, 1053 Jews and 154 Arabs; Italy, 999; Japan, 1321; Latvia, 1000; New Zealand, 1146; Norway, 1560; Philippines, 1200; Poland, 1221; Czech Republic, 1200; United Kingdom, 912; Russia, 2000; Slovenia, 1077; Switzerland, 1001; United States, 1149.

[21] For example, scientific citation databases can be used to reconstitute scientific networks in different fields.

tures. They act, forming and abandoning relationships, and in so doing they help to reproduce or modify these systems locally.

Acknowledgment

This chapter was translated from the French version by Harriet Coleman.

References

Barabási Albert-Laszlo, 2002. *Linked. The new science of networks*. Cambridge, MA, Perseus Publishing, 230 p.

Barnes John A., 1954. Class and committees in a Norwegian Island parish, *Human Relations*, no. 7, p. 39–58.

Bass Lee Ann and Stein Catherine H., 1997. Comparing the structure and stability of network ties using the social support questionnaire and the social network list, *Journal of Social and Personal Relations*, vol. 14(1), p. 123–132.

Blau Peter M., 1964. *Exchange and power in social life*. New York, John Wiley, 338 p.

Boissevain Jeremy, 1973. An exploration of two first-order zones, *in*: Jeremy Boissevain and J. Clyde Mitchell (eds.), *Network analysis. Studies in human interaction*, p. 125–148. The Hague, Mouton, 271 p.

Bonvalet Catherine, Gotman Anne, and Grafmeyer Yves, 1999. *La famille et ses proches. L'aménagement des territoires*. Paris, INED and PUF, 296 p. (Travaux et documents, cahier no. 143.)

Bonvalet Catherine, Maison Dominique, Le Bras Hérvé, and Charles Lionel, 1993. Proches et parents, *Population*, vol. 48(1), p. 83–110.

Bozon Michel, 1998. Amour, désir et durée. Cycle de la sexualité conjugale et rapports entre hommes et femmes, *in*: Nathalie Bajos, Michel Bozon, Alexis Ferrand, Alain Giami, and Alfred Spira (eds.), *La sexualité aux temps du sida*, p. 175–234. Paris, PUF, 494 p.

Burt Ronald S., 1980. Models of network structure, *Annual Review of Sociology*, vol. 6(6), p. 79–141.

Burt Ronald S., 1982. *Toward a structural theory of action*. New York, Academic Press, 381 p.

Burt Ronald S., 1984. Network items and the General Social Survey, *Social Networks*, vol. 6(4), p. 293–339.

Burt Ronald S., 1990. Kinds of relations in American discussion networks, *in*: Craig J. Calhoun, Marshall W. Meyer, and W. Richard Scott (eds.), *Structure of power and constraint*, p. 411–451. Cambridge, Cambridge University Press, 495 p.

Burt Ronald S., 1992. *Structural holes*. Cambridge, Harvard University Press, 313 p.

Burt Ronald S., 1995. Le capital social, les trous structuraux et l'entrepreneur, *Revue française de sociologie*, vol. 36(4), p. 599–628.

Campbell Karen E. and Lee Barrett A., 1991. Name generators in surveys of personal networks, *Social Networks*, vol. 13(3), p. 203–221.

Caplow Theodore, 1968. *Two against one, coalitions in triads*. Englewood Hills, NJ, Prentice Hall, 183 p.

Carrington Peter, Scott John, and Wasserman Stanley, 2005. *Models and methods in social network analysis*. Cambridge, Cambridge University Press, 344 p.

Chiesi Antonio M., 1980. L'analisi dei reticoli sociali: teoria e metodi, *Rassegna Italiana di Sociologia*, vol. 21(2), p. 291–310.

Chiesi Antonio M., 1981. L'analisi dei reticoli sociali: un itroduzione alle tecniche, *Rassegna Italiana di Sociologia*, vol. 22(4), p. 577–601.

de Federico de la Rúa Ainhoa, 2003. *Réseaux d'identification à l'Europe. Amitiés et identités d'étudiants Européens*. Lille and Pamplona, Université de Lille 1 and Universidad Pública de Navarra, 565 p. (Doctoral thesis.)

de Federico de la Rúa Ainhoa, 2004a. L'analyse longitudinale de réseaux sociaux totaux avec SIENA. Méthode, discussion et application, *Bulletin de méthodologie sociologique*, no. 84, p. 5–39.

de Federico de la Rúa Ainhoa, 2004b. Los espacios sociales de la transnacionalidad. Una tipología de los modos de integración relacional de los migrantes, *Araucaria. Revista Iberoamericana de Filosofía, Política y Humanidades*, vol. 6(12), p. 120–140.

Degenne Alain, 1978. La construction et l'analyse des réseaux sociaux, *L'Année Sociologique*, vol. 29(2), p. 283–310.

Degenne Alain and Forsé Michel, 1994. *Les réseaux sociaux*. Paris, Armand Colin, 288 p. (English translation: *Introducing social networks*. London, Sage, 248 p.)

Degenne Alain and Lebeaux Marie-Odile, 1991. L'entraide entre les ménages, un facteur d'inégalité sociale? *Sociétés contemporaines*, no. 8, p. 21–42.

di Nicola Paola, 1986. *L'uomo non è un'isola. Le reti sociale nella vita quotidiana*. Milan, Francesco Angeli, 166 p.

Doreian Patrick, Batagelj Vladimir, and Ferligoj Anuska, 2005. *Generalized Blockmodeling*. Cambridge, Cambridge University Press, 402 p.

Ferrand Alexis, 1989. Connaissances passagères et vieux amis. Les durées de vie des relations interpersonnelles, *Revue suisse de sociologie*, no. 2, p. 431–439.

Ferrand Alexis, 1997. La structure des systèmes de relations, *L'Année sociologique*, vol. 47(1), p. 31–54.

Ferrand Alexis, 2002. Las comunidades locales como estructuras meso, *REDES. Revista hispana para el análisis de redes sociales*, vol. 3(4).

Ferrand Alexis and Mounier Lise, 1996. Talking about sexuality: An analysis of relations between confidants, *in*: Michel Bozon and Henri Leridon (eds.), *Sexuality and the social sciences*, p. 265–288. Aldershot, Dartmouth, 357 p.

Ferrand Alexis and Mounier Lise, 1998. L'influence des réseaux de confidence sur les relations sexuelles, *in*: Nathalie Bajos, Michel Bozon, Alexis Ferrand, Alain Giami, and Alfred Spira (eds.), *La sexualité aux temps du Sida*, p. 255–304. Paris, Presses universitaires de France, 494 p.

Ferrand Alexis, Mounier Lise and Degenne Alain, 1999. The diversity of personal networks in France; social stratification and relational structures, *in*: Barry Wellman (ed.), *Networks in the global village*, p. 185–224. Boulder, Westview Press, 377 p.

Ferrand Alexis, Mounier Lise, and Groupe ACSF, 1993. Paroles sociales et influences normatives, *in*: Alfred Spira and Nathalie Bajos (eds.), *Les comportements sexuels en France*, p. 171–179. Paris, La documentation française, 352 p. (English translation in *Sexual behavior and AIDS*. Alfred Spira, Nathalie Bajos, and the ACSF Group. Aldershot, Avebury, 1994, p. 140–148.)

Fischer Claude S., 1982a. *To dwell among friends: personal network in town and city*. Chicago, University Press, 451 p.

Fischer Claude S., 1982b. What do we mean by friends? An inductive study, *Social Networks*, vol. 3(4), p. 287–306.

Flament Claude, 1963. *Applications of graph theory to group structure*. Englewood Cliffs, NJ, Prentice-Hall, 142 p.

Flament Claude, 1965. *Théorie des graphes et structures sociales*. Paris, Gauthier Villars, 167 p.

Flap Henk D., 1999. Creation and returns of social capital: a new research program, *La Revue Tocqueville*, vol. 20(1), p. 5–26.

Flap Henk D., 2002. No man is an island, *in*: Emmanuel Lazega and Olivier Favereau (eds.), *Conventions and structures*, p. 29–59. Oxford, UK, Oxford University Press, 256 p.

Flap Henk D. and de Graaf Nan D., 1986. Social Capital and attained occupational status, *Netherlands Journal of Sociology*, no. 2, p. 145–161.

Forsé Michel and Langlois Simon, 1995. Analyse structurelle comparée du changement social. Le cas de la France et du Québec, *in*: Michel Forsé and Simon Langlois (eds.), *Tendances comparées des sociétés post-industrielles*, p. 215–261. Paris, PUF, 270 p.

Frank Ove and Strauss David, 1986. Markov graphs, *Journal of the American Statistical Association*, vol. 81(3), p. 832–842.

Freeman Linton C., 2004. *The development of social network analysis. A study in the sociology of science*. Vancouver, Empirical Press, 205 p.

Freeman Linton C. and Ruan Danching, 1997. An international comparative study of interpersonal behavior and role relationships, *L'année sociologique*, vol. 47(1), p. 89–115.

Freeman Linton C. and Thompson Claire R., 1989. Estimating acquaintanceship volume, *in*: Manfred Kochen (ed.), *The small world*, p. 147–158. Noorood, NJ, Ablex, 382 p.

Freeman Linton C., White Douglas R., and Romney A. Kimball, 1992. *Research methods in social network analysis*. Brunswick, NJ, Transaction Publishers, 530 p.

Galaskiewicz Joseph, 1985. *Social organization of the urban grant economy: a study of business philanthropy and non profit organizations*. New York, Academic Press, 286 p.

Gokalp Catherine, 1978. Le réseau familial, *Population*, vol. 33(6), p. 1077–1096.

Granovetter Mark S., 1973. The strength of weak ties, *American Journal of Sociology*, vol. 78(3), p. 1360–1380.

Granovetter Mark S. (ed.), 1974. *Getting a job: a study of contacts and careers*. Cambridge, Harvard University Press, 251 p.

Granovetter Mark S., 1982. The strength of weak ties: a network theory revisited, *in*: Peter V. Marsden and Nan Lin (eds.), *Social structure and network analysis*, p. 105–130. Beverly Hills, CA, Sage, 319 p.

Gribaudi Gabriela, 1992. La metáfora della rete. Individuo e contesto sociale, *Meridiana*, vol. 15(1), p. 91–108.

Gurevich Michael, 1961. *The social structure of acquaintanceship networks*. Boston, MA, MIT. (Doctoral thesis.)

Harary Frank, 1969. *Graph theory*. Reading, Addison Wesley, 274 p.

Harary Frank, Norman Robert Z., and Cartwright Dorwin, 1965. *Structural models. An Introduction to the theory of directed graphs*. New York, John Wiley.

Héran François, 1987. Comment les français voisinent, *Economie et Statistiques*, no. 195, p. 43–59.

Héran François, 1988. La sociabilité, une pratique culturelle, *Economie et Statistiques*, no. 216, p. 3–22.

Héran François, 1990. Trouver à qui parler, le sexe et l'êge de nos interlocuteurs, *in*: INSEE (ed.), *Données sociales 1990*, p. 364–368. Paris, Institut National de la Statistique et des Études Économiques, 458 p.

Holland Paul W. and Leinhardt Samuel, 1976. Local structures in social networks, *in*: David R. Heise (ed.), *Sociological methodology*, p. 1–45. San Francisco, Jossey Bass, 320 p.

Holland Paul W. and Leinhardt Samuel, 1978. An omnibus test for social structure using triads, *Sociological methods and research*, vol. 7, p. 227–256.

Holland Paul W. and Leinhardt Samuel, 1979. *Perspectives on social network research*. New York, Academic Press, 532 p.

Holland Paul W. and Leinhardt Samuel, 1981. An exponential family of probability distributions for directed graphs (with discussion), *Journal of the American Statistical Association*, vol. 76, p. 33–65.

Killworth Peter D. and Bernard H. Russel., 1978. The reverse small world experiment, *Social Networks*, vol. 1(2), p. 159–192.

Knoke David, 1990. *Political networks: the structural perspective*. Cambridge, Cambridge University Press, 290 p.

Knoke David and Kuklinski James H., 1982. *Network analysis*. Newbury Park, CA, Sage, 96 p. (Quantitative applications in the social sciences, 28.)

Laumann Edward O., 1973. *Bonds of pluralism: the form and substance of urban social networks*. New York, John Wiley, 342 p.

Lazega Emmanuel, 1992. Analyze de réseau d'une organization collégiale: les avocats d'affaires, *Revue Française de Sociologie*, vol. 33(4), p. 559–589.

Lazega Emmanuel, 1998. *Réseaux sociaux et structures relationnelles*. Paris, PUF, 128 p.

Lazega Emmanuel, 2001. *The collegial phenomenon: the social mechanisms of cooperation among peers*. Oxford, Oxford University Press, 346 p.

Lazega Emmanuel and van Duijn Marijtje, 1997. Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data. *Social Networks*, vol. 19(3), p. 375–397.

Lin Nan, 1982. Social resources and instrumental action, *in*: Peter V. Marsden and Nan Lin (eds.), *Social structure and network analysis*, p. 131–146. Beverly Hills, CA, Sage, 319 p.

Lin Nan, 1995. Les ressources sociales: une théorie du capital social, *Revue Francaise de sociologie*, vol. 36(4), p. 685–704.

Lin Nan, 2001. *Social capital: a theory of social structure and action*. Cambridge, Cambridge University Press, 278 p.

Lin Nan and Dumin Marie, 1986. Access to occupations through social ties, *Social Networks*, vol. 8(4), p. 365–385.

Lozares Carlos, 1995. La teoría de redes sociales, *Papers, Revista de Ssociología*, no. 48, p. 103–126.

Marsden Peter V., 1987. Core discussion networks of Americans, *American Sociological Review*, vol. 52, p. 122–131.

Marsden Peter V., 1990. Network data and measurement, *Annual Review of Sociology*, no. 16, p. 435–463.

Marsden Peter V. and Lin Nan (eds.), 1982. *Social structure and network analysis*. Beverly Hills, CA, Sage, 319 p.

McCarty Christopher, Bernard H. Russell., Killworth Peter D., Selley Gene A., and Johnsen Eugene C., 1997. Eliciting representative samples of personal networks, *Social Networks*, vol. 19(4), p. 303–323.

McCallister Lynneet and Fischer Claude S., 1978. A procedure for surveying personal networks, *Sociological Methods and Research*, vol. 7(2), p. 131–148.

Mercklé Pierre, 2004. *Sociologie des réseaux sociaux*. Paris, La découverte, 128 p. (Repères no. 398.)

Milgram Stanley, 1967. The small world problem, *Psychology today*, vol. 1(1), p. 62–67.

Mills C. Wright, 1970. *L'imagination sociologique*. Paris, F. Maspero, 235 p. (French translation of the 1959 edition.)

Molina José Luis, 2001. *El análisis de redes sociales. Una introducción*. Barcelone, Edicions Bellaterra, 123 p.

Moreno Jacob L., 1934. *Who shall survive?* Washington, DC, Nervous and Mental Disease Publishing Company.

Newman Mark E. J., 2003. The structure and function of networks, *SIAM Review*, vol. 45(2), p. 167–256.

Nooy de Wouter, Mrvar Andrej, and Batagelj Vladimir, 2004. *Exploratory Network Analysis with Pajek*. Cambridge, Cambridge University Press, 334 p.

Petite Ségolène, 2002. *L'entraide: des règles sociales aux opportunités du réseau personnel*. Lille, Université de Lille 1, 409 p. (Doctoral thesis in sociology.)

Piselli Fortunata, 1995. *Reti. L'analisi di network nelle scienze sociali*. Rome, Donzelli Editore, 334 p.

Pizarro Narciso, 1998. *Tratado de metodología de las Ciencias Sociales*. Madrid, Siglo 471 p.

Pool Ithiel de Sola and Kochen Manfred, 1978. Contacts and influence, *Social Networks*, vol. 1(1), p. 5–51.

Price Frances V., 1981. Only connect? Issues in charting social networks, *Sociological Review*, vol. 29(2), p. 283–312.

Requena Santos Felix, 1991. *Redes sociales y mercado de trabajo*. Madrid, Centro de investigaciones sociológicas and Siglo, 243 p.

Requena Santos Félix, 1996. *Redes sociales y cuestionarios*. Madrid, Cuadernos Metodológios no. 18. Centro de investigaciones sociológicas and Siglo, 57 p.

Rodríguez Josep A., 1995. *Análisis estrucural y de redes*. Madrid, Cuadernos Metodológios no. 16. Centro de investigaciones sociológicas and Siglo, 86 p.

Rusbult Caryl E., 1980. Commitment and satisfaction in romantic associations, *Journal of Experimental Social Psychology*, vol. 16(2), p. 172–186.

Scott John, 1992. *Social network analysis: a handbook*. Newbury Park, CA, Sage, 208 p.

Simmel Georg, 1917. *Grundfragen der Soziologie*. Berlin, Göschen'sche Verlagshandlung, 103 p. (French translation.: *Sociologie et épistémologie*, Paris, PUF, 1981.)

Snijders Tom A. B., 1999. Prologue to the measurement of social capital, *La Revue Tocqueville*, vol. 20(1), p. 27–44.

Snijders Tom A.B., 2002. Markov Chain Monte Carlo estimation of exponential random graph models, *Journal of Social Structure*, vol. 3(2), p. 2–40 (http://www.cmu.edu/joss/index.html).

Snijders Tom A. B and Bosker Roel, 1999. *An introduction to basic and advanced multilevel modelling*. London, Sage, 266 p.

Snijders Tom A. B., Spreen Marinus and Zwaagstra Ronald, 1995. The use of multilevel models for analysing personal networks: networks of cocaine users in an urban area, *Journal of Quantitative Anthropology*, vol. 5(1), p. 85–105.

Sprecher Susan, 1988. Investment model, equity, and social support determinants of relationships commitment, *Social Psychology Quarterly*, vol. 15(4), p. 318–323.

Stokman Frans N. and van den Bos Jan M. M., 1992. A two stage model of policy making. With an empirical test in the US energy policy domain, *in*: Gwen Moore and J. Allen Whitt (eds.), *Research in Politics and society*, vol. 4. *The political consequence of social networks*, p. 219–253. Greenwich, CT, JAI Press.

Travers Jeffrey and Milgram Stanley, 1969. An experimental study of the small world problem, *Sociometry*, vol. 32(4), p. 425–443.

van der Gaag Martin P. J. and Snijders Tom A. B., 2004. Proposals for the measurement of individual social capital, *in*: Henk D. Flap and Beate Völker (eds.), *Creation and returns of social capital*, p. 199–228. London, Routledge, 282 p.

van Duijn Marijtje, 1995. Estimation of a random effects model for directed graphes, *in*: Tom A. B. Snijders (ed.), *SSS'95 Symposium Statistische Software, nr. 7. Toeval zit overal: programmatuur voor random coëfficiënt modellen*, p. 113–131. Groningen, ProGAMMA.

van Duijn Marijtje, Busschbach Joos, and Snijders Tom A. B., 1999. Multilevel analysis of personal networks as dependent variables, *Social Networks*, vol. 21(2), p. 187–209.

van Duijn Marijtje, Snijders Tom A. B., and Zijlstra Bonne H., 2004. *P2*: a random effects model with covariates for directed graphs, *Statistica Neerlandica*, vol. 58, p. 234–254.

Völker Beate, 1995. *Should an acquaintance be forgot? Institutions of communism, the transition to capitalism and personal networks: The case of East Germany*. Amsterdam, Thesis Publishers, 332 p.

Wasserman Stanley and Faust Katherine, 1996. *Social network analysis: methods and applications*. Cambridge, Cambridge University Press, 825 p.

Wasserman Stanley and Pattison Philippa, 1996. Logit models and logistic regression for social networks. I. An introduction to Markov graphs and *p**, *Psychometrika*, vol. 61, p. 401–425.

Watts Duncan J., 1999. *Small worlds. The dynamics of networks between order and randomness*. Princeton, NJ, Princeton University Press, 262 p.

Wellman Barry and Berkowitz Stephen D. (eds.), 1988. *Social structures. A network approach*. Cambridge, Cambridge University Press, 508 p.

White Harrison C., Boorman Scott A., and Breiger Ronald L., 1976. Social structure from multiple networks. I. Block models of roles and positions, *American Journal of Sociology*, vol. 81(6), p. 730–780.

P A R T

III

HISTORY OF DEMOGRAPHIC SCIENCE, TEACHING, AND RESEARCH

GRAZIELLA CASELLI, JACQUES VALLIN, AND GUILLAUME WUNSCH

Part III of Volume IV could equally have been entitled "Demography Viewed by Demographers" because the authors of its chapters, demographers themselves, were asked to consider demography in the making. It provides the expert views of the specialists whose careers have consisted of practicing demography from a perspective that is as critical as possible when required.

In Chapter 138, Eric Vilquin gives a more historical than demographic treatment of the history of demographic science, focusing more on its origins than on the most recent developments. These are considered at greater length in the later chapters on demographic research and teaching. Starting from the pioneers who invented the discipline without giving it a more specific name than *political arithmetic*, Vilquin also covers those who, from the 19th to the early decades of the 20th century, gave the discipline its name and structured it into a coherent corpus of issues and methods, while also developing the necessary sources of information and encouraging public awakening to the issues raised.

The next stages of the history are delivered in a very different form in the following chapters, because without addressing developments in the discipline in recent decades, there is no valid way of approaching demography teaching and the tools used (i.e., handbooks, treatises and other learning aids, software, and other educational programs) or demographic research and the centers that carry it out.

In Chapter 139, Annunziata Nobile and Alessandra De Rose focus on works published in the latter half of the 20th century by reference to developments in the topics addressed by authors whose choices largely reflect advances and shifts of direction in the discipline. The authors deliberately confined themselves to works in English, French, and Italian. Even though this sample is wide enough to cover the bulk of output in this field, it was useful to supplement this overview and its bibliography with short presentations of the teaching literature available in other languages that have had or still have some importance in the development of demography. Each is written by specialists with a background in the languages concerned: Russian, German, Spanish, Portuguese, Dutch, and Japanese. This argues that the same approach should have been taken for Arabic, Chinese, and other sources, but-, a limit had to be set, even if the one chosen may seem arbitrary.

In Chapter 140, Dionisia Maffioli mainly addresses developments in demographic research issues to trace what is the very recent history of demography teaching. The issues of science are not themselves wholly independent of the political, economic, and social issues that are the essence of demography, nor even the beliefs to which they give rise. However, in the educational sphere, the big issue of recent decades is still the introduction of demography into the university curriculum. It has been a sometimes eventful journey, and one that has taken very different pathways in different cultural areas, where there is a marked contrast between the way it has made inroads in a very practical way as a component of other disciplines in the English-speaking world and its much more assertive demands as a study in its own right in the French and Italian-speaking worlds.

Advances in the discipline and especially in technology are what prompted the emergence and development of demographic software, which is traced in Chapter 141 by Antonella Pinnelli and Anna Maria Birindelli. The case may be made that technologic development to some extent has supported the development of demography, because computers removed the restraints of computation times. Some approaches turned into real avenues of exploration only when computing power became sufficient to enable the development of software to implement appropriate methods.

In Chapter 142, Dionisia Mafioli examines centers of demographic research and teaching. Her assessment naturally reflects earlier propositions about the teaching, development, current state, and future prospects of demography.

Chapter 143, which is the end of Part III and of the entire book, gives a voice to demographers from the world over. Michel Loriaux and Tania Vishnevkaia summarize the results of a survey of the members of this small international scientific community on how they came to demography, how they approach and practice it, and how they see its future. It may not be a wholly representative sample, but it is sufficient to give a flavor of the main themes running through this microcosm and to indicate the scientific challenges of demography to be expected in the 21st century.

C H A P T E R

138

The Origins of Demographic Science

ÉRIC VILQUIN

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, Belgium

The birth certificate of demography (and of descriptive statistics and epidemiology) is generally considered to be a small book of roughly 100 pages published in London in 1662: *Natural and Political Observations Mentioned in a following Index, and made upon the Bills of Mortality. By John Graunt, Citizen of London. With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the Several Changes of the said City.* The author, John Graunt, was a prosperous London merchant who happened in the Parish Clerks' Hall on the bills of mortality compiled over decades by the roughly 130 parishes of London. Graunt's intellectual curiosity led him to undertake a useful, intelligent compilation of the disparate accumulated information. The set of methodologic tools he fashioned for the purpose was called *Political Arithmetick* by his friend (and possible collaborator) William Petty (1686, 1687). The term *démographie* was coined only 2 centuries later, by Achille Guillard in his *Elements of Human Statistics or Comparative Demography* (1855), after which demography and statistics would be increasingly recognized as separate disciplines.

Following John Graunt, the pioneers of scientific demography made significant advances in methods for the observation, measure, and analysis of demographic phenomena. Focusing primarily on mortality for over 2 centuries, they really ventured into the area of fertility only on the eve of the 20th century and into migration even later. However, the measurement of population size and the description of its sociodemographic structures have historical antecedents that considerably predate the advent of demographic analysis.

I. DEMOGRAPHY BEFORE 1662

Since ancient times, data collection has been undertaken using procedures that are part of the history of demographic science: enumerations and registers. Although the purely demographic objective of such operations was often secondary or even nonexistent alongside military, taxation, and legal purposes, they are nonetheless ancestors of the modern census and vital registration system.

1. Ancient Enumerations

The earliest known enumerations were carried out in Egypt, China, India, Japan, and the Near East. These were generally initiatives by central authorities that aimed to establish a quantitative and qualitative inventory of the country's human resources. Interpretation of the statistical results is difficult because we must always suspect that the figures were arranged for political or religious reasons. However, precise information on the enumeration methodology is sometimes available. In regions with relatively strong, stable central authority, periodical census taking is part of the art of governing. This generally involves establishing in each district of the territory an inventory of hearths, houses, or individuals, divided into a small number of simple categories based on their relative importance to the state. For instance, if the purpose is to raise an army, the enumeration could identify adult males, without counting women, children, and elderly males, and an enumeration for tax collection purposes could count heads of taxpaying families (ignoring family size), without counting paupers or rich members of tax-exempt categories. Ancient enumerations often cover only a specific subset of the population.

The best-documented case is China. Numerous Chinese texts describe the administrative structures and procedures devised to enumerate the population; these began at least 3000 years ago and were frequently changed. Combined with a system of continuous registration of certain categories of the population (e.g., conscripts, persons subject to corvée labor, various categories of taxpayers, families granted imperial land), the process generally involved a massive counting operation based on the registers, with the total results forwarded in steps along strictly hierarchical channels from the base to the center. At the local level, the registers sometimes included highly detailed sociodemographic information on individuals. From the early 7th century to the early 10th century, the census was based on a family register (i.e., a list of household members specifying their sex and age, as well as the family's social, economic, and tax status); the register was supposed to be exhaustive, updated once each year by the village or district chief, and sent to the central government once every 3 years. It is difficult to determine to what extent this system functioned properly, but it subsequently underwent considerable changes. The highest degree of refinement occurred under the Ming Dynasty (1368–1644), with the use of Yellow Registers. These were permanent, standardized lists of persons and property that were kept in duplicate; compilations were periodically carried out at every level of the tax administration (i.e., *li-chia* or *lijia* system). Every 10 years, the registers were reconstituted family by family and the results tallied to form the census. The imperial government attached great importance to homogeneity and synchronization of the count throughout the territory, as well as to error correction and proper storage of the basic documents. This extremely sophisticated, bureaucratic system has left plentiful statistical data, but critical analysis is problematic. Later transformations and degradations of the system made the Chinese census an instrument for police control and tax administration, but virtually unusable for demographic purposes, from the 17th through the 19th centuries. In China, as elsewhere, the tax function of registration and enumeration almost inevitably caused a deviation, because the exhaustive list of individuals gave way to an increasingly fictive count based on notional tax units (e.g., *hearths*, *men*, *mouths*).

The Roman Empire's quinquennial *census* has left few traces, and little information is available. What was surely a partial and relatively crude count disappeared with the collapse of the empire. Europe would then remain without a census for 1000 years. Christianity inherited Judaism's distrust of the census (mysteriously associated with terrible curses in the Biblical tradition), and the West would hesitate for a long time before implementing one. The rare attempts to reinstate a nation-wide census (e.g., the English *Domesday Book* in 1086 or the French *État des Paroisses et des Feux* in 1328) proved to be short-lived. Locally, there are abundant lists and inventories of all sorts, often precise and detailed,[1] but their lack of homogeneity and continuity generally precludes any useful statistical analysis.

During the Renaissance, virtually all theorists of the art of governing (including Niccolo Machiavelli, Jean Bodin, and Antoine de Montchrestien) argued energetically for an exhaustive, periodical, nominative enumeration, detailing its benefits, and generally arguing that accurate measurement of needs, resources and changes would allow the economy and society to be administered in an informed manner, rather than blindly. Rulers sometimes took these arguments seriously: the 16th century has left traces of numerous censuses, particularly in smaller city-states, but this was not a reinstatement of the tradition of the Roman census.

2. Recording Demographic Events

Nominative registration of births and deaths, in connection with ancestor worship, has existed for at least 2000 years in China within the most powerful clans. This may be a private vital register for strictly genealogic purposes, but it is a gold mine for historical demographers. Vestiges of similar registration systems exist in other parts of the world.

In the 14th century, the Christian West slowly introduced into its administrative and legal practices a change that was to have incalculable consequences: the replacement of oral testimony by written evidence. The Church prescribed that certain important events in the life of individuals, hitherto regulated by custom or by faith, should be recorded in writing. Henceforth, *parish registers* would provide documentary evidence that a person had received the sacraments of baptism, confirmation, or marriage; that the person was alive or dead; and evidence of his or her marital status. It took 2 or 3 centuries for the rules and practices of church

[1] One of the finest examples is the *catasto* compiled in Tuscany between 1427 and 1430 in connection with a reform of Florentine taxation. The 60,000 families of Florence and its dominions were surveyed, with details of their composition, relationships, and individual characteristics of their members, activities, assets, rights, and debts (Herlihy and Klapisch, 1978).

registration to move toward homogenization and generalization. Roughly, the idea was definitively established that each church would register the christenings, marriages, and burials of its faithful in the 16th century in some dioceses and in the 17th century throughout Christendom (both Catholic and Protestant).

Church and civil authorities then relayed one another in imposing registers and specifying their content. Royal ordinances in England (1538) and France (Villers-Cotterêts, 1539) requiring that parish registers be kept predated the first general legislation of the Roman Catholic Church (the Council of Trent, 1563). The mass of regulations attests to difficulties in implementing registration and hesitations by authorities. Gradually, after successive reforms, a rigorous, homogeneous system was put in place. In France, this was accomplished with the royal declaration in 1736. These turbulent changes, combined with historical events (marked by the disappearance and destruction of documents), justify the precautions now taken by historical demographers in the statistical analysis of this prodigious mass of individual data on births, marriages, and deaths.

From 1670 onward, the parish registers in some large towns were systematically counted and compiled into more-or-less detailed, periodical tables. As early as 1736, Sweden experimented with what would become in 1746 the *Tabellverket*, the first national system of demographic statistics and analysis based essentially on the exhaustive extraction of data from parish registers.

After the Revolution in 1789, France "secularized" vital registration by creating the *état civil* (1792), a civil register modeled on church registration, but which removed the selection criterion of religious affiliation. Several European countries followed suit in the 19th century.

II. THE BIRTH OF DEMOGRAPHIC ANALYSIS IN 1662

Alongside the parish registers, in the 16th century, a number of large European towns began occasionally to produce *bills of mortality*. This was a rudimentary system for local monitoring of plagues (the generic and Biblical term for major fatal epidemics); in London, for instance, every Thursday, the parish clerk posted on the church door a bill reporting the number of deaths in the parish during the previous week and the number of deaths attributed to the plague. This was done to inform the population; if the number and proportion of deaths attributable to the plague increased week after week, it was time to take protective measures.[2]

In London, starting early in the 17th century, the practice had become routine and continuous. The parish clerks gradually extended their bills to include the number of christenings, the breakdown between males and females, and the cause of death. (The age at death was added in 1728.) The Company of the Parish Clerks of London (representing 109 parishes in and adjoining London around 1600 and 130 around 1660) received and stored all the bills, and it published an annual report with a summary account at the end of each year.

John Graunt (1620–1674), a wealthy London merchant who served on the Common Council, was an exponent of the experimental science movement that was then in full expansion. Happening on those records in 1660 or 1661, he subjected them (apparently with no preconceived intent) to an enormous amount of critical examination, analysis, and synthesis. The most elementary mathematical tools (what he called his *Shop-Arithmetique*) proved sufficient for him to forge the tools needed for a synthetic, cautious, and pertinent description of a large mass of analogous information; in so doing, he invented the statistical approach and the principal techniques of descriptive statistics.

Faced with a long series of data on systematic, continuous observations of a natural and social phenomenon (mortality), he sought to identify *laws*, in much the same way as his astronomer and physicist friends. With a subtle combination of method and intuition, he identified statistical regularities in an area where previously one had seen only the unfathomable secrets of Providence, including the sex ratio at birth, differential mortality by cause and location, comparative demographic dynamics between London and Paris, and the ratio of annual births and deaths to the population. The ratios, proportions, and disproportions that abound in Graunt's work are the ancestors of our modern proportions, means, rates, and ratios. Despite his many computational errors and some mistaken assumptions,[3] he indicated for the first time the actual magnitude of demographic parameters.

One of his concerns was to evaluate the population of a town or country in the absence of any enumeration. After observing the stability of the numerical ratios between the population and the number of births and deaths per annum, he used the relationship to calculate the population from parish statistics. This

[2] The only effective measure at the time was to flee the city.

[3] He thought, for instance, that the persons who died in a given year formed a representative sample of the living population.

TABLE 138–1 John Graunt's Table, 1662

Of 100[a] there dies within the first six years	36
The next ten years, or *Decad*	24
The second *Decad*	15
The third *Decad*	9
The fourth	6
The next	4
The next	3
The next	2
The next	1
From whence it follows, that of the said 100 conceived there remains alive at six years end	64
″ At Sixteen years end	40
″ At Twenty six	25
″ At T[h]irty six	16
″ At Fourty six	10
″ At Fifty six	6
″ At Sixty six	3
″ At Seventy six	1
″ At Eighty	0

[a] These are 1000 "quick conceptions," indicating that some perinatal mortality is included in the 36 deaths for the first age group.

multiplier technique remained the principal method of evaluating population size until censuses were restored and continued to compete with censuses through the early part of the 19th century.

Graunt did not have what was needed to construct a life table, but he is nonetheless the inventor. With no information on the age at death, he adopted an arbitrary (and unrealistic) distribution of deaths by age groups[4] to develop the theoretical model of the age-specific distribution of deaths that he suspected (Table 138–1). Thirty years later, Edmund Halley (1693) relied on John Graunt's model to construct the first life table based on an actual distribution of deaths according to age.

Graunt was also a pioneer in the critical analysis of data. Every concept and figure was subjected to inquisitorial scrutiny, after which he corrected or rejected the data. With scrupulous caution and modesty, he always tried to check calculations by several independent means and gave only limited credit to the numeric value of his results—but that does not prevent him from being affirmative and categorical when certain of having disproved generally accepted ideas. Graunt the precursor left an impressive set of tools, which his successors perfected while preserving their initial spirit.

[4] John Graunt grouped the causes of death specific to early childhood and logically attributed them to the ages younger than 6 years, but he was unable to do this for the other age groups.

III. DIRECT MEASUREMENT OF POPULATION

Tax registers naturally provided a census frame that, although incomplete, was frequently brought up to date. Tax registers were the basis for most population estimates in the 16th and 17th centuries. There were significant methodologic problems, however, and the resulting estimates remained highly inaccurate because some levels of society were not included on the registers and because the registers often provided only a list of *hearths* (households), with the aggravating circumstance that in several regions the *hearth* was a conventional tax unit unconnected to demographic reality.

In the absence of a simple, homogeneous method, many attempts to conduct local or regional enumerations in the 16th and 17th century (in Spain, Italy, and France) proved to be short-lived. Around 1680, Sébastien de Vauban, the French administrator and military engineer, developed a "general and easy" (*générale et facile*) counting method, which he personally tested in numerous towns and regions and then placed at the disposal of the French government, which first applied it in the colonies (Vilquin, 1975; Vauban, 1686). Vauban's innovation was to print blank tables that the census agents completed by following strict instructions. There is one table per parish and one line per house; each line recorded the composition of the household based on the main categories of sex, age, and social status.[5]

The 17th and 18th century European governments were eager for qualitative and quantitative information, but they deluged local officials with such wide-ranging demands that the answers were inevitably vague, if not mere conjecture. One such enterprise, which was better prepared than most, was a genuine success. This was the French *Enquête des intendants pour l'instruction du duc de Bourgogne*[6] between 1697 and 1700. With collaboration from Vauban, the Duke of Burgundy's tutors undertook a vast survey with multiple objectives in every province of France. After considerable prodding for information and adjustments,

[5] The column headings can be translated as "Men," "Women," "Grown boys," "Grown girls," "Little boys," "Little girls," "Female servants," "Male servants," and "Number in family." The first two categories include married adults (or priests and nuns) who are not servants; grown boys are unmarried males older than 14 years, and grown girls are unmarried females older than 12 years (below those ages, they are little boys and little girls); the number in family is the total number in the household.

[6] Louis, Duke of Burgundy, grandson of Louis XIV, later became *dauphin* (i.e., heir apparent), but he died before acceding to the throne.

they obtained a complete series of reports that, despite enormous failings, were used pretty much as official statistical yearbooks (often without being updated) by French ministries throughout the 18th century. Their interest and utility were such that the bookseller Saugrain published them in 1709 (*Dénombrement du royaume par généralités, élections, paroisses et feux*) and even brought out a second edition in 1720 (*Nouveau dénombrement du royaume*) (Saugrain, 1709, 1720). For the purposes of the survey, a genuine direct census of the population was taken in some provinces and towns, more or less in keeping with Vauban's approach; however, in most cases, earlier estimates of uncertain provenance were reproduced, or a rapid count using the tax registers was performed.

Genuine nominative, comprehensive censuses have been conducted since the 17th century in nearly every European country, in limited territories. The first modern nationwide censuses were those taken in Iceland in 1703 and Sweden in 1755. The case of Sweden is distinctive. All Swedish parishes had maintained detailed, up to date registers of their population since the late 17th century. In the 1750s, the government ordered a periodic compilation and analysis of the registers. Pehr Wargentin and Edouard Runeberg (1772), who were in charge of the initial work by the *Tabellkommissionen*, set very high standards at the outset, and for a long time Sweden maintained a half-century lead over the rest of Europe in the area of collection and analysis of national demographic statistics (for Wargentin and Runeberg, see Dupâquier, 1977).

In France, the welter of censuses and surveys ordered by an array of Revolutionary bodies were typically carried out by incompetent, overworked personnel, and almost never reached the processing stage. The creation of a Statistical Bureau in 1799 introduced some order, but the Empire soon made statistics a matter of state secrecy and eliminated the Statistical Bureau, which would definitively reemerge only in 1834. Meanwhile, Austria, Prussia, Russia, the Netherlands and Belgium set up their own statistical bureaus, followed in the third quarter of the century by the other nations of Europe.

One hundred fifty years after Vauban, the ideal of a uniform, thorough, and simple census method was taken up again in the 19th century by the Belgian mathematician and astronomer Adolphe Quetelet. The censuses in the Netherlands (1829 to 1830) and Belgium (1846) allowed him to test his ideas regarding standardization of procedures for collection and analysis (Quetelet, 1847). From defining concepts to training census agents, his systematic approach worked wonders, so the organization of the Belgian census was seen as a model by neighboring countries. However, his dream of every country in the world simultaneously taking identical censuses would never be realized.

IV. POLITICAL ARITHMETIC

John Graunt's invention found its first enthusiast in his friend William Petty, who used the term *political arithmetic* to describe his resolutely quantitative approach to social phenomena (Petty, 1686, 1687). Petty used Graunt's formulas to build arguments in support of his political theories, but was largely unconcerned with methodologic improvements. Their successors, often astronomers or clergymen, would advance theory and techniques in two main areas: indirect population estimation and the life table.

1. Demographers

The first demographers were English and Dutch: Edmund Halley, Matthew Hale, Gregory King, Charles Davenant, Louis and Christian Huygens, and Jan De Witt. Their shared concerns included evaluating population worldwide and in selected countries and towns; the long-term population trend in Europe and worldwide; and the actuarial computation of annuities. Some demographers also addressed the distribution of population by sex or age group, and even ventured population projections. In 1696, Gregory King computed ratios for births, deaths, and nuptiality by dividing the population by the annual numbers of births, deaths, and marriages (King, 1696/1973); it was not until 80 years later that the corresponding rates were computed, when Jean-Louis Muret (1766) reversed the numerator and denominator. The Dutch actuaries Willem Kersseboom and Nicolas Struyck, starting in 1738, quarreled over the determination of multipliers; in doing so, they unwittingly laid the bases of stationary population theory,[7] and more importantly, they systematically compiled statistical surveys of demographic data from various sources (Kersseboom, 1970; Struyck, 1912, cited by Dupâquier and Dupâquier, 1985, p.162–166). Gregory King and Nicolas Struyck are doubtless the most important figures in the first half of the 18th century because they improved on the methods of their predecessors, and they methodically weighed the relevance and validity

[7] Kersseboom applied the life table to the average annual number of births to determine the population pyramid for the (stationary) population. The size of the (stationary) population is then equal to the annual number of births multiplied by the average lifetime.

of the data and techniques. Nicolas Struyck was the first to investigate the disturbing role of migrations in demographic estimates, to be wary of computations based on small numbers of subjects, and the first who did not set out to prove at all costs that all is for the best in Creation.

The German pastor and mathematician Johann Peter Süssmilch (1707–1767) was the first demographer to achieve fame and the first to popularize the new science. His principal work, *The Divine Order*, published in 1741 and revised in 1761–1762, enjoyed genuine success. The book provides a synthesis of the state of political arithmetic on the two publication dates, and its application to an impressive amount of data. Süssmilch was a judicious critic but not really an innovator. His purpose was to accumulate scientific evidence to support his religious convictions.

In France, Antoine Deparcieux (*Essay on the probabilities of the length of human life*, 1746, and *Addition to the essay . . .* , 1760) made a notable improvement to the measurement of the death rate. Louis Messance (1766) conducted a methodical demographic analysis of three French provinces. The Abbé Expilly (1762–1770) collected astronomic quantities of demographic data. Jean-Baptiste Moheau published a veritable treatise on demography (*Researches and Considerations on the population of France*, 1778), which can be compared with *The Divine Order*. His arithmetic, though brilliant, remained classic, but he stands out from most of his peers by devoting several chapters to interpreting his results, evidencing a systemic vision of the sociodemographic dynamics of a nation.

Each European country had its political arithmeticians. Among dozens of others, two who stand out were the Swede Pehr Wargentin (1717–1783) and Switzerland's Jean-Louis Muret (1715–1796). Wargentin carried out in the 1750s an exemplary analysis of the data produced by the admirable Swedish statistical system (see Dupâquier, 1977), and Muret (Muret, 1766), who performed many differential analyses (by sex, by region, and by period), can be called the true inventor of the crude rates of births, deaths, and marriages.

In the 18th century, the political arithmeticians knew and corresponded with one another, and submitted their discussions to a learned audience by publishing in the first scientific journals (i.e., the published proceedings of the academies). Their methodologic problems came to the attention of mathematicians, including the Bernouillis, d'Alembert, Euler, Lagrange, and Laplace. The mathematicians cautiously advanced the initial mathematical expressions for the political arithmeticians' empirical parameters, and used their observations (e.g., the stability of the sex ratio at birth or the probability of survival from one age to another) to formulate the standard problems in the calculus of probabilities. For instance, at the request of Johann Peter Süssmilch, Leonhard Euler developed around 1760 the empirical recipe for computing the doubling time of a population, which had hardly progressed since John Graunt), and established the algebraic relationships between population size, structure and movement. Pierre-Simon de Laplace in 1778 called on probability theory to demonstrate that the sex ratio at birth differs significantly from 1 and to measure the ranges of variation of certain parameters (Laplace, 1781).

2. The Life Table

Even before the construction of the first true life table by Edmund Halley (1693), the brothers Christian and Louis Huygens, in 1669 (Huygens, 1669/1920), were attracted by the model developed by John Graunt. They developed the concepts of life expectancy and the probable length of life at any given age, as well as the graphical representation of the parameters of a survival table. However, this was not reported to contemporaries. Around 1680, Gottfried Wilhelm Leibniz outlined a theoretical life table (Leibniz, 1680?/1866), but failed to follow through with his work. It therefore fell to Edmund Halley (1656–1742), the British astronomer, who used the first statistics of deaths according to age, from Breslau in 1687 and 1688, to construct the first table based on actual data. Taking some liberty with the crude data, he computed a table of survivors by age, starting with 1000 infants younger than 1 year. The ages are expressed in complete years at the start of the table, and subsequently by age at birthday. However, lacking information on the age structure of the population of Breslau, he had to assume that the distribution of deaths by age was the same as the age structure of the living population (the cumulative deaths method), which (as he was aware) is the case only for a closed, stationary population. He did not calculate death probabilities or life expectancies.

Life table methodology made slow, modest progress in the 18th century. Kersseboom, Struyck, Smart, and Simpson eliminated the need to assume a closed population and stationarity. Before introducing the technique into France, Antoine Deparcieux (1746) analyzed his predecessors' work in detail and definitively distinguished between the mean length of life and the median length of life (i.e., probable lifetime). Deparcieux gave the life table its current form, breaking with the cumulative deaths method. Disposing of good life history data on small populations of rentiers or religious communities, he could establish the ratio of deaths at any age to the total population exposed to the risk of dying at that age. Because he had no data on children, he surveyed mothers from various social backgrounds at the end of their reproductive lives regarding the survival of their children. Pehr Wargentin was the first to dispose of the age-specific distributions of the population and of deaths for the same period (see Dupâquier, 1977).

Early in the 19th century, Emmanuel Duvillard de Durand (1806) used an essay by Daniel Bernoulli (1760) as the basis for calculating the gain in life expectancy that would result from the elimination of a cause of death (i.e., smallpox).

As rulers and *savants* showed growing interest in statistics, the number of national life tables increased. There were the first timid steps in also establishing tables for subpopulations (e.g., males and females, town and country). Around 1850, on the strength of the exemplary work of Adolphe Quetelet (1854, 1872) in Belgium and William Farr (1885) in England, the life table was definitively anchored to the corresponding population pyramid derived from the census.

3. Indirect Methods of Population Estimation

The old procedure for population estimation based on an enumeration of hearths did not disappear quickly, and the *multiplier technique* discovered by John Graunt seems to have been used little before 1750. Gregory King tried both methods in 1696 but did not publish his results.[8] Spurred on by the first French demographers (e.g., Messance, Moheau), enthusiasm for Graunt's method spread in the second half of the 18th century. The political arithmeticians, focusing on small portions of the territory, conducted exhaustive enumerations of the population and counts of christenings in the parish registers: the ratio of these two numbers was the *birth multiplier*, which could be used in any town or region to compute the total size of the population on the basis of the average annual number of births registered. Similar multipliers based on the annual number of deaths or marriages were tested. Fascinated by the stability—a relative stability, but astonishing nonetheless—of the numerical ratios they found, demographers cherished the secret ambition of determining the *universal multiplier*, but the quest was gradually abandoned as they recognized the need to calculate specific multipliers for different contexts (e.g., town/country, plains/mountains, north/south). The most extravagant enterprise in this area was conducted by the Abbé Expilly, who spent years working on the whole of France. He had baptisms, marriages and burials in the registers of every parish in France counted for the periods 1690–1701 and 1752–1763 (to identify the direction of population change). Funds ran out as he neared the goal, but he still managed to publish statistics for more than 15,000 parishes, in six volumes; the seventh was never released (Expilly, 1762–1770).

French ministers (Turgot and then Necker) ordered that the computations be redone on regional samples, while scientists (Moheau, Laplace) were refining the theory. However, the French Revolution chose to return to a Roman census.

The rare rudimentary attempts at demographic projections, as early as the 17th century (e.g., Petty, King, Vauban), use exponential extrapolation and are based on fixed, crude assumptions: Vauban (no date/1845), for instance, assumed that marriage is universal, that each couple produces four children on average, and that generations succeed one another every 30 years.

[8] First published in 1802 (as an annex in George Chalmers, *An Estimate of the Comparative Strength of Great Britain*)

V. DEVELOPMENT OF MODERN DEMOGRAPHIC ANALYSIS IN THE 19TH CENTURY

1. Death Rate

To construct his hypothetical life table, John Graunt implicitly used a mathematical function expressing the age-specific risk of dying,[9] but because this was not indicated in his book, the idea was not taken up immediately. After several attempts—more complicated but not more convincing—to mathematically model mortality by age in the 18th century (De Moivre, Euler, Duvillard, Laplace), the Englishman Benjamin Gompertz, in 1825, made the assumption that the number of survivors decreases exponentially as age increases; this was corrected in 1860 by the German William Makeham, who incorporated in the equation a term to represent deaths that were not dependent on age. This initial mathematical law provides a close fit to the death rate for adults, but not for infants and children. In the meantime, William Farr (1843/1885) provided the definitive formulation of the concepts of age-specific rates (including the infant mortality rate), death probabilities, the probability of survival, and standardized rate. Louis-Adolphe Bertillon (1866) discovered how the crude death rate is dependent on the age structure of the population. In

[9] In modern scientific notation: ${}_{10}q_x = 3/8$.

1868, Georg Friedrich Knapp, in analyzing the classification of deaths by years of age and year of birth, distinguished between longitudinal analysis and cross-sectional analysis.

The age-old dream of a mathematical law expressing mortality as a function of age was shattered in 1897 by Karl Pearson, who showed that the statistical curves of mortality by age must be decomposed into several mathematical laws that are juxtaposed or superposed.

2. Fertility

If the political arithmeticians focused more on the factors of fertility than on measuring fertility, it was because the necessary data were totally unavailable. Births were first classified by mother's age in Sweden in 1775, and the example was not imitated elsewhere until after 1850. Before then, considerations on differential fertility (between town and country, between monogamy and polygamy, and between seasons) were based on economic or moral arguments; the rare calculations were based on rough assumptions regarding "prolificacy" (e.g., that an average marriage produces four children). The nearly universal "populationism" among 18th century Europeans, including demographers, provided less incentive to measure fertility than to seek ways to maximize fertility. This explains the considerable persistence of extremely crude assumptions.

Even before it became possible to study age-specific fertility, a number of significant steps forward were made. The first fertility estimator was the birth multiplier (number of inhabitants per birth). The political arithmeticians all devoted numerous pages to this, as they sought to determine whether it was stable or variable, and the best way of evaluating it. After Gregory King, but independently, Jean-Louis Muret transformed the multiplier into a crude birth rate by reversing the ratio (number of births per 100 inhabitants). This use would not become widespread before the 19th century.

Some considered it more logical to take the ratio of births to marriages, and this parameter enjoyed considerable favor. The general legitimate and illegitimate fertility rates were developed in 1815 by Joshua Milne (1815). Even toward the middle of the 19th century, Quetelet believed the best measure of a population's fertility was the ratio of births to marriages and that the analysis could be further refined by classifying couples based on the age difference between spouses and age at marriage (Quetelet, 1835).

The breakthrough came from Scotland. Dr. James Matthews Duncan (*Fecundity, Fertility, Sterility, and Allied Topics*, 1866) took advantage of an exceptional statistical source, which decomposed births in Edinburgh and Glasgow by the mother's age and the duration of marriage; he used this to measure, among other things, the intervals between successive births, parities by age and by duration of marriage, and completed fertility. However, these methodologic innovations would be disseminated 30 years later, when the national statistical offices grasped the need to collect highly detailed data on fertility. Similarly, the clear distinction he introduced between fertility and fecundity was not entirely understood at the time, and was generally accepted only 50 years later. Because he conducted a cohort analysis of married women who had neither died nor become widows before the age of 45, mortality did not interfere with his estimates of completed fertility.

3. General Models

Adolphe Quetelet had met Thomas Malthus in 1833, and the two had briefly corresponded. Quetelet admired the English clergyman's theory but considered that it remained to be cast in the bronze of mathematics. He assigned a young colleague, the brilliant mathematician Pierre-François Verhulst, the task of finding the best mathematical model for Malthus' theory of population growth over time. The main difficulty was finding an exponential function (Malthus' *population increases in a geometric ratio*) that progressively generates a slowdown in the rate of growth (i.e., that incorporates Malthus' *checks*). After an initial attempt left Quetelet unsatisfied, in 1847 Verhulst discovered that the *logistic curve* was an excellent way to model Malthus' theory and that it provided the theory with a complement, if not the crowning touch, in that the curve tended toward a horizontal asymptote: the maximum population. Verhulst's work was published but passed unnoticed, and the logistic model was rediscovered only in 1920 by the Americans Raymond Pearl and Lowell Reed.

Advances from the concept of marital fertility to that of population reproduction required the modeling of nuptiality. This was done by Louis Adolphe Bertillon (1872). The Germans Richard Böckh (1886) and Robert Kuczynski (1931) could then define the crude and net reproduction rates. However, the concept of population replacement was to become operative through another channel. In 1907 and 1911, Alfred Lotka adopted a purely theoretical, abstract approach to study the renewal of a population that evolves as a function of inputs and outputs (Sharpe and Lotka, 1911; Lotka, 1907). Laying the bases for mathematical demography, he developed the theory of stable populations (Lotka, 1934, 1939) that would revitalize demographic analysis after World War I.

4. Graphical Tools

The first graphs in the history of demography were drawn by Christian Huygens (1669), who used curves to depict the numbers of survivors at various ages in Graunt's table, and the corresponding life expectancies. However, he showed them only to his brother, and they remained unknown. Setting aside three or four curious but largely impractical attempts at geometric illustration, which failed to gain acceptance, the first to make systematic, frequent use of graphical representations in statistics were Jean-Baptiste Fourier (1821) and Adolphe Quetelet (starting in 1827). Quetelet was virtually obsessed by means and variations, and privileged the representation of deviations between observations and their mean. Through his educational abilities and the international statistics conferences he was the first to organize and to which he contributed indefatigably, he led all European statisticians to use a broken line or curve to represent variations in the phenomena they observed.

The second half of the 19th century saw several ingenious but virtually indecipherable attempts at three-dimensional graphical representation (i.e., stereograms), which was soon abandoned. Another invention that proved more successful was the population pyramid. The initial idea came from General Francis Walker (1874), the Superintendent of the 1870 Census in the United States, who used it liberally to illustrate his analysis of the U.S. censuses of 1850, 1860, and 1870. In 1878, Heinrich Schwabe brought the shape into line with the principles of the histogram, and it has remained unchanged ever since.

In 1874, following up on the intuition of several precursors, three German mathematicians, Georg Friedrich Knapp, Karl Becker, and Wilhelm Lexis, looking for the most efficient way of plotting biographic events in a system with three coordinates (i.e., time, age, and date of birth), devised two alternative versions of what is now called the Lexis diagram. The modern version was developed by Roland Pressat around 1953 (Personal communication. See Vandeschrick, 1992, p. 1249.).

CONCLUSIONS

The origins of demographic science are recent. Only 350 years ago, it could seriously be claimed without incurring scientific refutation that men outnumber women six to one worldwide or that gout is a hundred times more lethal than influenza. John Graunt, in his simple but inspired constructions, planted the seed that has grown into a full array of instruments for measurement and analysis of demographic phenomena. Some developed faster than others. Progress in life table techniques was significantly stimulated by the quest for accurate calculation of annuities. At the same time, the political arithmeticians, eager to use their *multipliers* to prove that the population was growing (or declining), thought they could eliminate the need for complicated census taking. They also evaluated *the constants* of what Adolphe Quetelet would call *social physics*. Mathematicians then joined the fray, testing the concepts and theorems of probability theory on the only continuous statistical series available, vital registration statistics. Fertility had long been the poor relation to demographic research—not for lack of interest, but for lack of the data required to analyze fertility. The development of concepts and tools for demographic analysis has always been closely associated with changes in the production of population statistics, but quite often, the conceptual developments preceded the data. From its primitive, artisan, and utilitarian origins, the progress and results achieved by demography between the late 17th century and the early part of the 20th were remarkable. The considerable work by Alfred Lotka, before and after World War I, would unify demographic analysis and impart a new impetus, supported and amplified by fears of population explosion.

Acknowledgment

This chapter was translated from the French version by Bernard Cohen.

References

Affichard Joëlle (ed.), 1977. *Pour une histoire de la statistique.* Paris, INSEE and Economica, 2 vol., 593 + 969 p.

Bardet Jean-Pierre and Dupâquier Jacques (eds.), 1997. *Histoire des populations de l'Europe.* Paris, Librairie Arthème Fayard, 660 + 647 + 792 p.

Behar Cem Lazare, 1977. Des tables de mortalité aux XVII^e et XVIII^e siècles. Histoire. Signification, *Annales de démographie historique*, p. 173–200.

Behar Cem Lazare, 1985. La mesure de la fécondité chez les premiers démographes, *Annales de démographie historique*, p. 173–196.

Behar Cem Lazare and Dupâquier Jacques, 1991. L'invention des taux de fécondité légitime: J. Matthews Duncan, *Annales de démographie historique*, p. 297–335.

Bernouilli Daniel, [1760]. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et les avantages de l'inoculation pour la prévenir, *in: Histoire de l'Académie royale des sciences. Mémoires.* Année 1760, p. 1–45.

Bertillon Louis-Adolphe, 1866. *Des diverses manières de mesurer la durée de la vie humaine.* Strasbourg, Vve Berger-Levrault, 20 p. (Reprinted from *Journal de la Société de statistique*, March 1866.)

Bertillon Louis-Adolphe, 1872. Mariage, *in*: Amédée Dechambre (ed.), *Dictionnaire encyclopédique des sciences médicales, 2e série, tome V*, p. 7–83. Paris, G. Masson and P. Asselin, 732 p.

Böckh Richard, 1886. Statistik des Jahres 1884, *Statistisches Jahrbuch der Stadt Berlin*, vol. 12.

Caselli Graziella and Lombardo Enzo, 1990. Graphiques et analyse démographique, *Population*, vol. 45(2), p. 399–413.

Deparcieux Antoine, 1746. *Essai sur les probabilités de la durée de la vie humaine.* Paris, H.L. Guérin and L.F. Delatour, 208 p.

Deparcieux Antoine, 1760. *Addition à l'Essai sur les probabilités de la durée de la vie humaine.* Paris, H.L. Guérin and L.F. Delatour, 36 p.

Droesbeke Jean-Jacques and Tassi Philippe, 1990. *Histoire de la statistique.* Paris, PUF.

Duncan James Matthews, 1866. *Fecundity, fertility, sterility, and allied topics*. Edinburgh, A & C Black.

Dupâquier Jacques, 1977. Les tables de mortalité de Wargentin, *Annales de démographie historique*, p. 385–420.

Dupâquier Jacques, 1996. *L'invention de la table de mortalité.* Paris, PUF, 178 p.

Dupâquier Jacques and Dupâquier Michel, 1985. *Histoire de la démographie. La statistique de la population des origines à 1914.* Paris, Librairie Académique Perrin, 462 p.

Duvillard de Durand Emmanuel, 1806. *Analyse et tableaux de l'influence de la petite vérole sur la mortalité à chaque âge, et de celle qu'un préservatif tel que la vaccine peut avoir sur la population et la longévité.* Paris, Imprimerie nationale, 210 p.

Euler Leonhard, 1760. Recherches générales sur la mortalité et la multiplication du genre humain, *Mémoires de l'Académie Royale des sciences et des Belles Lettres*, no. 16, p. 144–165.

Expilly (Abbé) Jean-Joseph, 1762–1770. *Dictionnaire géographique, historique et politique des Gaules et de la France.* Paris, Dessaint & Saillant, 6 vol. (The seventh and last volume was never published.)

Farr William, 1843. Construction of life tables, *British Parliamentary Papers*, sess. 1843, vol. 21, p. 191–177. (First published in: George Graham (ed.), *Fifth annual report of the Registrar-General of births, deaths and marriages in England and Wales.* London, Her Majesty Stationary Office, 1843, p. 342–367.)

Farr William, 1885. *Vital statistics. A memorial volume of selections from the reports and writings of William Farr.* London, The Sanitary Institute of Great Britain, 259 p.

Fourier Jean-Baptiste, 1821. *Recherches statistiques sur la ville de Paris et le département de la Seine*. . . . Paris, Ballard Imprimeur du Roi, 3 vol.

Gompertz Benjamin, 1825. On the nature of the function expressive of the law of human mortality, *Philosophical Transactions*, vol. 27, p. 513–585.

Graunt John, 1662. *Natural and political observations made upon the bills of mortality*. . . . London, printed by T. Roycroft for J. Martin, J. Allestry, and T. Dicas, 150 p. (American edition by W.F. Wilcox, Baltimore, Johns Hopkins Press, 1939.)

Guillard Achille, 1855. *Éléments de statistique humaine ou démographie comparée*. Paris, Guillaumin, 376 p.

Halley Edmund, 1693. An estimate of the degrees of the mortality of mankind, *Philosophical Transactions*, vol. 17(196), p. 579.

Hecht Jacqueline, 1977. L'idée de dénombrement jusqu'à la Révolution, in: Joëlle Affichard (ed.), *Pour une histoire de la statistique*, vol. 1, p. 21–81. Paris, INSEE and Economica, 2 vol., 593 + 969 p.

Herlihy David and Klapisch-Zuber Christiane, 1978. *Les Toscans et leurs familles: une étude du Catasto florentin de 1427*. Paris, Fondation nationale des sciences politiques and École des hautes études en sciences sociales, 702 p. (English abridged edition: *Tuscans and their Families: A Study of the Florentine Catasto of 1427*. New Haven: Yale University Press, 1985.)

Huygens Christiaan and Huygens Louis, [1669] 1895. Correspondance, *in*: Société Hollandaise des Sciences (ed.), *Oeuvres complètes de Christiaan Huygens*, vol. VI. The Hague, Martinus Nijhoff, 653 p.

Kersseboom Willem, [1738–1742] 1970. *Essais d'arithmétique politique contenant trois traités sur la population des provinces de Hollande et de Frise occidentale . . . par Willem Kersseboom et Kersseboom et son œuvre, la table de mortalité par M. Van Haaften.* Paris, INED, 172 p.

King Gregory, [1696] 1973. *Natural and Political Observations and Conclusions upon the State and Condition of England* (Manuscript [1696], reprinted *in* Peter Laslett (ed.), *The earliest classics: John Graunt . . . Gregory King . . .* , p. 1–292. Westmead, Farnborough, Hants (England), Gregg International Publishers Limited, 1973, 457 p.)

Knapp Georg Friedrich, 1868. *Ueber die Ermittlung der Sterblichkeits*. Leipzig.

Kuczynski Robert R., 1931. *Fertility and reproduction. Methods of measuring the balance of birth and death*. New York, McMillan, 150 p. (New edition: Berlin, Akademic Verlag, 1982, 147 p.)

Laplace (de) Pierre Simon, [1781] 1893. Mémoires sur les probabilités, *in*: Académie des sciences (ed.), *Œuvres complètes de Laplace, Tome neuvième*, p. 383–485. Paris, Gauthier-Villars et Fils, 485 p. (First published in *Mémoires de l'Académie royale des sciences, Année 1781*, p. 227–332.)

Leibniz Gottfried Wilhelm, [1680?] 1864–1884. Essay de quelques raisonnements nouveaux sur la vie humaine et sur le nombre des hommes, *in*: Onno Klopp (ed.), *Die Werke von Leibniz gemass seinem hanschriftlichen Nachlasse in der Koniglichen Bibliothek zu Hannover, tome V*, p. 326–337. Hanover, Klindwoth. (Manuscript, ca. 1680.)

Lexis Wilhelm, 1875. *Einleitung in die Theorie der Bevölkerungsstatistik*. Strasbourg, Trübner, 137 p.

Lotka Alfred J., 1907. Studies on the mode of growth of material aggregates, *American Journal of Science*, vol. 24, p. 199–216.

Lotka Alfred J., 1934. *Théorie analytique des associations biologiques. Première partie: principes*. Paris, Herman, 150 p. (Actualités scientifiques et industrielles, no. 187.)

Lotka Alfred J., 1939. *Théorie analytique des associations biologiques. Deuxième partie: analyse démographique avec application particulière à l'espèce humaine*. Paris, Herman, 139 p. (Actualités scientifiques et industrielles, no. 780.)

Makeham William M., 1860. On the law of mortality and the construcion of annuity tables, *Assurance Magazine*, no. 8, p. 301–310.

Messance Louis, 1766. *Recherches sur la population des généralités d'Auvergne, de Lyon, de Rouen, et de quelques provinces et villes du Royaume, avec des réflexions sur la valeur du blé tant en France qu'en Angleterre, depuis 1674 jusqu'en 1764.* Paris, Durand, 330 p.

Milne Joshua, 1815. *A treatise on the valuation of annuities and assurances on lives in survivorships; on the construction of tables of mortality; and the probabilities and expectations of life.* London, Longman, Hurst, Rees, Orme, and Brown, 2 vol., 783 p.

Moheau Jean-Baptiste, [1778] 1994. *Recherches et considérations sur la population de la France.* Paris, INED, 652 p. (New critical edition by Éric Vilquin.)

Muret Jean-Louis, 1766. *Mémoire sur l'état de la population dans le pays de Vaud. [followed by] Tables servant de pièces justificatives pour le mémoire.* Yverdon, 130 + 128 p. (Reprinted by Bibliothèque Nationale in 1977.)

Pearl Raymond and Reed Lowell J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation, *Proceedings of the National Academy of Sciences*, vol. 6(6), p. 275–288.

PEARSON Karl, 1897. *The chances of death, and other studies in evolution.* London, E. Arnold, 2 vol.
PETTY William, 1686. *An essay concerning the multiplication of mankind*, 2nd ed. London, Printed for M. Pardoe, 50 p.
PETTY William, 1687. *Two essays in political arithmetick, concerning the people, housing, hospitals, &c, of London and Paris.* London, J. Lloyd, 8 + 21 p. (Printed for J. Lloyd in the Middle Exchange next Salisbury-House in the Strand.)
QUETELET Adolphe-Lambert-Jacques, 1827. *Recherches sur la population, les naissances, les décès, les prisons, les dépôts de mendicité, etc., dans le royaume des Pays-Bas*. Bruxelles, H. Tarlier, 90 p.
QUETELET Adolphe-Lambert-Jacques, 1835. *Sur l'homme et le développement de ses facultés ou Essai de physique sociale.* Paris, Bachelier, 2 vol., 327 + 327 p.
QUETELET Adolphe-Lambert-Jacques, 1847. *Sur les anciens recensements de la population belge.* Bruxelles, 38 p. (Reprinted from *Bulletin de la Commission centrale de statistique de Belgique*, vol. III.)
QUETELET Adolphe-Lambert-Jacques, 1854. *Sur les tables de mortalité et spécialement sur les tables de mortalité de la Belgique.* Batignoles, Impr. de Hennuyer, 12 p. (reprinted from *Journal des économistes*, 15 November 1854.)
QUETELET Adolphe-Lambert-Jacques, 1872a. *Sur la statistique morale et les principes qui doivent en former la base.* Bruxelles, F. Hayez, 112 p. (*Mémoires de l'Académie royale des sciences, des lettres et des arts de Belgique*, vol 21[7].)
QUETELET Adolphe-Lambert-Jacques, 1872b. *Tables de mortalité et leur développement.* Bruxelles, F. Hayez, 40 p. (Reprinted from *Bulletin de la Commission centrale de statistique de Belgique*, vol. XIII.)
RUNEBERG Edouard, 1772. Mémoire sur la population et la force du royaume de Suède, *in*: *Collection Académique*, *partie étrangère*, vol. 11, Paris, Panckoucke, p. 325–329.
SAUGRAIN Claude-Marin, 1709. *Dénombrement du royaume par généralitez, élections, paroisses et feux*. . . . Paris, Charles Saugrain, 2 vol. 408 + 342 p.
SAUGRAIN Claude-Marin, 1720. *Nouveau dénombrement du royaume par généralitez, élections, paroisses et feux*. . . . Paris, Saugrain, 2 vol. 486 + 370 p.
SHARPE F. R. and LOTKA Alfred J., 1911. A problem in age distribution, *London, Edinburgh and Dublin Philosophical Magazine*, 6th series, no. 21, p. 435–438.
STRUYCK Nicolas, [1740–1753] 1912. *Les oeuvres de Nicolas Struyck (1687–1769) qui se rapportent au calcul des chances, à la statistique générale, à la statistique des décès et aux rentes viagères.* Amsterdam, Société générale néerlandaise d'assurances sur la vie et de rentes viagères, 432 p. (Translation by Johan A. Vollgraff.)
SÜSSMILCH Johann Peter, 1741. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechtes aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen*. Berlin, Kultur-Verlag, 360 p. (Reprinted by J.C. Spener, Berlin, 1741.)
SÜSSMILCH Johann Peter, 1761. *Die göttliche Ordnung, vol. 1.* Berlin, Buchladen der Realschule, 722 p.
VANDESCHRICK Christophe, 1992. Le diagramme de Lexis revisité, *Population*, vol. 47(5), p. 1241–1262.
VAUBAN (Le Prestre, chevalier de) Sébastien, 1686. *Méthode generalle et facile pour faire le dénombrement des Peuples.* Paris, Impr. Vve d'A. Chrestien, 15 p.
VAUBAN (Le Prestre chevalier de) Sébastien, [16??] 1845. Moyen de rétablir nos colonies de l'Amérique et de les accroître en peu de temps, *in*: Cel Antoine-Marie Augoyat (ed.), *Oisivetés de M. de Vauban, vol. IV*, p. 1–37. Paris, J. Corréard, 2 vol. (Undated manuscript.)
VERHULST Pierre-François, 1847. *Deuxième mémoire sur la loi d'accroissement de la population*. Bruxelles, Académie royale des sciences et belles lettres, 32 p. (Nouveaux mémoires de l'Académie royale des sciences et belles lettres de Belgique, vol. 20[3].)
VILQUIN Éric, 1975. Vauban inventeur des recensements, *Annales de démographie historique*, p. 207–257.
VILQUIN Éric, 1976. La naissance de la démographie, *Population et Famille*, vol. 3(39), p. 145–164.
WALKER Francis Amasa, 1874. *Statistical Atlas of the United States, based on the results of the ninth census 1870.* Washington, Government Printing Office, 14 + 13 + 8 + LIV p.
WESTERGAARD Harald, 1932. *Contributions to the history of statistics.* London, Paris, The Hague and Wakefield, P.S. King and Son, 280 p. (New edition: The Hague, Mouton, 1969.)

C H A P T E R

139

Demographic Textbooks and Treatises: Some Considerations

ANNUNZIATA NOBILE AND ALESSANDRA DE ROSE

Dipartimento di Istituzioni politiche e Scienze sociali, Università degli Studi Roma Tre, Italy

Dipartimento di Studi geoeconomici, linguistici, statistici e storici per l'analisi regionale, Università degli Studi di Roma "La Sapienza," Italy

Interest in the didactics of demography has been growing intensely. This topic is covered in Chapter 140. Here, we wish only to emphasize that interest is demonstrated by the numerous occasions this topic appears in the programs of major international conferences and the growing proliferation of articles in demographic journals that tackle various aspects of the content, methods, and the semantics of demography. This interest is accompanied by an increasing awareness of the epistemologic implications, particularly of the much-felt need to delve deeper into the causal analysis of demographic behavior (Boulanger, 1980; Véron, 1997).

The *vexata quæstio* of interdisciplinary discussion and research, or rather problems relating to the interplay with other disciplines interested in demographic phenomenology, demands our consideration. Such an approach, which is in keeping with an awareness of the limits of any endeavors toward self-sufficiency in demography, must focus on content and on the ways in which this information is transmitted. The fear of loss of identify and autonomy felt throughout the scientific community, not to mention the subsequent demotion of demography to a subsidiary role, provides further impetus to redefine how and what to teach students who attend a course on demography. To what extent does the educational substance that remains the basis of training in this subject as contained in the popular textbooks continue to be suitable? Would it be preferable to pursue an interdisciplinary approach to teaching that would promote an understanding of the causal mechanisms underlying demographic phenomenology? How should we respond to the demands made by a society that is increasingly aware of demographic topics and that compels research in the direction of analysis of topical issues? Some analysts believe this trend should be resisted (Livi Bacci, 1994), but others are convinced that it should be fully embraced (Caselli, 1997; Pressat, 1996), and the ultimate decision will no doubt influence future research and teaching content.

Beyond the traditional problems of how and what to teach, education in demography must address where and whom to teach. The academic environments in which demographic courses are offered and the students' didactic itineraries have an impact on teaching content and methods and how textbooks are written and targeted.

Although there is increasing awareness among demographers that the future of demography depends on the training of its practitioners, the same cannot be said for current textbooks. In terms of quantity, the situation is quite satisfactory. A glance at *Population Index* for the past 20 years, under the heading of "Textbooks and Teaching Programs," produces, counting handbooks, textbooks on demographic methodology, and exercises, a total of 162 titles (in different languages, but mainly English and French), many of them republished in the meantime (some on their seventh edition, and updates concern more than the documentary section). Most of these titles are solely destined for teaching and their substance is built to measure around a specific course. Other textbooks, aimed at a larger public or as a teaching aid, tackle demographic issues on a much more general level, without going into the methodology. Too few are textbooks that offer a wide overview, that probe the links between analyses and syntheses, and that to a fairly large extent reflect demographic concerns, trends, and progress in the field.

An analysis of these textbooks (listed in Appendix 139–1, which does not pretend to be exhaustive) lies far beyond our abilities with regard to read the many languages used and to assess the widely differing pedagogic approaches and aims covered. We instead consider only textbooks published in Europe, North America, and Australia that are printed in English, French, and Italian[1] and that are mainly from the more renowned schools.[2] In an attempt to compensate for this, brief comments are provided on texts published in Russian (Appendix 139–2), German (Appendix 139–3), Spanish (Appendix 139–4), Portuguese (Appendix 139–5), Dutch (Appendix 139–6), and Japanese (Appendix 139–7). Focus is on textbooks printed after World War II, when the production of pedagogic texts really took off.

A quantitative comparison of the content of educational textbooks published in this time span is not our goal. We instead focus as much as possible on the various approaches practiced and the differences among cultures and over time in some of the major and more popular textbooks available.[3]

I. EUROPEAN TEXTBOOKS DURING "WILD DEMOGRAPHY'S" YEARS

Before dealing in detail with the content of teaching textbooks in the fifty years considered it is worth taking a glance at the situation in the interim war years. In Europe, these mark the break from statistics and the point when demography begins to acquire a certain autonomy in the teaching sphere. This hastened the desire to fix the boundaries of the field of research, still the subject of discussion today. Despite expressing different positions, academics unanimously agree that demography as meant by Achille Guillard (1855, p. XXVI) to be "the mathematical knowledge of populations, of their general movements and their physical, civil, intellectual, and moral status,"[4] is too wide a definition as research could apply different approaches other than the descriptive approaches practiced by pioneers in the field.

Different positions engender different itineraries that are also determined by various cultural contexts where demography has begun to play a role. Their only point of contact, perhaps, is the naturalist concept underlying the subject, where the biologic aspects of demographic behavior prevail, at times disclosing eugenic concerns. It is during this phase that demography, to quote Alfred Sauvy (Sauvy, 1946), is a "science sauvage, sans maîtres, ni élèves, cultivée seulement par des 'amateurs,'" a savage science without masters or pupils, cultivated only by amateurs. The air of "savagery" that marked these years was by no means an obstacle but generated a series of treatises, not particularly numerous but certainly wide ranging, a lack of which is keenly felt today.

This is the golden age of the Italian school. Between 1935 and 1937, edited by Corrado Gini and crowning his career as organizer and coordinator of research in demography, his *Trattato elementare di statistica* is published, with the two tomes of the second volume entirely devoted to demography (Gini, 1935–1937). The topics covered, involving the collaboration of leading Italian academics of the day, range from historical demography to population policy, testify to the wealth of research underway in those years. Shortly after, Livio Livi published the two volumes of his *Trattato di demografia* (Livi, 1941–1942). This treatise offers an image of demography that is refined and open to addressing other social disciplines, which is clearly in the tradition of Rodolfo Benini, and the book is a compilation of his research. Livio Livi accords this subject a scientific systematization that totally differs from that of his contemporaries, assigning demography the task of finding the "biologic basis of the social structure," anticipating by many decades the sociobiologic thread of thought. A particularly fascinating hypothe-

[1] Our analysis risks underestimating the role played by textbooks published in emerging countries or for this market. However, this would require a separate section that is beyond our capacity. Teaching and training itineraries are closely linked to employment opportunities or to the training needs specific to each country and tend to lie beyond the scope of this study.

[2] This does not undermine the importance of other schools. Publications in German in the 19th century were particularly important, not to mention more recent publications in Spain. Russian demography also has a long tradition and has produced important scientific studies, particularly throughout the first half of the 20th century. Ideologic censorship after World War II led to isolation from the rest of the scientific community, but since the collapse of the USSR, demographic studies have enjoyed a new launch in Russia, with the publication of a substantial number of teaching textbooks. Regarding publications in these and other languages, see the notes included in the appendices.

[3] Only the title and authors are cited. A complete bibliography is provided at the end of the chapter.

[4] "La connaissance mathématique des populations, de leurs mouvements généraux, de leur état physique, civil, intellectuel et moral."

sis proposed in volume one of the *Trattato* (*I fattori biodemografici dell'ordinamento sociale*) considered the biodemographic matrices from which the social order may be derived.

The *Traité de démographie*, by Adolphe Landry (Landry, 1945), is a major theoretical work. Alongside a classic analysis of natural and migratory trends, which offers a description of methods and focuses on causal factors, Adolphe Landry gives ample coverage to what he calls "pure demography," the essence of which is the stable population theory. What interests the author is not the mere description of population processes, but rather the need for general truths, of a "population law." Although he never specifically tackled the question, Adolphe Landry drew attention to the question of a qualitative demography to be achieved by the introduction of qualitative methods.

A more classic text that devotes major space to methodology is Michel Huber's, *Cours de démographie et de statistique sanitaire*, widely distributed and well known in the French-speaking world (Huber, 1938). Michel Huber emphasizes the importance of a knowledge of both sources and statistical methods in demography, and in his introduction, he stresses the fundamental role of choosing the correct methodology to ensure the success of the research.

II. THE 1950s AND 1960s: DEMOGRAPHY AS A GLOBAL SCIENCE

Demography's pioneering years came to a close at the end of the Second World War. This discipline, albeit differing in timing and modality depending on the country's idiosyncrasies, became more institutionalized, with the creation of various research centers, both public and private, and offered as part of the university syllabus. This gradual process of institutionalization was hastened by the need to be acquainted with and understand the demographic trends that became prominent in the after war years, such as the drop in fertility in the Western countries and the rapidly burgeoning world population, a source of interest for academics and of concern for governments.

The academic acknowledgment of demography as a discipline unto itself proceeded quite slowly and differed in pace from country to country. North America was one of the first countries to award full academic status to the new discipline right from the beginning of World War II. However, demography was often included in the department of sociology or sometimes the economics department, which tended to limit its autonomy (Lee, 1997; Wargon, 1997). There was also a tendency for teaching staff to come from the ranks of sociologists or similar disciplines. The demographer as a professional figure was not the objective. Focus instead was given to providing the means to understand demographic themes, given further impetus by the international interest it attracted as well as by the work of the expanding Population Division of the United Nations. In France, the importance of training was clear from the outset, as witnessed by the founding in 1957 of the Institut de Démographie of the University of Paris (IDUP) to cater to the needs that arose from the intense research carried out by INED. An emblematic figure in this synergy between demographic research and university teaching was Louis Henry, a leading founder of the French School of Demography, which tended to favor a quantitative approach.

In Italy and in Germany demography was slow to be accorded academic recognition. Interest in demographic issues, both at a political level and among public opinion, was marginal and looked at with suspicion given the political use to which it had been put in both countries in the between-war years. The first 'chair' in demography in Italy was appointed as late as 1961 and the teaching of demography in the following years tended to be fairly restricted both quantitatively (small number of courses) and qualitatively (the training goals were unclear).

Hence pedagogic projects and training itineraries are quite different as they very much depend on the cultural context of the country in question. The conceptual foundations of demography follow two main threads that are quite different from each other. The most popular deals with demographic questions in a broad sense and researches into the causes and consequences of demographic processes. The other approach is more focused on quantifying population dynamics and on how different phenomena interacted. In English, a distinction is made between *population study* and *formal demography* (Bogue, 1969, p. 4), whereas in France, the distinction is between *démographie large* and *démographie pure*, which Alfred Sauvy (1976, p. 16) defines as a "comptabilité d'hommes" and a quantitative events' analysis, which gives rise to demographic analysis that has become another part of the university syllabus. Italy continued to pursue a more global approach to demography, and it was only in the 1960s that the mathematical measurement of demographic events and their interplay were recognized and that demographic analysis was first introduced as an autonomous area of study (Federici, 1968, p. 7).

Textbooks tended to apply a unique framework to demography, while taking into account the different approaches used. The largest and most widely available textbooks may be considered as true and proper treatises that provide a state-of-the-art of formal and conceptual study of population and offer a synthesis of current knowledge in the field and urge new research.

One of the most popular textbooks of the day in the English-speaking world was *The Study of Population*, published in 1959 and edited by Philip Hauser and Otis Duncan. At 800 pages written by various collaborators, this is a monumental work, of which only a section may be considered a textbook of demography. The rest is dedicated to illustrating the state of the art of research and is more suitable for specialists than for students.

Principles of Demography by Donald Bogue was published 10 years later. It proved to be very popular and became a classic in its field. This also may be classified as a treatise, and like *The Study of Population*, it is particularly long (about 1000 pages), but the language used is more suitable for students. The approach taken is encyclopedic; the introductory chapter offers a description of the foundations of demography and its links across other sciences. The approach used is not very technical with a strong emphasis on the principles involved. *Principles of Demography* was largely inspired by Warren Thompson's *Population Problems*, which was one of the more outstanding demographic textbooks in the 1930s and 1940s and which enjoyed numerous revised editions.[5]

Among the textbooks that give a secondary role to more technical and methodologic aspects is *Population Dynamics: Causes and Consequences of World Demographic Change*, by Ralph Thomlinson, which even in the title emphasizes the more problematic aspects of demographic processes and shows an early interest in the question of urbanization. In *Population*, by William Petersen, methodology is reduced to a minimum, understandable as this work is mainly geared toward sociology students. The author, however, devotes ample space to the social context in which demographic processes occur.

Compared with English-language textbooks, continental European textbooks tend to be more well defined regarding theory, albeit still of a comprehensive nature. A key role is played by the French and Italian schools. In France, Alfred Sauvy's *Théorie générale de la population* was a real milestone. This treatise is composed of two volumes. Volume I covers economic aspects of demographic trends, and volume II addresses sociologic aspects. Sauvy also deals with technique, which is dedicated only limited space because the overall aim is not methodologic, and he encourages research into the underlying causes and consequences of demographic processes, providing theoretical deliberations and a cross-discipline approach. Although firmly placed in the tradition of Adolphe Landry, a number of innovative viewpoints are explored, with Sauvy being among the first to draw attention to issues such as aging, which was taken up by most demographers many years later.

In Italy, the leading contribution made in this period was the work by Nora Federici, *Lezioni di demografia*. This volume, modestly defined by the author as a textbook, is a treatise on the state of the art of demography in Italy. Nora Federici firmly places demography at the center of the social sciences. She describes how demographic behavior, particularly in modern times, is the result of the interplay between natural and socioultural factors, the latter having a crucial role to play. Migration, with a special focus on internal migration, is another key topic covered, and seen to be a keystone in demographic and social change. This work is valuable in pedagogic terms as can be evinced from the importance given to technical aspects of this discipline, in the search for patterns and trends, and in the study of the interrelationships found when analyzing demographic processes. This work embraces a global line of research that Corrado Gini called *integral demography*. The volume is permeated throughout by references to the biologic aspects of population and related phenomena that is reflected in the idea of demography as a bridge-discipline between the natural and social sciences. This focus on biologic factors is typical of the Italian school of demography at that time and is again found in the work by Marcello Boldrini, *Demografia*, which tackles demographic issues from both micro and macro viewpoints.

A more formal and quantitative approach to demography is found in a limited number of works that tend to be almost purely pedagogic in their intent. In the English-speaking world Mortimer Spiegelman's *Introduction to Demography* proved to be quite popular. This is a more technical work, published in 1955, and targeted at teaching actuaries. It provides an accurate systematization of basic concepts, definitions and methods, as well as their application to North American and Canadian data, whose sources are discussed in depth in a critical introduction. In 1957, Peter Cox published another textbook, *Demography*, which has enjoyed a number of revised editions over the years. The author tackles methodologic issues using an analytic approach, but the language used is accessible even for beginners, and the book covers some aspects of demographic history. The year 1958 saw the publi-

[5] The fifth edition, in 1965, was updated in collaboration with David T. Lewis.

cation of *Techniques of Population Analysis*, by George Barclay; this was part of a project by the Population Council, which lamented that no technical manual existed based on the logic of demography rather than methods and that a text was needed that was more accessible to a broader public with fewer mathematical skills than actuaries. This volume is a classic pedagogic introduction to demography, covering topics such as population definition, rates and ratios, life tables, fertility measures, population growth, and migration.

Formal demography, defined as the mathematical description of human populations is first spoken of in the United States after the publication in 1968 of the path-breaking work by Nathan Keyfitz, *Introduction to the Mathematics of Population*. This constitutes a methodologic reference framework and provides the foundations for successive publications, including *Applied Mathematical Demography*, published in 1977, and *Mathematics through Problems*, which appeared in 1982. These volumes complete Keyfitz's first work, which contains a systemic and rigorous overview of the basic theories of demographic processes, although the language would be unfamiliar to nonmathematicians, and it offers a series of examples of applications and contextual references to historical and geographic realities.

In Europe, the quantitative approach to the analysis of demographic processes is dominated by the *démographie pure* of the French school as in *L'analyse démographique. Concepts, méthodes, résultats*, by Roland Pressat. The approach taken is quite different from that in the Anglo-Saxon culture, where the endeavor is to identify the mathematical models underlying demographic processes. In Europe, the focus on perfecting ways to measure the phenomena; identifying interrelationships; and distinguishing the effects of each demographic variable from that of others so that a pure measure can be taken of both longitudinal and cross-sectional analyses. It is not by chance that demographic analysis was born in a country with a long tradition of statistics and an efficient system of data collection and surveys. In this context a series of refined measures was possible, totally inconceivable in other contexts where the statistics system was more backward. This work enjoyed an enormous success with a number of translations into different languages.[6]

[6] Translations into other languages include Spanish, El Analisis Demografico, Metodos, Resultados, Applicaciones (Mexico, 1967); Russian, Narodonaselenie i Ego Izoutcheni, (Moscow, 1966); Polish, Analiza Demograficzna (Varsaw, 1966); English, Demographic Analysis: Methods, Results, Applications (Chicago, 1972), and Italian, Analisi demografica. Concetti, metodi, risultati (Milano, 1976).

III. THE 1970s: DEMOGRAPHY IN THE LIMELIGHT

Although time limits are fairly arbitrary and difficult to fix, especially, as in our case, when dealing with a cultural trend, no one can deny that the 1970s was a period of major change that merits particular attention.

The rate of growth of the world population increased, and the north-south socioeconomic divide became even more pronounced. During this time of social upheaval, when public opinion was increasingly vigilant and well informed, demography came to occupy a pivotal position on the international scene. The first World Population Conference was held in Bucharest in 1974.

University courses on demography increased by the score, along with the production of textbooks, both new and revised editions, many of which tackled recent advances in the discipline in methods and content.

One offshoot of the cultural climate of the time was the different approach applied to demographic trends, now interpreted in terms of social processes, moving away from the previous tendency in favor of biologic factors. Another crucial element, closely linked to this and vital for the future of the discipline, was the cross-pollination between demography and the other social sciences. This exchange among disciplines facilitated grasping the existing causal links and was fundamental in interpreting demographic processes. Although it had to be admitted that demography alone could not provide all the answers, demographers failed to defend their field's specific merits regarding measuring and analyzing demographic processes. Other disciplines did not hesitate to step in to teach questions of a demographic nature (Federici 1977, p. 394). A veritable explosion occurred in the publication of texts regarding population geography. In France, in the wake of the classic work by Pierre George,[7] Daniel Noin's *Géographie de la population* appeared. In the English-speaking world, particularly popular were *Human Geography: Theories and theIr Applications*, by Michael Bradford and Ashley Kent; *Population Geography: Problems, Concepts and Prospects*, by Gary Peters and Robert Larkin; *Population Analysis in Geography*, by Robert Woods; and Geoff Dinkele's deft handbook, *Population Geography*. Occasional incursions, accompanied by the publication of a minimum number of textbooks, were made by sociologists and economists. We should mention here Judah Matras' *Introduction to Pop-*

[7] Population geography first developed in France in the 1950s, mainly thanks to his publications, including the well-known Questions de géographie de la population, (George, 1959).

ulation: A Sociological Approach, in which basic demographic techniques are described in the second part of the volume. It is not until the 1980s that demography begins to recuperate some of the ground it had lost to other disciplines. A process of integration does not see it play a subordinate role, but rather furthers the proliferation of specialized courses.

As the number of university courses multiplied, so did the publication of textbooks, along with teaching aids and other volumes aimed at a more general public. These provided simple analytic tools in accessible language to cover the burning political, economic, and ethical issues triggered by the upheaval of world events. This resulted, particularly in the United States, in a shift from analysis to "policy." This change is reflected in the work by Ian Thomas, *Population Growth*, and in *Comprendre la démographie*. *Méthodes d'analyse et problèmes de population*, by Hubert Gérard and Guillaume Wunsch. Within the same category, we find *Démographie sociale*, by Roland Pressat, which met the growing need of a society that was becoming increasingly attuned to issues of a demographic nature and in which the author showed quite clearly that the scope of demography goes far beyond the "laborious accounts of men and events."[8]

A more generalized, comprehensive book is that by Lynn Smith and Paul Zoff, *Demography: Principles and Methods*, which describes classic population analytic techniques, conducts international comparisons of the main demographic indicators, and focuses on U.S. population trends.

Mention should also be made of *Population: An Introduction to Concepts and Issues*, by John Weeks. Although the layout is quite traditional in scope, coverage of basic methods in demography is accompanied by other, more unusual themes, such as the demographic basis of economic development, the environmental impact of demographic trends, and virtually untouched in other textbooks, women's emancipation.

In France, *Démographie: analyse et modèles*, by Louis Henry, was a major work born of many years' experience with teaching at the University of Paris. In the first section of the book, the author describes demographic analytic techniques and the underlying logic, whereas the second section focuses on trends in models development, particularly those regarding fecundability, a field in which Louis Henry made great strides. The latter approaches are the result of his vast experience in dealing with historical data. Louis Henry played a pioneering role ensuring that historical demography became an independent field of research in the 1950s, with pedagogic publications following somewhat later. Worth citing during the next decade is Jacques Dupaquier's synthetic and thorough *Introduction à la démographie historique*, which appeared in 1967 and which follows in the footsteps of Henry's *Manuel de démographie historique*.

More directly aimed at training the demographer we find the two volumes by Henry Shryock and Jacob Siegel, *The Methods and Materials of Demography*, considered the "red bible" (from the color of the cover) of demographic methods. Using a somewhat technical approach, in 25 chapters this volume covers the range of demographic issues, using updated data and a language accessible to all, making this textbook suitable for specialist training and for institution based teaching. Sharing a rather similar training goal, Alfred Pollard, Farhat Yusuf, and G. Pollard edited the volume *Demographic Techniques*, that offers a more synthetic description of the basic techniques as well as a series of pertinent exercises on the themes covered.

During the 1970s the number of textbooks that appeared in the branch known as *formal demography* also grew quite considerably. Nathan Keyfitz and Wilhelm Flieger published *Population Facts and Methods of Demography* (1971), describing the most commonly used models, with a particular emphasis on theory, without which it is impossible to interpret data. A novel aspect of this volume is the inclusion of computer programs needed to apply the models. A few years later the already cited *Applied Mathematical Demography* appears, the second volume in the trilogy by Nathan Keyfitz on formal demography.

In the field of demographic models, we should cite *Introduzione ai modelli demografici*, by Paolo De Sandre, and *Mathematical Models for the Growth of Human Populations*, by John Pollard. Pollard concentrates particularly on stochastic models and presents a series of models that had only appeared in scientific journals. Although the author makes every attempt to temper the mathematical language used, expertise in this field is necessary to be able to appreciate this work.

A new branch in demography, multiregional demography, begins to appear in these years and continues to expand successively. Early crucial contributions in this field were made by Andrei Rogers, publishing in 1975 *Introduction to Multiregional Mathematical Demography*. This offers a mathematical description of a certain aspect of trends in the human population in time and in space. More specifically, to quote the author, multiregional demography ". . . analyses simultaneously the spatial dynamics of a system of several interdependent populations linked by directional migration flows" (Rogers, 1995, p. ix).

[8] "Laborieuse comptabilité d'hommes et d'événements" (Pressat, 1971, p. 7).

In the Francophone world, but written in English, the painstaking and comprehensive text by Guillaume Wunsch and Marc Termote, *Introduction to Demographic Analysis: Principles and Methods* appeared in 1978. This is not merely a collection of techniques used in demographic analysis, presented quite analytically, but rather the authors focus on the demographic process itself. First of all the issues at stake are outlined and then the techniques to tackle them are described.

IV. DEMOGRAPHY AND DEMOGRAPHIES: TEXTBOOKS SINCE THE 1980s

Throughout the 1980s demography, making the most use of the many stimuli and opportunities that were created in the previous decade, enjoyed an unforeseen scientific expansion that however had to wait until the next decade before this materialized in the publication of textbooks.

Very briefly the multifaceted developments that occurred in demographic research fall under three main intrinsically linked headings: the emergence of new lines of research, hand in hand with new methodologies and interpretative approaches; greater experience in the use of statistical analytic techniques and models; the increased availability of a rich and "multivaried" documentary material and ever more rapid and accessible technology.

More focus on emerging countries and endeavors to understand demographic processes in an underdeveloped economic and social context demanded proactive demographic research. This produced more refined methods of demographic analysis, such as defining tools to deal with data from more backward statistical systems—for example the empirical application of population models and the construction of model life tables—as well as in the theoretical field and the elaboration of explanatory frameworks describing behavior, particularly reproductive behavior. Fundamental steps were taken at this stage in research, and new analysis strategies were forged, constituting a veritable store of experience for scientific research in the following years: the shift from observation to interpretation; from macro to micro, in other words from an analytic approach to aggregate data to the exploration of individual behavior; from focusing on the impact of behavior to seeking the underpinning motivations; from cross-sectional to longitudinal analysis applying a *life course* approach. All of this occurred in a context that generated a wealth of new data, the result of sample surveys on individuals and families that were internationally comparable (the *World Fertility Surveys* and the subsequent *Demographic and Health Surveys*, for example) and fostered greater technical skills in applying suitable statistical methods for processing enormous quantities of complex and multivariate data and for meeting the need to come up with satisfactory explanations dictated to by new research strategies. These ranged from the application of exploratory techniques to multidimensional data, to the development of generalized linear models to verify explanatory hypotheses, *event history analysis* models, and simultaneous equation models to identify the latent sides of behavior as well as the treatment of unobserved heterogeneity.

During what was a particularly stimulating moment for all the social sciences, demography encountered comparable sciences, particularly sociology and economics, with the intent of pinpointing conceptual frameworks of reference, and defining new analytic approaches alongside their peers in statistics. Although there is much interplay between the various sciences, different from the immediate past, an interdisciplinary approach is not pursued, but rather a process of specialization is favored in which demography profits from experiences in other fields while acquiring and elaborating specific competencies in determined sectors of research and scientific development. There is also a shift away in teaching from an interdisciplinary dominated approach toward the specific training of the demographer. Other sciences are borrowed from with regard to reading and interpreting population issues and emphasis is given, not to the need to merge with other disciplines but, rather, to the need to pursue a multidirectional approach to research and teaching in demography. In this period a series of specialist branches of demography were developed.

An offshoot of these developments during this 20-year span is the emergence of a number of "demographies" that generate in different cultural and geographic contexts autonomous university courses along with specialized textbooks in each sector.[9]

Great strides are made by *economic demography*, not only with regard to contents and the publication of textbooks but also due to the expansion it enjoyed in academic sectors as a "frontier discipline" (De Santis, 1997). Although sometimes labeled differently (e.g., *population economics*) economic demography focuses on the interplay between demographic and economic processes. A micro approach is favored in the United

[9] A vast number of textbooks were published during this period. There is not enough space to describe all of them, nor can the choice be limited to one of merit. Preference is given to texts that introduce an element of novelty in content or because they reflect progress in teaching methods or orientations.

States, one harking back to the theories of the economist Gary Becker, and it is based on the microeconomic analysis of individual, and especially family behavior. Authors of texts in the United States tend to be economists. In the book by Assaf Razin and Efraim Savka, *Population Economics*, ample space is given to analyzing the microeconomic implications of demographic behavior. European authors favor a more macroeconomic attitude, and textbooks, essentially those written by demographers, are classic demographic manuals that deal mainly with methods and data to study the population, with a limited number of sections near the end dedicated to selected aspects of links between demography and economics. Significant publications in the field are the French *Démographie. Analyse des populations et démographie économique*, by Gérard-François Dumont, and *Analyse dynamique des populations: Les approches démographiques et économiques*, by Marie Christine Challier and Michel Philippe. Another interesting work is that by Annie Fouquet and Annie Vinokur, *Démographie socio-économique*, which introduces the concept of "social populations," such as the school population or the employed population, for whom mobility (i.e., entry and exit flows) determines social renewal.

In Italy, the work by Gustavo De Santis, *Demografia ed economia*, offers a more global vision of links between demography and economic trends by applying micro and macro analytic approaches. Some of the issues tackled include family microeconomic theory, the cost of children, and the impact of demographic aging on welfare systems. Other books that are useful as teaching aids and that offer a collection of writings on aspects linking demography and economics are the work by Marcello Natale, *Economia e popolazione*, and that edited by Paul Schulz, *Economic Demography*.

With a special focus on the impact of demographic processes on the market, first of all in the United States and more recently in Europe, a new branch known as *business demography* has emerged that particularly favors scientific and professional collaboration with the world of business. The Population Association of America (PAA) founded the Business Demographic Interest Group (BDIG), which brings together business professionals and academic researchers. In 1987, Louis Pol published the first textbook in the field, *Business Demography*, followed 10 years later by *Demography for Business Decision Making*, co-authored with Richard Thomas. The explicit aims of the books are to describe demographic trends and their economic impact and to assess how demographic reasoning and analytic techniques can be effectively applied to business problem solving. Pivotal to this approach is the focus on dynamic-structural interplay and population forecasts.

Business demography falls under the broader heading *Applied demography* that entails the ". . . application of demographic methods, data, and theory to a host of real world problems" (Pol and Thomas, 1997, p. 3). Textbooks such as *Applied demography: an introduction to basic concepts, methods and data* by Steve Murdock and David Ellis or the more substantial and recent work by Jacob Siegel, *Applied demography: application to business, government, law and public policy* (686 pages, published in 2002), may be considered useful consultation books regarding basic methods, an overview of demographic and socioeconomic data sources, the possible implications of population trends, but also regarding the real application of certain demographic techniques, such as forecasts, in the business world, in government, the social reality as it is, often with clear reference to subnational geographic entities.

From this point of view, applied demography is not confined but can embrace very specific areas of study, producing very focused teaching textbooks, such as *Housing Demography: Linking Demographic Structure and Housing Markets*, by Dowell Myers, or *Principes de démographie politique: population, urbanization et développement*, by Eric Weiss-Altaner.

Many authors are of the opinion that applied demography may be considered an extension of *social demography*, defined as "the study of the determinants and consequences of population size, distribution and composition and of the demographic processes of fertility, mortality, and migration that determine them" (Murdock and Ellis, 1991, p. 4). *Population: An Introduction to Social Demography*, by Paul Zopf is a successful example of texts in this field. The first section covers the main issues in current demographic trends, including zero growth in developed countries, the population explosion in emerging countries, and the changed population age structure. The second section offers a moderately technical examination of the various components of the demographic process. The third section deals with the interplay between the population and social systems, such as urban-rural tension, ethnic conflict, family and marriage, and economic and educational systems. The fourth and final section assesses the impact of various possible measures and policies on demographic behavior.

Among the many topics tackled under the heading social demography, particular emphasis must be given to the family that generated a specific subdiscipline, *family demography*, namely the study of family composition and structure, as well as explicit aspects of family life. An important publication in this area is *Family Demography: Methods and Their Applications*, by John Bongaarts, Thomas Burch, and Kenneth Wachter.

The family and the identification of suitable research strategies to cast light on transformations underway was given a fresh impulse in Europe in the 1990s thanks to the new paradigm known as the Second Demographic Transition. The volume, *Household Demography and Household Modeling*, edited by Evert van Imhoff, Anton Kujsten, Pieter Hooimeijer, and Leo van Wissen is a witness to current trends, for the first time examining this field of study from a variety of perspectives: macro and micro, cross-sectional and longitudinal, static and dynamic.

Social demography, particularly in academic circles, has been traditionally seen to be in contrast with the *formal demography* of mathematical measurement of basic populations (Murdock and Ellis, 1991, p. 4), which is given a new momentum in this period in the Anglo-Saxon school of demography and in the French and Italian schools.

New introductory manuals to mathematical demography begin to appear or revised editions of texts published previously, such *Formal demography* by David Smith or the second edition of *Introduction to the mathematics of demography* by Robert Brown. These offer a thorough, moderately technical description of quantitative analysis methods, with a particular focus on mortality and population models. Although these books do not require specialist knowledge of mathematics or statistics, they do address students who must be "familiar with algebra and at least not intimidated by logarithms" (Smith, 1992, p. ix). There is no lack of other fairly elementary but broader ranging texts that are more inclined toward a description of analytic techniques, while also offering examples of applications. *Demographic Methods*, by Andrew Hinde, deserves particular mention. The author endeavors to show how fertility, mortality, and nuptiality analytic methods involve the application of a few basic but common principles, and he introduces more sophisticated topics, such as parity progression ratios, survival analysis, and birth interval analysis. Each chapter contains practical examples, and an appendix is included with sample exercises, using an electronic sheet and population data that can be downloaded from the Internet.

A more formal mathematical approach to demographic processes analysis, which provides a superior description of demographic interrelationships and of population dynamics, is not an easy task. This is not easy for students or authors who strive to maintain a similar approach to a material that is highly specialized according to the processes studied and that over the past decades has been enriched by methodologic progress in statistics and the increasing availability of data. This difficulty is witnessed by the work by Krishnan Namboodiri, *Demographic Analysis: A Stochastic Approach*, which, after an introduction to the basic concepts and to single- and multiple-decrement life tables, presents a markovian approach to multistate processes, compositional analysis, and spatial distribution, as well as a thorough introduction to the topic of heterogeneity and how to deal with it. This book is exemplary with regard to the meticulous treatment it offers, but it is rather tough going for nonexperts. A more recent publication (2001), simpler to use, is the book by Samuel Preston, Patrick Heuveline, and Michel Guillot, *Demography: Measuring and Modelling Population Processes*. The text illustrates much mathematical material in demography through clear and concrete explanations, even though the material is targeted at students who have mathematical skills. It deals with basic concepts, rates and probabilities, life tables, single- and multiple-decrement processes, fertility and reproduction, population models, indirect estimation models, and increment-decrement life tables (by Alberto Palloni). Other texts describe certain methodologies adopted from statistics and then apply them to demographic processes. A pertinent example is the book by Shiva Halli and Vaninadha Rao, *Advanced Techniques of Population Analysis*, which is aimed at students and researchers and which covers, among other topics, age-period-cohort (APC) modeling, covariance structure models (e.g., LISREL), logit and probit models, and proportional hazards models.

In manuals from the French school, the approach to formal demography remains strictly anchored to the "elegant mathematics"[10] of Roland Pressat, who with the publication of *Eléments de démographie mathématique* in 1995, draws to a conclusion a lengthy itinerary dating from the early 1960s, when he proposed methods to quantify demographic processes. This itinerary began with *L'analyse démographique* and was followed in the 1980s with *Manuel d'analyse de la mortalité* and *Les méthodes en démographie.* Of particular prominence is the more specialized book by Jean Bourgeois-Pichat, *La dynamique des populations: populations stables, semi-stables et quasi-stables*. This is the fruit of 11 years' teaching experience, and it remained for a long time unfinished; it was finally completed by his students. It offers comprehensive coverage of the theory of stable populations and related topics.

Among the introductory texts published mention must be given to Alain Hillion's, *Les théories*

[10] "Mathématique élégante," Keyfitz, 1995, p. 1691.

mathématiques des populations, published in the *Que sais-je?* series. Special mention goes to the book by Julien Amegandjin, *Démographie mathématique*, which is intended as an introduction to the work of Nathan Keyfitz and is particularly suited for students that never tackled the study of demography in a classic sense in that it proposes a simple reading of demographic processes directly in continuous time, making it easier to pass from a stepwise to a continuity approach, essential for an understanding of more complex demographic processes and models.

Mathematical demography texts were also published in other European countries, such as Janina Jozwiak's *Mathematical Models of Population*, published by the Netherlands Interdisciplinary Demographic Institute (NIDI), or that edited by Giuseppe Micheli and Piero Manfredi, *Matematica delle popolazioni*.

Italy successfully contributes during this period to the teaching of demographic methods with the appearance of Antonio Santini's *Analisi demografica. fondamenti e metodi*. The result of many years' teaching and research, the book, which is divided into two volumes, pursues mainly pedagogic aims, although it reaches beyond this as it is also directed at research and makes use of techniques and data. The approach is more theoretical than pragmatic and supports the idea that analysis and measurement must adapt to the reality at hand rather than to the strictures posed by the data available. Another volume in the field of demographic analysis, published in Belgium, is Christophe Vandeschrick's *Analyse démographique*, which places greater emphasis on interpreting demographic concepts and the results of analyses than on methods and techniques.

Population geography, a discipline that was born in the field of geography but with strong demographic connotations, so as to be defined as "the study of population by geographers" (Schnell and Monmonier, 1983, p. 7), pays increasing attention to aspects of application in this period. Alongside more classic texts such as *The Study of Population. Elements, Patterns, Processes*, by George Schnell and Mark Monmonier, and the numerous revised editions of older texts, such as *Population Geography: Problems, Concepts, and Prospects*, by Gary Peters and Robert Larkin, a number of innovative texts begin to appear. These include *The Geographic Analysis of Population, with Applications to Planning and Business*, by David Plane and Peter Rogerson. The text, divided into two sections, deals mainly with fundamental notions of geography and population analysis, with a particular emphasis on spatial aspects and mobility. The second section offers concrete examples from the United States regarding the implications of population trends on territorial planning (urban infrastructure, transport, building) and locating a business (e.g., market segmentation).

During the last 20 years of the 20th century, interest was still very much alive among population experts regarding historical demography, which is a restricted field of research but one that enjoys expertise from other fields such as history, geography, and genetics and that avails itself of specific analytic techniques suited to the particular nature of the sources of information used. Two texts that stand out are Louis Henry and Alain Blum's *Techniques d'analyse en démographie historique*, which offers a re-reading of classic analytic techniques using historical data and, in this case, French data, and the work by Lorenzo Del Panta and Rosella Rettaroli, *Introduzione alla Demografia Storica*, which in Italy fills a gap because there are no other comprehensive texts on this material in Italian and existing texts do not adequately describe the wealth and specific nature of the documentary material available. The text also contains references to more recent trends in methodology and techniques to tackle complex data, such as *record linkage*. Researchers and students with a good background in mathematics and statistics who are attracted to more advanced methods in the field of historical demography can consult the book edited by David Reher and Roger Schofield, *Old and New Methods in Historical Demography*, which brings together the most significant addresses made at the IUSSP Committee Seminar on Historical Demography in 1991 and that ranges from the application of methods of time series analysis, family reconstitution techniques, to event history analysis and simulation procedures applied to historical data series.

There was during these years a considerable increase in the number of introductory demographic texts; alongside revised editions of previously published texts, a series of first editions also appeared. This editorial dynamism was generated by the need to adapt contents to the different educational training of the student and to the different teaching goals of demography courses. In some instances, the task is to furnish the student with a knowledge of complex techniques, whereas in others, they need an introduction to specific issues covered by a particular course. Two types of introductory textbooks are available; one type is more focused on techniques, and the other is focused on issues related to demographic change.

Training needs and the approach used are also reflected in the very definitions the authors provide of demography. In these definitions, we can detect an attempt to define the boundaries of demography that had become somewhat blurred because of the multiplication of the different natures of the "demographies" that had been generated. Elsewhere, a certain reticence was expressed regarding establishing restrictions, opting more for affinities with other disciplines than defining the content, methods, and aims of demography as such. There were also the "minimalists," who do not perceive the need for definition and are happy with the definition provided by the multilingual demographic dictionary published by the IUSSP at the beginning of the 1980s.[11]

Another aspect that deserves attention is the attitude on the part of authors to interpretative paradigms that are only rarely covered in the texts mentioned. We could expect, particularly when a major effort is made to define demography, to find some trace of discussion, even critical, about possible "laws" underlying demographic change. Although this is the object of scientific discussion, any attempt to elaborate a general model or paradigm, a theoretical interpretation to contain partial explanations of single processes, is rare in teaching and usually only refers to the first demographic transition theory. Other major theories, such as that of John Hajnal on forms of nuptiality and economic trends or that of Richard Easterlin regarding cohort size alternation as the theory of fertility fluctuations, are never covered.

For the reasons previously given, we address only the more popular textbooks or those that offer certain novelties compared with previous publications. With regard to Italy, mention should be made of the book by Luciano Petrioli, *Demografia. Fatti e metodi di studio della popolazione*. This work tackles traditional aspects of demography along with in-depth analyses of the main mortality and fertility models that once again implies a good grounding in mathematics. A diskette is also supplied with the book that contains programs created by the author to analyze themes covered in the text, a novelty as far as Italy is concerned. The most popular textbook available in Italy is Massimo Livi Bacci's *Introduzione alla demografia*. This is a text on methodology that also describes to what extent demography is intertwined with other biosocial processes and that critically steers the interpretation of the results obtained by the application of various methods. The store of analytic techniques supplied is particularly rich, and the methods presented are sometimes quite sophisticated, but they are accessible to anyone who is approaching demography for the first time.

[11] The *Multilingual Demographic Dictionary* defines demography as "the scientific study of human populations primarily with respect to their size, their structure and their development; it takes into account the quantitative aspects of their general characteristics" (van de Walle and Henry, 1982, p. 15).

Textbooks in English destined for a wider public, even beginners because only a limited knowledge of mathematics is requested, include *Demographic Methods and Concepts*, by Donald Rowland, which contains a wealth of exercises; *A primer of Population Dynamics*, by Krishnan Namboodiri, which has an unusual layout, with each chapter containing a number of questions and answers on the topics covered; and *Beginning Population Studies*, edited by David Lucas and Paul Meyer, which is widely used by Australian universities.

Other textbooks ascribe a secondary role to methodology privileging themes regarding demographic processes. One such work is that by Helen Daugherty and Kenneth Kammayer, *An Introduction to Population*, which is targeted at students from different educational backgrounds. Demography is described as an interdisciplinary science, and ample room is awarded to how this relates to economic, political, and sociologic factors. Charles Nam and Susan Philliber's work, *Population. A Basic Orientation*, also focuses more on issues than on techniques, particularly emphasizing the causes and effects of population change and population policies. The same approach is applied in French publications, and most demography primers are quite comprehensive, not mere introductions to demographic methods. Examples include is *La Science de la population*, by Gabriel Poulalion, which also covers environmental policies; the work by Georges Tapinos, *Eléments de démographie. Analyse, déterminants socio-économiques et histoire des populations*, in which population change is explained within an economic context; and *Démographie*, by Michel Dupaquier, which largely concentrates on an analysis of the family and its current crises.

A more innovative text is *Démographie: approche statistique et dynamique des populations*, by Henri Leridon and Laurent Toulemon. The authors successfully combine a methodologic approach with the treatment of demographic issues. The work concentrates on methodology (covering recent developments and purely demographic techniques, little known among the other social sciences, such as life course transition or event history analysis), while ensuring that the reader is introduced to the socioeconomic causes and consequences of population trends. This guarantees that it is a valuable volume even for the expert.

Alongside primers and textbooks an increasing number of books and publications that also act as teaching aids are published. Furthering an encyclopedic style, on behalf of UNFPA, *Readings in Population Research Methodology*, edited by Donald Bogue, Eduardo Arriaga, Douglas Anderton, and George Rumsey, appeared in 1993. This consists of eight volumes that systematize the great bulk of methodologic research that was conducted in the period in question, offering a state-of-the-art work that is updated to January 1 of the year of publication. It contains a collection of selected reprints from journals, treatises, textbooks, and research reports that were mainly focused on the methodologic content of specific demographic topics. Each volume is dedicated to a single topic: *Basic Tools*; *Mortality Research*; *Fertility Research*; *Nuptiality, Migration, Household and Family Research*; *Population Models, Projections and Estimates*; *Advanced Basic Tools*; *Contraception and Family Planning*; *Environment and Economy*. Although the authors aim "to promote decentralization and widespread diffusion of population research and population planning" (vol. 1, p. iv), this reader is targeted at experts, even though it could be used in introductory courses as an aid "to cover methodologic facets not covered in their favorite textbook" (vol. 1, p. iv).

Another aid to teaching is the *Handbook of Population and Family Economics*, a two-volume work with a total of 1344 pages, edited by Mark Rosenzweig and Oded Stark. This is a collection of readings on classic themes (e.g., family, fertility, mortality, and migration), this time constituting a state-of-the-art work regarding theory and the most recent empirical evidence.

Besides these rather weighty volumes, a substantial number of smaller and more manageable textbooks were published. They cover demographic topics and provide elementary analytic techniques, and they are clearly designed as teaching aids or designed for the layman, in line with a tradition going back to the 1970s. Among the former, there is Shirley Hartley's *Comparing Populations*, which fosters collaborative learning in which each student chooses an emerging and developing country and then compares it with a developed country using a selection of sociodemographic indicators. Among the latter, is the particularly effective (already at its fourth edition) *The Population Reference Bureau's Population Handbook*, by Arthur Haupt and Thomas Kane. It is a quick guide to population dynamics for journalists, politicians, teachers, students, and the general public.

Pursuing a similar approach is the French series *Que sais-je?* that published *La démographie*, by Jean-Claude Chesnais, and the series *Repères* that published *La démographie*,[12] by Jacques Vallin, each of which have been republished several times. Other pocket books that enjoyed a certain success are *Déchiffrer la démographie*, by Michel Lévy; *La démographie. Population, économie et sociétés*, by Georges Tapinos; *Population et développement*, by Jacques Véron, which uncovers links between demography and economics; and *La Popolazione del pianeta*, by Antonio Golini, offering a competent synthesis of population issues in the modern world.

V. FUTURE HORIZONS: TEXTBOOKS AND MANUALS IN THE NEW MILLENNIUM

As we begin a new millennium, we understand that demography has grown steadily as a science and as a teaching discipline. In the United States alone, there were as many as 250 research and training centers at the end of the past century (Roussel and Chasteland, 1996, p. 436). In Europe, numerous university syllabi now offer demography as a subject—although this is often a complementary course included in other departments, such as statistics, economics, political science, sociology, not to mention architecture, urbanism, natural sciences—and various prestigious nonacademic research centers and postdoctoral training bodies closely tied to more traditional centers, such as the Institut National d'Études Démographiques (INED) in France, the Istituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS) in Italy, or the Netherlands Interdisciplinary Demographic Institute (NIDI) and the International Institute for Applied Systems Analysis (IIASA) in Austria. Special mention must be given to the Max Planck Institute for Demographic Research in Rostock, set up in 1995, whose activity is emblematic of the recognition from the other sciences that demography now enjoys (the Max Planck Society has founded more than 80 research and training organizations in the natural and social sciences). Thanks to a substantial budget and the caliber of their highly qualified staff, the Max Planck Institute has added further impetus to the international nature of the training of professional demographers. Doctoral students come from all over the world to study within the framework of an international doctorate program, and postdoctoral fellows and junior researchers participate in ongoing research activities, achieving outstanding results and actively promoting international collaboration and the global dissemination of knowledge, aided in this by the online journal *Demographic Research.*

In the near future, the global diffusion of knowledge will make great strides, and opportunities of access offered by the Internet will become central to research and to professional training and teaching. Two aspects

[12] This series also contains two other publications by the same author: *La Population mondiale* and *La population Française*, each of which are complementary, and describe the concepts and analytic approaches pursued in the first volume.

are particularly important. First, there is the increasing availability of free, easily accessible databases that offer resources that cut across various branches but of assistance for teaching purposes in that they offer concrete examples of issues treated in the traditional classroom, as well as direct applications of statistical analytic techniques on real data and case studies, an alternative to the more simplified invented situations used to date by the teacher. It is not by chance that data provider institutes, often also doubling as research centers, offer support aids to facilitate the use of their databases, which are increasingly consulted for pedagogic purposes. An example of this is the U.S. Bureau of Census; it is possible to download from their site data to construct population pyramids for populations throughout the world and the pyramids themselves. A number of national statistic bureaus also provide resources for use by university students, while others, such as ISTAT (Italy), make material available for schools with a special online course on statistics—more or less in line with Power from Data, Statistics Canada, with frequent reference to population data. Another aspect of the Internet revolution is the availability of electronic journals, online catalogs, virtual libraries—in other words, immediate access to scientific information around the world, a valuable tool for research and for teaching purposes that is easy to access and cheap, cutting through barriers of time and space and offering a multitude of stimuli to further knowledge and establish unexpected links right across disciplines.

This brief comment on the opportunities provided on the web regarding access to data and information across the globe creates awareness of the new horizons looming in teaching in terms of content and method. Generally, demographic didactics lends itself to a multimedia approach, using interactively a variety of tools, both real and virtual and global, capable of tackling issues at a theoretical level and their application, grasping their implications regarding international comparisons and multifactorial explanations.

What are the texts to be for teaching demography in the new millennium? We do not pretend to provide an exhaustive list nor propose new "typologies" of teaching aids. Our real scope is to offer some suggestions regarding the book material available today within an information framework and teaching context that is continually evolving.

Classic texts, used primarily for basic institutional courses, will continue to cover basic demographic concepts and fundamental analytic techniques, but they will become less essential for providing examples of the issues covered and applications of the techniques studied, replaced in this by resources available on the web. Easy-to-use textbooks, which are on the increase but by no means exhaustive in their coverage of the complexity of the population system, are a useful teaching aid within a multimedia course, particularly where demography is a complementary examination.

Less and less use will be made of anthologies, such as "readers," containing reprints of scientific articles or parts of books on specific topics, still valid learning tools yet destined to be replaced by electronic resources and virtual libraries. With the web, a dream may come true: global sharing of all science and culture, thanks to the growing online access to research, proceedings, and scientific forums, a witness to the desire on the part of the scientific community to make their work available even through the use of nontraditional channels of communication.

Textbooks on methodology remain fundamental, offering a systematic and thorough description of the techniques of formal demography and statistical analysis and their potential application to demographic data and issues. The web offers sources and data for the description and interpretation of events, but the correct use and scientific accuracy of the application of the logical principles and analytic techniques depends on the culture and skills of the individual scientist.

Although method is still pivotal in forming a demographer, trends focusing on population issues applying a global approach tend toward the "construction" of a whole researcher, one not isolated in a methodologic ivory tower but one who is more attentive, using the techniques of the trade, and is ready to explain the processes underway and assess their many possible impacts.

There is a real need for a comprehensive work that covers the underlying principles of demographic analysis, including mathematical and statistical techniques, and that offers detailed illustration of the sources, a theoretical guide to understanding population processes and a conceptual framework to assist in researching causes and consequences. In an increasingly global society, a treatise such as this one, which was lacking during the last 20 years of the 20th century, becomes a necessity. The promoters of this enterprise, who have pursued the ambitious endeavor of gathering for the future the rich heritage of demographic knowledge and skills, have ensured that this gap is filled.

Acknowledgment

This chapter was translated from the Italian version by Mary Joan Crowley.

References

Bogue Donald J., 1969. *Principles of Demography.* New York, John Wiley & Sons, 917 p.

Boulanger Paul-Marie, 1980. *Introduction à l'usage de quelques conceptsépistemologiques dans le champ de la démographie.* Louvain-la-Neuve, Département de Démographie, Université Catholique de Louvain, 19 p. (Working paper, no. 82.)

Caselli Graziella, 1997. Réflexions sur l'avenir de la démographie en Italie, *in:* Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la Démographie au seuil du XXI*e *siècle*, p. 295–397. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

De Santis Gustavo, 1997. *Demografia e economia.* Bologna, Il Mulino, 354 p. (Studi e ricerche.)

Federici Nora, 1968. Le tendenze della demografia e le prospettive del suo sviluppo, *Genus*, vol. 24(1–4), p. 1–25.

Federici Nora, 1977. Rimeditando sulle scienze statistiche e sulle scienze demografiche, *Rivista italiana di Economia*, *Demografia e Statistica*, vol. 31, "Studi in memoria di Livio Livi e Lanfranco Maroi," p. 387–400.

George Pierre, 1959. *Questions de géographie de la population.* Paris, INED and PUF, 231 p. (Travaux et documents, cahier no. 34.)

Gini Corrado (ed.), 1935 and 1937. *Trattato elementare di statistica. Vol. II: Demografia.* Milan, Antonino Giuffré Editore, 2 vol., 307 + 306 p.

Guillard Achille, 1855. *Éléments de statistique humaine ou Démographie comparée.* Paris, Guillaumin et Cie Libraires, 376 p.

Huber Michel, 1938–1941. *Cours de Démographie et de Statistique sanitaire.* Paris, Hermann et Cie Éditeurs, 6 vol., 678 p.

Keyfitz Nathan, 1995. Ecrire la démographie mathématique: un nouveau manuel de R. Pressat, *Population*, vol. 50(6), p. 1685–1692.

Landry Adolphe, 1945. *Traité de Démographie.* Paris, Payot, 646 p.

Livi Livio, 1941–1942. *Trattato di demografia.* I. Le leggi naturali della popolazione. II. I fattori bio-demografici nell'ordinamento sociale. Padova, Cedam, Casa Editrice dott. Antonio Milani, 2 vol., 662 p. (Reprint, 1974.)

Livi Bacci Massimo, 1994. *La demografia*, *in:* Massimo Livi Bacci, Gian Carlo Blangiardo and Antonio Golini (eds.), *Demografia*, p. 3–23. Torino, Fondazione Giovanni Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Murdock Steve H. and Ellis David R., 1991. *Applied demography: an introduction to basic concepts*, *methods and data.* Boulder/San Francisco/Oxford, Westview Press, 300 p.

Pol Louis Georges and Thomas Richard K., 1997. *Demography for business decision making.* Westport (Connecticut) Quorum Books, 270 p.

Pressat Roland [Пресса Роллан], 1966. *Народонаселение и его изучение [Population et méthodes d'études].* Moscow, Статистика [Statistika], 443 p. (Перевод с французского В.Э.Шпринка под редакцией Б.Ц.Урланиса. [Translated from French by V. E. Shprink, edited by B. Cz. Urlanis].)

Pressat Roland, 1966. *Analiza demograficzna.* Warsaw, Pantswove Wydawnictwo Naukowe, 523 p.

Pressat Roland, 1967. *El analisis demografico*, *metodos*, *resultados*, *applicaciones.* Mexico, Fondo de Cultura Economica, 440 p.

Pressat Roland., 1971. *Démographie sociale.* Paris, Presses Universitaires de France, 168 p.

Pressat Roland, 1972. *Demographic analysis: methods*, *results*, *applications.* Chicago, Aldine Atherton, 498 p.

Pressat Roland, 1976. *Analisi demografica. Concetti*, *metodi*, *risultati.* Milan, Etas Libri, 320 p.

Pressat Roland, 1996. Recherche et enseignement en démographie: réflexions diverses, *in:* Graziella Caselli (ed.), Demografia: analisi e sintesi. Cause e conseguenze dei processi demografici, Proceedings of the international seminar. Demography: analysis and synthesis. Siena 22–24 April 1996, vol. 1, p. 241–251. Rome, Università di Roma "La Sapienza," Dipartimento di Scienze Demografiche, 252 p. (Materiali di studi e ricerche, special number.)

Rogers Andrei, 1995. *Multiregional demography: principles*, *methods and extensions*, Chichester, John Wiley & Sons, 236 p.

Roussel Louis and Chasteland Jean-Claude, 1996. Un demi-siècle de démographie dans les pays industriels. Quelques réflexions sur un bilan, *Population*, vol. 51(2), p. 417–438.

Sauvy Alfred, 1946. Éditorial. *Population*, vol. 1(1), p. 5–8.

Sauvy Alfred, 1976. *Eléments de démographie.* Paris, Presses Universitaires de France, 391 p.

Schnell George A. and Monmonier Mark S.,1983. *The study of population: elements*, *patterns*, *processes.* Columbus, OH, Charles E. Merril, 371 p.

Smith David P., 1992. *Formal Demography*, New York, Plenum Press, 330 p.

Thompson Warren S., 1930. *Population problems.* New York, McGraw-Hill, 500 p.

Véron Jacques, 1997. *Pour un enseignement d'épistémologie de la démographie.* Paris, INED, 14 p. (Paper presented at the XXIII International Population Conference, IUSSP, Beijing, 11–17 Oct. 1997.)

APPENDIX

139–1

Demographic Textbooks and Treatises in English, French, and Italian[13]

A SELECTION OF TREATISES AND TEXTBOOKS PUBLISHED BETWEEN THE TWO WORLD WARS

Boldrini Marcello, 1928. *Lezioni di demografia.* Milan, Tipografia Mariani, 215 p.

Gini Corrado (ed.), 1935, 1937. *Trattato elementare di statistica. Vol. II. Demografia.* Milan, Antonino Giuffré Editore, 2 vol., 307 + 306 p.

Gini Corrado, Boldrini Marcello, De Berardinis Luigi, and Zingali Gaetano, 1930. *Trattato italiano d'igiene pubblicato per monografie. Vol. 18: Demografia, antropometria, statistica sanitaria, dinamica delle popolazioni.* Turin, Unione Tip. Edit. Torinese, 740 p.

Huber Michel, 1938–1941. *Cours de Démographie et de Statistique sanitaire.* Paris, Hermann et Cie Éditeurs, 6 vol., 678 p.

Landry Adolphe, 1945. *Traité de démographie.* Paris, Payot, 646 p.

Livi Livio, 1934. *Lezioni di demografia.* Padova, CEDAM, 327 p.

Livi Livio, 1941–1942. *Trattato di demografia.* I. *Le leggi naturali della popolazione.* II. *I fattori bio-demografici nell'ordinamento sociale.* Padova, CEDAM, Casa Editrice dott. Antonio Milani, 2 vol., 662 p. (Republished in 1974.)

Mazzilli Beniamino, 1943. *Corso di demografia.* Bari, Città di Castello, 289 p.

Niceforo Alfredo, 1924. *Lezioni di demografia.* Naples, G. Majo, 466 p.

Parenti Giuseppe, 1935. *Lezioni di demografia: anno accademico 1934–35.* Florence, Casa editrice poligrafica universitaria, 328 p.

Savorgnan Franco, 1936. *Corso di Demografia.* Pisa, Nistri-Lischi, 226 p.

Tagliacarne Guglielmo, 1940. *Elementi di demografia.* Milan, Giuffré, 223 p.

Thompson Warren S., 1930. *Population problems.* New York, McGraw-Hill, 500 p.

Vinci Felice, 1939. *Problemi demografici.* Bologna, Nicola Zanichelli Editore, 228 p.

Major Textbooks and Treatises Published between 1946 and 1980[14]

Audroing Jean-François, 1978. *Démographie. Le cas français.* Paris, Economica, 158 p.

Barberi Benedetto, 1975. *Elementi di demografia.* Rome, La Goliardica, 238 p.

Barclay George W., 1958. *Techniques of population analysis.* New York, John Wiley & Sons, 311 p.

Beaujeu-Garnier Jacqueline, 1956 and 1958. *Géographie de la population.* Paris, Génin, 2 vol., 435 + 574 p.

Benjamin Bernard, 1968. *Demographic analysis.* London, G. Allen and Unwin Ltd., 162 p.

Blanc Robert, 1962. *Manuel de recherche démographique en pays sous-développé.* Paris, INSEE, Service de coopération, 227 p.

Bogue Donald J., 1969. *Principles of demography.* New York, London, Sydney and Toronto, John Wiley & Sons, 917 p.

Boldrini Marcello, 1956. *Demografia.* Milan, Giuffré, 527 p. (Republished: 1st edition, 1938.)

Bonaguidi Alberto, 1978. *Corso di demografia: principi elementari di analisi demografica: appunti ad uso degli studenti.* Pisa, Opera universitaria, 121 p.

Bourcier de Carbon Luc, 1976. *Démographie géo-économique.* Paris, Ed. Montcrestien, 2 vol., 417 + 421 p. (Les Cours de Droit.)

Bradford Michael G. and Kent W. Ashley., 1978. *Human geography: theories and their applications.* Oxford, Oxford University Press, 180 p.

[13] We are grateful to Martine Deville-Velloz and Marie-Claude Lunazzi, archivists at INED, for the additional references provided by them that have enabled us to improve this list. For reasons already mentioned, the publications included here are only those published in Europe, North America, and Australia. The later appendices complete this list through a specific focus on books and manuals in Russian, Spanish, Portuguese, German, Dutch, and Japanese.

[14] Where there are several editions, the most recent edition is cited and, when possible, the date of the first edition is also mentioned.

BUQUET Léon, 1974. *Démographie.* Paris, Masson et Cie, 186 p.

CAMPBELL H., KOLLER S., HANSLUWKA Harald, and MANDEL S., 1977. *Manual of mortality analysis. A manual on methods and analysis of national mortality statistics for public health purposes*. Geneva, WHO, 245 p.

CHIANG Chin Long, 1978. *Life Table and Mortality Analysis.* Geneva, WHO/OMS, 399 p.

COX Peter R., 1976. *Demography*, 5th ed. Cambridge, London, New York and Melbourne, Cambridge University Press, 393 p. (1st edition, 1950.)

CROZE Marcel, 1965. *Cours de démographie.* Paris, Institut national de la statistique et des études économiques, 232 p. (1st edition, 1960.)

DAJOZ Roger, 1974. *Dynamique des populations.* Paris, Masson, 301 p.

DE SANDRE Paolo, 1974. *Introduzione ai modelli demografici.* Padova, CLEUP, 281 p.

DINKELE Geoff, 1980. *Population geography.* London, G.G. Harrap, 32 p.

DUPÂQUIER Jacques, 1974. *Introduction à la démographie historique.* Paris, Tournai and Montreal, Gamma, 126 p.

FEDERICI Nora, 1968. *Lezioni di demografia*, 3rd ed. Rome, De Santis, 856 p. (1st edition, 1955.)

FEDERICI Nora, 1979. *Istituzioni di demografia, parte I.* Rome, Elia, 789 p.

GÉRARD Hubert and WUNSCH Guillaume, 1973. *Comprendre la démographie. Méthodes d'analyse et problèmes de population*. Verviers, Gérard & Cie, 184 p.

GEORGE Pierre, 1959. *Questions de géographie de la population.* Paris, INED and PUF, 231 p. (Travaux et documents, cahier no. 34.)

GRANELLI Benini Luciana, 1974. *Introduzione alla demografia storica.* Florence, La nuova Italia, 159 p.

HAUSER Philip M. and DUNCAN Otis Dudley (eds.), 1959. *The study of population: an inventory and appraisal.* Chicago, University of Chicago Press, 864 p.

HEER David M., 1968. *Society and population.* Englewood Cliffs, Prentice Hall, 127 p.

HENRY Louis, 1970. *Manuel de démographie historique*, 2nd ed. Geneva and Paris, Droz, 146 p. (1st edition, 1967.)

HENRY Louis, 1972. *Démographie: analyse et modèles.* Paris, Larousse, 341 p.

HENRY Louis, 1973. *Perspectives démographiques*, revised 2nd ed. Paris, INED, 115 p. (1st edition, 1964.)

HENRY Louis, 1976. *Population: analysis and models.* London, Edward Arnold, 301 p. (Translated by Etienne van de Walle and Elise F. Jones.)

HUGON Paul, 1971. *Démographie.* Paris, Dalloz, 337 p.

KELSALL Keith, 1979. *Population*, 4th ed. London and New York, Longman, 136 p. (1st edition, 1967.)

KEYFITZ Nathan, 1977. *Applied mathematical demography.* New York, John Willey and Sons, 388 p. (Republished in 1985 by Springer-Verlag, New York, 442 p.)

KEYFITZ Nathan, 1977. *Introduction to the mathematics of population with revisions.* Reading Menlo Park, MA, Addison-Wesley Publishing, 490 p. (1st edition 1968.)

KEYFITZ Nathan and FLIEGER Wilhelm, 1971. *Population facts and methods of demography.* San Francisco, W.H. Freemen, 613 p.

LASORSA Giovanni, 1962. *Demografia*, 7th ed. Bari, Cacucci, 301 p.

LUZZATTO FEGIS Paolo, 1968. *Statistica demografica e economica*, 4th ed. Turin, Utet, 701 p. (1st edition, 1940.)

MATRAS Judah, 1977. *Introduction to population: a sociological approach.* Englewood Cliffs, NJ, Prentice-Hall, 452 p.

MAUCO Georges and GRANDAZZI Maurice, 1948. *La démographie à l'école, manuel à l'usage des maîtres.* Paris, Alliance nationale contre la dépopulation, 121 p.

MCARTHUR Norma, 1964. *Introduction aux statistiques démographiques.* Paris, Dunod and Commission Pacifique sud, 160 p. (Translated by J. Saloma.)

MORI Alberto, 1968. *Lineamenti di demografia economica.* Pisa, Lib. Goliardica, 261 p.

MORTARA Giorgio, 1960. *Economia della popolazione. Analisi delle relazioni tra fenomeni economici e fenomeni demografici.* Turin, Utet, 514 p.

MOUCHEZ Philippe, 1968. *Démographie*, 2nd ed. Paris, Presses universitaires de France, 260 p. (1st edition, 1964.)

NAM Charles B. and GUSTAVUS Susan O., 1976. *Population: the dynamics of demographic change.* Boston, Houghton Mifflin, 363 p.

PENOUIL Marc, 1977. *Démographie.* 2nd ed. Paris, Dalloz, 90 p. (1st edition, 1970)

PETERSEN William, 1969. *Population*, 2nd ed. New York and London, The MacMillan Company, 186 p.

POLLARD John H., 1973. *Mathematical models for the growth of human populations.* Cambridge, Cambridge University Press, 186 p.

PRESSAT Roland, 1961. *L'analyse démographique: méthodes, résultats, applications*. Paris, PUF, 402 p.

PRESSAT Roland, 1966. *Principes d'analyse: cours d'analyse démographique de l'IDUP*. Paris, INED, 158 p.

PRESSAT Roland, 1967. *Pratique de la démographie: trente sujets d'analyse.* Paris, Dunod, 302 p.

PRESSAT Roland, 1971. *Démographie sociale.* Paris, Presses universitaires de France, 168 p.

PRESSAT Roland, 1971. *Principi di analisi demografica.* Florence, Cooperativa libraria Universitatis studii Florentini, 170 p.

PRESSAT Roland, 1971. *Demografia applicata: trenta soggetti di analisi.* Rome, La Goliardica, 396 p.

PRESSAT Roland, 1972. *Démographie statistique.* 2nd ed. Paris, Presses universitaires de France, 321 p. (2nd ed.)

PRESSAT Roland, 1972. *Elementi e problemi di demografia.* Milan, Franco Angeli, 149 p.

PRESSAT Roland, 1972. *Demographic analysis: methods, results, applications.* Chicago and New York, Aldine Atherton, 498 p.

PRESSAT Roland, 1973. *L'analyse démographique: concepts, méthodes, résultats*, 3rd ed. Paris, Presses universitaires de France, 295 p. (1st edition 1969.)

PRESSAT Roland, 1976. *Analisi demografica. Concetti, metodi, risultati.* Milan, Etas Libri, 320 p.

ROGERS Andrei, 1975. *Introduction to multiregional mathematical demography.* New York, London, Sidney and Toronto, John Wiley and Sons, 203 p.

ROUSSEL Louis and GANI Léon, 1973. *Analyse démographique. Exercices et problèmes.* Paris, Colin, 219 p.

SAUVY Alfred, 1963. *Théorie générale de la population, vol. I: Economie et croissance.* 3rd ed. Paris, Presses universitaires de France, 373 p. (1st edition, 1952.)

SAUVY Alfred, 1966. *Théorie générale de la population. Volume II. La vie des populations*, 3rd ed. Paris, PUF, 403 p. (1st edition, 1954, *Biologie Sociale.*)

SAUVY Alfred, 1974. *General theory of population.* London, Methuen, 550 p. (Translated by Christophe Campos, Preface by E.A. Wrigley.)

SAUVY Alfred, 1976. *Éléments de démographie.* Paris, Presses universitaires de France, 391 p.

SHRYOCK Henry S., SIEGEL Jacob S. and LARMON Elizabeth A. (eds.), 1975. *The methods and materials of demography*, 3rd ed. Washington, US Bureau of the Census, US Government Printing Office, 2 vol., 369 + 519 p. (1st edition, 1971.)

Shryock Henry S., Siegel Jacob S., and Stockwell Edward G. (eds.), 1976. *The methods and materials of demography*, condensed ed. New York, Academic Press, 577 p. (1st edition, 1971.)

Smith T. Lynn and Zopf Paul E., 1976. *Demography: principles and methods*, 2nd ed. Port Washington, Alfred Publishing, 615 p.

Somogyi Stefano, 1977. *Introduzione alla demografia.* Palermo, Arti grafiche siciliane, 429 p.

Spiegelman Mortimer, 1980. *Introduction to demography.* London, Harvard University Press, 514 p. (1st edition, 1955, Chicago, The Society of Actuaries.)

Thomas Ian, 1980. *Population growth.* London, McMillan Education, 60 p.

Thomlinson Ralph, 1965. *Population Dynamics: Causes and Consequences of World Demographic Change.* New York, Random House, 575 p.

Thompson Warren S. and Lewis David T., 1965. *Population problems*, 5th ed. New York, McGraw-Hill, 450 p. (1st edition Thompson Warren S., 1930.)

Villey Daniel, 1958. *Leçons de démographie.* Paris, Montschrestien, 2 vol., 417 + 356 p.

Weeks John R, 1978. *Population: an introduction to concepts and issues.* Belmont (California), Wadsworth Publishing, 366 p.

Woods Robert, 1979. *Population analysis in geography.* London and New York, Longman, 278 p.

Wrong Dennis H., 1961. *Population*, revised 2nd ed. New York, Random House, 134 p. (1st edition, 1956.)

Wunsch Guillaume J. and Termote Marc G., 1978. *Introduction to demographic analysis: principles and methods.* New York, Plenum Press, 274 p.

Textbooks, Popularized Publications, and Treatises Published after 1980

Amegandjin Julien, 1989. *Démographie mathématique.* Paris, Economica, 265 p.

André Robert, 1981. *Démographie.* Brussels, Presses Universitaires de Bruxelles, 3 vol., 390 p.

Arafat Ibtihaj S. and Allen Donald E., 1995. *Thinking about population: an introduction to modern demography.* Dix Hills (New York), General Hall, 377 p.

Balario Grégoire, 1990. *Dynamique des populations.* Louvain-la-Neuve and Brussels, CIDEP, United Nations and Artel, 66 p. (cahiers du CIDEP no. 6.)

Biswas Suddhendu, 1988. *Stochastic processes in demography and applications.* New York, Chichester and Brisbane, John Wiley, 359 p.

Blangiardo Gian Carlo, 1997. *Elementi di demografia.* 2nd ed. Bologna, Il Mulino, 224 p. (1st edition, 1987.)

Bogue Donald J., Arriaga Eduardo E., Anderton Douglas L., and Rumsey George W. (eds.), 1993. *Readings in population research methodology.* Chicago, UNFPA and Social Development Center, vol.1: *Basic tools*, pag. mult., total: 357 p.; vol.2: *Mortality research*, pag. mult., total: 479 p.; vol.3: *Fertility research*, pag. mult., total: 359 p.; vol.4: *Nuptiality*, *migration*, *household and family research*, pag. mult.; vol.5: *Population models*, *projections and estimates*, pag. mult., total: 471 p., vol.6: *Advanced basic tools*, pag. mult., total: 414 p.; vol.7: *Contraception and family planning*, pag. mult., total: 364 p.; vol.8: *Environment and economy*, pag. mult., total: 285 p.

Bongaarts John, Burch Thomas, and Wachter Kenneth (eds.), 1987. *Family demography: methods and their applications.* Oxford, Clarendon and IUSSP, 365 p. (International Studies in Demography.)

Bonneuil Noël, 1997. *Introduction à la modélisation démographique: 26 problèmes corrigés*. Paris, Armand Colin, 128 p. (Collection U.)

Bourgeois-Pichat Jean, 1994. *La dynamique des populations. Populations stables*, *semi-stables*, *quasi-stables.* Paris, INED and PUF, 296 p. (Travaux et documents, cahier no. 133.)

Brown Robert L., 1997. *Introduction to the mathematics of demography*, 3rd ed. Winsted, CT, ACTEX Publications, 287 p. (1st edition, 1991.)

Calot Gérard, Livi-Bacci Massimo, Blayo Chantal, Sardon Jean-Paul, Peron Yves, Caselli Graziella, Prioux France, Courgeau Daniel, Valkovics Emil, Hoem Jan, Henry Louis, Blayo Yves, Monnier Alain, Rychtarikova Jitka, Pinnelli Antonella, Mancini Paola, Legaré Jacques, Desjardins Bertrand, Festy Patrick, and Rallu Jean-Louis, 1991. Méthodes et applications de l'analyse démographique. Hommage à Roland Pressat, *Population*, vol.46(6), p. 1361–1760.

Campbell Arthur A., 1983. *Manual of fertility analysis.* Edinburgh, Churchill Livigstone, 100 p.

Carrol Glen R. and Hannan Michael T., 2004. *The demography of corporations and industries.* Princeton, Princeton University Press, 520 p.

Caswell Hall, 1989. *Matrix population models: construction*, *analysis*, *and interpretation.* Sunderland, MA, Sinauer Associates, 328 p.

Challier Marie-Christine and Michel Philippe, 1996. *Analyse dynamique des populations. Les approches démographiques et économiques.* Paris, Economica, 216 p.

Chesnais Jean-Claude, 1998. *La démographie*, 4th ed. Paris, Presses Universitaires de France, 127 p. (1st edition, 1990.)

Christiansen Freddy B. and Feldman Marcus W., 1986. *Population genetics.* Palo Alto, CA, Blackwell Scientific Publications, 196 p.

Chu Cyrus C.Y., 1998. *Population dynamics: a new economic approach.* New York, Oxford University Press, 226 p.

Corna-Pellegrini Giacomo, dell'Agnese Elena, and Bianchi Elisa, 1991. *Popolazione*, *società e territorio: manuale di geografia della popolazione.* Milan, Edizioni Unicopli, 282 p.

Courgeau Daniel, 1988. *Méthodes de mesure de la mobilité spatiale: migrations internes*, *mobilité temporaire*, *navettes.* Paris, INED, 301 p.

Courgeau Daniel and Lelièvre Éva, 1989. *Analyse démographique des biographies.* Paris, INED, 268 p. (Also published in English: *Event history analysis in demography*, Oxford, Clarendon Press, 1992, 226 p.)

Crook Nigel and Timaeus Ian M., 1997. *Principles of population and development: with illustrations from Asia and Africa.* Oxford, New York and Toronto, Oxford University Press, 226 p.

Daugherty Helen Ginn and Kammeyer Kenneth C.W., 1995. *An introduction to population*, 2nd ed. New York and London, Guilford Press, 343 p. (1st edition 1986, Chicago, The Dorsey Press.)

De Bartolo Giuseppe, 1997. *Elementi di analisi demografica e demografia applicata*, 2nd ed. Rende, Centro Editoriale e Libraio, 253 p. (1st edition, 1996.)

De Blij Harm J., 1982. *Human geography: culture*, *society and space*, 2nd ed. New York, John Wiley, 656 p.

De Rose Alessandra, 2001. *Introduzione alla demografia.* Rome, Carocci Editore, 125 p.

De Santis Gustavo, 1997. *Demografia ed economia.* Bologna, Il Mulino, 354 p. (Studi e Ricerche.)

De Sarno Prignano Anna and Natale Marcello, 1992. *I concetti di base della metodologia demografica.* Rome, La Goliardica, 304 p.

Del Panta Lorenzo and Rettaroli Rosella, 1994. *Introduzione alla demografia storica.* Bari and Rome, Editori Laterza, 314 p.

Di Comite Luigi and Chiassino Giuseppe, 1990. *Elementi di Demografia.* Bari, Cacucci Editore, 184 p.

Dittgen Alfred and Lamy-Festy Marlène, 1989. *Travaux pratiques d'analyse démographique.* Paris, Masson, 223 p.

Duchêne Josianne, 1992. *Mathématiques pour démographes: rappels théoriques-exercices résolus.* Louvain-la-Neuve, Institut de Démographie, Université catholique de Louvain and Academia, 224 p.

Dumont Gérard-François, 1992. *Démographie. Analyse des populations et démographie économique.* Paris, Dunod, 244 p.

Dupâquier Michel, 2001. *Démographie.* Paris, Presses universitaires de France, 241 p.

Findlay Allan and Findlay Ann, 1987. *Population and development in the third world.* New York and London, Methuen, 84 p.

Fleury Michel and Henry Louis, 1985. *Nouveau manuel de dépouillement et d'exploitation de l'état civil ancien*, 3rd ed. Paris, Editions de l'Institut national d'études démographiques, 202 p. (1st edition, 1956.)

Fouquet Annie and Vinokur Annie, 1996. *Démographie socioéconomique.* 2nd ed. Paris, Dalloz, 194 p. (1st edition, Fouquet Annie, 1990, 184 p.)

Franck Cadier Christiane, 1990. *Démographie. 1, Les phénomènes démographiques. 2, La population.* Paris, Economica, 168 p. (Exercices et cas, no. 8.)

Gani Léon, 2001. *Démographie expliquée: méthodes d'analyse etétudes de cas.* Paris, Nathan, 181 p.

Gendreau Francis, 1993. *La population de l'Afrique*, *manuel de démographie.* Paris, Karthala and CEPED, 463 p.

Gendreau Francis, Gubry Françoise, Lohlé-Tart Louis, van de Walle Étienne and Waltisperger Dominique (eds.), 1985. *Manuel d'estimations indirectes en démographie africaine: le manuel de Yaoundé.* Liège, Ordina, 276 p.

Golini Antonio, 1999. *La popolazione del pianeta.* Bologne, Il Mulino, 127 p.

Halli Shiva S. and Rao Vaninadha, 1992. *Advanced techniques of population analysis.* New York, Plenum, 226 p.

Hartley Shirley F., 1982. *Comparing populations: projects in demography.* Belmont, CA, Wadsworth Publishing, 272 p.

Haupt Arthur et Kane Thomas T., 1990. *Guide de démographie du Population Reference Bureau.* 2nd ed. Washington, D.C., Population Reference Bureau, 80 p.

Haupt Arthur and Kane Thomas T., 1997. *The Population Reference Bureau's population handbook*, 4th ed. Washington, Population Reference Bureau, 80 p. (1st edition, 1978.)

Heer David M. and Grigsby Jill S., 1992. *Society and population*, 3rd ed. Englewood Cliffs, NJ, Prentice Hall, 127 p. (1st edition, 1968.)

Henry Louis and Blum Alain, 1988. *Techniques d'analyse en démographie historique*, 2nd ed. Paris, éditions de l'Institut national d'études démographiques, 180 p. (1st ed., 1985.)

Henry Jean-Pierre and Gouyon Pierre-Henry, 1998. *Précis de génétique des populations: avec 42 exercices et problèmes corrigés*. Paris, Masson, 180 p. (Enseignement des sciences de la vie.)

Hillion Alain, 1986. *Les théories mathématiques des populations.* Paris, Presses Universitaires de France, 128 p. (Que sais-je?)

Hinde Andrew, 1998. *Demographic methods.* New York, Oxford University Press, 305 p.

Hornby William F. and Jones Melvyn, 1993. *An introduction to population geography*, 2nd ed. New York, Cambridge University Press, 168 p. (1st edition, 1980.)

Jacobsen Judith E., 1996. *Population growth.* Sausalito, CA, University Science Books, 43 p.

Jones Huw, 1981. *Population geography*, 2nd ed. London, Paul Chapman, 321 p. (1st edition, 1981.)

Jozwiak Janina, 1992. *Mathematical models of population.* The Hague, Netherlands Interdisciplinary Demographic Institute, 134 p.

Kammeyer Kenneth C.W. and Ginn Helen L., 1986. *An introduction to population.* Chicago, The Dorsey Press, 336 p.

Keyfitz Nathan, 1985. *Applied mathematical demography*, 2nd ed. New York and Berlin, Springer-Verlag, 442 p.

Keyfitz Nathan and Beckman John A., 1984. *Demography through problems.* New York and Berlin, Springer-Verlag, 141 p.

Land Kenneth C. and Rogers Andrei (eds.), 1982. *Multidimensional mathematical demography.* New York, Academic Press, 604 p.

Le Bras Hervé, 2005. *La démographie.* Paris, O. Jacob, 455 p.

Lecaillon Jean-Didier, 1992. *Démographie économique. Analyse des ressources humaines*, 2nd ed Paris, LITEC, Librairie de la Cour de Cassation, 289 p. (1st edition, 1990.)

Leridon Henri and Toulemon Laurent, 1997. *Démographie. Approche statistique et dynamique des populations.* Paris, Economica, 440 p. (Économie et statistiques avancées.)

Lévy Michel-Louis, 1997. *Déchiffrer la démographie*, 2nd ed. Paris, Syros, 207 p. (1st edition, 1990.)

Livi Bacci Massimo, 1999. *Introduzione alla demografia*, 3rd ed. Turin, Loescher Editore, 439 p. (1st edition, 1981.)

Lucas David and Kane Penny (eds.), 1985. *Asking demographic questions.* Canberra, Australian National University, National Centre for Development Studies, 220 p.

Lucas David and Meyer Paul (eds.), 1994. *Beginning population studies*, 2nd ed. Canberra, Australian National University, National Centre for Development Studies, 199 p. (1st edition, 1984.)

Menard Scott W. and Moen Elizabeth W. (eds.), 1987. *Perspectives on population. An introduction to concepts and issue.* Oxford, Oxford University Press, 487 p.

Micheli Giuseppe A. and Manfredi Piero (eds.), 1996. *Matematica delle popolazioni.* Milan, Angeli, 256 p.

Minguzzi Assunta, 1983. *Initiation à la démographie.* Bologna, Pitagora Editrice, 204 p.

Murdock Steve H. and Ellis David R., 1991. *Applied demography: an introduction to basic concepts*, *methods and data.* Boulder San Francisco and Oxford, Westview Press, 300 p.

Myers Dowell (ed.), 1990. *Housing demography: linking demographic structure and housing markets.* Madison, University of Wiscounsin Press, 317 p.

Nam Charles B., 1994. *Understanding population change.* Itasca (Illinois), F.E. Peacock Publishers, 471 p.

Nam Charles B. and Philliber Susan G., 1984. *Population. A basic orientation*, 2nd ed. Englewoods Cliffs, Prentice-Hall, 383 p.

Namboodiri Krishnan, 1991. *Demographic analysis: a stochastic approach.* San Diego, CA, Academic Press, 370 p.

Namboodiri Krishnan, 1991. *A primer of population dynamics.* New York and London, Plenum Press, 368 p.

Natale Marcello, 1983. *La ricerca demografica a fini operativi. Fonti*, *metodi*, *applicazioni.* Rome, La Goliardica Editrice universitaria, 2 vol., 252 + 222 p.

Natale Marcello, 2003. *Economia e popolazione*, revised ed. Milan, Angeli, 535 p. (1st edition, 1990.)

Newell Colin, 1988. *Methods and models in demography.* London, Belhaven Press, 217 p.

Newman James L. and Matzke Gordon E, 1984. *Population: patterns*, *dynamics and prospects.* Englewood Cliffs, Prentice-Hall, 306 p.

Noin Daniel, 1996. *Géographie de la population*, 4th ed. Paris, Armand Colin, 280 p. (1st edition, 1979.)

Overbeek Johannes, 1982. *Population: an introduction.* New York and London, Harcourt Brace Jovanovich, 278 p.

Pacione Michael, 1986. *Population geography: progress and prospect.* Dover and London, Croom Helm, 322 p.

Palmore James A. and Gardner Robert W., 1994. *Measuring mortality*, *fertility and natural increase: a self teaching guide to elementary* *measures*, 5th ed Honolulu, East West Center, 169 p. (1st edition, 1983.)

Peters Gary L. and Larkin Robert P., 2002. *Population geography: problems, concepts, and prospects*, 7th ed. Dubuque (Iowa), Kendall and Hut, 300 p. (1st edition, 1979.)

Petersen William, 2000. *From birth to death: a consumer's guide to population studies.* New Brunswick, Transaction Publishers, 197 p.

Petrioli Luciano, 1982. *Elementi di demografia.* Siena, Libreria editrice Ticci, 450 p.

Petrioli Luciano, 1998. *Demografia. Fatti e metodi di studio della popolazione.* Milan, Angeli, 600 p.

Plane David A. and Rogerson Peter A., 1994. *The geographical analysis of population with application to planning and business.* New York, Chichester, Brisbane, Toronto and Singapore, John Wiley & Sons, Inc., 417 p. (Printed in Singapore.)

Pol Louis Georges, 1987. *Business Demography.* New York, Quorum Books, 289 p.

Pol Louis Georges and Thomas Richard K., 1997. *Demography for business decision making.* Westport, (Connecticut), Quorum Books, 270 p.

Pollard Alfred H., Yusuf Farhat, and Pollard G. N., 1990. *Demographic techniques*, 3rd ed. Elmsford, NY, Pergamon Press, 186 p. (1st edition, 1974.)

Poulalion Gabriel, 1984. *La science de la population.* Paris, LITEC, 333 p.

Pressat Roland, 1983. *L'analyse démographique: concepts, méthodes, résultats*, 4th ed. Paris, Presses Universitaires de France, 295 p. (1st edition, 1969.)

Pressat Roland (ed.), 1985. *Manuel d'analyse de la mortalité.* Paris, Geneva, INED, WHO, 164 p.

Pressat Roland, 1986. *Les méthodes en démographie.* 2nd ed. Paris, Presses universitaires de France, 127 p. (1st edition, 1981.)

Pressat Roland, 1995. *Éléments de démographie mathématique.* Paris, AIDELF, 280 p.

Preston Samuel H., Heuveline Patrick, and Guillot Michel, 2000. *Demography: measuring and modelling population processes.* Oxford, Blackwell, 290 p.

Pupion Georges and Poulalion Gabriel, 1993. *Analyse démographique: les concepts et les indicateurs*, Paris, L'Hermès, 116 p.

Razin Assaf and Sadka Efraim, 1995. *Population economics.* Cambridge, MIT Press, 275 p.

Reher David and Schofield Roger (eds.), 1993. *Old and new methods in historical demography.* Oxford, Oxford University Press, 426 p. (International Studies in Demography.)

Rives Norfleet W. and Serow William J., 1984. *Introduction to applied demography: data sources and estimation technics.* Beverly Hills, CA, Sage Publications, 96 p.

Rogers Andrei, 1995. *Multiregional demography: Principles, methods and extensions.* New York, Chichester, 236 p.

Rollet Catherine, 1995. *Introduction à la démographie.* Paris, Nathan, 128 p.

Rosenzweig Mark and Stark Oded, 1997. *Handbook of population and family economics.* Amsterdam, Elsevier, 2 vol., 644 + 700 p.

Rowland Donald T., 2003. *Demographic methods and concepts.* Oxford, Oxford University Press, 546 p.

Santini Antonio, 1992. *Analisi demografica: fondamenti e metodi.* Scandicci (Florence), La Nuova Italia, 2 vol., 422 + 176 p.

Saunders John, 1988. *Basic demographic measures: a practical guide for users.* Lanham, University Press of America, 100 p.

Schnell George A. and Monmonier Mark S., 1983. *The study of population: elements, patterns, processes.* Columbus, OH, Charles E. Merril, 371 p.

Schultz T. Paul (ed.), 1998. *Economic Demography.* Northampton (Maryland), Edward Elgar, 2 vol., 562 + 621 p.

Siegel Jacob S., 2002. *Applied demography. Application to business, law, and public policy.* New York, Academic Press, 686 p.

Siegel Jacob S. and Swanson David A., 2004. *The methods and materials of demography*, 2nd ed. San Diego, Academic Press, 819 p.

Smith David P., 1992. *Formal demography.* New York, Plenum Press, 330 p.

Spiegelman Mortimer, 1980. *Introduction to demography.* 2nd ed. London, Harvard University Press, 514 p. (1st edition, 1955, Chicago, The Society of Actuaries.)

Srinivasan Krishnamurthy, 1998. *Basic demographic techniques and applications.* Thousand Oaks, CA, Sage Publications, 224 p.

Stockwell Edward G. and Groat H. Theodore, 1984. *World population: an introduction to demography.* London, Franklin Watts, 524 p.

Tapinos Georges, 1996. *Éléments de démographie: analyse, déterminants socio-économiques et histoire des populations*, 3rd ed. Paris, Armand Colin, 367 p. (1st edition, 1985.)

Tapinos Georges, 1994. *Elementi di demografia.* Milan, EGEA, 525 p.

Tapinos Georges P., 1996. *La démographie. Population, économie et sociétés.* Paris, Fallois Le Livre de Poche, 255 p. (Inédits.)

Tapinos Georges, 2000. *La demografia.* Milan, Franco Angeli, 220 p.

Vallin Jacques, 1986. *La population mondiale.* Paris, La Découverte, 128 p. (Repère, no. 45.)

Vallin Jacques, 1989. *La population française.* Paris, La Découverte, 127 p. (Repère, no. 75.)

Vallin Jacques, 2002. *La démographie*, 3rd ed. Paris, La Découverte, 128 p. (Repères, no. 105, 1st edition, 1992.)

Vallin Jacques, D'Souza Stan, and Palloni Alberto (eds.), 1988. *Mesure et analyse de la mortalité: nouvelles approches.* Paris, INED and PUF, 460 p. (Travaux et documents, no. 119.)

Vallin Jacques, D'Souza Stan, and Palloni Alberto (eds.), 1990. *Measurement and analysis of mortality: new approaches.* Oxford, Clarendon, 415 p. (International Studies in Demography.)

Vallin Jacques, Pollard John, and Heligman Larry (eds.), 1984. *Methodologies for collection and analysis of mortality data.* Liège, IUSSP, Ordina Éditions, 530 p.

Vandeschrick Cristophe, 1995. *Analyse démographique.* Louvain-la-Neuve, Academia-Bruylant and L'Harmattan, 183 p.

van Imhoff Evert, Kuijsten Anton, Hooimeijer Pieter, and van Wissen Leo (eds.), 1995. *Household demography and household modeling.* New York, Plenum Press, 369 p. (Plenum Series on Demographic Methods and Population Analysis.)

Véron Jacques, 1991. *Démographie.* Paris, Armand Colin, 128 p. (Flash U.)

Véron Jacques, 1996. *Population et développement.* Paris, Presses Universitaires de France, 128 p. (Que sais-je?)

Vidal Annie, 1994. *Démographie: eléments d'analyse et évolution du peuplement humain.* Grenoble, Presses Universitaires de Grenoble, 154 p.

Vidal Annie, 2001. *Démographie. Les outils: exercices corrigés.* Grenoble, Presses universitaires de Grenoble, 140 p. (L'économie en plus.)

Weeks John R, 1999. *Population: an introduction to concepts and issues*, 7th ed. Belmont, CA, Wadsworth Publishing, 673 p. (1st edition, 1978.)

Weiss-Altaner Eric, 1992. *Principes de démographie politique: population, urbanization et developpement.* Geneva, Economica, 247 p.

Weinstein Jay and Pillai Vijayan K., 2000. *Demography. The science of population.* Needham Heights (Massachussetts), Allyn and Bacon, 450 p.

WELLER Robert H. and BOUVIER Léon F., 1981. *Population: demography and policy.* New York, St. Martin's Press, 365 p.
WITHERICK Michael E., 1990. *Population geography.* London and New York, Longman, 177 p.
WUNSCH Guillaume, 1984. *Techniques d'analyse des données démographiques déficientes.* Liège, Université de Louvain, Département de Démographie and Ordina Éditions, 221 p.
WUNSCH Guillaume, 1987. *Méthodes d'analyse démographique.* Louvain-la-Neuve, CIDEP, 49 p. (Document Pédagogique CIDEP.)
YAUKEY David, 1985. *Demography: the study of human population.* New York, St. Martin's Press, 352 p.
ZOPF Paul E., 1984. *Population: an introduction to social demography.* Palo Alto, CA, Mayfield Publishing, 501 p.

APPENDIX

139–2

Teaching Demography in the Russian Language

ALEXANDRE AVDEEV AND MARIA AVDEEVA

Université de Moscou, Moscow, Russia, and Université Marc Bloch, Strasbourg, France

Departement de bibliographie et de l'information, Centre d'etudes demographiques, Faculte d'economie, Université de Moscou, Moscow, Russia

The Russian school of demography has its origins in the conjunction of a strong mathematical tradition with the first sociologic population studies. Very early on, Russian demography adopted a mathematical direction in the wake of the work of mathematicians and statisticians from Russia, or who had studied in Russia, such as Daniel Bernoulli (1700–1782), Leonhardt Euler (1707–1783), Viktor Bouniakovski[15] (1804–1889), Vladislav Bortkewicz[16] (1868–1930), Alexandre Tchouprov[17] (1874–1926), and Aaron Boiarski[18] (1906–1985). The sociologic direction appeared more recently, from the mid-20th century, influenced by the ideas of Adolphe Quetelet, Achille Guillard, and Jacques Bertillon. One of the promoters of these ideas, a professor at the University of Saint Petersburg, Iuliy Ianson[19] (1835–1893), published a book entitled *Compared Population Statistics* in 1895, which can be considered a demographic textbook.[20] According to Iuliy Ianson, the word "demography" was synonymous with "compared population statistics." Within this discipline, it included:

"Population statistics," or "anatomy of society," based on numbers, spatial distribution and population age and sex structure, as well as family composition

Population dynamics, or "physiology of the social organism," including nuptiality, fertility, and mortality

Several Russian demographers and statisticians were trained by Ianson, including the brilliant Vladislav Bortkewicz, who went on to teach at Berlin University.

At the end of the 19th century and at the beginning of the 20th, as in other countries, population statistics and health statistics were closely linked in Russia. Most demographic research at that time was the work of physicians and many statisticians had a medical training. As a result, the first post-Revolutionary Russian demographic textbook was a translation of George Whipple's manual,[21] *Vital Statistics: An Introduction to the Science of Demography*, published twice in the United States, in 1919 and 1923. However, this basic manual had been considerably improved and completely revised by the Russian demographer Serguey Novoselski[22] and his publication in Russian in 1929 was jointly signed by Serguey Novoselski and

[15] Виктор Яковлевич Буняковский.

[16] Владислав Иосифович Борткевич.

[17] Александр Александрович Чупров.

[18] Аарон Яковлевич Боярский.

[19] Юлий Эдуардович Янсон.

[20] Юлий Эдуардович Янсон. This author can be considered the founder of statistical teaching in Russia. In 1885, he published the first statistics manual, which was reprinted four times.

[21] George Chandler Whipple (1866–1924), American demographer, is known for his method of smoothing the population pyramid, which eliminates the attraction effects of round ages (see Chapter 4).

[22] Сергей Александрович Новосельский (1872–1953), Russian demographer, co-author (with V. Paevski) of the first national life tables calculated from census data. From 1930 to 1934, he directed the Institute of Demography of the USSR's Academy of Sciences. This institute was closed in 1934.

George Whipple under the title *Bases of Demography and Health Statistics*.[23] Despite its origins, this publication can be considered as the first Soviet textbook. It was aimed at medical students and, according to the administrative practices of the period (which lasted until the end of the Soviet era), it was recommended as essential for health-related teaching.

Two years later, a new textbook entitled Elements of Statistics, a Textbook for Agents of the Maternity Protection System, Infanthood, and Other Branches of Social Hygiene,[24] was published for those working in the health sector by the assistant director of the USSR's Academy of Science Institute of Demography, Vladimir Paevski.[25] Although the term *demography* does not appear in the title, this textbook included a description of all the basic methods of demographic analysis. These two textbooks were used as a basis for teaching demography in universities until the mid-1940s.

In 1945, a new demography textbook, *Lessons in statistical demography*,[26] by the brilliant demographer and statistician Aaron Boiarski[27] was published. This textbook was aimed at statisticians. It was mainly devoted to the methods of formal demographic analysis and the application of mathematics to population studies. Although it was an excellent textbook that included all the demographic contributions of the period, it did not remain on the agenda for a long time. A short time after its publication, ideologic persecutions began in the USSR that were aimed at suppressing the use of mathematical and statistical methods, not only in the social sciences, but also in the hard sciences, such as biology. The law of large numbers was a particular target for criticism; the Stalin administration considered it as a threat to its social development theory. According to Stalinist doctrine, statistics cannot and should not interfere with the law of social development as Adolphe Quetelet stated, as all the principles are already recognized and defined by the Marxist-Leninist and Stalinist theory of social development. Statistics should merely provide an illustration of the success of the Socialist system and the crisis in Capitalist society. Within this atmosphere of reinforced ideologic pressure, Aaron Boiarski was obliged to rewrite his textbook to a considerable extent and to the detriment of the mathematical aspect, and was also obliged to add ideologic banalities on the successes of Socialism. The new version was published in 1951 in the form of a textbook entitled *Demography Statistics*,[28] which was jointly signed by Aaron Boiarski and Pietr Shousherin.[29] In the preface, the authors apologized for their exaggeration of the role of mathematical analysis in demography. The book was republished in 1955. Despite its shortcomings, it was used as the basis for training Russian statisticians and demographers until the end of the 1960s.

The first half of the 1960s saw somewhat of a rebirth in Russian Soviet demography (Avdeev, 1997). A statistical research institute was created by the State Statistics Committee with Aaron Boiarski as its chief. A laboratory and a Chair of Demography Studies were created within the Faculty of Economics of Moscow State University and placed under the direction of Dimitri Valentei.[30] Today, they are known as the Center for the Study of Population Problems of the University of Moscow, where students acquire a specialization in demography over the course of five years of advanced education and three years of doctoral studies. These new needs required new textbooks. One of the first was the Russian translation of Roland Pressat's *Demographic Analysis, Methods, Results, and Applications*, published in Russian in 1966 under the title *Population and the Study of Population*.[31] The translation was carried out by V. E. Shprink under the supervision of Boris Ourlanis,[32] who himself was fluent in French. However, during the following year, in 1967, a 100% Russian textbook was published, which resulted from a considerable collective effort, carried out under the supervision of Aaron Boiarski: *Demography Lessons*,[33] republished in 1974 and 1985. This publication discussed all the aspects of demographic analysis (e.g., fertility, mortality, migration), the stable and stationary population model in matrix and continuous form, methods of building and applying demographic tables, and population forecasts. There were also ele-

[23] С.А.Новосельский и Дж.Уиппль: Основы демографической и медицинской статистики.

[24] Элементы статистики. Пособие для работников охраны материнства и младенчества и других отраслей социальной гигиены.

[25] Владимир Владиславович Паевский (1893–1934), Russian demographer, co-author (with Serguey Novoselski) of the first national life tables calculated from census data. From 1930 to 1934, he was assistant director of the Institute of Demography of the USSR's Academy of Sciences. He died of a heart attack on the same day as the Institute closed.

[26] Курс демографической статистики.

[27] НАарон Яковлевич Боярский (1906–1985), Russian and Soviet statistician and demographer, holder of the Chair of Statistics at Moscow State University, and founder and first director of the State Statistical Committee's Research Institute.

[28] Демографическая статистика.

[29] Петр Павлович Шушерин (1895–1958).

[30] Дмитрий Игнатьевич Валентей (1922–1994).

[31] Народонаселение и его изучение.

[32] Борис Цезаревич Урланис (1906–1981).

[33] Курс демографии.

ments of world population history, the history of demography, a description of the demographic situation in the USSR, the history of demographic doctrine and population policies, as well as the birth of the encyclopedic tendency that has since prevailed in the works used in Russia for teaching demography. At the time, this textbook was possibly one of the best in the world. However, its use required that students had a solid knowledge of mathematical analysis. A high level of mathematical teaching was provided in the Faculties of Economics and Geography of Moscow State University and the Institute of Statistics and Economics, but for students from social (e.g., history, philosophy) and administrative (e.g., medicine, protection of health) backgrounds, this textbook was difficult. The textbook by Aaron Boiarski, Dimitri Valentey, and Alexandre Kvasha,[34] *Bases of Demography*,[35] published in 1980, was a compromise; all the aspects of demographic analysis and population theory were also included, but the mathematical aspect was reduced and the Lexis diagram for the classification of demographic events was introduced in the version proposed by Roland Pressat. In 1971 and 1977, two textbooks by Ilya Venetzki[36] were published, in which the mathematical (*Mathematical Methods in Demography*,[37] 1971) and statistical aspects (*Statistical Methods in Demography*,[38] 1977) of demographic analysis were developed in detail.

The necessity of teaching students the qualitative aspects of demography, demographic doctrines, and population policies led to the publication in 1973 of a specific textbook, written under the supervision of Dmitri Valentey, *Foundations of the Theory of Population*,[39] which was republished in 1977 and 1986. In its structure and ideology, this textbook had many similarities with Alfred Sauvy's *General Theory of Population*, which was itself translated into Russian in 1977 and used as a textbook by both students and teachers.

At the end of the 1970s, a work of colossal proportions was undertaken, involving around one hundred authors, and completed in 1985 with the publication of the *Encyclopedic Demographic Dictionary*,[40] supervised by Dmitri Valentey and Dmitri Shelestov.[41] This work is made up of approximately 2000 articles. It describes demographic terms and methods of analysis in detail (including calculation formulas) and includes biographic articles, information on all national and international organizations dealing with population issues, and information on all the specialized journals in the world. This dictionary is without a doubt the best textbook in the Russian language. It was republished in 1994 under the title *Encyclopedic Dictionary of Demography*.[42]

Since the end of the 1990s, many demographic textbooks have been published, most have been written by professors for their own courses and are aimed at specific categories of students. In contrast to earlier textbooks, which were of a general nature and were used in general preparatory courses for demographers, the textbooks of this new wave, although they provide a general introduction to demography, demographic indicators, and some analysis methods, are far from being exhaustive, although this does not diminish their educational qualities. (Zvereva *et al*., 2004; Borisov, 2001, 2003; Medkov, 2003;).

However, *Introduction to Demography*,[43] under the supervision of Vladimir Iontsev and Alexandre Sagradov (1st edition, 2002; 2nd edition, 2003) is an example of a successful old-type encyclopedic general textbook; it was written by a group of authors teaching at the Faculty of Economics at Moscow State University (MGU). The presence in the title of the word *introduction* must not be misunderstood; it is a voluminous, essential textbook. It includes all the contemporary knowledge of demography (e.g., analyses, models, forecasts); it describes the current problems of fertility, mortality, and migration, as well as the role of demographic variables in development models and biometric fertility models. It also includes a small Russian-English-French dictionary of demographic terms and a bibliography at the end of each chapter.

References

Avdeev Alexandre, 1997. Avenir de la démographie en Russie, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI^e^ siècle*, p. 369–396. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Whipple George C., 1919. *Vital statistics, an introduction to the science of demography*, 1st ed. New York, John Wiley & Sons, 517 p.

Whipple George C., 1923. *Vital statistics, an introduction to the science of demography*. New York, John Wiley & Sons, 579 p. (2nd ed.)

[34] Александр Яковлевич Кваша.

[35] Основы демографии.

[36] Илья Григорьевич Венецкий (1914–1981).

[37] Математические методы в демографии.

[38] Статистические методы в демографии.

[39] Основы демографии.

[40] Демографический Энциклопедический Словарь.

[41] Дмитрий Кузьмич Шелестов (1927–2000).

[42] Энциклопедический демографический словарь.

[43] Введение в демографию.

Demographic Treatises and Textbooks in Russian

SELECTED BY ALEXANDRE AVDEEV,[44] MARIA AVDEEVA,[45] AND MAÏTÉ ELY[46]

Basic Textbooks (Chronologic Order)

Ianson Iuliy E. [Янсон Юлий Эдуардович], 1892. *Сравнительная статистика населения ((Состав населения,, брачность,, рождаемость,, смертност, механическое движение населения)) [Compared Population Statistics (Composition of Population, Nuptiality, Fertility, Mortality, Mechanical movement of the Population)].* Saint-Petersburg, Типография Дома призрения малолетних бедных (Typography of the Poor Minor Charity House), 443+3 p.

Whipple George Ch. [Уиппль Джорж] and Novosselski Serguei A. [Новосельский Сергей. . Александрович], 1929. Основы демографической и санитарной статистики *[Bases of Demography and Health Statistics].* Moscow, Гос. мед. изд-во. [State Medical Publications], 683 p. (Supplemented and revised translation of the publication by Wipple, 1926.)

Paievski Vladimir. V. [Паевский Владимир Владиславович], 1931. *Элементы статистики. Пособие для работников охраны материнства и младенчества и других отраслей социальной гигиены [Elements of statistics, a textbook for agents of the maternity protection system, infanthood, and other branches of social hygiene].* Moscow and Leningrad, ОгизГос.мед.изд., тип.«Коминтерн» в Лнг. [OGIZGos.med.izdat], 304 p.

Boiarski Aaron Ia. [Боярский Аарон Яковлевич], 1945. *Курс демографической статистики [Lessons in Statistical Demography].* Moscow, ГосПолитИздат [State Political Publications], 263 p.

Boiarski Aaron Ia. [Боярский Аарон Яковлевич] and Shousherin Petre P. [Шушерин Петр Павлович], 1951. *Демографическая статистика [Demographic Statistics].* Moscow, Государственное статистическое издательство [State Statistical Publications], 343 p. (2nd ed, Moscow, 1955, 332 p.)

Boiarski Aaron Ia. [Боярский Аарон Яковлевич] (dir.), 1967. *Курс демографии [Demography lessons].* Moscow, Статистика [Statistika], 399 p. (Допущено в качестве учеб. Пособия для студентов экономических специальностей вузов [Textbook for students specializing in economics in higher education establisment].)

Boiarski Aaron Ia. [Боярский Аарон Яковлевич] (dir.), 1974. *Курс демографии [Demography lessons].* Moscow, Статистика [Statistika], 453 p. (2nd ed, Moscow.)

Boiarski Aaron Ia. [Боярский Аарон Яковлевич], Valentey Dimitri I. [Валентей Дмитрий Игнатьевич] and Kvsha Alexandre Ia. [Кваша Александр Яковлеви] (dir.), 1980. *Основы демографии.. Учебное пособие для экономических вузов [Bases of demography: Textbook for Teaching Economics].* Moscow, Статистика [Statistique], 295 p.

Boiarski Aaron Ia. [Боярский Аарон Яковлевич] (dir.), 1985. *Курс демографии [Demography lessons]*, 3rd ed revised and supplemented. Moscou, Финансы и Статистика [Finance and Statistics], 389 p.

Boiarski Aaron Ia. [Боярский Аарон Яковлевич], Bakhmetova Ghuldjan Sh. [Бахметова Гюльджан Шариповна.] and Khartchenko Lia P. [Харченко Лия Петровна, 1985. *Практикум по демографии [Practical Demography].* Moscow, Мысль [Mysl], 141 p.

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич] and Kvasha Alexandre Ia. [Кваша Александр Яковлевич] (dir.), 1989. *Основы демографии [Bases of demography].* Moscow, Мысль [Mysl], 286 p.

Kvasha Alexandre Ia. [Кваша Александр Яковлевич] and Iontcev Vladimir A. [Ионцев Владимир Алексеевич] (dir.), 1995. *Современная демография [Contemporary Demography].* Moscow, Изд-во Московского университета [Moscow State University edition], 272 p.

Iontcev Vladimir A. [Ионцев Владимир Алексеевич] and Sagradov Alexandre A. [Саградов Александр Альбертович] (dir.), 2002. *Введение в демографию [Introduction to Demography].* Moscow, Экономический факультет МГУ, ТЕИС [Faculty of Economics, Moscow State University, TEIS], 636 p.

Iontcev Vladimir A. [Ионцев Владимир Алексеевич] and Sagradov Alexandre A. [Саградов Александр Альбертович] (dir.), 2003. *Введение в демографию [Introduction to demography]*, 2nd ed. Moscou, Экономический факультет МГУ, ТЕИС [ТЕИС [Faculty of Economics, Moscow State University, TEIS], 636 p.

Mathematical and Statistical Application Textbooks

Venetskiy Iliya G. [Венецкий Илья Григорьевич], 1971. *Математические методы в демографии [Mathematical Methods in Demography].* Moscow, Статистика [Statistics], 295 p.

Venetskiy Iliya G. [Венецкий Илья Григорьевич], 1977. *Статистические методы в демографии [Statistical Methods in Demography].* Moscow, Статистика [Statistics], 208 p.

Textbooks Based on Population Doctrine, Theory, and Policies

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич], 1961. *Проблемы народонаселения [Population Problems].* Moscow, Высшая школа [Advanced college], 160 p.

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич], 1967. *Теория и политика народонаселения [Population theories and policies].* Moscou, Высшая школа [Advanced college], 184 p.

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич] (dir.), 1976. *Система знаний о народонаселении [Population Knowledge System].* Moscow, Статистика [Statistics], 367 p.

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич] (dir.), 1973. *Основы теории народонаселения [Bases of population theory]*, 1st ed. Moscow, Высшая школа [Vys'shaia Shkola], 367 p. (Supplemented and revised 2nd edition, 1977, 269 p.; 3rd edition, 1986, 376 p.)

Encyclopedias

Valentey Dimitri I. [Валентей Дмитрий Игнатьевич], Shelestov Dimitri K. [Шелестов Дмитрий Кузьмич.], Volkov Andrei G. [Волкова Андрей Гаврилович.] and Khorev Boris S. [Хорева Борис Сергеевич] et al. (dir.), 1985. *Демографический Энциклопедический Словарь [Encyclopedic Demographic Dictionary].* Moscow, Советская энциклопедия [Soviet Encyclopaedia], 608 p.

Mélikian Guennadi G. [Меликьян Геннадий Георгиевич], Kvasha Alexandre Ia. [Кваша Александр Яковлевич.], Tkachenko Alexandre A. [Ткаченко Александр Александрович], Shapovalova Natalia N. [Шаповалова Наталия Николаевна] and Shelestov Dimitri K. [Шелестов Дмитрий Кузьмич], 1994. *Народонаселение.. Энциклопедический словарь [Population. Encyclopaedi Dictionary].* Moscow, Большая Российская Энциклопедия [Comprehensive Encyclopaedia of Russia], 640 p.

[44] Strasbourg University and Moscow University.
[45] Moscow University.
[46] Institut national d'études démographiques, Paris.

Textbooks for Demography as an Additional Discipline

Medkov Viktor M. [Медков Виктор Михайлович], 2002. *Демография для студентов вузов [Demography for students of higher education]*. Rostov-on-Don, Феникс [Phoenix], 448 p. (Учебное пособие Серия «Учебники и учебные пособия» ["Textbooks and Practical Guides" series].)

Medkov Viktor M. [Медков Виктор Михайлович], 2003. *Демография [Demography]*. Moscow, ИНФРА-М [INFRA-M], 544 p. (Учебник. Серия «Высшее образование» ["Higher Education" series].)

Borisov Vladimir A. [Борисов Владимир Александрович], 2001. *Демография. Учебник для вузов [Demography. Textbooks for higher education establishments]*, corrected 2nd. Moscow, Nota Bene, 272 p. (Updated 3rd edition, 2003, 344 p.)

Zvereva Natalia V. [Зверева Наталья Викторовна], Elizarov Valeri V. [Елизаров Валерий Владимирович] and Veselkova Irina N. [Веселкова Ирина Николаевна.], 1997. *Основы демографии [The Bases of Demography]*. Moscow, Высшая школа [Vys'shaia Shkola], 374 p.

Translations

Pressat Roland [Пресса РРоллан], 1966. *Народонаселение и его изучение [Population and study methods]*. Moscow, Статистика [Statistika], 443 p. (Перевод с французского В. Э. Шпринка под редакцией Б. Ц. Урланиса [Translated from French by V. E. Shprink, under the supervision of B. Cz. Urlanis].)

Pressat Roland, 1961. *L'analyse démographique: méthodes, résultats, applications*. Paris, PUF, 402 p.

Sauvy Alfred [Сови Альфред], 1977. *Общая теория населения [Théorie générale de la population]*. Moscow, Прогресс [Editions du Progrès], 2 vol.: 501 + 518 p. (Сокращенный перевод с фр. Ф. Р. Окуневой, предисловие А. Я. Боярского [Synopsis, translated by F. R. Okouneva, preface by A. Ia. Boiarski].)

Sauvy Alfred, 1963. *Théorie générale de la population, vol. I: Economie et croissance*, 3rd ed. Paris, Presses universitaires de France, 373 p. (1st edition, 1952.)

Sauvy Alfred, 1966. *Théorie générale de la population. Volume II. La vie des populations*, 3rd ed. Paris, PUF, 403 p. (1st edition, 1954, Biologie Sociale.)

APPENDIX

139–3

Demographic Treatises and Textbooks in German

ALEXIA FÜRNKRANZ-PRSKAWETZ
Vienna Institute of Demography, Vienna, Austria

In the wake of the *Divine Order*, by the great Johann Peter Süssmilch (1775), demography and the production of treatises and textbooks relating to the field experienced a significant upturn in German-language countries during the 19th century, notably with the works of Christof Bernoulli (1841, 1843), Georg Knapp (1874), Wilhelm Lexis (1875, 1903), Arthur von Fircks, and Kuno Frankenstein and Max von Heckel (1898), and again at the beginning of the 20th century with the works of Paul Mombert (1929), and Alfred Lotka (1932). However, the tendencies of the Nazi period discredited the discipline and it is only very recently that it has regained scientific recognition in German-language universities.

It was not until the 1970s that new research centers began to emerge,[47] again stimulating the teaching of demography in universities and the production of textbooks in German. Neighboring disciplines such as statistics, economics, sociology, geography, anthropology, biology, and medicine had nurtured the flame by including a little demography in their *curricula* and demographers working within these various disciplines were able to establish a certain corpus of basic demographic concepts, methods, models, and techniques for the analysis of demographic processes and structures.

[47] Notably the creation of the Bundesinstitut für Bevölkerungsforschung (BIB) in 1973 in Wiesbaden, the Max Planck Institute for Demographic Research/Max-Planck-Institut für demografische Forschung in 1995 in Rostock, and the Vienna Institute of Demography/Institute für Demographie in Vienna in 1978.

However, because of the absence of specific demographic teaching, demographic textbooks in German remained few up to the 1970s. Among the most recent textbooks, many are very clearly similar to statistical textbooks, even though they are less mathematical (e.g., Feichtinger, 1973; Esenwein-Rothe, 1982; Rinne, 1996). Of these, three works of reference dealing formally with demography, its concepts, and developments should be singled out—those of Hans Reiner Dinkel (1989), Gustav Feichtinger (1979), and Flaskämper (1962). Among allied disciplines, in German-language countries as is true elsewhere, sociology has the closest links with demography. Other important contributions include the work of Gerhard Mackenroth (1953), published at the beginning of the 1950s, and the population sociology textbooks of Josef Schmid (1976), Jürg Hauser (1982), and François Höpflinger (1997). The range of demographic methods extends to many other fields. Some are covered by German literature such as, for example, population geography (Bähr, Jentsch and Kuls, 1992), historical demography (Ehmer, 2004), and the demography of health (Mueller, 1993). The recent demography textbook by Ulrich Mueller, Bernhard Nauck, and Andreas Diekman (2000) is far more comprehensive and provides an extensive and thorough overview of various applications of demography. The first part of the handbook is devoted to methods and models of demography, and the second part provides an extensive and comprehensive overview of the different applications of demography.

Low fertility rates and continued increases in life expectancy, as well as their implications for population aging, have brought demography back into the forefront of political, economic, and scientific debate in German-language countries. It is hoped that this increasing interest in demography will be accompanied by the development of the discipline. New generations of German-speaking demographers have already appeared within prestigious centers of demographic research and increase confidence for such a development. However, the teaching of demography in German-language countries relies increasingly on English-language textbooks. Within the context of this trend, the need for demographic textbooks in Germany must be discussed.

Acknowledgment

I am grateful to Richard Gisser, Wolfgang Lutz, and Jacques Vallin for their valuable contribution.

DEMOGRAPHIC TREATISES AND TEXTBOOKS IN GERMAN[48]

SELECTED BY ALEXIA FÜRNKRANZ-PRSKAWETZ, HENRIETTE ENGELHARDT,[49] GUSTAV FEICHTINGER,[50] RICHARD GISSER,[49] WOLFGANG GRIESER,[51] AND MARTINE DEVILLE-VELLOZ[52]

Works Published before 1950

Bernoulli Christoph, 1841. *Handbuch der Populationistik über der Völker und menschenkunde nach statistischen Ergebnisse*. Ulm, 612 p.

Bernoulli Christoph, 1843. *Neuere Ergebnisse der Bevölkerungsstatistik. Zugleich als Nachtrag zum Handbuch der Populationistik*. Ulm, Verlag der Stettin'schen Buchhandlung, 80 p.

Fircks (von) Arthur, Frankenstein Kuno, and von Heckel Max, 1898. *Bevölkerungslehre und Bevölkerungspolitik*. Leipzig, Hirschfeld, 492 p.

Knapp Georg F., 1874. *Theorie des Bevölkerungswechsels: Abhandlungen zur angewandten Mathematik*. Braunschweig, Friedrich Vieweg und Sohn, 139 p.

Lexis Wilhelm, 1875. *Einleitung in die Theorie der Bevölkerungsstatistik*. Strasburg, Trübner, 137 p.

Lexis Wilhelm, 1903. *Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik*. Jena, Fischer, 253 p.

Lotka Alfred J., 1932.*Zur Dynamik der Bevölkerungsentwicklung* Allgemeines Statistisches Archiv, 23(1932). S587–588 plus Ergänzung S98.

Mombert Paul, 1929. *Bevölkerungslehre*. Jena, Fischer, 490 p.

Süssmilch Johann Peter, 1775–1776. *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode und der Fortpflanzung desselben*. Berlin, Verlag der Buchhandlung der Realschule, vol. 1 (1775): 130 p.; vol. 2 (1775): 79 p.; vol. 3 (1776): 735 + 68 p.

Works Published since 1950

Bähr Jürgen, Jentsch Christoph, and Kuls Wolfgang, 1992. *Bevölkerungsgeographie*. Berlin, Walter de Gruyter, 1158 p. (Lehrbuch der Allgemeinen geographie, published by Erich Obst, vol.9.)

Bolte Karl Martin, Dieter Kappe, and Schmid Josef, 1980. *Bevölkerung: Statistik, Theorie, Geschichte und Politik des Bevölkerungsprozesses*. Opladen, Leske und Budrich, 212 p.

Cromm Jürgen, 1988. *Bevölkerung—Individuum—Gesellschaft: Theorien und soziale Dimensionen der Fortpflanzung*. Opladen, Westdeutscher Verlag, 247 p. (WV Studium, no. 152.)

Dinkel Hans Reiner, 1989. *Demographie*. *Band I, Bevölkerungsdynamik*. Munich, Vahlen, Vahlens Handbücher der Wirtschaftswissenschaften, 455 p.

Ehmer Josef, 2004. *Bevölkerungsgeschichte und historische Demographie 1800–2000*. Munich, Oldenbourg, Enzyklopädie deutscher Geschichte, 71 p.

Esenwein-Rothe Ingeborg, 1982. *Einführung in die Demographie: Bevölkerungsstruktur und Bevölkerungsprozess aus der Sicht der Statistik*. Wiesbaden, Steiner, 400 p.

Feichtinger Gustav, 1977. *Stationäre und schrumpfende Bevölkerungen. Demographisches Null- und Negativwachstum in Österreich*. Berlin, Springer, 262 p. (Lecture Notes in Economics and Mathematical Systems, vol.149.)

Feichtinger Gustav, 1973. *Bevölkerungsstatistik*. Berlin and New York, Walter de Gruyter.

Feichtinger Gustav, 1979. *Demographische Analyse und populationsdynamische Modelle: Grundzüge der Bevölkerungsmathematik*. Vienna and New York, Springer, 367 p.

Flaskämper Paul, 1962. *Bevölkerungsstatistik*. Hamburg, Verlag Richard Meiner, 496 p.

Haupt Arthur and Kane Thomas T., 1999. *Handbuch Weltbevölkerung: Begriffe, Fakten, Konzepte*. Stuttgart, Balance-Verl, 112 p.

Hauser Jürg, 1982. *Bevölkerungslehre*. Stuttgart, Haupt, 359 p.

Höpflinger François, 1997. *Bevölkerungssoziologie. Eine Einführung in bevölkerungssoziologische Ansätze und demographische Prozesse*. Weinheim and Munich, Juventa Verlag, 232 p. (Grundlagentexte Soziologie.)

Imhof Arthur E., 1977. *Einführung in die historische Demographie*. Munich, Beck, 149 p. (Beck'sche Elementarbücher.)

Khalatbari Parviz, 1989. *Handbuch der Demographie in der DDR*. Berlin, Akad. D. Wiss d. DDR, 185 p.

Khalatbari Parviz, 1981. *Bevölkerungstheorie und Bevölkerungspolitik*. Berlin, Akademie-Verlag, 274 p.

Kuls Wolfgang and Kemper Franz-Josef, 2000. *Bevölkerungsgeographie: eine Einführung*. Stuttgart and Leipzig, Teubner. (Teubner Studienbücher der Geographie.)

Mackenroth Gerhard, 1953. *Bevölkerungslehre: Theorie, Soziologie und Statistik der Bevölkerung*. Berlin (u.a.), Springer, 531 p. (Enzyklopädie der Rechts- und Staatswissenschaft: Abteilung Staatswissenschaft.)

[48] Translations from English or French textbooks to German have been excluded from this list.

[49] Vienna Institute of Demography, Vienna.

[50] Vienna Institute of Demography, Vienna, and Vienna University of Technology, Vienna.

[51] Max Planck Institute for Demographic Research, Rostock.

[52] Institut National d'Études Démographiques, Paris.

Marschalck Peter, 1984. *Bevölkerungsgeschichte Deutschlands im 19. und 20. Jahrhundert*. Frankfurt am Main, Suhrkamp, (Neue Historische Bibliothek. Edition Suhrkamp), 203 p.

Mueller Ulrich, Nauck Bernhard, and Diekmann Andreas, 2000. *Handbuch der Demographie.* Vol. 1: *Modelle und methoden*. Vol. 2: *Anwendungen*. Berlin, Springer, 1426 p.

Mueller Ulrich, 1993. *Bevölkerungsstatistik und Bevölkerungsdynamik: Methoden und Modelle der Demographie für Wirtschafts-, Sozial-, Biowissenschaftler und Mediziner*. Berlin, Walter de Gruyter, 306 p. (Gruyter Lehrbuch.)

Rinne Horst, 1996. *Wirtschafts- und Bevölkerungsstatistik: Erläuterungen, Erhebungen, Ergebnisse*. Munich, Oldenbourg, 635 p. (Lehr- und Handbücher der Statistik.)

Rosenstiel (von) Lutz, Nerdinger Friedmann W., Oppitz Günther, SpieB Erika, and Stengel Martin, 1986. *Einführung in die Bevölkerungspsychologie*. Darmstadt, Wissenschaftliche Buchgesellschaft, 188 p.

Schmid Josef, 1976. *Einführung in die Bevölkerungssoziologie*. Reinbek bei Hamburg, Rowohlt Taschenbuch Verlag, 332 p. (Rororo-Studium; 98: Sozialwissenschaft.)

Ungern-Sternberg (von) Roderich and Schubnell Hermann, 1950. *Grundriss der Bevölkerungswissenschaft*. Stuttgart, Piscator Verlag, 602 p.

Vogel Friedrich and Grünenwald Werner, 1996. *Kleines Lexikon der Bevölkerungs- und Sozialstatistik*. Munich and Vienna, Oldenburg, 424 p.

Wingen Max, 1975. *Grundfragen der Bevölkerungspolitik*. Stuttgart, Kohlhammer, 155 p. (Sozioökonomie, no. 9.)

Winkler Wilhelm, 1952. *Typenlehre in der Demographie (Reine Bevölkerungstypen.)* Vienna, Rohrer in Komm, 160 p. (Sitzungsberichte / Österreichische Akademie der Wissenschaften, Philosophisch-historische Klasse.)

Winkler Wilhelm, 1969. *Demometrie*. Berlin, Duncker & Humblot, 447 p.

Zimpel Heinz-Gerhard, 2001. *Lexikon der Weltbevölkerung: Geographie, Kultur, Gesellschaft*. Berlin, Gruyter, 615 p.

APPENDIX

139–4

Demographic Treatises and Textbooks in Spanish

ROSA GÓMEZ REDONDO AND MARÍA EUGENIA COSÍO-ZAVALA

Universidad Nacional de Educación a Distancia (UNED), Facultad de Ciencias Políticas y Sociología, Madrid, Spain

Center of Research and Documentation on Latin America (CREDAL), Paris, and Université de Paris X-Nanterre, Paris, France

Demographic literature in Spanish is abundant, and even in the context of this short article on the production of treatises and textbooks (i.e., the works that make up the major part of the discipline), we can mention only the main publications, those that have had the most influence on teaching and research. It is also quite diverse, and although the same publications are frequently used on both sides of the Atlantic, the teaching and practice of demography has developed differently, and we thought it desirable to divide this article into two sections, one dealing with Spain and the other with the Latin American countries.

INTRODUCTION TO TREATISES IN SPAIN[53]

In Spain, the first treatises and handbooks appeared in the middle of the 20th century,[54] both original Spanish publications (Ros Jimeno *et al*., 1945) and translations from Italian (Gini, 1952). After the barren period that the years between 1950 and 1960[55] constituted for all scientific disciplines, the publication of demographic treatises and textbooks was somewhat delayed. Population studies really began to take shape at the beginning of the 1970s, mainly in two particular fields: historical demography, with the translation of works by Anthony Wrigley (1969), and Louis Henry (1971, 1976, 1983), and demography as a supplement to socioeconomic studies in relation to the modernization process. However, it was not until the 1970s that the publication of a comprehensive demography textbook written by a Spanish author occurred (Leguina, 1973); this textbook was used by successive generations of students and researchers as a result of several republications. Juan Lasterra's textbook was also published in 1976, and Spanish versions of Hubert Gérard and Guillaume Wunsch's small textbook (1975) and the more substantial demographic analysis by Roland Pressat, *La Pratique de la Démographie* (*The Practice of Demography*) (1977a), were translated and supplemented by an appendix on the Spanish population (Pressat, 1977b).

Toward the end of the 1970s and, particularly during the 1980s, demography developed rapidly in Spain. The period is characterized by consideration of data sources, analysis methods and techniques, and the basic mechanisms of population dynamics. Researchers became aware of American works published and translated into Spanish, such as those of Mortimer Spiegelman (1972) and Nathan Keyfitz (1975; 1979), as well as those published by international organizations, such as the United Nations (1975, 1980, 1986), the Population Reference Bureau (Haupt and Kane, 1980), and CELADE.

[53] The collaboration in this section of Yolanda Morcillo, librarian of the UNED, is acknowledged.

[54] With the exception of some minor works such as Elementos de Demografía by Felipe Paz, dated 1915.

[55] During that period, Spain was isolated from contact with international organizations.

In response to the needs of the period, population forecasts were of undeniable interest at the time. With the democratic transition and setting up of autonomous regional governments, the demand for demographic analyses became pressing from local authorities and autonomous communities and as from national governments. More works by Roland Pressat were translated (1983, 1985) that to this day have a major influence on the teaching of demographic analysis in Spain.[56] In the same decade, the Spanish version of the publication of John Weeks (1984) was published, which had a very educational sociologic approach making demography accessible to students of social science. During the same period, the work of Julio Vinuesa, Ana Olivera, Antonio Abellán, and Antonio Moreno was published (1982, republished 1988), which emphasized spatial dynamics and which is very useful for public administration. At the same time, surveys on fertility, migration, disability, and health were developed, but this was not reflected by an increase in the production of textbooks.

During the 1990s, textbooks written by Spanish authors were published in addition to new translations that were much appreciated for teaching, such as those of Georges Tapinos (1990); Massimo Livi Bacci (1993), an author who has great knowledge of Spanish demography; and the short book by Jacques Vallin (1995).[57] As for Spanish authors, they notably published two recent textbooks: the collective and multidisciplinary publication of Julio Vinuesa, Francisco Zamora, Ricardo Génova, Pedro Serrano, and Joaquín Recaño (1994), and the book by Graciela Sarrible (1996) on population theories.

At the beginning of the 21st century, demography is consolidated in Spain, both as a field of research and of education, and it is often included in social science courses in universities under various names. Within most public and private universities and research centers, the study and research of population are conducted within a multidisciplinary framework. However, demography is still not a specific university degree and there is no official qualification in demography, which explains why more specialized works are published rather than real demographic treatises or textbooks that provide an overall view of the discipline.

[56] A first little-known Spanish edition was published in 1967.

[57] Information on the Spanish population by Rosa Gómez-Redondo were added to this version.

INTRODUCTION TO LATIN AMERICAN TREATISES

In Latin America, the first publications or Spanish translations of textbooks appeared in the context of a training program set up by the UNFPA[58] in CEPAL,[59] for which during the 1950s in Santiago (Chile) the *Centro latinoamericano de Demografía* (CELADE) was created, with Carmen Miró as director. This program also had a considerable influence on national publications of demographic treatises and textbooks when, from the 1960s, the first graduates of CELADE returned to their respective countries (e.g., Argentina, Brazil, Colombia, Costa Rica, Cuba, Ecuador, Mexico, Peru) to set up Master's programs in demography in various Latin American universities.[60]

The first Spanish textbooks used within this context were translations of United Nations publications (Naciones Unidas, 1968, 1975, 1986), and then new textbooks were prepared by CELADE (1984) or translated into Spanish by the same organization (Bourgeois-Pichat, 1985; Brass, 1972; Carrier, 1975; Keyfitz, 1975, 1979; Granados, 1986; Ortega, 1987; Pinto, 1973; Somoza, 1975; Tabah, 1962; Vieira, 1973). At the same time, Spanish translations of North American treatises were used as textbooks in different Latin American demographic training (Bogue, 1973; Haupt, 1980; Hauser and Duncan, 1975; Keyfitz and Flieger, 1975; Potter, 1983; Spiegelman, 1972; Weeks, 1984). Spanish versions of the works of European authors, mainly French, Italian, Spanish, and English, were also widely diffused, some of them very recently (Courgeau and Lelièvre, 2001; Vallin, 1994; Livi-Bacci, 1993; Wunsch, 1992; Tapinos, 1990; Henry, 1971, 1976, 1983; Leguina, 1973; Pressat, 1963, 1977, 1981, 1983a, 1983b, 1985, Wrigley, 1969, 1985; Gini, 1952). Treatises and textbooks written by Latin Americans were published in countries where demographic research and teaching is very active: Argentina (Recchini de Lattes, 1978; Bottinelli, 1963), Cuba (Bueno, 1993, 1997; Catasus and Rodriguez, 1978), Chile (Castillo Muñoz *et al*., 1990; Mattelard, 1964), Colombia (Gonzalez, 1998; Carvajal, 1978), Mexico (Canales and Lerner, 2003; Rodriguez Cabrera, 1992; Welti, 1997; Corona *et al*., 1978; Navarrete and Vera, 1964), Nicaragua (Ocón, 1990), Venezuela (Ramírez, 1994).

The various treatises and textbooks published in Spanish and widely used in Latin America have

[58] United Nations Fund for Population Activities.

[59] United Nations Economic Commission for Latin America (Santiago, Chile).

[60] Notably the Maestría de Demografía of the Colegio de México (Mexico) was created in 1964.

several things in common. Initially, the objective was to provide tools and methods to become acquainted with and measure the main demographic phenomena explaining the very rapid population growth (e.g., mortality, fertility) and to make population forecasts. It was necessary to estimate these phenomena on the basis of data available in the 1950s, which were frequently partial or imperfect, mainly from censuses, because in Latin America, civil registration data were at the time practically unusable. The textbooks had to take this into account.

Then the objective moved toward a comparative and more detailed study of Latin American populations as a result of the development of international survey programs: PECFAL surveys,[61] World Fertility Survey (WFS), Demographic and Health Surveys (DHS), and numerous national and local sociodemographic surveys. The tools used were diversified to take the greatest advantage of the new data collected and the development of computing. During the 1970s and the 1980s, the emphasis was placed on fertility determinants and family planning.

Knowledge of demographic phenomena has been extended to nuptiality, internal and international migration, and aging. There is an increasing number of sociodemographic surveys and they have benefited from new methods of biographical surveys, such as is the case in Mexico (la Encuesta demográfica retrospectiva [EDER], 1998; Mexico City survey, 1970; Monterrey survey, 1968). The problems of urbanization and socioeconomic development (1960 to 1980) have been gradually replaced by the current issues of poverty and migration (1990 to 2000).

There is now an active Latin American school of demography, because Latin American demographers are very aware of their role with regard to the problems posed by the acceleration of population growth in the construction of modern states from the second half of the 20th century and were very much involved in the Third World movement of the period. However, the Latin American approach attached to the historical-structural mindset that predominated between 1960 and 1990, analyzing social relations of power at every level and seeking to define family strategies as units of analysis, has not produced any real demographic treatises, with authors preferring to express themselves in other types of publications, frequently in critical essays on particular issues.[62]

[61] A program of comparative surveys on fertility in Latin America, organized by CELADE during the 1960s and aimed at studying fertility in different countries in urban and rural areas.

[62] More attached to the sociology of populations than demography.

In 2005, this specific Latin-American movement of thought is barely represented in demographic literature, and Latin American demographers have been largely integrated into North American and European movements and increasingly work directly in English or French.

DEMOGRAPHIC TREATISES AND TEXTBOOKS IN SPANISH

SELECTED BY
DOMINIQUE CHAUVEL-MERKMAN,[63]
MARIA-EUGENIA COSIO-ZAVALA, AND
ROSA GÓMEZ-REDONDO

References

Alexeev Alexandr, 1987. *Geografía de la población con fundamentos de demografía*. Moscow, Progreso, 234 p.

AMIDEP, 1980. *Seminarios sobre metodología de la investigación científica y su aplicación al estudio de los problemas poblacionales*. Lima, Asociación Multidisciplinaria de Investigación y Docencia en Población (AMIDEP).

Bogue Donald J., 1973. *Técnicas demográficas de análisis de fecundidad*. Chicago, University of Chicago, Centre for Community and Family Studies, 122 p.

Bottinelli Pedro, 1963. *Nociones de demografia*. Buenos Aires, Dirección de estadística y economia sanitaria, 221 p.

Bourgeois-Pichat Jean, 1985. *Nuevas fronteras de la demografía*. Santiago (Chile), Centro Latinoamericana de Demografía (CELADE), 163 p.

Brass William, 1972. *Crítica de métodos para estimar el crecimiento de la población en los paises con datos limitados*. San José, CELADE, 26 p. (Serie DS, no. 11.)

Bueno Sánchez Eramis, 1993. *Los estudios de poblacion y su metodologia*. Havana (Cuba), Universidad de La Habana, Centro de Estudios Demográficos (CEDEM), 98 p. (Draft edition.)

Bueno Sánchez Eramis, 1997. *Estudios de población*. Havana (Cuba), Universidad de La Habana, Centro de Estudios Demográficos (CEDEM), 155 p.

Canales I. Alejandro and Lerner Sigal Susana, 2003. *Desafíos teórico-metodológicos en los estudios de población en el inicio del milenio*. Mexico, El Colegio de México, Universidad de Guadalajara and SOMEDE, 246 p.

Carrier Norman and Hobcraft John, 1975. *Estimaciones demográficas para sociedades en desarrollo: manual de técnicas para la detección y corrección de errores en la información demográfica*. San José (Costa Rica), Centro Latinoamericano de Demografía (CELADE), 232 p. (Serie D, no. 1026.)

Carvajal S. Alfonso, 1978. *La demografía: su objeto y método*. Medellín (Bolivia), Universidad Pontificia Bolivariana, 36 p.

Castillo Muñoz Berta, Mardones Azócar Graciela, and Solís F. Fresia, 1990. *Composición de la población; estudios demográficos*, Santiago (Chile), Universidad de Chile, Facultad de Medicina, Escuela de Salud Pública, 67 p. (Serie A, no. 2.)

Catasús Sonia and Rodríguez Mazorra Oscar, 1978. *Curso de demografia: programa analítico y laboratorios*. Havana, Universidad

[63] Archivist at Institut National d'Études Démographiques, Paris.

de La Habana, Centro de estudios demográficos, 77 p. (Economia. Serie 1, Estudios demográficos, no. 19.)

CELADE, 1960. *Proyeccion de la poblacion por sexo y grupos de edades*. Santiago (Chile), CELADE. (Series A and C: studies by country.)

CELADE, 1984. *Métodos para proyecciones demográficas*. San José (Costa Rica), Centro Latinoamericano de Demografía (CELADE), 254 p.

CEPAR, 1990. *Compendio basico poblacional*. Quito (Equateur), CEPAR, 6 + 110 + 23 p.

CGSNAGI, 1982. *Sistema de estadísticas sociales y demográficas: bases téoricas y metodológicas: aplicaciones generales*. Mexico, Coordinación General de los Servicios Nacionales de Estadística, Geografía e Informática (CGSNAGI.)

Corona V. Rodolfo, Minujin Z. Alberto, and Vera F. Gabriel, 1978. *Manual de técnicas de evaluación y ajuste de información estadística*. Mexico, Centro Nacional de Información y Estadísticas del Trabajo, 293 p.

Courgeau Daniel and Lelièvre Éva, 2001. *Análisis demográfico de las biografías*. Mexico, El Colegio de México, 305 p.

DPP, 1996. *Formulario de los principales indicadores demográficos*. Mexico, Dirección de Política Poblacional (DPP), 32 p.

Gérard Hubert and Wunsch Guillaume, 1975. *Demográfia*. Madrid, Ediciones Pirámide, 191 p. (Translated by Teresa Alameda Puig.)

Gini Corrado, 1952. *Teorías de la población*. Madrid, Aguilar, 278 p. (Spanish translation of *Teorie della popolazione*.)

Gonzalez Alejandro, 1998. *Conceptos y técnicas básicas de análisis demográfico*. Bogotá, Universidad Externado de Colombia, Cuadernos del CIDS, 123 p. (Serie II (3).)

Granados María del Pilar, 1986. *Técnicas de proyección de población de áreas menores: aplicación y evaluación*. Santiago (Chile), CELADE, 100 p. (Programa de Maestría en Demografía y Estudios Sociales de la Población, 1985–1986. Maestría en Demografía.)

Haupt Arthur and Kane Thomas T., 1980. *Guía rápida de población del Population Reference Bureau*. Washington (D.C.), Population Reference Bureau, 77 p. (Spanish translation of the *Population Reference Bureau's population handbook*.)

Hauser Philip M. and Duncan Otis Dudley, 1975. *El estudio de la población*. Santiago (Chile), Centro latinoamericano de demografia (CELADE), 3 vol., 1209 p.

Henry Louis, 1971. *Perspectivas demográficas*. Barcelona, Vicens-Vives, 139 p. (Translator: Armando Sáez.)

Henry Louis, 1976. *Demografía*. Barcelona, Editorial labor, 349 p. (Colección Labor Universitaria.)

Henry Louis, 1983. *Manual de demografía histórica: técnicas de análisis*. Barcelona, Editorial Crítica, 279 p. (Colección Crítica/Historia.)

Keyfitz Nathan and Flieger Wilhelm, 1975. *Demografía: métodos estadísticos*. Buenos Aires, Ediciones Marymar, 609 p.

Keyfitz Nathan, 1979. *Introducción a las matemáticas de la población*. Santiago (Chile), CELADE, 456 p. (Serie E, no. 18.)

Lasterra Juan, 1986. *Estudio de población*. Madrid, Alhambra, 111 p. (Biblioteca de recursos didácticos Alhambra, no. 8.)

Le Roy y Cassá Jorge Eduardo, 1908. *Algo sobre demografía*. Havana, Impr. Avisador comercial.

Leguina Joaquín, 1973. *Fundamentos de demografía*, 1st ed. Madrid, Siglo Veintiuno de España, 372 p. (4th edition, 1989.)

Livi Bacci Massimo, 1993. *Introducción a la demografía*. Barcelona, Editorial Ariel, 480 p. (Ariel Historia.)

Mattelard Armand, 1964. *Manual de analisis demográfico: un ejemplo de investigación en un pais latino-americano*. Santiago (Chile), Centro de investigaiones sociológicas, 626 p.

Mertens Walter, Przeworski Adam, Zemelman Hugo, and Mora y Araujo Manuel, 1982. *Reflexiones teórico-metodológicas sobre investigaciones en población*. Mexico, El Colegio de México, Consejo Latinoamericano de Ciencias Sociales, 191 p.

Naciones Unidas, 1968. *Métodos para establecer mediciones demográficas fundamentales a partir de datos incompletos*. New York, United Nations, Department of Economic and Social Affairs, 135 p. (Estudios sobre población, no. 42. Manuales sobre métodos de cálculo de la población, no. 4. ST/SOA/A/42.)

Naciones Unidas, 1975. *Hacia un sistema de estadísticas sociales y demográficas*. New York, United Nations, Department of Economic and Social Affairs, 209 p. (Estudios de métodos. Serie F, no. 18.)

Naciones Unidas, 1986. *Manual X: técnicas indirectas de estimación demográfica*. New York, United Nations, Department of International Economic and Social Affairs, 318 p. (Estudios de población, no. 81.)

Navarrete Emma Liliana and Vera Bolaños Marta G., 1964. *Población y sociedad*. México, El Colegio mexiquense, Consejo estatal de Población, 442 p.

Ocón Abaunza Jaime, 1990. *Conceptos elementales de técnicas demográficas*. Managua (Nicaragua), CIPRES (Centro para la Promoción, Investigación y el Desarrollo Rural y Social), 278 p.

Ortega Antonio, 1987. Metodos de construccion de tablas abreviadas, *in: Tablas de mortalidad*, p. 129–168. San José (Costa Rica), Naciones Unidas, CEPAL, CELADE, 297 p. (Serie E, CELADE [San José], no. 1004.)

Paz Felipe Santiago, 1915. *Elementos de demografía*. Madrid, Librería "Fernando Fe," 267 p. (Preface by Guillaume de Greef.)

Pinto Vieira Alvaro, 1973. *El pensamiento crítico en demografía*. Santiago (Chile), Centro Latinoamericano de Demografía (CELADE), 454 p.

Potter Joseph, 1983. *Métodos para la estimación demográfica: logros y perspectivas*. México, Population Council, 20 p. (Documento de trabajo, Population Council, Oficina Regional para América latina y el Caribe, no. 15.)

Pressat Roland, 1963. *Ubicación de los fenómenos demográficos en el tiempo*. Santiago (Chile), Centro Latinoamericano de Demografía (CELADE), 18 p. (Synopsis. *L'analyse démographique*, Paris, Presses universitaires, 1961.)

Pressat Roland, 1967 (republished 1983.) *El análisis demográfico*. Madrid, Fondo de Cultura Económica, 353 p.

Pressat Roland, 1977. *La práctica de la demografía: 30 problemas*. Mexico, Fondo de Cultura Económica, 368 p. (Sección de Obras de Sociología.)

Pressat Roland, 1981. *Demografía estadística*. Barcelona, Ariel, 211 p. (Translated by Manuel Delgado Cabeza.)

Pressat Roland, 1983a. *El análisis demográfico*. México, Fondo de Cultura Económica, 359 p. (Sección de obras de sociología.)

Pressat Roland, 1983b. *Los métodos en demografía*. Barcelona, Oikos-Tau, 143 p. (Coleccion ¿ Qué sé? no. 150.)

Pressat Roland, 1985. *Introducción a la demografía*. Barcelona, Ariel, 210 p. (Ariel Quincenal no. 34.)

Ramírez Erasmo, 1994. *Demografía general: teoría, metodos y comportamientos*. Caracas, Academia Nacional de Ciencias Económicas, 311 p.

Recchini de Lattes Zulma L., 1978. *Reflexiones críticas sobre estudios demográficos del crecimiento urbano: una propuesta metodológica*. Buenos Aires, Centro de Estudios de Población (CENEP), 23 p.

Rodríguez Cabrera Aída, 1992. *Demografía: conceptos y técnicas fundamentales/Aída Rodríguez Cabrera, Pedro Maldonado Cruz*. Oaxaca (Mexico), Instituto Tecnológico de Oaxaca, 159 p.

Ros Jimeno José, Villar Jesús, Ruiz Javier, Varón Rodolfo, Vallejo Antonio, and de la Quintana Primitivo, 1945. *Estudios demográficos*. Madrid, Consejo Superior de Investigaciones

Científicas, Instituto “Balmes” de Sociología, 301 p. (prólogo de Severino Aznar. Diana. Artes gráficas.)

Sarrible Graciela D., 1996. *Teoría de la población*. Barcelona, Edicions de la Universitat de Barcelona, 55 p. (Textos docents, no. 73.)

Sarrible Graciela D., 1998. *Teoría de la población*. Barcelona, Edicions de la Universitat de Barcelona (EUB), 148 p. (Revised 2nd ed, Textos docents, no. 73.)

Sireau Alberto, 1982. Instrumento de demografia, *in*: UNESCO (ed.), *Programa Regional de Educacion en Poblacion. Estudio de referencia sobre educacion en poblacion para America Latina: enfoque interdisciplinario y sistemico*, p. 405–420. Santiago (Chile), UNESCO, Oficina Regional de Educacion para America Latina y el Caribe.

Somoza Jorge L., 1975. *El método de encuesta demográfica de CELADE*. Chapel Hill (North Carolina), University of North Carolina, Department of Biostatistics, Carolina Population Center, International Program of Laboratories for Population Statistics, 50 p. (Laboratories for Population Statistics. Scientific Reports, no. 18.)

Spiegelman Mortimer, 1972. *Introducción a la demografía*. México, Fondo de Cultura Económica, Sección de obras de sociología, 492 p.

Tabah Léon and Arevalo J., 1962. *Algunos modelos teóricos y numéricos de población*. Santiago (Chile), United Nations, Centro latinoamericano de demografía (CELADE) and Universidad de Chile, Centro latinoamericano de demografía.

Tapinos Georges, 1990. *Elementos de demografía*. Madrid, Espasa Calpe, 410 p. (preface by Antonio Izquierdo.)

Vallin Jacques, 1994. *La demografía*. Santiago (Chile), CEPAL, Centro Latinoamericano de Demografia (CELADE), 164 p. (Spanish translation of *La démographie*, Paris, La Découverte, 1991.)

Vallin Jacques, 1995. *La demografía*. Madrid, Alianza Universidad, 142 p. (Translated from French by Maria Hernandez and Rosa Gómez Redondo.)

Vegara Alfonso and Entrenout (d’) Alban, 1988. *Introducción al análisis demográfico. Un enfoque para urbanistas*. Burlada (Navarra), Centro de Estudios Urbanos, 142 p. (Taller de Ideas.)

Vinuesa Julio, Zamora Francisco, Génova Ricard, Serrano Pedro, and Recaño Joaquín (eds.), 1994. *Demografía: análisis y proyecciones*. Madrid, Síntesis, 366 p. (Espacios y sociedades, no. 9.)

Vinuesa Angulo Julio, Olivera Ana, Abellán Garcia Antonio, and Moreno Jiménez Antonio, 1982. *El estudio de la población*. Madrid, Instituto de Estudios de Administración Local, 234 p. (Espacios y sociedades, no. 9) (Republished by the Centro de estudios para la Administración local, in 1988.)

Weeks John R., 1984. *Sociología de la población: introducción a los conceptos y cuestiones básicas*. Madrid, Alianza, D. L., 471 p. (Spanish version by María del Mar Terán and José Juan Toharia, Alianza Universidad. Textos; 84.)

Welti Carlos, 1997. *Demografia I*. Buenos Aires (Argentina), Programa Latinoamericano de Actividades en Poblacion (PROLAP), 237 p.

Wrigley E. Anthony, 1969. *Historia y población: introducción a la demografía histórica*. Madrid, Ediciones Guadarrama, 255 p. (Colección Biblioteca para el hombre actual.)

Wrigley E. Anthony, 1985. *Historia y población: introducción a la demografía histórica*. Madrid, Editorial Crítica, 252 p. (Serie General Estudios y Ensayos.)

Wrong Dennis H., 1971. *La población*, 3rd ed. Buenos Aires, Paidós, 135 p. (Biblioteca del hombre contemporáneo, no. 62.)

Wunsch Guillaume, 1992. *Técnicas para el análisis de datos demográficos deficientes*. Mexico, El Colegio de México, Centro de Estudios Demográficos y de Desarrollo Urbano, 244 p. (Translation by Daniel Hiernaux Nicolas.)

APPENDIX

139–5

Demographic Treatises and Textbooks in Portuguese

ELZA BERQUÓ

Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil

TREATISES, TEXTBOOKS, AND EDUCATIONAL PUBLICATIONS

The number of demographic textbooks published in Brazil is quite small because the systematic teaching of this discipline has only recently developed. The study of demography at university level began 20 years ago. It began with postgraduate courses in 1985 in CEDEPLAR,[64] the Regional Center for Development and Planning at the Federal University of Minas Gerais (UFMG[65]). Thereafter, a doctoral course was introduced in 1993 at the Institute of Philosophy and Human Sciences[66] at Campinas State University[67] (UNICAMP), in partnership with the National Center for the Study of Demography (NEPO[68]), which was recently supplemented by a masters course. The National School of Statistics, created in 1953, in collaboration with the Brazilian Institute of Geography and Statistics (IBGE[69]) has provided a masters course in demography since 1998.

The abundant international demographic literature, notably American and French, plays a very important role in teaching in Brazilian universities, as a result, fluency in at least one of these languages is essential.

[64] Centro de Desenvolvimento e Planejamento Regional.
[65] Universidade Federal de Minas Gerais.
[66] Instituto de Filosofia e Ciências Humanas.
[67] Universidade Estadual de Campinas.
[68] Núcleo de Estudos de População.
[69] Instituto Brasileiro de Geografia e Estatística.

In addition, the Brazilian Association for the Study of Population (ABEP[70]) has channeled, over almost 30 years, the important Brazilian scientific production in the area of population in its journal *Revista Brasileira de Estudos de População* (*Brazilian Journal of Population Studies*), created in 1982. Through its 22 issues, it reflects the country's evolution with regard to scientific knowledge of demography. The following paragraphs provide the few Brazilian works published since 1980 that can be included in the category of demography textbooks.

Dinâmica da População: Teoria métodos e Técnicas de Análise, published in 1980 and edited by Jair Santos, Maria-Stella Levy, and Tamas Szmreczanyi, is a collection of articles on demography. It discusses the objectives and the field of demography; the sociohistorical process of the evolution of the population; the factors related to population dynamics, notably mortality and fertility; the nature of demographic data and their sources, the main methods of analysis and projection; nuptiality, fertility, mortality, and migration; the relation between population and economic development; the contribution of demography to the issues of development, education, public health, and population policies.

Pequeno Histórico do Desenvolvimento da Demografia no Brasil: A História, o Ensino, a Pesquisa, as Instituições e o Financiamento para Demografia (Berquo and Rocha, 1982) is a retrospective study (1970 to 1980) of the

[70] Associação Brasileira de Estudos Populacionais.

development of demography. The work is based on a compilation of 587 articles dealing with population studies and demographic analyses, listed according to several criteria, and the organization responsible for each article.

José Alberto Magno de Carvalho, Diana Oya Sawyer, and Rodriguez Nascimento published *Introdução a Alguns Conceitos Básicos e Medidas em Demografia* in 1992, which introduced the fundamental bases of demography and the calculation of the principal indicators of the components of population dynamics by presenting applications and illustrating their potential. Population growth, fertility, mortality, and reproduction are discussed in a simple, critical, and objective manner with examples of applications to real data. The publication aims at providing a practical and realistic overview of population trends in Brazil, in the light of the major changes that have affected the components of its dynamics over past decades.

Fontes de Dados Demográficos, by Ralph Hakkert (1986), provides an analysis and a discussion of demographic data sources, distinguishing stock data (population) and flow data (births, deaths, migrations).

Introdução a Métodos de Estimativas Interpolações Populacionais, written by Gustavo Naves Givisiez (2004), discusses the interpolation of demographic data by presenting different graphical and mathematical methods useful for adjusting the most frequently used demographic indicators, thereafter it discusses the most common methods for estimating the distribution of a population by age and sex from a total number or on the basis of different age groups to that which is required.

References

Berquó Elza and Rocha Maria Isabel Baltar (da), 1982. Pequeno histórico do desenvolvimento da demografia no Brasil: a história, o ensino, a pesquisa, as instituições e o financiamento para demografia, *in*: *Demografia: avaliação e perspectivas*, p. 438–448. São Paulo, SEPLAN/CNPq. (SEPLAN/CNPq, no. 1.)

Carvalho (Magno de) José Alberto, Sawyer Diana Oya, and Nascimento Roberto Rodriguez (do), 1992. *Introdução a alguns conceitos básicos e medidas em demografia*. Belo Horizonte, CEDEPLAR, 63 p. (Didactic Texts, 1.)

Hakkert Ralph, 1996. *Fontes de dados demográficos*. Belo Horizonte, ABEP, 71 p. (Didactic Texts, 3.)

Navez Givisiez Gustavo Henrique, 2004. *Introdução a métodos de estimativas interpolações populacionais*. Campinas, ABEP, 45–70 p. (Didactic Texts, 1.)

Santos Jair L. F., Levy Maria Stella Ferreira, and Szmrecsanyi Tamás, 1980. *Dinâmica da população: teoria métodos e técnicas de análise*. São Paulo, T. A. Queiroz, 362 p. (Biblioteca Básica de Ciências Sociais, Série 2ª, Textos v. 3).

TREATISES, TEXTBOOKS, AND EDUCATIONAL PUBLICATIONS IN PORTUGUESE

SELECTED BY ELZA BERQUO[71] AND DOMINIQUE DIGUET[72]

References

Arroteia (de Carvalho) Jorge, 1998. *Demografia escolar: teoria e métodos*. Aveiro, Universidad, 87 p. (Cadernos de Análise Sócio-Organizacinal da Educação, 11.)

Arroteia (de Carvalho) Jorge, 1986. *emografia escolar: relatório sobre o programa, o conteúdo e os métodos de ensino da disciplina da demografia escolar*. Aveiro, [s. n.], 73 p.

Bandeira Mario Leston, 2004. *Demografia: objecto, teorias e métodos*. Lisbon, Escolar Editora, 340 p.

Bandeira Mario Leston, 1996. *Demografia e modernidade: família e transição demográfica em Portugal*. Lisbon, Imp. Nac.-Casa da Moeda, 572 p. (Análise Social.)

Beltrão Pedro Calderan, 1972. *Demografia ciençia da população: análise e teoria*. Porto Alegre, Sulina, 335 p.

Berquó Elza and Rocha (Balta da) Maria Isabel, 1982. Pequeno histórico do desenvolvimento da demografia no Brasil: a história, o ensino, a pesquisa, as instituições e o financiamento para demografia, *in*: *Demografia: avaliação e perspectivas*, p. 438–448. São Paulo, SEPLAN/CNPq. (SEPLAN/CNPq, no. 1.)

Campos Ezequiel de and Grei A, 1915. *Subsídios para a demografia portuguesa*. Porto, Renascença Portuguesa, 273 p.

Cardoso Ciro Flamarion S. and Perez Brignoli Héctor, 1983. *Métodos da história: introdução aos problemas, métodos e técnicas da história demográfica, econômica e social*. Rio de Janeiro, Graal. (Translation by João Maia.)

Carvalho (Magno de) José Alberto, Sawyer Diana Oya and Nascimento Rodriguez (do) Roberto, 1992. *Introdução a alguns conceitos básicos e medidas em demografia*. Belo Horizonte, CEDEPLAR, 63 p. (Didactic Texts, 1.)

Hakkert Ralph, 1996. *Fontes de dados demográficos*. Belo Horizonte, ABEP, 71 p. (Didactic Texts, 3.)

Henry Louis, 1977. *Técnicas de análise em demografia histórica*. Curitiba, Universidade Federal do Paraná, 165 p.

Henry Louis, 1988. *Técnicas de análise em demografia histórica*. Lisbon, Gravida, 231 p. (Translation and notes by J. Manuel Nazareth.)

Machado J. M. de Sá C., 1998. *Análise demográfica*. Braga, [s.n.], 2 vol. p.

Naves Givisiez Gustavo Henrique, 2004. *Introdução a métodos de estimativas interpolações populacionais*. Campinas, ABEP, 45 + 70 p. (Didactic Texts, 1.)

Nazareth J. Manuel, 1996. *Introdução à demografia: teoria e prática*. Lisbon, Presença, 194 p. (2nd ed, Coleção Fundamentos, 8.)

Nazareth J. Manuel, 1998. *Princípios e métodos de análise da demografia portuguesa*. Lisbon, Presença, 323 p.

Rodrigues Helia M. and Duarte Jonas M., 1987. *Alguns elementos estatísticos de demografia, emprego e produtividade*. Lisbon, M. T.-G. P, 112 p. (Série Estatística, 31 / Análise, 2.)

Santos Jair L. F., Levy Maria Stella Ferreira, and Szmrecsanyi Tamás, 1980. *Dinâmica da população: teoria métodos e técnicas de análise*. São Paulo, T. A. Queiroz, 362 p. (Biblioteca Básica de Ciências Sociais, Série 2ª, Textos v. 3).

Torres Adelino, 1996. *Demografia e desenvolvimento: elementos básicos*. Lisbon, Gradiva, 168 p. (Trajectos, 31.)

[71] Cento Brasileiro de Análise e Planejamento (CEBRAP), Sao Paulo, Brazil.

[72] Institut National d'Études Démographiques, Paris.

APPENDIX

139–6

Demographic Treatises and Textbooks in Dutch

GIJS BEETS AND JOLANDE SIEBENGA

Nederlands Interdisciplinair Demografisch Insituut (NIDI), Den Haag, Pays-Bas

The Netherlands has a long-standing demographic tradition. After the famous study by John Graunt (1620–1674) of the London "bills of mortality," the Dutch mathematician Christiaan Huygens (1629–1695) reflected further on life table theory. During the 18th century, Nicolaas Struyck (1687–1769) and Willem Kersseboom (1690–1771) published some basic demographic studies. "Kersseboom had already pictured the methodologic outline of demographic forecasting in the 18th century" (De Gans, 1999, p. 93). As a tribute to him, the NIDI staff planted a cherry tree (*Kersseboom* in old Dutch) in the NIDI garden.

The 20th century was a particularly thriving period for Dutch demography. The Netherlands Demographic Society (Nederlandse Vereniging voor Demografie [NVD]) was created in 1928 at the same time as the creation of the IUSSP. Almost immediately, annual meetings were convened with presentations and discussions on a wide variety of topics, such as aspects of rapid population growth, fertility rates, and population projections. Some famous names from the first half of the 20th century are Angenot (1901–1979), Delfgaauw (1905–1984), Holwerda (1887–1944), Methorst (1868–1955), 't Hooft (1896–1941), Van Lohuizen (1890–1956), Van Zanten (1874–1944), Verrijn Stuart (1865–1948), and Wiebols (1895–1960). In Belgium, the statisticians Quetelet (1796–1874) and Verhulst (1804–1849) in the 19th century merit special attention.

This overview briefly summarizes the most important contributions to demography made by native Dutch speakers over the past 100 years. In the wake of the above-mentioned demographers, several others contributed to questions such as: What is demography? What are demographic tools? How does one become a demographer? What makes demography different from other scientific disciplines?

From the NIDI library's online documentation system, we have selected 25 demographic treatises and textbooks published before 1980 in the Dutch language and 12 that were published during or after 1980. At first glance, it may be surprising that the number of publications decreased while the field of population research increased significantly after 1980. Many of the most basic treatises are already available, some in Dutch but mostly in other languages, and understanding other languages is not a problem for native Dutch speakers, Moreover, Dutch researchers mainly publish in English rather than their native language. To judge whether Dutch-language demographers have also published treatises in other languages we made another search on the online documentation system, and as a consequence added 7 more publications to the list; all of which were published in 1982 or later. In total, our appendix includes 44 publications, more than half of which were published before 1980.

The 25 demographic treatises published before 1980 were mainly published in the Netherlands, apart from one title published in Belgium and two in South Africa. Most of these treatises are basic introductions to demography as a scientific field in its broadest sense, although a large minority can be classified as focusing mainly on specific demographic topics, such as mathematical and analytic demographic issues, demometrics, population forecasting, and some have a specific geographic or sociologic perspective. Also of interest is the fact that six of these treatises are translations into Dutch. All of them were originally written in French—

one by Beaujeu-Garnier, one by Pressat, and four by Sauvy—and published between 1960 and 1975 in a series of Dutch pocket books devoted to popularized science (i.e., very accessible to a wider audience, including students). It is remarkable that during this period no publications in English or any other language were translated into Dutch.

After 1980, two of the twelve demographic treatises were published in Belgium, the others in the Netherlands. Again, most of these treatises are basic introductions to demography. Three were produced at NIDI and are specifically aimed at students and the wider public (NIDI, 2003; Van Imhoff *et al.*, 1995; Bronsema, 1990). Some treatises look forward to the 21st century, another focuses on life course issues. Lesthaeghe and Van de Kaa launched their well-known theory of the "second demographic transition" originally in Dutch (Van de Kaa and Lesthaeghe, 1986). During the following year, a translation of a practical population handbook (from English to Dutch) was produced at NIDI (van de Kaa, 1987).

Increasingly, we see that demographic output from the Dutch-speaking world is published in English-language journals (not included in the appendix). Dutch native speakers have written books in English as well. Among these are nine demographic treatises, originally published in English but co-produced by Dutch-speaking scientists (i.e., Cruijsen, De Gans, Dykstra, Keilman, Lesthaeghe, Overbeek, Van de Kaa, Van Imhoff, and Van Wissen). Three of these focus specifically on forecasting and modeling issues. Because these Anglo-Dutch publications are missing in this chapter, we have listed them separately in the appendix.

Acknowledgment

The authors would like to thank Frans van Poppel, Nico van Nimwegen, and Frans Willekens (all from the Netherlands Interdisciplinary Demographic Institute) for their valuable contributions.

DEMOGRAPHIC TREATISES AND TEXTBOOKS IN DUTCH

Books Published before 1980

In the Netherlands

Bak L, 1964. *Volk in veelvoud*. Meppel, Boom, 192 p. (Gamma reeks, no. 1.)

Bakels L., 1978. *Over vooronderstellingen van migratietheorieën: bijdrage tot een maatschappij-kritische theorie van de ruimte*. Amsterdam, Sociaal-Geografisch Instituut Universiteit van Amsterdam, 134 p. (ASVS-publicatiereeks, no. 2.)

Brand Willem, 1961. *Bevolkingsexplosie*. Amsterdam, Querido, 160 p. (Speciale Salamander.)

De Froe Arie, 1966. *Het vraagstuk van de wereldbevolking*. Baarn, Het Wereldvenster, 84 p. (Anatomie van de toekomst.)

De Gans Henk and Korzec Michel, 1978. *Mathematische demografie*. Groningen, Wolters-Noordhoff, 100 p.

De Jong Andries M., 1946. *Inleiding tot het bevolkingsvraagstuk*. 's-Gravenhage, Nijhoff, 224 p.

Guffens Theodorus M. G., Nelissen Nicolaas J. M. and Nas Peter J. M., 1972. *Beginselen van de kwantitatieve demografie: [videocursus]*. Nijmegen, Sociologisch Instituut Universiteit Nijmegen, 110 p. (Urbane en rurale sociologie, no. 21.)

Heeren Hendrik and van Praag Philip (eds.), 1974. *Van nu tot nul: bevolkingsgroei en bevolkingspolitiek in Nederland*. Utrecht, Het Spectrum, 357 p. (Aula boeken, no. 531.)

Hofstee Evert W., 1978. *De demografische ontwikkeling van Nederland in de eerste helft van den 19e eeuw: een historisch-demografische en sociologische studie*. Deventer, Van Loghum Slaterus, 231 p. (NIDI-publikaties, no. 2.)

Kooy Gerrit A., 1967. *Het modern-westers gezin: een inleidende gezinssociologische beschouwing*. Hilversum, Brand, 264 p. (Samen mens zijn, no. 6.)

Methorst Hendrikus W. and Sirks Marius J., 1948. *Het bevolkingsvraagstuk*. Amsterdam, Scheltema en Holkema, 230 p. (Van eigen erf, no. 3.)

Speckman Johan, 1978. *De bevolkingsproblematiek in de Derde wereld*. Assen, van Gorcum, 104 p. (Terrein verkenningen in de culturele antropologie, no. 5.)

't Hooft Floris W, 1929. *Het bevolkingsvraagstuk: waar wij heengaan in West Europa*. Amsterdam, Paris, 111 p.

van de Kaa Dirk J. and van der Windt Kees, 1978. *Minder mensen, meer welzijn? Bevolking en welzijn in Nederland*. Utrecht, Het Spectrum, 284 p. (Aula-boeken, no. 615.)

van Hinte E, 1946. *Inleiding in het bevolkingsvraagstuk*. Amsterdam, Wereldbibliotheek, 147 p. (Encyclopedie in monografieën.)

Yntema Luitzen, 1977. *Inleiding tot de demometrie*. Deventer, Van Loghum Slaterus, 281 p. (NIDI-publikaties, no. 3.)

In Belgium

van Mechelen Frans, 1963. *Dynamiek van over- en onderbevolking*. Antwerpen, Standaard-Boekhandel, 126 p.

In South Africa

Franzsen D, 1950. *Inleiding tot die bevolkingsvraagstuk*. Stellenbosch, Die Universiteits-uitgevers en Boekhandelaars, 186 p.

Sadie J, 1949. *Westerse bevolkingsneiginge*. Stellenbosch, Die Universiteits-uitgevers en Boekhandelaars, 131 p.

Translated from French into Dutch

Beaujeu-Garnier Jacqueline, 1965. *Mensen in miljarden: het probleem van de bevolkingstoename*. Utrecht, Het Spectrum, 479 p. (Aula-boeken, no. 382. [vert. uit het Frans]. Oorspr. titel: Trois milliards d'hommes.)

Pressat Roland, 1975. *Inleiding tot de demografie*. Utrecht, Het Spectrum, 160 p. (Aula-boeken, no. 550. [vert. op basis van de Engelse bewerking van het Franse origineel]. Vert. van: Population, 1970.)

Sauvy Alfred, 1960. *Het bevolkingsvraagstuk in de wereld*. Utrecht, Het Spectrum, 259 p. (Aula-boeken, no. 41. [vertaald uit het Frans]. Oorspr. titel: De Malthus à Mao Tsé-Toung.)

Sauvy Alfred, 1963. *Het probleem van de overbevolking*. Utrecht, Het Spectrum, 316 p. (Aula-boeken, no. 272. [vertaald uit het Frans]. Oorspr. titel: Malthus et les deux Marx.)

Sauvy Alfred, 1975. *Nulgroei?* Utrecht, Het Spectrum, 255 p. (Aula-paperback, no. 29. [vertaald uit het Frans]. Oorspr. titel: Croissance zéro?)

Books Published since 1980

In the Netherlands

Barnett Larry D. and Leeuw Frans L. (eds.), 1981. *Bevolking als vraagstuk: dimensies van bevolkingsbeleid in Nederland en de Verenigde Staten*. 's-Gravenhage, Vuga, 320 p.

Bronsema Harry, 1990. *Over bevolking: een wegwijzer voor de praktijk*. The Hague, Nederlands Interdisciplinair Demografisch Instituut, 134 p. (Herziene uitgave. Een NIDI publikatie.)

Frinking Gerard, 1998. *Het bevolkingsvraagstuk in de eenentwintigste eeuw*. The Hague, NWO, 60 p.

Groot Jéjé J. and Tiessen Hans M., 1998. *Statistische gegevensverwerking en demografie*, 1st ed. Utrecht, Lemma, 192 p. (2nd edition, 2002.)

Liefbroer Aart C. and Dykstra Pearl A., 2000. *Levenslopen in verandering: een studie naar ontwikkelingen in de levenslopen van Nederlanders geboren tussen 1900 en 1970*. The Hague, SDU, 256 p.

NIDI (Nederlands Interdisciplinair Demografisch Instituut), 2003. *Bevolkingsatlas van Nederland: demografische ontwikkelingen van 1850 tot heden*. Rijswijk, Elmar, 176 p. (samenstelling Nederlands Interdisciplinair Demografisch Instituut: Peter Ekamper *et al*.)

van de Kaa Dirk J. and Lesthaeghe Ron, 1986. *Bevolking: groei en krimp*. Deventer, Van Loghum Slaterus, 115 p. (Boekaflevering Mens en Maatschappij, jaargang 61.)

van Imhoff Evert, Moors Hein, and Bronsema Harry *et al*., 1995. *PopTrain: een introductie in de demografie: een computerprogramma voor gebruik in het demografieonderwijs op middelbare scholen*. The Hague, NIDI, diskette version 1.0 + guide de 23 p.

van Nimwegen Nico and de Jong Gierveld Jenny (eds.), 1992. *De demografische uitdaging: Nederland in Europa op weg naar de 21ste eeuw. Over aspecten van het bevolkingsvraagstuk in Nederland en Europa*. Houten, Bohn Stafleu Van Loghum, 147 p.

In Belgium

van Mechelen Frans and Vandamme Annemie, 1980. *Demografie: algemene inleiding tot de bevolkingswetenschap; nota's bij de colleges van Frans Van Mechelen, vol. 1. 3de herziene uitgave*. Leuven, Acco, 175 p.

van Mechelen Frans and Vandamme Annemie, 1980. *Demografie: algemene inleiding tot de bevolkingswetenschap; nota's bij de colleges van Frans Van Mechelen, vol. 2. 3rd herziene uitgave*. Leuven, Acco, 149 p.

Translations into Dutch

Haupt Arthur and Kane Thomas T., 1981. *Over bevolking: een practische handleiding in de demografie voor allen die denken en schrijven over het proces van bevolkingsontwikkeling [vertaald en bewerkt uit het Amerikaans door H. G. Moors en J. C. van den Brekel]*. Voorburg, NIDI, 95 p. (Oorspronkelijke titel: The Population Reference Bureau's Population handbook, 1978.)

English Publications by Dutch-Speaking Authors

de Gans Henk, 1999. *Population forecasting 1895–1945: the transition to modernity*. Dordrecht, Kluwer Academic Publishers, 290 p. (European studies of population, no. 5.)

Keilman Nico and Cruijsen Harri (eds.), 1992. *National population forecasting in industrialized countries*. The Hague, Swets en Zeitlinger, 365 p. (NIDI-CBGS publications, no. 24.)

Overbeek Johannes, 1982. *Population: an introduction*. New York and London, Harcourt Brace Jovanovich, 278 p.

Lesthaeghe Ron (ed.), 2002. *Meaning and choice: value orientations and life course decisions*. The Hague and Brussels, NIDI, 335 p. (NIDI-CBGS publications; 37.)

van de Kaa Dirk J., 1987. Europe's second demographic transition, *Population Bulletin*, vol. 42 (1), p. 1–59.

van Imhoff Evert, Kuijsten Anton, Hooimeijer Pieter and van Wissen Leo (eds.), 1995. *Household demography and household modeling*. New York, Plenum Press, 369 p. (Plenum series on demographic methods and population analysis.)

van Wissen Leo J. G. and Dykstra Pearl A. (eds.), 1999. *Population issues: an interdisciplinary focus*. Dordrecht, Kluwer Academic, 287 p. (The Plenum series on demographic methods and population analysis.)

APPENDIX

139–7

Demographic Textbooks in Japanese

SHIGEMI KONO
Faculty of International Economics, Reitaku University, Chiba-Ken, Japan

There are not too many demographic textbooks in Japanese. This is partly because demography is not often taught in Japanese universities. Moreover, although demography is usually taught within sociology departments in the United States, in Japan, it is usually taught in economics and uncommonly in geography departments.

The first comprehensive demographic work in Japanese was published in 1960 by Minoru Tachi: *Keishiki Jinkogaku: Jinko Gensho no Bunseki Hoho* or *Formal Demography: Methods of Analysing Population*. Tachi was regarded as a leader by Japanese demographers from the 1950s to the 1970s. With this 829-page publication, he succeeded in producing practically all the basic data and indicators regarding population, methods of analyzing components of population dynamics (e.g., fertility, mortality, and migration), and population models, from model life tables to stable populations. However, Tachi's main objective in writing this volume was not merely to present an illustration of demographic methods, but rather to establish empirical regularities and even some sort of population laws for the interaction between fertility, mortality, and migration, and population structure and its geographic distribution.

Jinko Tokeigaku, or *Demographic Statistics* (1980), by Yoichi Okazaki, is easy to read and introduces practical calculations. The book deals with basic population growth indicators, and population composition, methods for analyzing fertility, mortality, and migration, the calculation of life tables and their application, as well as population estimates and projections. This work, which to a certain extent resembles George Barclay's *Techniques of Population Analysis* (see Appendix 139–1), has been very widely used in Japan for introducing students to basic demographic methods.

Another book that is easy to read and use in specific situations is *Jinko Busnseki Nyumon*, or *An Introduction to Population Analysis*, edited by Kiichi Yamaguchi in 1989. To a certain extent, it is an update of Tachi's work, due to the incorporation of new methodologic developments. However, it is also simpler to understand and use for applications to real data, and it includes some very useful illustrations. Similar to Tachi's book, and in contrast to most American and European textbooks, it focuses more on analyzing migration and spatial distribution. However, there are several new features in this book. It introduces a wide range of tools for analyzing fertility, and then it demonstrates functional projections for the labor force, families, and households. It also paves the way for the analysis of multistate demography.

Suuri Jinkogaku, or *Mathematical Models for Demography and Epidemiology*, by Hisashi Inaba (2002), explains mathematical models that make it possible to describe population dynamics. According to Inaba, technical or formal demography is mainly inductive, as it attempts to find empirical regularities in demographic phenomena and to approximate them by mathematical functions, while mathematical demography is essentially deductive, focusing on the process of human reproduction and population replacement. The main chapters concentrate on the development of stable population models and the development of two-sex models.

Because life tables are an essential tool for demographic analysis, reference should be made to *Seimei-hyo Kenkyu*, or *Life Table Methods*, edited by Kiichi Yamaguchi, Zenji Nanjo, Takao Shigematsu, and Kazumasa Kobayashi (1995). The book begins with the general concept of life tables and then discusses simple-decrement tables and their specific applications, such as life tables by cause of death or by marital status. It then discusses multiple-decrement tables, such as working life tables and nuptiality tables, and it concludes with a discussion of multistate or multiregional life tables. The book includes interesting methods for creating tables for small populations, based on a Bayesian approach and other approaches.

Sekai no Jinko, or *World Population*, by Shigemi Kono (2000), provides an overview of demography based on the demographic situation in the world and its regions. It includes chapters on world population growth, methods for demographic estimates and projections, mortality and life expectancy, fertility trends and determinants, population composition and aging, migration and urbanization, the relations among population, resources, and the environment, and population policies.

Demography is most frequently taught within economics departments in Japan and it would therefore seem pertinent to mention a textbook on population economics recently published (2001) by Hisakazu Kato: *Jinko Keizaigaku Nyumon*, or *An Introduction to Population Economics*. This very useful publication discusses the development of population economics from Malthus to Keynes, macroeconomic and microeconomic theories of fertility behavior, the economics of nuptiality, the relation between female labor force participation and fertility, as well as population aging and social security problems.

Jinko Dai Jiten, or *Encyclopaedia of Population*, edited by the Editorial Committee of the Population Association of Japan (PAJ, 2002) is not precisely a demographic textbook, but it has been extensively used as an essential reference book in public and university libraries. It contains 999 B5 format pages and includes eight major sections and 192 articles. These eight sections are world population, its history and its geography; population issues of the world and Japan; population ideas and theories; population statistics and methods of analysis; family and population replacement; spatial mobility and population distribution; socioeconomic aspects of population, and population and family policies.

A SELECTION OF DEMOGRAPHIC TEXTBOOKS IN JAPANESE

References

Atoh Makoto, 2000. *Gendai Jinkogaku: Shoshi-Korei Shakai no Kiso Chishiki [Modern Demography: Basic Information on Ageing Societies with Low Fertility]*. Tokyo, Nihon Hyoron-sha, 258 p.

Hama Hidehiko and Yamaguchi Kiichi (eds.), 1997. *Chiyiki Jinko Bunseki no Kiso [Bases of Regional Demographic Analysis]*. Tokyo, Kokon Shoin, 236 p.

Hayami Akira (ed.), 2003. *Rekishi Jinkogaku to Kazokushi [Historical Demography and History of the Family]*. Tokyo, Fujiwara Shoten, 548 p.

Inaba Hisashi, 2002. *Suuri Jinkogaku [Mathematical Models for Demography and Epidemiology]*. Tokyo, Tokyo University Press, 412 p.

Kato Hisakazu, 2001. *Jinko Keizaigaku Nyumon [Introduction to Population Economics]*. Tokyo, Nihon Hyoron-sha, 285 p.

Kono Shigemi, 2000. *Sekai no Jinko [World Population]*. Tokyo, University of Tokyo Press, 250 p. (2nd ed.)

Ohbuchi Hiroshi, 1988. *Shusshoryoku no Keizaigaku [The Economics of Fertility]*. Hachioji-shi (Tokyo), Chuo University Press, 239 p.

Okazaki Yoichi, 1999. *Jinko Tokeigaku [Demographic Statistics]*. Tokyo, Kokon Shoin, 248 p. (revised and enlarged).

PAJ (Population Association of Japan) Editorial Committee (ed.), 2002. *Jinko Dai Jiten [Population Encyclopaedia]*. Tokyo, Baifukan, 999 p.

Tachi Minoru, 1960. *Keishiki Jinkogaku: Jinko Gensho no Bunseki Hoho [Formal Demography: Methods of Analysing Population]*. Tokyo, Kokon Shoin, 819 p.

Yamaguchi Kiichi (ed.), 1989. *Jinko Busnseki Nyumon [An Introduction to Population Analysis]*. Tokyo, Kokon Shoin, 250 p.

Yamaguchi Kiichi, Nanjo Zenji, Shigematsu Takao, and Kobayashi Kazumasa (eds.), 1995. *Seimei-hyo Kenkyu [Life Table Methods]*, 332 p.

APPENDIX

139–8

Demographic Treatises and Textbooks Published in Other Languages or Other Countries

MARTINE DEVILLE

Institut national d'études démographiques (INED), Paris, France

References[73]

Europe

Greece

Siampos Georges S., 1983. *Demographia*. Athens, Athens Graduate School of Economics and Business Science, 481 p.

Tsaouse D. G., 1986. *Koinonike demographia [Sociological Demography]*. Athens, Gutenberg, 221 p. (Koinoniologike Vivliotheke, no. 15.)

Hungary

Hooz Istvan, 1988. *Demografia*. Budapest, Tankonyvkiado, 319 p.

Valkovics Vladimir, 2001. *Demografia*. Budapest, Osiris Kiado, 415 p.

Norway

Dyrvik Stale, 1983. *Historisk demografi: ei innforing i metodane._[An Introduction to the Methods of Historical Demography]*. Bergen, Universitetsforlaget, 213 p.

Poland

Cieslak Marii, 1984. *Demografia: metody analizy i prognozowania. [Demography: Analysis and Projection Methods]*. Warsaw, Panstwowe Wydawnictwo Naukowe, 273 p.

Latuch Mikolaj, 1985. *Demografia spoleczno-ekonimiczna. [Social and Economic Demography]*. Warsaw, Panstwowe Wydawnictwo Ekonomiczne, 451 p. (Revised 2nd ed.)

Pressat Roland, 1966. *Analiza demograficzna*. Warsaw, Pantswove Wydawnictwo Naukowe, 523 p.

[73] This bibliography is not exhaustive and is only a partial outline.

Czech Republic

Pavlik Zdenek, Rychtarikova Jitka and Subrtova Alena, 1986. *Zaklady demografie. [The Principles of Demography]*. Prague, Academia Praha, 732 p.

Romania

Pescaru Alexandru, 1968. *Elemete de demografie*. Bucharest, Editura stiinfica, 216 p.

Pressat Roland, 1974. *Analiza demografica*. Bucharest, Edutira stiintifica, 320 p.

Tarca Mihai, 1997. *Demografie: informatie, metode, analiza, prognoza*. s.l., Editura Economica, 352 p.

Trebici Vladimir, 1982. *Ce este demografia? [What is Demography?]*. Bucharest, Editura Stiintifica si Enciclopedica, 130 p. (Stiinta Pentru Toti, no. 186.)

Trebici Vladimir, 1979. *Demografia*. Bucharest, Editura stiinfica i encyclopedia, 781 p.

Slovenia

Malacid Janez, 1993. *Demografija: teorija, analiza, metode in modeli*. Ljubljana, Ekonomska fakultete, 366 p. (Zbirka Maksime.)

Asia

China

Liu Zheng, Wu Cangping and Cha Shuichuan, 1983. *[Population statistics]*. Peking, Chinese People's University, 406 p. (College Teaching Materials in the Humanities.)

Tsai Hong-Chin and Liao Cailian, 1987. *Renkou xue [Principles of demography]*. Taipei (Taiwan), Chu Liu Book Company, 429 p.

India

Premi Mahendra K., Ramanamma Adimadhyam and Bambawale Usha, 1983. *An introduction to social demography*. New Delhi, Vikas Publishing House Pvt, 214 p.

Ghosh B. N., 1985. *Fundamentals of population geography*. New Delhi, Sterling Publishers Private, 252 p.

Chandna R. C., 1986. *Geography of population: concepts, determinants and patterns*. New Delhi, Kalyani, 243 p.

Mahadevan Kuttan and Krishnan Parameswara (eds.), 1993. *Methodology for population studies and development*. New Delhi and London, Newbury Park and Sage, 470 p.

Mandal Ram Bahadur, Uyanga Joseph and Prasad Hanuman, 1989. *Introductory methods in population analysis*. New Delhi, Concept Publishing, 267 p.

Iran

Amani Mehdi, 1964. *[Méthodes d'analyse démographique]*. Tehran, Institute of Social Studies and Research, University of Tehran, 128 p.

Amani Mehdi, 1973. *Principes de la démographie / Principles of demography*. Tehran, Faculty of Social Sciences and Comparative Studies, University of Tehran, 124 p.

Africa

Cameroon

Brouard Nicolas, 1989. *Mouvements et modèles de population*. Yaoundé, IFORD, 184 p. (Les Documents pédagogiques de l'Iford.)

Algeria

Des Forts Jacqueline, 1990. *Initiation à la démographie dans le contexte algérien*. Oran, Direction des Activités Pédagogiques et médicales, CHU, 153 p.

Ghana

Nuama Nicholas N., 1994. *Statistical and demographic measures for population studies*. Legon, RIPS, 266 p.

Venkatacharya K., 1994. *Elements of mathematics for demographers*. Legon, RIPS, III + 184 p. (RIPS Monograph Series, no. 9.)

Morocco

Rachidi Mohamed, 1992. *Démographie: analyse, pratique et politiques. 1, Analyse*. Casablanca, Impr. Najah El Jadida, 136 p.

CERED, 1994. *Manuel d'analyse démographique: guide pratique*. Rabat, CERED, 282 p.

Mali

Ignegongba Keumaye, 1996. *Manuel d'initiation à la démographie*. Bamako, CERPOD, 149 p.

Togo

RIADEP, 1994. *Questions de population pour l'Afrique*. Lomé, Coordination générale du RIADEP, Dossiers pédagogiques du RIADEP, no. 1.

Tunisia

Seklani Mahmoud, 1999. *Traité d'analyse démographique approfondie. 1, Sources et ajustements des données, instruments de mesure, mortalité, survie, charge de morbidité globale*. Tunis, Centre de publication universitaire, 354 p. (Sciences économiques.)

Australia

Lucas David and Meyer Paul (eds.), 1994. *Beginning population studies*, 2nd ed. Canberra, Australian National University, National Centre for Development Studies, 199 p. (1st edition, 1984.)

C H A P T E R

140

Demography Teaching and Research

DIONISIA MAFFIOLI

Università degli Studi di Bari, Bari, Italy

I. THE CHALLENGES TO BE MET

Teaching and research in demography must be closely linked. Knowledge transmission must be regularly updated as progress is achieved in the discipline. If training is meant to prepare the students for ever-renewed social demand, particularly through evolution of the labor market, scientific research must contribute to providing the instruments to meet that objective. It is essential to understand whether, in concrete terms, those principles are actually applied, and how. Several significant initiatives have been taken in the past few decades, aimed at appraising the teaching of demography across the world, the latest of which was conducted by a Union for the Scientific Study of Population (IUSSP)[1] working group. This made available a large mass of updated information, the need for which had been acutely felt. However, it remains difficult to pass judgment on such a complex, multifarious, or even contradictory situation, which varies from one country to another due to differences in the background and the organization of university training, as well as demands on the labor market. The following sections discuss certain points that, at the international level, seem crucial for the development of the discipline in terms of scientific acquisition and academic strength and positioning; the consequences for demographic and population training, in view of the various teaching objectives, are also examined.

Demographers have good reasons for feeling gratified with the way their discipline, although comparatively recent, has progressed. While its objectives and frontiers could—and should—be questioned, the identity of the discipline, as far as its central core is concerned, is well defined and characterized by strong elements, the pride of all demographers—a sound methodologic system, specifically designed for the needs of understanding demographic systems, and which has grown more sophisticated and consolidated in the past few decades; the empirical approach and the refusal to build or accept theories unless they are based on objective observations on whole populations or statistically representative samples; the numeric measurements that can be statistically processed and make it possible to formulate mathematical models that can explicitly and clearly represent how events are linked between themselves and to their causes, as opposed to qualitative descriptions (sometimes perceived as vague and potentially contradictory); the abundance of available data derived from official sources and large-scale surveys, which makes it easier to meet quantification needs. The very first aim of demography is to describe "demographic systems,"[2] along with their changes in time and space and to analyze the ways in which the various components of a system relate to one another and interact. In many cases, particularly in the past, description was the main core in the work of many demographers, but it

[1] The work of the Group, which was coordinated by Graziella Caselli, culminated in the Demographic training in the Third Millennium seminar, which was held in Rabat, Morocco, on May 15 to 18, 2001, under the aegis of the IUSSP. For information on earlier initiatives, see Chapter 142.

[2] "Il sistema demografico" (Livi Bacci, 1994, p. 11).

was never perceived to be a sufficient or exclusive objective. As shown by the definitions of the discipline, expressed at various periods in time by different authors with different scientific priorities, demography has always aimed to identify the "general principles that determine the evolution of populations,"[3] highlighting the cause and effect relations between demographic dynamics and social dynamics in a broad sense. In other words, the aim is to understand the *causes* of the behaviors and processes that belong to its field of study, and the *consequences* of population dynamics on all other aspects of social life. In its quest for causes and consequences, the approach of demographic research is typically inductive—based on measurable facts that have been verified as rigorously as possible, to build constructions and theories that, once formulated, must always be verified again, through the appropriate statistical procedures, so that their degree of generality can be established. All that is part of the *genetic code* of demography and a strong feature of identity, not only for the discipline, but also for demographers, inasmuch as they are really direct heirs to Comte's positivism, as stated by John Caldwell (1996, p 311). As a matter of fact, as the same author writes (1996, p. 310), demographers are usually intolerant about the practices of other social disciplines that, in their opinion, "seem to build one unsubstantiated assumption upon another, . . . which can be tied to the observable world indeed without foundation which can be in some way measured or shown to be real and have ascertainable magnitude" (Caldwell, 1996, p. 310).

However, criticism of demography may also stem from the very features that make its strength. The same tradition of quantitative construction of reality, which contributes so much to clarity of observation and constructions, has limits because, as Guillaume Wunsch (1987) remarked, "while facts can generally be expressed in numbers, ideas cannot"[4]; however, ideas are quite often an important factor in demographic changes. Actually expressing *facts* in numbers implies a reduction of meanings and the risks of *variabilization* (Micheli, 2000) (i.e., of restricting construction of realities only to those aspects that can be processed along demographic canons). An urge is widespread to ignore the confines between disciplines—appropriate in university organization, but not required by scientific research—and to bring together all the abilities necessary to analyze complex phenomena. There are causes for frustration and uneasiness, which are related to the (in)adequacy of the tools and approaches commonly used in demographic research, if the latter, rather than simply describing demographic systems, is to address the crucial nodes of the causes and consequences of the changes and evolution that occur in the systems. The matter becomes all the more serious if the results of research, beyond the interest in scientific progress, also is meant to have operational scope by providing social policies with the information needed. This seems inevitable in view of Guillaume Wunsch's statement (1987, p. 411) that "demographers, any more than other scientists, cannot remain aloof from the social usefulness of the results of their research."[5] The *causes* and *consequences* of demographic processes are not (only) demographic: the causes are rooted in societies, cultures, economies, histories; the consequences are widely spread across the various sectors of the social system. Studying them consequently implies extending the field of research to new issues, on the borders of other social sciences that are also focused on humans and their activities. This implies probing into the very roots of human behavior reconstructing decision-making processes and linking them to the context in which they were shaped. This means that demographers must adopt contributions in knowledge, awareness, or tools that were developed within other disciplines or invent new ones for themselves.

An aspiration to examine relations between societies and demographic systems has always been discernible in population studies. That is where the demographic transition theory, the main paradigm in the interpretation of modern demographic upheavals, originated. However, despite many significant exceptions, that aspiration had largely remained theoretical, and hardly materialized in the daily research work of population specialists. But it seems to have conquered now a status as a priority requirement, to avoid impoverishment and fruitlessness of the discipline. In the last ten or fifteen years, many authors gave evidence about that phase in the history of the development of demography. A few are mentioned, although no full list of the most significant contributions can be offered here. Massimo Livi Bacci (1994, p. 11), describing the progress of demography, has shown how in the past few decades most of the progress achieved was "on the construction of a formal structure of how the

[3] "Principi generali che reggono l'evoluzione delle popolazioni" (Savorgnan, 1936).

[4] "Se i fatti possono essere generalmente tradotti in numeri, altrettanto non si può fare per le idee" (Wunsch, 1987).

[5] "Il demografo, come ogni scienziato, non può disinteressarsi dell'utilità sociale dei risultati delle sue ricerche" (Wunsch, 1987, p. 411). The statement was made by Guillaume Wunsch in Rome in 1985 during a meeting held in honour of Nora Federici.

demographic system works"[6] and how, although further improvement remains desirable, the time has come for focusing on the research of the causes of phenomena variability, which is now possible because of the extraordinary increase in available data, particularly at a micro level. According to John Caldwell (1996), "It is appropriate to return to theory, because the ability to construct theory from empirical data, and then to test its potential for correctly predicting future or past but unexamined events, is the hallmark of both social and other sciences" (p. 332). Giuseppe Micheli (1997) suggests that views be widened and new roads explored, asserting that "no further progress could be achieved in understanding complex phenomena such as demographic processes unless we give up that oversimplifying method that consists in solely quantitative verification and sterile, if reassuring, reference to ourselves."[7] Bart de Bruijn (1999, p. 1) was referring to Frans Willekens and Guillaume Wunsch when he wrote, "Although the mathematical interest is a distinctive and indeed an indispensable feature of demography, it is not sufficient for a mature science capable of interpreting, understanding, forecasting demographic reality; even less for conceiving conditions for behavioral intervention strategies or policies related to population change. Such capabilities can only emerge from a body of explanatory theory: a set of interrelated concept and propositions that specify relevant entities and events, the relations between them, and the underlying causal mechanisms" (de Bruijn, 1999, p. 1). Paolo De Sandre (2000, p. 10) observed a trend toward "interpretative and explanatory objectives superseding descriptive ones,"[8] whereas Graziella Caselli (2000, p. 142) sees new developments of the discipline "in the specialization phase being outgrown and the integration phase emerging. The field for investigation has been significantly widened and multidisciplinary research is now offering new, exciting experiences."[9] Many more examples could be quoted, showing that such a view is widely shared.

[6] "La costruzione di un'architettura formale del funzionamento del sistema demografico" (Livi Bacci, 1994, p. 11).

[7] "Passi in avanti nella comprensione di fenomeni complessi come i processi demografici potranno esser fatti solo abbandonando la strada di troppo semplificate verifiche empiriche di tipo esclusivamente quantitativo e di rassicuranti ma sterili certezze autoreferenziali" (Micheli, 1997).

[8] "Recupero degli obiettivi di interpretazione e spiegazione su quelli descrittivi" (De Sandre, 2000, p. 10).

[9] "Nel superamento di una fase di specializzazione e nel delinearsi di una fase di integrazione. In effetti il campo di indagine si dilata notevolmente e la ricerca multidisciplinare propone nuove interessanti esperienze" (Caselli, 2000, p. 142).

The effort to reach the objective of comprehension, although mainly motivated by the aim of knowledge, also supports another fairly common aspiration among population specialists—to help policy-makers and public opinion reach better understanding of how useful demographic research can be for good governance, considering the multiple interrelations identified between the population dynamic and the evolution of society. Among other things, clearer perception of the social usefulness of demography must have positive repercussions on the demand for demography qualifications on the labor market and therefore on the ability of the university structures to attract students and on the academic position of the discipline, which would be reinforced. Making more room in academic training and in research for the creation and development of such competence as may be required in the productive sector may launch such a process, although it must be admitted that in many countries there is little hope of sensitizing the business world, as is the case in the North American context. Although the place of demography in the context of university curricula may seem a minor academic concern, it is crucial for the very survival of population studies. As Frans Willekens emphasized (referring to the European case, but his remarks are valid everywhere), the rather marginal position of demography in departments or structures that have more general training objectives becomes dangerous when, as is currently the case in many European countries, competition to attract more students and resources leads to refocusing every training program onto what is considered its central core, which usually is not demography: "Demography, being small in scale but great in scope, is often situated in university departments that do not consider demography their core business. . . . To boost productivity and the competitive edge, departments retract to what they consider is their core business. Demography is often the victim of that process because the field has been unable to demonstrate that either it is part of the core business of the department in which it is located or has a core business of its own" (Willekens, 2001). It seems desirable to pursue the strategy of interdisciplinary collaboration that, besides promoting scientific understanding of complex phenomena, remains the best way of rousing more interest in demography and its scientific perspectives.

The main challenges, which are partly interrelated, that demography needs to meet can be summarized in three points: develop the capabilities of understanding demographic phenomena; develop the capacity to participate in *social management*; and develop the expertise that can be used in the productive sector. Meeting those challenges requires in-depth re-thinking about teaching orientations (methods and contents) and about the potential target of demographic training supply.

II. UNDERSTANDING DEMOGRAPHIC BEHAVIORS

The first point refers to a widespread need that has been asserted by many authors in the past 15 or 20 years: the need to find the means of shifting from the *description* of demographic phenomena (which are traditionally placed at macro level) to *understanding* demographic behaviors and hence population dynamics. On the contemporary scene the need to understand has clearly influenced the evolution of concepts, observation methods, theoretical approaches and methodologic tools.

Understanding (or accounting for) demographic behaviors means that individuals, their motivations, and the decisions they take along their life course are taken into consideration. This requires a much greater wealth of information than can be provided by official statistics, on which traditional demographic analysis has been based. Individual decisions and behaviors, even though they imply various degrees of freedom and are based on psychological preferences, must be placed within the contexts in which they were generated. Contexts are extremely difficult to delineate and embrace in all their complexity, because they are made up of features such as institutions, standards, values, and economic constraints. Eventually, it becomes necessary to see inside the mechanisms through which ideas are transmitted, values are disseminated, and cultural and ideologic changes occur. Demographers soon find themselves stepping into the areas of other disciplines, adopting research strategies and analysis tools typical of other scientific fields. In some cases or to a certain extent, it also becomes useful to leave the familiar, safe field of quantitative analysis and take up the tools necessary to process qualitative information.

However, it must be admitted that significant progress has been achieved in the past few years toward *understanding* demographic phenomena. The possibilities to understand behaviors (at individual level) and processes (at aggregate level) have fast improved, because of the extraordinary accumulation of new data (particularly those derived from retrospective surveys at individual level, which made it possible to study aspects beyond the scope of official statistics, such as attitudes, motivations, personal behaviors), and of the huge improvement of computer processing capacities. Both factors, but especially the first one, were highly stimulated by the researchers' determination to systematically explore new research paths. Thanks to the availability of individual data, demographic analysis is no more confined to the macro level, where observation is largely supplied by data from official statistics.[10] In particular, reliable individual data are available on reproductive behavior and the family; such information is largely comparable among many countries, including developing ones, and makes it possible to go far beyond the macro level of usual demographic analysis and get access into the field of individual behaviors, sometimes even in a longitudinal perspective. The longitudinal observation of demographic events and processes, through which analysis can be framed round the notion of life course, represents the emergence of a high quality tool, which is sometimes hailed as a revolutionary one, because it affords an opportunity to comprehend the succession of facts in their chronologic order, which is essential information for approaching causal processes.

Other, equally significant progresses have been possible thanks to the advancement in statistical sciences and the development of a large array of new statistical methods. These were suitable for processing large masses of data and individual variables, both with merely descriptive purposes and with the aim to interpret relations and detect causalities, in a statistical sense, through the particular way in which information is organized or modeled (Coppi and Pinnelli, 1990): multivariate analysis (see Chapters 132 and 133), event history analysis (see Chapter 23), path analysis and causal modeling (see Chapter 130), contextual analysis, multilevel analysis, and latent structure analysis (see Chapter 24).

In short, the discipline has experienced an evolution along a general axis, moving from the macro-period to the micro-longitudinal (according to Antonio Santini's diagram, 1999) through two phases, a macro-longitudinal and a micro-period one, which chronologically more or less coincided. This substantially improved the capacity to explain demographic changes, by taking into consideration the behaviors and motivations of individuals. However, a real causal explanation is still out of reach,[11] because as it is appearing ever more clearly, it would also be necessary to understand the mechanisms through which the environment (in a broad sense and at all levels) influences individual decisions, and how individual behavior and decisions

[10] It is precisely the wealth of information that may have caused "overdevelopment of observation . . . at the expense of theoretical thinking and analysis" at a certain phase of the development of that discipline (Santini, 1999, p. 5).

[11] The suggestion (Billari and Rosina, 1999) that the event history analysis should be used in that way also deserves mention.

then combine into a collective result, in a circular, ever renewed process; in other words, it is necessary to understand how "interaction between the individual sphere and social transformation processes works" (Salvini and Santini, 1999, p. 54).

As they grew aware of the need for causal research, demographers were confronted with the limits of their tools of analysis. This difficulty, which is inherent to the cultural training of most population specialists,[12] urged them to seek for the methodologic and conceptual instrumentation necessary to explore causality and build interpretative theories of demographic processes which, in the past few decades, have been formulated, if not within other disciplines, at least with the contribution of knowledge drawn from other disciplines—to mention only Richard Easterlin's and Gary Becker's well-known demographic-economic theories or John Caldwell's theory on inter-generation flows, which is rooted in anthropologic conceptions (see Chapter 37).

The need for interdisciplinary cooperation stems from the very nature of the issues to be analyzed, the complexity of which cuts across disciplinary divisions that only reflect institutional needs. Surveys of actual problems require extremely varied abilities. Studies on the evolution of fertility imply a variety of knowledge—economic and econometric (when the economic factors related to individual behavior are to be brought out), sociologic (to understand the dissemination of the ideas and values that influence choice), and anthropologic (to explain the way the type of social organization determines specific normative responses in reproductive behaviors). That extensive base of knowledge must converge toward an explanatory framework:

> Marital fertility decline emphasizes the multiplicity of potential alternatives. At one time or another, every feature that is thought to distinguish societies with high from those with low fertility has been advocated as a cause. Most of the central ideas have at least superficial plausibility and some intuitive appeal. Moreover, they may be inextricably intertwined. Structural modernization and transmission of new ideas go hand with hand, and their impact on the economic role of children, on material aspirations, and acceptance of the principle of the birth control are likely to change in concert. As always in the social sciences, the separation of cause, from concomitant or consequence is difficult. Plausible post factum explanations for particular trends abound, but an understanding of causal mechanisms, such as would allow confident prediction is out of reach. Nevertheless, it is possible to confront these central ideas with the empirical mass of historical and contemporary research findings in an illuminating way" (Cleland and Wilson, 1987, p. 9).

Many more examples could be given by reviewing all the other areas of demographic studies. The study of mortality, and its corollary, morbidity, requires knowledge in health statistics, epidemiologic science, medicine, and biology. The study of various aspects of migration, excluding the measurement of flows and migrant characteristics, largely draws on economics and political science—even more so for subjects such as marriage, the family, the consequences of demographic aging, and the relations between population and development or between population and environment.

It appears that the results, as far as causal research and the production of interpretative theories on population dynamics are concerned, are still far from satisfactory. The thinking necessary to improve the situation must probably be done in several directions, seizing all the opportunities derived from the methodologic progress in the area of demography and statistics and "using the methods and survey protocols that the results of related disciplines may suggest, to break through the complexity and reach the multiform roots of demographic change."[13]

Most crucial is the exploration of possibilities of acquiring new abilities inspired from other disciplines, which could make it possible to build more powerful and relevant frameworks for interpretation of population dynamics. The capacity to manage nonquantitative information must also be improved, among other things, to reach better understanding of the part played by ideas in behavior changes, which can hardly be measured and expressed in terms of numbers. That extension of conceptual tools and survey methods, precisely because it would correspond to demographic issues, would not entail any distortion of demography, but it would provide the discipline with elements to reinforce its identity.

However, the issue of multidisciplinary competence is not easy to address. There are too many disciplines likely to provide contributions to demographic research (e.g., economics, sociology, anthropology, political science, epidemiologic science, medicine) for demographers to take up all the knowledge that may be useful to them. If researchers must allow the specific needs of their research to guide them in widening their abilities, the problem is in the design for training demographers and future researchers, particularly

[12] Those who, as John Caldwell wrote (1996, p. 319), seem "in their element when there are massive but limited data which must be used skillfully to throw light on a matter which really needs much more information to explain it fully."

[13] "Affiancando loro metodologie e protocolli di indagine che i risultati di discipline contigue possono suggerire, per 'ingabbiare' la complessità e le multiformi radici del mutamento demografico" (Micheli, 1997).

about the choice to be made concerning complementary training in other disciplines—which fields, with what divisions, and with what contents?

Providing such training, open onto the contributions of other disciplines while keeping its completeness in the specific methods of demographic analysis and the statistical knowledge, may help eliminate or at least reduce what has sometimes been denounced as the *intellectual isolation* of demography compared with other social sciences. This would have positive effects on population research and may contribute, through better communication with other researchers, to improving the prospects of demography in universities and the perception of its usefulness in developing social policies.

III. PARTICIPATING IN SOCIAL MANAGEMENT

Demographic changes are likely to have significant repercussions or even cause major upsets in social organization at local, country, or global levels. The dynamics of birth and mortality and the migration flows that partly result from them have triggered off mechanisms that are likely to produce deep changes in the structures of current societies and cultures. The unequal growth of the population in the various parts of the world and the large migration movements are affecting the cultural, political, and economic balance on a global scale. In developed countries, the decline of birth and death rates, as well as the resulting unprecedented aging of the population, are the preconditions of deep changes in the perception of the relations between generations and even in the concept of life and death. The first short-term repercussion is that the retirement system and the whole structure of social welfare are being questioned. In developing countries, although fertility rates are fast declining, the rapid population growth phase has not come to an end due to populations that are still comparatively young. The sustainable development issues resulting from that dynamic still remain to be addressed. The consequences of world population growth on the environment are still largely ignored, whether it is because of short-sightedness, the selfishness of the current generations, or faith in the future, although thoroughly unjustified. The problem cannot be removed by simply stating that environmental degradation results far more from the consumption level in rich countries than from the growth of the populations in the poor ones.

Beyond the major demographic changes and their large range of repercussions on social reality, the importance of knowing demographic trends at the local level is also evident, particularly in terms of numbers and age structures, knowledge of which are essential for sound administration (e.g., schools, hospitals, housing, labor). For all those reasons, it would be highly desirable for political decision-makers to better appreciate the competences of population experts and fully consider the findings of demographic research. However, the relationships between demography and policy-making have proved to be very difficult.

It is generally admitted among demographers (although much less so among policy-makers and the general public) that the sound management of a country's affairs requires political attention to the demographic situation, because it is necessary to find the means through which unwanted consequences of the population dynamic can be prevented or curtailed by making the necessary adjustments on the social level. Many also believe it is possible and useful to try (using permissible democratic means) to contain or redirect demographic dynamics to check undesirable trends—in practice, this would mostly consist of policies aimed at controlling or encouraging fertility and at controlling migration flows.

Social and demographic policies are admittedly never quite neutral, because they inevitably include a normative view of social relations. The extreme complexity of demographic causality often allows more or less ideologic interpretation of reality, and political interventions often are grounded on controversial theories. That is what happened when fertility control policies in developing countries enlisted international support, which was inspired by neo-malthusian conceptions on population and development relations that had never been fully proved, although it could have been justified on non-economic grounds. The delicate relationship with policy-making—which demography has in common with other disciplines such as economics—is made still more complicated by the bad reputation of demography in some countries, where in the past it became more or less involved with racist and nationalistic policies of fertility support. This induced many demographers to reject interventionism and instead engage in pure research, casting wary looks at population policies. Even in the best of cases and with the most honest purposes, decision-makers and population experts have to work in a minefield, striving to reconcile democratically made choices and efficient interventions, collective and individual interests, social justice and individual rights—squaring the circle.

Many examples could be cited to illustrate how difficult collaboration can be between population researchers and decision-makers (Colombo, 2000; Keyfitz, 1985). This problem reflects policy-makers' well-known reluctance to deal with unpopular issues on which consensus can hardly be secured and a set of reasons intrinsically related to differences about the ideas and objectives of decision-makers and demographers.

Political action is often demographically shortsighted because it considers only times far shorter than those needed for demographic processes. It is very difficult for a politician who wants to be re-elected in the short term to implement unpopular measures that would bring no perceptible benefit at individual or family level until decades had passed, whereas the sacrifices to be made would be immediate. The painful process of reforming pension systems is a good illustration of this problem.

The problem partly originates in a certain policymakers' cultural imperviousness to population issues, which can probably be accounted for by lack of training and information. Demographers must never neglect the popularization and dissemination of basic demographic culture through school curricula. At the university level, specific attention should be focused on the integration of population issues into the various channels of future administrators' training.

There are causes other than blindness, opportunism, and lack of cultural sensibility to politicians' difficult relations with demographers. The perspectives and objectives of politicians and scientists are very different. Scientists seek knowledge and may focus their research on issues of no immediate practical interest. Freedom in research is vital and is probably the best guarantee of the social usefulness of research in the long term. This does not preclude that some research topics may be of operational interest. However, serious research takes time, and the findings are often uncertain and provisional, whereas action requires certainties, or at least convictions, and short times. Researchers and political decision-makers have their own reasons. A demographer who has decided to study an issue that is important for social life must gather the necessary statistical documentation, design an appropriate methodology, develop indicators, analyze the results, and draw conclusions. This may take a long time, and the conclusions may be assumptions that require further verification. This is thoroughly normal because the truth about demographic behaviors and population processes, barely detected in the jumble of interactive relations that surround and condition it, is never simple, unambiguous, or conclusive. However, action requires safe knowledge to be made available at the dates imposed by political circumstances, which are usually much earlier than it is possible for scientific investigation. Otherwise, decisions would be taken (inaction itself is a form of decision) regardless of the demographic knowledge or by using it very badly, based on opportunity considerations that are sometimes quite foreign to the elements of scientific knowledge. Such considerations inevitably also have ideologic and ethic bases. This is no reason for demographers to shut themselves away in their ivory tower and refrain from expressing their opinions. In such cases, the issue of clarity and intellectual honesty must be addressed, as emphasized by Paolo De Sandre (2000) who, remarking that "as the means used for research and action are not akin or logically linked enough for them to be automatically integrated,"[14] recommended that "sharp attention be paid so that no ideologic or political principles seep into the scientific field, or "acquire scientific value"[15] (De Sandre, 1994, p. 453), to avoid a disturbing confusion of the different points of view and criteria of judgment.

However, it should also be remarked that opinions are not always optimistic about the possibilities of efficiently contributing to the solution to social problems. Among the most skeptic authors is Nathan Keyfitz, who mentioned several convincing cases of failed or even counterproductive policies to illustrate his point: "As a result, our discipline, demography, does not have enough means to account for the past, anticipate the future or imagine the choice and lines of action likely to ward off unpredictable, pernicious effects"[16] (Keyfitz, 1987, p. 399). A few years later, in a long article published in *Population Studies* (1996), Keyfitz explained his opinion as follows: "When social scientists make policy recommendations they often forget that life in society is more complex than any of the disciplines that profess to interpret it and that offer policy recommendation based on their interpretation. Legislation recommended by economists and other social scientists and passed by governments with the best intentions can lead to results that are the very opposite of those aimed at" (Keyfitz, 1996). He concluded: "I cannot provide a constructive answer in such cases. What I do know is that our knowledge as well as our

[14] "Le strade della ricerca e dell'azione non sono così contigue e consequenziali da essere automaticamente integrabili" (De Sandre, 2000).

[15] "La vigilanza affinché non si abbia un'intrusione surrettizia di principi di preferenza ideologico-politica nell'ambito scientifico o, viceversa, non si dia valenza scientifica indebita ai principi di preferenza" (De Sandre, 1994, p. 453).

[16] "Risulta così che la nostra disciplina, la demografia, non ha mezzi sufficienti per spiegare il passato, per anticipare il futuro e per escogitare scelte e linee d'azione in grado di evitare effetti imprevisti e pervers i" (Nathan Keyfitz, 1987, p. 399).

powers to do good are bounded and that moral dilemmas exist we cannot buy our ways off" (Keyfitz, 1996).

Such remarks, however pessimistic, seem truthful and convincing. Still, they simply suggest that there is a limit to human knowledge in every field, particularly in social sciences. They call on us to be careful by emphasizing how uncertain and provisional our knowledge is, but I do not believe that they provide enough arguments for us to neglect our duty to contribute to social policy as best as we know. Above all, they remind us that we should never forget that doing that duty necessarily implies a trade-off between the disadvantages of inaction and those of action based on imperfect knowledge. If necessary, the unfortunate experiences described by Nathan Keyfitz also remind us that it is necessary to invest more in designing procedures to assess the results of policies, so that their implementation can be monitored and gradually reshaped in a continuous process of adjusting them to the real necessities.

The aim to participate in social management has important fallbacks on the orientation of research, because in such a perspective, it is essential to take causal processes into consideration. This further stresses the necessity of progressing in identifying the causes and consequences of demographic phenomena. This also provides further motivation for paying attention to the dissemination of demographic research findings. Concerning teaching aspects, the objective of social participation requires a sound orientation and particular contents for the training courses for future demographers, who must get ready to contribute to that endeavor with scientific and professional abilities at various levels. In particular, for the needs of social policies, not only the *causes* of population dynamics should be taken into consideration, but also their *consequences*. This is still an issue of causal research, although presented in opposite terms. In a broader sense, all kinds of interactions between the individual sphere and social changes are of concern, and hence the teaching of the methods of integrating micro data into their context deserve a place of primary importance in the training programs. For issues that are of interest for social policies, although research on very specific subjects may sometimes prove useful, what is mainly at stake is research with interdisciplinary borders on complex issues, such as gender and population, the relations between population, development, poverty, resources and environment, population aging and the welfare state, mortality by cause and morbidity, migration issues, and forecasting techniques.

Eventually, although communication with policymakers and managers usually requires reorientation of research toward the current issues and sometimes demands results within shorter periods than are suitable for research, the problem is not insurmountable. Research must be carried out between the two opposite ends of a continuum, skirting the perils of either—that of the ivory tower, where it is perfectly free but in danger of being sterile, and that of applied research, which may become an obstacle to real progress of the discipline if too closely linked to current issues.

IV. DEVELOPING ABILITIES OF INTEREST FOR BUSINESS

The market evolution, the enlarging sizes of business firms, and increasing complexity of their environment make the need for statistical and demographic information, beyond the traditional economic and financial data, more pressing, in order to design more efficient production strategies and policies of business. The idea is to get to know the characteristics and dynamics of the potential pools of customers for their products and the possible means to prepare to seize opportunities or ward off possible setbacks. It is also important to keep informed on the evolution of the labor force in quantitative and qualitative terms for better management of staff, the human capital of businesses. Demographic methods (e.g., analyses by cohort, life tables) have been offered and disseminated for a sound management of the personnel of various categories or levels. Similar methods, derived from demographic research, can be useful to address other problems in business, such as those concerning assessment of the quality and longevity of products. This type of approach is of primary interest to big business, especially to firms with large numbers of workers and operating in large markets. However, they seldom show enough interest in such approaches to recruit experts in demography for permanent positions, and they instead usually call on external consultants.

The set of abilities useful for business is usually gathered under the concept of "business demography" and is part of a particular branch of demography called *applied demography*. The latter, according to the pragmatic definition by one of its most eminent supporters (Siegel, 2002),

> explores the practical issues with which demographers working for government agencies, private non profit organizations, and businesses deal. . . . Applied demographers may be called on to advise in planning the operations of local governments and businesses to the degree that demographic skills, knowledge, and perspectives are viewed as useful for such planning (p. xvii–xviii). . . . Accordingly we can characterize applied demography as a decision-oriented science concerned with aiding managers, administrators, and government officials in making practical decisions congruent with the policy, or goals of the organization, or of the institution for which they work (Siegel, 2002, p. 2).

In the area of applied demography, as well as that of business demography, the American experience remains most advanced; it is even unique in the world. In the United States, the demand for applied demographers is such that (Casacchia and Racioppi, 1992, p. 3) this branch of the discipline has become a kind of private industry, the activity of which consists of providing other lines of business with its abilities and products. The activity of the sector is also intense in scientific research and academic training and includes the setting up of scientific committees within the prestigious *Population Association of America*, organizing the biennial *North American Conference on Applied Demography*, producing *ad hoc* academic texts, establishing targeted programs in the major universities, publishing various specialized journals, including *American Demographics*, to which thousands of firms have subscribed.

In many other national contexts, the current conditions of the markets do not seem to encourage the emergence of a demand in business demographic abilities comparable with the American case, but knowledge and competences in business demography must undoubtedly be expanded, as part of a demographer's training, so that all opportunities for employment can be seized. Moreover, it is important to develop information programs in order to sensitize the productive sector on the abilities that will then be available. Concerning the field of public administration-oriented applied demography, I previously mentioned the themes that should be developed and showed the necessity to cover that field in academic training and in research.

V. TEACHING DEMOGRAPHY

To highlight the teaching consequences of the challenges that demography as a discipline must meet, an overview of the main characteristics of the current organization of population studies, as far as research and teaching are concerned, is given. The form and contents of the teaching at different levels depend on the present situation, which therefore must be taken into consideration, if only to question it.

1. A Brief Survey of the Current Situation

From the information available in the literature,[17] it appears that universities are the main institutions for demographic research and teaching. Although there are also external institutions and bodies of all kinds that cover a wide range of scientific and teaching needs, the training courses offered by extra-university institutions are usually short ones, and they deal with specific subjects—mostly as part of summer schools or retraining courses.

The organization and objectives of demographic training are closely related to the position of demographic research and teaching centers within universities. There are very few courses of study that are wholly on demography, unlike sociology or economics. It is widely agreed that the demand for demographers' specific professional abilities is too low for them to constitute the core business of any undergraduate university degree. As a result, the teaching of demography, particularly in universities[18] of the developed countries, aims at providing highly specialized professional training (i.e., graduate or postgraduate' level) or supplementing broader training curricula. In other words, the training is intended first to produce future researchers and teachers and second to impress awareness of population issues on the general culture of other professionals who are likely to work in one of the many fields in which the influence of demographic processes is to be felt. However, it seems that professional training for demography specialists, which was rare until recently, is being consolidated, as exemplified by the United States, where demand from the market exists, and by developing countries, although with different characteristics and objectives, where intermediate-level skills are in demand for data collection and management; techniques in the use of imperfect statistics; and for the implementation, monitoring, and evaluation of population policies.

This situation affects the place and status of academic centers dealing with demography, and it conditions their organization and, to a large extent, the orientation of their research. The departments and other autonomous university institutions that deal exclusively with the study of population issues are comparatively few. It is much more common to have the discipline nurtured and taught by researchers who belong to other departments—sociology, statistics, economics, geography, history, epidemiology, genetics, and others, depending on the alliances between disciplines that were formed and consolidated in the history of the different countries or universities.

However, the current world population trends, their influence on public opinions and political decision-makers, particularly through the action of interna-

[17] The contents of this paragraph are based on a general review of the literature conducted a few years ago (Maffioli, 1997) and updated for the Rabat Seminar on Demographic Training in the Third Millennium (May 15–18, 2001).

[18] Jane Menken (2001) emphasizes the existence of a few demographic courses leading to undergraduate degrees in the United States.

tional agencies, and the resulting wealth of studies and writings have given prestige to many demographic research and teaching institutions. The latter consequently acquired significant independence, particularly in universities whose organization is most flexible. The number of demography departments kept increasing, and many population centers, although they were only subdivisions within larger autonomous structures, then became able to issue university degrees at graduate or postgraduate level.

Besides the traditional forms of demographic teaching, a more flexible type of training is fast emerging, with a variety of courses to offer for different university degrees, and it is intended to meet very diversified needs. Undergraduate university training is being offered in some developing countries, although with rather variable curricula and contents, often depending on the teachers' interests and abilities. Short courses are also being developed; they do not necessarily lead to formal university degrees but aim at providing persons who are already involved in government activities or planning with abilities in various areas of population policy management.

2. Diversifying the Courses Offered

Teaching must keep in line with the developments and acquisitions of the discipline, but if it is also to meet the demands emerging from society, particularly from the labor market, it must also examine the profiles of the students that it is meant for, so that the methods and contents can be targeted more accurately, in function of their occupational objectives. I think that the main training needs, involving currently existing or emerging demographic knowledge, with more or less wide prospects and comparatively good chances of consolidation, could be summarized in four groups:

Training staff for scientific research and teaching
Training population policy specialists
Training professional demographers
Training policy-makers or administrators to make them aware of population issues (with demography as a training supplement)

Each of these points deserves considerations in terms of their characteristics, methods, and the main contents of demography teaching and how research can contribute to improving curricula. However, it is impossible to be specific about their concrete implementation, because the organization of courses and teaching programs is too closely related to the multiform structures of school and university systems in the different countries and to the demands on the labor market, which are exceedingly different from one country to another.

3. Training Future Researchers and Teachers

Future researchers or teachers, who are bound to guarantee the development of the discipline and continuity in the faculties, should be given comprehensive training at doctorate level. Although they may initially come from various backgrounds, doctorate students should be required to become perfectly conversant with the conceptual perspectives, methodologic instruments, and concrete issues specific to basic demography, but they should have in-depth study only in the sector in which they plan to conduct their future research activities. The gradual segmentation of demography, the mass of knowledge accumulated on a very wide range of heterogeneous issues, and the enrichment of the array of methodologic tools that are increasingly specific to the themes to be addressed make it impossible to cover all demographic knowledge in a single doctorate course.

Such specialization is all the more necessary as a multidisciplinary approach correspond to the very nature of many population issues overlapping *(at the cutting edge)* with other disciplines, that are bound to emerge and account for the most significant future progess of demographic knowledge. So that it is highly desirable for doctorate students to get the conceptual and methodologic knowledge from one or several related disciplines, which would be identified in view of their research interests. Besides the doctorates related to some single sector of demography (e.g., economic demography, historical demography, social demography, anthropologic demography), which will still be necessary, there must also be doctorates on in-depth study of complex thematic cores (e.g., population and development; aging and society; population, health, and health systems), and doctorates organized on different overlapping disciplinary sectors (such is the case of the sociology and demography doctorate created at the University of Berkeley).

Despite the huge diversity of possible doctorate training proposals, a few common characteristics can be identified. One of them consists of sharing a knowledge package that would include the methodologic core of demographic phenomena analysis, as well as in-depth knowledge of the demographic processes that historically have had a strong influence on humankind and their relations to social reality in the broadest sense. Another common characteristic should be the requirement of sound statistical abilities, particularly those commonly used in a wide range of research situations, such as individual data analysis (i.e., multivariate analyses), implementation of the longitudinal approach (i.e. life course studies), or the study of causal nodes (i.e., latent structure analysis); others are useful mostly for particular curricula (i.e., techniques for evaluation of social policy effects, population policy interventions, and development policy adequateness). One of the main keys to identifying the statistical abilities that should be common knowledge to all demographic doctorates lies in the strategic objective of the exploration of causal nodes hindering the comprehension of the relations between population and social dynamics.

The specialized training needed for areas of overlap with various other disciplines (e.g., economics, sociology, anthropology, epidemiology) must be tailored to the specific research interests of the students. On no account must the teaching of elements of related sciences give mere generalities, with no practical impact on demographic research; it must be focused on issues that are of interest for investigation in the area of population studies.

Because of the training of future researchers, most decisive progress has been achieved in the past few years. According to the comprehensive review by Jane Menken (2001), in the major universities a set of highly desirable circumstances exist—a high degree of specialization on themes that require vast multidisciplinary abilities; training courses that meet an extremely diversified range of needs; and a high degree of integration with related disciplines on themes of common interest (including the creation of joint doctorates, such as in demography and sociology at Berkeley). This set of circumstances is likely to guarantee subsequent disciplinary developments and is the precondition for further improvement in the academic status of demography. The conditions and constraints imposed on research and teaching programs by donors' political and cultural orientations, although not necessarily an element of restraint to disciplinary development, establish a link with the needs of society. In her analysis, however, Jane Menken wonders if sufficient attention has been devoted to fields of study that attract students from developing countries to American universities, and she deplores the focus of teaching and research on subjects that reflect only the teachers' cultural interests, possibly rendering American demography "quite individualistic and even, perhaps, idiosyncratic" (Menken, 2001).

European universities (Willekens, 2001) are going through a special phase of uncertainty because of the ongoing general reorganization of courses and reduction of public funding. Many new doctorate programs have just been created, and it is too early to assess their chances of success. Moreover, the number of students has decreased, as a result of population trends, and there is fierce competition between the different courses of study. From the rather special viewpoint of Italian universities, however, it is possible to measure to what extent the academic training of younger researchers has been more specific and comprehensive on the level of disciplinary acquisition than the self-taught type of training of previous generations. Various factors about the organization of training courses have contributed to that evolution, including intensified—if still insufficient—relations with foreign universities, which helps make the different types of training more homogeneous, at least at European level. However, much still remains to be done—and no wonder, considering that the contents of the teaching change along with research acquisitions and the abilities available on the side of the teaching staff.

In developing countries, where material and human resources for the training of population specialists, although still largely insufficient, have substantially increased in the course of time, there are many excellent centers that have the capacity to train researchers, particularly in the field of development. The main causes for concern is related to continuity of the resources necessary to keep up the close international relations that promote teachers' and students' mobility and to ensure constant adequateness between training programs and needs, by gradually incorporating new scientific acquisitions, new tools of analysis, and updating interpretation frameworks in the wake of disciplinary evolution.

4. Training Population Policy Specialists

Training in population policy is a type of professional specialization consisting of abilities that are in high demand in the international environment or in the developing countries that are implementing population policies but that are neglected in traditional demographic university training. The themes of population policies, particularly in their relations to the themes of development, involve practical needs such as pedagogic ones that are not the sole concern of academics. On the contrary, the major international organizations (e.g., FAO, ILO, UNESCO, UNFPA, and other specialized United Nations agencies, as well as other institutions) have experience and abilities that are far superior to those of universities (even taking in account the frequent cases of collaboration between academics). The training that such organizations provide on those themes to their staff or to agents in developing countries is probably more specific and specialized, more soundly rooted in the experience of concrete situations, and more effectively transferable into operational contexts than that likely to be provided through university training.

It is highly desirable for universities to take greater interest in the training of such types of professionals at various levels, as well as in scientific research on the difficult themes of policy evaluation, which in certain aspects are akin to causal research. In that area, it would be profitable to strengthen collaboration with the world of international organizations, in that it would make scientific research more sensitive to the various aspects of social demand, while greatly reducing the risk of improvisation, which occurs when, for market-related reasons, agencies, nongovernmental organizations, or consultants work without well-established knowledge or skills.

Ideally, this type of professionals should be able to do the following:

- Design development and social policies incorporating the population variables (numbers, structures, geographic distribution) and reckoning the interrelations between demographic processes and social dynamics
- Formulate, implement, monitor, and evaluate population policies that may have a wide variety of contents
- Prepare sensitization campaigns and dissemination programs on issues concerning population dynamics

To meet these objectives, specialists in population issues and demographic policies must have varied and extensive abilities in different areas. Their basic preparation concerning demographic knowledge and statistics must be complete and combined with the knowledge of facts and methods from other disciplinary sectors. In particular, they must have skills in socioeconomic and anthropologic areas, although skills in epidemiology may also be useful, considering the importance of health policies. The methods of causal analysis and all the techniques useful in estimating the effects of policies must be given primary attention.

Training in that area may be conducted at very diversified levels, depending on the specific objectives of professional location, ranging from the doctorate level, which is necessary for anyone who intends to deal with policy formulation and evaluation systems development, to the Master's level or annual specialization courses for those who intend only to work in implementation, control, and monitoring, and on to the levels of short courses with specific, limited aims, courses meant for those already active in a particular sector and who wish to acquire, update or improve special abilities.

To complete this overview of the type of training useful in what I have referred to as participation of demographers in social management, I must mention that professional demographers who work in collaboration with public bodies and local administrations for better management of public affairs must be given the abilities specific to the sector of applied demography, which is discussed in the next section.

5. Training Professional Demographers

Training professional demographers is better justified in the countries where the nonacademic labor market offers more opportunities for this type of professional qualification. However, with the globalization of markets, the growing complexity of the business environment, and the increasing average size of firms, the demand for professional demographers is likely to expand. A similar evolution can be expected concerning public administrators. It is therefore advisable that universities become prepared to meet this emerging demand in research, by reinforcing the study areas relevant to applied or business demography, and in available training courses. An appropriate campaign to popularize the professional values in demography, conducted by defining purposes and reinforcing linkages with the world of work, could also help in meeting the demographers' claim to better consideration for population variables in many fields of action, ranging from political and administrative ones to productive activities.

In the current training systems, the professional aspects of demography learning do not always seem to be clearly identified or addressed in any coherent way. As a result of a lack of specificity about the professional profiles of nonacademic demographers, demography is virtually never the backbone of any training course leading to an autonomous undergraduate university degree. This occurs even in the United States, although Americans have long made a habit of making professional use of demographic skills. More than 15 years ago, Nathan Keyfitz stated in one of his analyses that in his country "using demographic planning and data is a constant practice in all planning activities, to build structures, sell products, provide public and private services"[19] and that "for demographers from American universities . . . most job opportunities come from that sector."[20] The situation is very

[19] "L'uso di proiezioni e dati demografici è prassi costante in tutte le attività di progettazione per costruire strutture, vendere prodotti, fornire servizi pubblici e privati" (Keyfitz, 1987, p. 384).

[20] "Per i demografi sfornati dalle università americane . . . l'occupazione in questo settore costituisce uno degli sbocchi principali" (Keyfitz, 1987, p. 384).

much the same in other countries, particularly developing ones, where the lack of undergraduate or graduate training courses in the sector is all the more acutely felt because they could provide precious abilities that are very much in demand in ministries and public administrations.

Training of professional demographers could include all three university levels, depending on the objectives of their professional insertion, but the second one (e.g., Master's, specialty degrees) is most crucial. The contents of the learning should be the same as those of applied demography, the scope of which is very large (see Siegel, 2001), ranging from production of demographic estimates and forecasts for small areas to migration forecasts, to applications of demography to marketing, and to family demography. Comprehensive manuals have been published at the international level that could be useful in defining study programs.

The training for professional demographers, only recently introduced in the university system and still not well established, deserves special attention. Even in the United States, where demographers' professional abilities are more in demand in the labor market and their training more widely provided for than in any other countries, there is heated debate about how adequate the training provided by the university systems is in that sector of studies.

> It is not clear to what extent many of the programs are emphasizing new skills that are demanded, especially for students who will enter the nonacademic world. One such skill is carrying out high quality surveys—from design to implementation to analysis (quantitative methods at the cutting edge for the social sciences, qualitative methods, and the complementarities between quantitative and qualitative approaches). Another is the evaluation of intervention programs. A third is demographic forecasting. For the social science programs, there is almost no mention of biology or disease or morbidity, or reproduction, and surprisingly little on migration or urbanization, forces that surely will shape much of population change in the coming decades" (Menken, 2001, p. 13).

6. Training Policy-Makers and Administrators for Awareness of Population Issues

University training in demography is sometimes very short and summarized in a single course or a limited number of credits, when it constitutes a mere supplement to studies that are meant to train specialists in other disciplines. This occurs quite frequently because demography often has marginal standing in departments of economics, sociology, or political science. We may wonder if any appropriate thinking has ever been done to define the contents of demography teaching in such courses, which sometimes are wrongly considered of minor importance.

The complementary demographic training for specialists in other sectors deserves peculiar attention, because it plays a crucial part in emphasizing the importance of demographic factors, which is often neglected, in the various aspects of human activity. It is essential in such cases to avoid giving an image of demography as autonomous and unconnected with the teaching and scientific context in which it dwells. If demography were to keep aloof from the students' cultural interests, it would become marginalized from the academic mainstream.

The transmission of demographic knowledge in other teaching contexts, although limited to a few classes, requires special efforts in embracing and adjusting to the issues peculiar to the studies in which it is inserted. While selecting the contents, priority should be given to those that highlight interdisciplinary linkages, emphasize the interest in a demographic key to the understanding of events, and provide methodologic tools that are useful to the students' vocational training. That approach alone can guarantee that demography is harmoniously incorporated into various teaching contexts and that its teaching potentialities are emphasized on a professional and cultural level.

To sensitize professionals, policy-makers, and administrators to population issues, short training courses could be organized in which the linkages with the specific aspects of the beneficiaries' activities would be emphasized. The range of topics may be very wide, but a typical case of short training courses in developing countries is on the relations between development and population policies. Such training courses, which are still few but quickly expanding, have been promoted and organized by non-university institutions, ranging from the major international organizations to nongovernmental organizations, which have been carrying on significant work (in quantitative and qualitative terms) and have acquired greater abilities in the area than exist in universities.

VI. PRE-UNIVERSITY TRAINING AND POPULARIZATION

Before closing these remarks on the teaching of demography, it seems appropriate to include a brief survey on whether a demographic component is included in primary and secondary school curricula.

In general, demography is not a subject matter in pre-university training,[21] any more than the other social sciences, which are virtually never included in programs that are traditionally established in conformity with an essentially humanistic conception of human sciences. However, considering the pedagogic value of knowledge of population issues, it would be advisable to make some space for them in school curricula.

It seems that the high ethical value of training in population issues is more easily recognized in countries where in the past few decades the need for population policies intended to affect individual behaviors has been felt to reconcile society's general objectives with those of families and individuals. As stated by African researcher Daniel Mpembele Sala Diakanda (1997, p. 11), the main objective of studying population issues in schools

> is not only to raise awareness . . . of the problems and challenges with which one's community and the world at large are faced (particularly those relating to social production and reproduction), but also to give rise to firm commitment to contributing, individually and collectively, as far as possible and within the limits of everyone's sphere of responsibility, to solving those problems. The interest for population and development issues in higher education (whether as a subject for specialization or a complement to other areas) and in all the spheres of society, including the higher level of decision taking, as well as the expected impact on society at large, will be all the greater as sound population education has been provided at the lower levels of the school systems.

While recognizing that issues that do have ideologic aspects are extremely sensitive, I believe that the students of primary and secondary schools should discover (with proper objectiveness) the major themes of global population explosion, responsible procreation, sustainable development, respect for the environment, gender issues, fall in birth rates, and aging of the population.

Although not wishing to give the discipline a dominant place in pre-university studies,[22] in which it would have to compete with the equally important contents of other disciplines, the introduction of population issues may take place through allied disciplines such as history, geography, or biology, whose position in school curricula is established by a centuries-old tradition. The modern disciplinary orientations of history and geography, the programs of natural sciences that often highlight the importance of populations, and their structures and dynamics in relation to the environment, available resources, and sustainable development, are a good prerequisite for insertion of demographic issues into school curricula. Moreover, classroom work ignoring the disciplinary borders may also allow a better use of the pedagogic potential of demography, because its object (population and its evolution) is suitable for studies requiring contributions from many disciplines, including human studies and natural sciences.

Little information is available on the teaching of demography at pre-university level. Although there are great differences from one country to another, it appears that the possibilities, however limited, offered by school curricula are not always fully exploited, because population issues are hardly included in the teachers' training, awareness, or cultural interests.[23] If the possibilities of partial transmission of demographic knowledge through other disciplines are to be properly tapped, demographers and population specialists must actively undertake to update the training of pre-university teachers and produce the relevant teaching documents.

The need to recognize the formative value of demography in school curricula is only one aspect of the larger question of sensitizing public opinion in general as the basis for accurate appraisal of the social consequences of individual demographic behaviors and the necessary presupposition for socially agreed and democratically based population policies (Sonnino, 1987). Scientific popularization, which is not much practiced by demographers, should receive greater attention through more frequent and polished use of mass media and through publication of books in which scientific accuracy would be combined with accessibility and gratifying reading for nonspecialist readership.

Acknowledgment

This chapter was translated from the French version by Mireille Rabenoro.

[21] There are exceptions, particularly in certain African countries.

[22] Concerning Italy, the area that is more familiar to me (Lombardo and Maffioli, 1985), the debate on the question of teaching demography in schools was particularly heated in the 1970s and 1980s, leading to substantial convergence of opinions on the appropriateness and methods of transmitting demographic knowledge through adequate inclusion of demography in history and social sciences teaching. The approach was selected as preferable and more realistic than independently introducing the discipline, which was thought difficult to implement.

[23] An analysis conducted more than 20 years ago on curricula and manuals (Lombardo and Maffioli, 1985) showed that on that point, the situation in France and Belgium was outstandingly good in many respects, probably as a consequence of the more firmly asserted historiographic orientation of the *Annales* and a tradition of advanced human geography, into which demographic contexts and themes can be inserted more easily.

References

Publications Cited in this Chapter

Billari Francisco C. and Rosina Alessandro, 1999. L'utilizzo dell'event history analysis a fini esplicativi in demografia. Per un approccio tramite meccanismi e variabili, *in*: Francisco C. Billari, Alberto Bonaguidi, Alessandro Rosina, Silvana Salvini, and Antonio Santini (eds.), *Quadri concettuali per la ricerca in demografia: quattro saggi*, p. 55–91. Florence, Dipartimento di Statistica G. "Parenti", Università degli Studi di Firenze, 101 p. (Ricerche Teoriche, no. 15.)

Caldwell John C., 1996. Demography and social sciences, *Population Studies*, vol. 50(2), p. 305–333.

Casacchia Oliviero and Racioppi Filomena, 1992. La demografia per il settore pubblico e privato nell'esperienza internazionale: contenuti e formazione, *in*: Società Italiana di Statistica (ed.), *Atti della XXXVI Riunione Scientifica della Società Italiana di Statistica*, p. 249–260. Rome, Società Italiana di Statistica, 519 p.

Caselli Graziella, 2000. Elaborazione teorica e sviluppi metodologici: gli aspetti rilevanti della demografia di fine millennio, *in*: Paolo De Sandre and Fausta Ongaro (ed.), *Demografia: presente e futuro*, p. 141–154. Padova, CLEUP, 174 p.

Cleland John and Wilson Christopher, 1987. Demand transitions of the fertility transition: an iconoclastic view, *Population Studies*, vol. 41(1), p. 5–30.

Colombo Bernardo, 2000. Demografia per conoscere e per intervenire, *in*: Paolo De Sandre and Fausta Ongaro (eds.), *Demografia: presente e futuro*, p. 89–94. Padova, CLEUP, 174 p.

Coppi Renato and Pinnelli Antonella, 1990. Trattamento dei dati individuali in demografia: una prospettiva di nuova collaborazione fra demografi e statistici, *in*: Società Italiana di Statistica (ed.), *Atti della XXXV Riunione scientifica della Società Italiana di Statistica, Padova, Vol. 1*, p. 37–59. Rome, Società Italiana di Statistica, 338 p.

De Bruijn Bart J., 1999. *Foundations of demographic theory: choice, process, context*. Amsterdam, Thela Thesis, 298 p.

De Sandre Paolo, 1994. Demografia, politica ed etica, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 452–482. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

De Sandre Paolo, 2000. *Lettera di collegamento del Gruppo di Coordinamento per la Demografia*. Padova, Università di Padova. (Unpublished.)

Keyfitz Nathan, 1987. La demografia accademica negli Stati Uniti: situazione attuale e prospettive, *in:* Eugenio Sonnino, Antonella Pinnelli, Dionisia Maffioli and Annunziata Nobile (ed.), *Demografia: scienza, insegnamento, professione*, p. 383–401. Rome and Milan, Dipartimento di Scienze Demografiche and Franco Angeli, 566 p.

Keyfitz Nathan, 1996. Population growth, development and the environment, *Population Studies*, vol. 50(3), p. 335–359.

Livi Bacci Massimo, 1994. La demografia, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 3–23. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.").

Livi Bacci Massimo, Blangiardo Gian Carlo, and Golini Antonio (ed.), 1994. *Demografia*. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Lombardo Enzo and Maffioli Dionisia, 1987. L'insegnamento preuniversitario della demografia. Alcune osservazioni, *in*: Eugenio Sonnino, Antonella Pinnnelli, Dionisia Maffioli, and Annunziata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 307–374. Rome and Milan, Dipartimento di Scienze Demografiche and Franco Angeli, 582 p.

Maffioli Dionisia, 1997. Centri di ricerca e di insegnamento della demografia, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Démographie: analyze et synthèse. Causes et conséquences des évolutions démographiques. Actes du Séminaire de San Miniato (Pisa)*, 7–19 December 1997, vol. 1, p. 291–310. Rome, Paris, Louvain-la-Neuve, Dipartimento di Scienze Demografiche, INED, SPED, 318 p.

Menken Jane, 2001. *Current training in the US: is it meeting the demand for demographers?* Paris, IUSSP/UIESP. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Micheli Giuseppe A., 1997. *Due sfide da raccogliere per gli studi di popolazione*. *Quaderni di Discussione del Gruppo di Coordinamento per la Demografia della Società Italiana di Statistica*. Rome, Società Italiana di Statistica, no. 1, 72 p.

Sala-Diakanda Daniel Mpembele, 1997. *L'enseignement des questions de population: progrès accomplis, nouvelles approches et méthodes*. Addis Abeba, United Nations, Economic Commission for Africa, Food Security and Sustainable Development Division, 18 p. (a paper read at the International Population Conference, session on "The teaching of demography: new approaches and methods", IUSSP, Beijing, 11–17 October 1997.)

Salvini Silvana and Santini Antonio, 1999. Dalle biografie alle coorti, dalle coorti alle biografie, *in*: Francisco C. Billari, Alberto Bonaguidi, Alessandro Rosina, Silvana Salvini, and Antonio Santini (eds.), *Quadri concettuali per la ricerca in demografia: quattro saggi*, p. 41–54. Florence, Dipartimento di Statistica G. "Parenti", Università degli Studi di Firenze, 101 p. (Ricerche teoriche, no. 15.)

Santini Antonio, 1999. Quadri concettuali per l'analisi demografica, *in*: Francisco C. Billari, Alberto Bonaguidi, Alessandro Rosina, Silvana Salvini, and Antonio Santini (eds.), *Quadri concettuali per la ricerca in demografia: quattro saggi*, p. 1–26. Florence, Dipartimento di Statistica G. "Parenti", Università degli Studi di Firenze, 101 p. (Ricerche Teoriche, no. 15.)

Savorgnan Franco, 1936. *Corso di Demografia*. Pisa, Nistri-Lischi, 226 p.

Siegel Jacob S., 2002. *Applied demography. Application to business, law, and public policy*. New York, Academic Press, 686 p.

Sonnino Eugenio, Pinnelli Antonella, Maffioli Dionisia, and Nobile Anunciata (ed.), 1987. *Demografia: scienza, insegnamento, professione*. Milan, F. Angeli, 574 p.

Willekens Frans, 2001. *Demographic training: Europe*. Groningen, University of Groningen, Population Research Centre, 32 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Wunsch Guillaume, 1987. L'insegnamento della demografia in Europa: dalla teoria all'azione, *in*: Eugenio Sonnino, Antonella Pinnelli, Dionysia Maffioli, and Anunciata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 383–401. Milan, F. Angeli, 574 p.

Other Useful Publications

Baldwin Stephen C., 2001. *The "Uganda approach" made easy: a universal guide to assessing current population and development training needs*. New York, Prefix Pre-dispute Solutions LLC, 34 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Billari Francisco C., Bonaguidi Alberto, Rosina Alessandro, Salvini Silvana, and Santini Antonio (eds.), 1999. *Quadri concettuali per la ricerca in demografia: quattro saggi*. Florence, Dipartimento di Statistica G. "Parenti,", Università degli Studi di Firenze, 101 p. (Ricerche Teoriche, no. 15.)

Blanc Ann K. and Ezeh Alex C., 2001. *Demographic training in Africa: past, present and future*. Nairobi, African Population and Health

Research Centre, 21 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Bonaguidi Alberto and Santini Antonio, 1999. Informazione statistica e ricostruzione dei processi demografici, *in*: Francisco C. Billari, Alberto Bonaguidi, Alessandro Rosina, Silvana Salvini, and Antonio Santini (eds.), *Quadri concettuali per la ricerca in demografia: quattro saggi*, p. 27–54. Florence, Dipartimento di Statistica G. "Parenti", Università degli Studi di Firenze, 101 p. (Ricerche Teoriche, no. 15.)

Bose Ashish, 2001. *India's expanding horizon of demographic training, teaching and research: evolution of population sciences*. Delhi, Institute of Economic Growth 18 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Burch Thomas K., 1993. Theory, computers and parametrization of demographic behavior, *in*: IUSSP/UIESP (ed.), *International Population Conference, Montreal 1993/Congrès international de la population, Montréal 1993, Volume 3*, p. 377–388. Liège, IUSSP/UIESP, 483 p.

Caselli Graziella, 1997. Réflexions sur l'avenir de la démographie en Italie, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 295–327. Paris, INED et PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Caselli Graziella, 2002. Teaching demography: a number of priority from the IUSSP Working Group, *Genus*, vol. 58(3–4), p. 11–17. (Special issue on "Teaching demography in the early 21st century," ed. by Graziella Caselli.)

Chasteland Jean-Claude and Roussel Louis (ed.), 1997. *Les contours de la démographie au seuil du XXIe siècle*. Paris, INED and PUF, 434 p.

De Sandre Paolo and Ongaro Fausta (ed.), 2000. *Demografia: presente e futuro*. Padova, CLEUP, 174 p.

De Sandre Paolo and Ongaro Fausta, 2000. Spunti per un quadro di riferimento, *in*: Paolo De Sandre and Fausta Ongaro (eds.), *Demografia: presente e futuro*, p. 7–19. Padova, CLEUP, 174 p.

De Sandre Paolo and Santini Antonio, 1987. Appunti per una nuova demografia istituzionale e per l'individuazione di un processo formativo nell'area demografica, *in*: Eugenio Sonnino, Antonella Pinnelli, Dionysia Maffioli, and Anunciata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 209–235. Milan, F. Angeli, 574 p.

Golini Antonio, 2000. Quale formazione per il demografo del 2000, *in*: Paolo De Sandre and Fausta Ongaro (eds.), *Demografia: presente e futuro*, p. 95–120. Padova, CLEUP, 174 p.

Labourie-Racapé Annie, and Locoh Thérèse, 1999. Genre et démographie, nouvelles problématiques ou effet de mode? *in*: Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant, and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie, Actes de la Chaire Quetelet 1997*, p. 65–92. Louvain la Neuve, Academia-Bruylant, l'Harmattan, 669 p.

Lutz Wolfgang, Scherbov Serguei, Prskawetz Alexia, Winkler-Dworak Maria, and Feichtinger Gustav, 2002. Population, natural resources and food security. Lessons from comparing full and reduced-form models, *Population and Development Review*, vol. 28, Suppl., p. 199–224. (Special issue on "Population and environment: methods of analysis," edited by Wolfgang Lutz, Alexia Prskawetz, and Warren Sanderson.)

Maffioli Dionisia, 1994. Organizzazione accademica, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 63–104. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Makhlouf Hesham, 2001. *Demographic training in Egypt and the Middle East*. Cairo, Cairo Demographic Centre, 18 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Micheli Giuseppe A., 1993. Modèles: logique, sémantiques, cycles de vie, *in*: IUSSP/UIESP (ed.), *International Population Conference, Montreal 1993/Congrès international de la population, Montréal 1993, Volume 3*, p. 361–376. Liège, IUSSP/UIESP, 483 p.

Micheli Giuseppe A., 2000. L'esercizio del comprendere in demografia. Cinque regole pratiche di ricerca, *in*: Paolo De Sandre and Fausta Ongaro (eds.), *Demografia: presente e futuro*, p. 33–59. Padova, CLEUP, 174 p.

Palloni Alberto, 2000. Demographic analysis: new theories, new models and new data, *in*: Paolo De Sandre and Fausta Ongaro (eds.), *Demografia: presente e futuro*, p. 61–87. Padova, CLEUP, 174 p.

Poirier Jean and Piché Victor, 1999. Trente ans de recherche explicative en démographie. Réflexions autour des dangers du cloisonnement, *in*: Dominique Tabutin, Catherine Gourbin, Godelieve Masuy-Stroobant, and Bruno Schoumaker (eds.), *Théories, paradigmes et courants explicatifs en démographie, Actes de la Chaire Quetelet 1997*, p. 41–64. Louvain la Neuve, Academia-Bruylant, l'Harmattan, 669 p.

Pressat Roland, 1996. Recherche et enseignement en démographie: réflexions diverses, *in*: Graziella Caselli (ed.), *Demografia: analisi e sintesi. Cause e conseguenze dei processi demografici,* Proceedings of the seminar *Demography: analysis and synthesis. Sienna, 22–24 April 1996*, vol. 1, p. 241–251. Rome, Università di Roma, Dipartimento di Scienze Demografiche, 252 p. (Materiali di studi e recerche, no. spécial.)

Preston Samuel H., 1993. The contours of demography: estimates and projections, *Demography*, vol. 30(4), p. 593–606.

Racioppi Filomena, 1996. *Demografia professionale e opportunità formative: un'indagine fra i demografi italiani*. Rome, Dipartimento di Scienze Demografiche dell'Università degli Studi di Roma "La Sapienza." (Materiali di Studi e Ricerche.)

Rowland Donald, 2001. *Specialist and generalist goals: changing emphases in demographic training in North America and Australasia*. Canberra, Australian National University, 20 p. (Paper presented at the IUSSP seminar "Demographic training in the third millennium," Rabat, Marocco, 15–18 May 2001.)

Santini Antonio, 1994. *Il contributo dell'analisi demografica alla gestione delle risorse umane*. Florence, Dipartimento Statistico dell'Università di Firenze, 14 p.

Tabutin Dominique, Gourbin Catherine, Masuy-Stroobant Godelieve, and Schoumaker Bruno (eds.), 1999. *Théories, paradigmes et courants explicatifs en démographie*. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 672 p. (Chaire Quetelet, 1997.)

Trivellato Ugo, 1990. Modelli di comportamento e problemi di misura nelle scienze sociali: alcune riflessioni, *in*: Società Italiana di Statistica (ed.), *Atti della XXXV Riunione scientifica della Società Italiana di Statistica, Padova*, vol. 1, p. 11–34. Rome, Società Italiana di Statistica, 338 p.

Wunsch Guillaume, 1993. The foundation of demographic models: theories or empirical exercises? *in*: IUSSP/UIESP (ed.), *International Population Conference, Montreal 1993/Congrès international de la population, Montréal 1993, Volume 3*, p. 355–376. Liège, IUSSP/UIESP, 483 p.

Wunsch Guillaume, 1995. "God has chosen to give the easy problems to the physicists" or why demographers need theory, *in*: EAPS and IUSSP (eds.), *Evolution or revolution in European population/évolution ou révolution dans la population européenne. European Population Conference/Congrès européen de démographie, Milano, 1995, Plenary sessions*, vol. 1, p. 201–224. Milan, Franco Angeli, 266 p.

C H A P T E R

141

Advantages and Limitations of Demographic Software and Simulation Programs

ANTONELLA PINNELLI AND ANNA-MARIA BIRINDELLI

Department of Demographic Science, Università degli Studi di Roma "La Sapienza," Rome, Italy

Dipartimento di Statstica, Università degli Studi Milano-Bicocca, Milan, Italy

I. COMPUTING AND DEMOGRAPHY

The advent of micro-computers revolutionized working conditions in demography, as well as in other fields. However, the extraordinary spread of this device and the assiduity with which it is employed could undoubtedly not have happened without the development of software to facilitate its use and increase the performance of users, not only at the level of general computing tools but also at that of more specialized professional tools. Over the past 2 decades, a certain number of software packages that facilitate the handling of demographic data and phenomena have been developed.

In the beginning, it was a question of developing procedures to increase the speed of calculations or to make operations more user-friendly, such as facilitating the application of demographic analysis methods. Quite early, it became accepted that the computer could facilitate the teaching of demography due to the development of teaching software. Software producers immediately took this route, and software packages for teaching demography quickly came on the market. For example, the first version of Pyramids, one of the best-known programs, was produced in 1988. It is necessary to distinguish two types of demographic software: that which is mainly aimed at research or the practice of demography and that which facilitates the work of teachers.

The demographic software developed for the application of methods used well-known statistical and mathematical packages as reference models. There is more experience in these fields, and the packages integrated already have a solid position in the market of essential tools for researchers and practitioners. On the basis of this success, demographers are attempting to develop a new field of education and research, *computational demography*, which combines computational expertise with demographic skills, breaking with a past of considerable amateurism and little professionalism.

However, developing software for teaching and training in demography poses more problems than writing programs that improve the speed of calculations or that help in the application of demographic analysis. In addition to skills in computing and demography, it requires a solid knowledge of education. Because teaching and the production of manuals are an integral part of the work of many demographers, it could be said that teaching is a natural element of the culture of demographers. Although this statement is partly acknowledged, producing teaching software poses very specific educational problems. It opens onto a field of specific research that aims at establishing the reference framework for teaching software, whether in demography or in other disciplines. It is because of this specificity that we begin by devoting the first section of this chapter to software for teaching demography, before making as comprehensive an assessment as possible of the demographic software packages that are available.

II. USING SOFTWARE FOR TEACHING DEMOGRAPHY

Is computing useful for teaching demography? The transmission of knowledge requires a method of communication. There are different types of media for this, and each medium opens the way to a specific type of training (Olson, 1979). Each one makes a specific contribution to the development of human knowledge and intelligence. It is essential for a teacher to be capable of using the different media available and to evaluate what each new medium can provide in specific or additional functions compared with the traditional media already in use.

Each communication tool, if it is used correctly, can be a learning opportunity. A *multimedia* tool, capable of using several media that each make a different contribution to the development of multiple aspects of the intellect can enable students to approach the same issue from different perspectives. The computer largely provides these advantages, and it should therefore be possible to use it to the advantage of teaching, especially because it is dynamic, programmable, and interactive (Greenfield, 1984). As a result of these assets, it is possible to design software packages that

Are adapted to the various levels of knowledge and ability of students
Enable students to consider and test different possible solutions to a given question
Enable users to play an active role in the learning process
Allow a greater personalization of the teaching or learning process than traditional media (Pontecorvo *et al.*, 1989)

To take full advantage of the specific resources of a medium, it must be used correctly. The computer itself does not present any disadvantages, but its potential can be considerably diminished and even destroyed if its use is not adapted to its objective. Much therefore depends on the software. Great efforts have been made to produce quality teaching packages capable of promoting and guiding self-teaching and to create and facilitate the accumulation of personal knowledge in users. It is therefore necessary for teams of multidisciplinary experts to be involved at different stages in the creation of a teaching software package: experts in demography, experts in software development, and what is most frequently lacking, experts in educational communication. The best software benefits from comprehensive knowledge of educational issues and uses resources provided by computing in the most creative manner.

We have reviewed the opinion of demographers on the usefulness and the reasons for using educational software. Six different, but recurring, points of view have been identified.

1. No Need for Software Specific to Demography

A first point of view is that specific software is unnecessary for teaching demography, apart from some very explicit exceptions. However, students must learn to use statistical or traditional mathematical packages, because these normally make it possible to apply nearly all the methods of demographic analysis (Burch, 1992).

The use of computers and data processing software should be included in the training of demographers. Essentially, this is sufficient for rapid calculations, to transform results into graphs, to rapidly change entries (data or parameters), and to observe the influence of these changes on results.

The preference for a particular package depends on the type of demography taught. More mathematical tools are necessary for teaching formal demography, but statistical software and spreadsheets are essential for practical applications. Spreadsheets are a category of software that is particularly appreciated by demographers because it can fulfill different needs for a low investment in terms of training. According to this approach, using software is advantageous when theoretical teaching is extended with practical applications and when the appropriate tools available on the market are sufficient for this. Students can learn to use software themselves or during *ad hoc* lessons, and the demography teachers need only recommend the applications.

2. Using Nondemographic Software to Teach Demography

Demography teachers themselves can also learn how to use software and create examples of demographic applications for the students to carry out. This is what Antonella Pinnelli and Mario Porri (1989) proposed in a small handbook that enables students with little or no computing knowledge to start using a computer and Lotus-123 for carrying out the operations necessary for applying basic methods generally taught during the first demography lessons.

From a slightly different point of view, nondemographic software can also be used for teaching demography by using these packages to "change the demography to be taught" (Burch, 1992). The use of mathematical software, for example, can be made extremely easy and pleasant, with very few problems, by transforming the methods taught, by changing the approaches, and by making this first contact with demographic analysis a point of entry into more formal and mathematical demography, which has existed for a long time but which is not widely taught (Burch, 1992; Hakkert, 1985).

3. The Usefulness of Software Specific to Demography

Many analysts focus more attention on the use of software that is specifically demographic, thinking, for example, that it is preferable to illustrate the teaching of methods by using the software designed for their application. The pioneers of this idea were Nathan Keyfitz and Wilhelm Flieger (1971). It was developed further by Eduardo Arriaga (1992), who proposed a text accompanied by 45 worksheets, with which the application algorithms are developed. The theory is learned, and the means for its application are immediately available, avoiding the phase of algorithm development, which can cause problems and errors, while clearly specifying the stages included in the procedure and the formulas used in the worksheet.

Other demographic packages can be used to the same end, as long as their use is explicitly associated with the theoretical course (Srinivasan, 1992). However, we must avoid those that are based on a kind of black box, which allow students only to enter the data and to recover the results without any knowledge of the formulas or the algorithms used (Burch, 1992). Such packages should be excluded from teaching; they are useful only for tasks of execution. As stated by Peter Holmes (1988), the difference between teaching software and application software lies in the fact that in the former, the procedure remains obvious, whereas the latter aims for speed of execution and precision of results.

Whereas the first point of view was to invite students to take themselves from the theory to its application by experiencing the development of the necessary algorithms, this third point of view consists of immediately placing the students in a situation where they are able to carry out applications without risking error. The emphasis is placed on the teaching of the practical aspects and the main underlying concepts of the methods, the required data, the advantages and limitations of each method, the possible errors when the data do not match the hypotheses made during method design, and the analysis of results (Arriaga, 1992), instead of simply presenting the theoretical aspects and asking students themselves to transform the theory into operational algorithms.

These two points of view are based on different educational objectives. One of them is typical of the university approach and the training of researchers, and the other is oriented more toward professional applications. There is no doubt that this second type of teaching uses demographic software to enable decision-makers to undertake demographic analyses, including very sophisticated ones, and to obtain useful and correct results (Burch, 1992). In reality, this type of approach should not be relegated to second place in university courses because it introduces educational perspectives whose benefits are not limited to training practitioners and that can be just as useful for training researchers and teachers, without being exclusive of the first strategy.

4. The Usefulness of Complex Demographic Analysis Software

Regardless of opinion on this debate, most demographers are in agreement with regard to the specific usefulness of demographic software that enables the application of certain very complex methods of demographic analysis, which would otherwise necessitate the development of algorithms whose construction is well beyond the ordinary abilities of mere students. Students would never have the opportunity of applying these methods if *ad hoc* packages did not exist. The same applies to multistate models, micro-simulation, and complex models of the interrelations between population and development (Burch, 1992).

Moreover, such software is useful to a large number of researchers and scientists as well as to teachers and students. Regardless of the user, the undeniable advantage of these packages is that they discharge the user from dealing with vast series of calculations and from programming the implementation of the necessary algorithms. The user can then concentrate more on the input and output, on the necessary data, and on the results. If users gain all the qualifications required to work with the same rigor as professionals and to interpret the results, they could probably learn to apply more methods in less time than if they followed a traditional teaching method, which is a considerable advantage for their general training.

5. Simulation Software, a Pioneer of Teaching Software

By using a population projection algorithm, it is possible to understand demography better than through studying the issue in a purely theoretical manner. For example, it is possible to test the effects of the implementation of different hypotheses of fertility, mortality, and migration trends on population size and structure using real or imaginary populations (Burch, 1992). By using different simulation models that link population projections to different variables of economic and social development (e.g., concerning education, the labor market, savings, investment, health), it is possible to test the consequences of demographic trends on the problems that economic and social policies would have to resolve. With micro-simulation family models, it is possible to take into account the effects of different demographic changes on family structure and size. Simulation is a very useful tool for answering the "what if?" question, and the software packages that today facilitate its use contribute to the teaching of theory.

This was possibly the idea that most excited demographers as soon as they first had access to the computing tool, although at the time, it was still crude, difficult to access, and hard to use (as described in the pioneering studies of Étienne van de Walle and John Knodel in 1970), and this is possibly the area in which demographic software has developed the most.

6. Software Instead of Books

Some demographers have suggested that educational software packages are useful in teaching demography and could partially or totally replace books and teachers' handbooks. It is perhaps necessary to date this idea back to historical packages such as Popshow or Pyramids, which aimed at explaining one or more demographic issues by using only computer science. Going to the extreme of this reasoning, some authors have since suggested replacing the entire text of a traditional handbook with a package that explains and demonstrates all the demographic theories and methods and that provides all the useful aspects of online help. For example, the concept of *demographic hypermedia* (Leete, 1996) is associated with this idea. With this approach, we enter feet first into the world of so-called educational software.

However, the desire to replace traditional didactics with teaching software has caused skepticism and perplexity: "The best software is not a substitute for a good course in demography . . . it only facilitates the application of methods, techniques, and models" (Willekens and Hakkert, 1996), and "No software or hardware will replace well-trained demographers" (Gerland, 1992).

It is not completely unreasonable to consider that some subjects could benefit more than others from the possibilities provided by computers and that some of them could be dealt with completely or largely with the help of specially designed software, provided that it has clear advantages over traditional courses.

It seems necessary to repeat that traditional mathematical or statistical software such as spreadsheets and worksheets, as useful and sometimes even essential as they are for teaching demography, do not form part of what is generally known as demographic software. Demographic software must be specially designed for the application of demographic methods or for the illustration of theoretical courses in demography. Moreover, to be considered as educational, it must have, in addition to the characteristics that make up an appropriate demographic software for research or professional use, the characteristics that make it an educational tool fulfilling educational requirements, or it must be a substitute for a handbook or even for a teacher.

III. DEMOGRAPHIC SOFTWARE: INVENTORY AND CHARACTERISTICS

Despite the fragmentary and sporadic nature of the information available, it can safely be said that the development of demographic software has increased significantly since the late 1980s. At the end of the 1980s, the International Union for the Scientific Study of Population (IUSSP) considered that the time had come to make a first evaluation, and in 1990, it set up the Working Group on Demographic Software and Micro-Computing (WGDSMC). One of the tasks of this Working Group was to make an inventory of demographic software (Willekens and Hakkert, 1996). To make a list of the existing packages, the WGDSMC used the small number of inventories and reviews already available in gray literature or in published literature. This enabled it to collect information on more than 200 demographic software packages. Between August and October 1991, the Working Group sent a questionnaire by mail (followed, if necessary, by a reminder) to all producers and distributors of the demographic software packages identified. Of these, 90 replied by providing more detail and information on their software or packages that the group did not have (Willekens and Hakkert, 1996). In reproducing the inventory here, we excluded programs concerned with data collection, storage, and publication and with cartography or presentation of data that, although useful for analyzing demographic data and in some specific cases created to that end (e.g., Integrated System for Survey Analysis [ISSA] undertook to memorize and analyze the data from the Demographic and Health Surveys [DHS]) are not really demographic packages. Seventy-two packages remained that could be listed in the category of interest. With the impetus of the Working Group, updates of the initial inventory by the WGDSMC were then published in *Population Index* up to the year 2000, describing the new versions (updates or additions) of these packages or of new programs. Since the second half of the 1990s, the production of demographic software seems to have left its first phase of amateurism. The tendency to place demographic packages in circulation outside of the circle of initiated scientists was consolidated at this time. Two important factors contributed significantly to acknowledgment of a wider circle of potential users: first, the near generalization of a work station, Windows, now largely considered a standard, and second, the increasing readiness to disseminate advertisements or the products themselves (i.e., freeware) at the global level on the Internet. As is the case in other fields of study, the tendency to provide information online on the production of new programs or on updated versions of existing programs has spread. The Internet makes it possible to visit sites that include sections dedicated to demographic software (e.g., www.nidi.nl, www.un.org) and to use all kinds of hypertext links that generally appear in the numerous sites of national and international organizations and universities working in the field of population studies. Twenty-eight additional programs have been added to the inventory made by the WGDSMC, which is used as a base here.

Table 141–1 provides the list of these programs and some information on the main characteristics of the demographic software. The information included is the acronym of the software or package, names of the authors, version, year of publication, type of distribution, type of user interface (i.e., guidance by menu, use of commands, or a sequential command procedure called *batch*), the availability of online assistance, main type of use (i.e., education, research, and application), the main fields of application (e.g., projection, macro- and micro-simulation, indirect estimates, life tables). Table 141–2 summarizes all of this information.

1. Production Rhythm

The chronology of the appearance of these different packages on the market has been subject to considerable discontinuity. The number of packages increased rapidly between 1986 and 1991, when the WGDSMC was established, only to fall again quite significantly. Because of the survey, it was possible to verify the existence of a far greater number of software packages than were detected by the later updates published in *Population Index*. This highlights a certain lack of dynamism on the part of producers to promote the distribution of their products among potential users, perhaps due to the amateurism of many of the packages. Nevertheless, almost one-half of the packages have been updated once or several times, a fact that bears witness to a certain commitment to maintaining and improving these tools, probably linked to the fact that the number of users was not all that low.

2. The Producers

The largely proven amateurism that reigns in this sector can be seen in the copyright status and the distribution method. Most of the packages listed are in the public domain (43%), a low percentage (7%) of them are shareware, and only 25% are commercialized. Moreover, no information at all is provided on this subject in 23% of cases. However, the fact that a package is commercialized does not provide any information on its real cost because numerous international organizations are committed to providing technical assistance and professional training to staff in many developing countries, and they provide the software free of charge or at a low cost. The fact that some software is available for sale confirms the professional nature of the product.

3. Technical Characteristics

The technical characteristics provide indirect information on the level of update and sophistication of the software. With regard to communication between the program and its user, weighty procedures such as the sequential command system (i.e., batch) linked to programming within the computer are rare (7%) and generally only relate to old versions. Most programs (84%) are installed in interface with a menu, which makes it possible to select commands on the screen.

Online help is missing for 47% of the programs listed here. Among those that have it, the context-sensitive system is the prevailing one, usually as online help, or default responses are displayed with the help of a prompt from a menu of help subjects (usually by pressing the F1 button). Some packages provide help windows that can be opened at will. This is currently considered as the most modern form of help. It enables great flexibility in navigation and is generally organ-

TABLE 141–1 Demographic Software for Microcomputers

Software name or acronym and author	Version and year	Distribution	User interface	Help	Major use	Main objective
ABACUS, United Nations	v. 1, 1991	C	B	No	T, R, A	Population projections
APPLAUSI, M. Breschi, G. De Santis	v. 1, 1994	C	MD	No	R, A	Own-children method
BERGSON, S. Bertino, D. Grassi, E. Sonnino	v. 1, 1994	S	MD	No	R, A	Macro-simulation marriage models by cohort and life cycle
BOH/PROBOII, Micro Ingeniersbureau	NA	C	MD	1	A	Schooling projections and building requirements
CAIRO, M. Sivamurti	v. 1, 1986–91	PD	CL (Basic)	No	T, A	Indirect estimates and population projections
CAPPA, ESPT-FAO	v. 1.2, 1990	PD	MD	1	T, A	Demographic scenarios for agriculture
CCMLT, Jiang Zhengua	v. 1, 1991	PD	MD	2	R, A	Model life tables
DDMSLT, C. Calhoun	1991	C	CL (Gauss)	2	R, A	Multistate and duration life tables
DEMOGRAPHICS, G. Heilig (IIASA)	v. 1, 1995	C	MD	1	T	Demographic development and socioeconomic conditions (world)
DEMOMETRIC, A. A. Klementiev	v. 1.2, 1991	NA	MD	2	R	Mortality by cause and projection
DEMPROJ, Glastonbury Futures Group	v. 3, 1991	PD	MD	No	R, A	Population projections
DGRAPH, K. Wunner, N. Fickel	v. 1, 1991	C	MD	1	R, A	Population projections
EDFISMO, M. Gymelman	v. 1, 1987	C	MD	No	T, A	Financial consequences of education policies
EDSIM (DEMPROJ), Glastonbury Futures Group	v. 1, 1988	PD	MD	No	A	Projection of school-age children and consequences for the education system
ESCAPPOP, R. Hanenberg (ESCAP)	v. 1, 1986	PD	MD	1	T, R, A	Population projections
EVACOV, E. Lelièvre (INED)	v. 1, 1987	NA	MD	No	R, A	Estimate of semi-parametric models
FAMSIM, S. Bower	v. 1.0	NA	MD	No	T	Micro-simulation: families
FAMY, Zeng Yi, Wang Zheng Lian	v. 1, 1990	NA	MD	1	T, R, A	Simulation: families
FASIM, A. Pinnelli, M. Vichi	v. 1, 1988	S	MD	No	T	Micro-simulation: families
FERT, Applied Population Laboratory	1991	PD, C	MD	1	T	Education: fertility
FERTILITY ESTIMATE PROGRAMS, East West Center	v. 1.1, 1987	PD	MD	1	R, A	Indirect estimates: fertility
FIVFIV, F. Shorter, R. Sendek (Population Council)	v. 11, 1994	NA	MD	No	R, A	Population projections, stable populations
FUNFIT, R. Hakkert	v. 1, 1989	PD	MD	No	T, A	Adjustment of demographic curves
Futpop, R. R. Sell	v. 3.2, 1991	C	B	No	T, R	Population projections
FADEP, SES Modeling Group (Statistics Canada)	β test, 1991	NA	MD	No	A	Micro-simulation: families and households
GENESIS, J. Johnson, D. A. Wolf	v. 1.21, 1990	PD	CL (C)	No	T, R	Micro-simulation: discreet stochastic events
HALLEY, N. Levine	v. 3.2, 1988	NA	MD	No	T, R, A	Population projections (Lotus)
HoMes, A. Mason, W. Shima	v. 1, 1989	C	MD, CL (C)	1	R, A	Household projection
HIVAIDS90, M. Artzrouni	v. 1.1, 1990	PD	MD	No	A	Projection of the HIV-AIDS epidemic
HRP (NPROJ), Research Triangle Institute	v. 1, 1991	NA	MD	2	T, R, A	Projection of the resources and costs of health systems

(continues)

TABLE 141–1 *(continued)* **Demographic Software for Microcomputers**

Software name or acronym and author	Version and year	Distribution	User interface	Help	Major use	Main objective
INDIRECT, R. Hakkert (PNUD/UNDP)	v. 3, 1988	PD	MD	No	T, A	Indirect estimates
IMPECC, R. Wery (OIT/ILO)	1992	PD	MD	1	A	Integrated projection and economic planning
LF-EMP (NPROJ), Research Triangle Institute	v. 1, 1991	NA	MD	2	T, R, A	Projection of workforce and employment
LIFE, G. Broström	v. 4,1992	C	CL	2	T, R, A	Analysis of survival data
LIFEHIST, F. Rajulton, K. Shaburov, E. Andreev	1991	NA	CL, B	No	T, R, A	Biographic analysis in the perspective of multistate tables
LIFETIME, J. H. Pollard	v. 1.21, 1989	C	MD	No	T, R, A	Model life tables (by cause)
LIPRO, E. van Imhoff	v. 2, 1991	NA	MD	No	R, A	Population, multistate, and planning projections
LRPM/PC, CELADE	v. 1, 1989	PD	MD	1	T, R	Integrated population and planning projections
LTCD, E. Andreev	1991	NA	MD	1	R	Model life tables (by cause)
LUDZIK, J. Kuropka, M. Cieslak, B. Radikowska	v. 2, 1990	PD	MD	2	R	Population projections, stable populations
MCPDA, Institute of Resource Development at Westinghouse	1986	NA	B	No	R, A	Indirect estimates, population projections, stable populations
MEDS, F. T. Denton, C. H. Feaver, B. G. Spencer	v. 5, 1990	PD	MD	2	R, A	Integrated projections and economic development in Canada
MODELS, B. Vaughan	1986	NA	MD	No	R, A	Model life tables
MORTAL, J. M. Costes, D. Waltisperger	v. 1, 1988	PD	MD	No	T, A	Model life tables, indirect estimates
MORTPAK-LITE, United Nations	v. 3, 1990	C	MD	1	T, R, A	Model life tables, indirect estimates, population projections, stable populations
MULTIREG, R. Hakkert (Undp)	v. 1, 1990	PD	MD	No	T, A	Multiregional projections (matrix)
NPROJ, Research Triangle Institute	v. 1, 1991	NA	MD	2	T, R, A	Population projections (host)
NSFA, N. Dworak, S. Kirmeyer	NA	NA	MD	2	T	Teaching: fertility
PANDEM, CELADE	1988	PD	MD	No	R	Demographic analysis, indirect estimates
PaRity, N. B. Dworak	v. 1.21, 1991	PD	MD	2	T, R	Fertility analysis (CPA)
PAS/RUP, E. Arriaga	1994	C	MD	2	T, R, A	Teaching: basic multitask course
PDPM/PC, United Nations	v. 1, 1992	C	MD	1	T, R, A	Integrated population and planning projections
PEOPLE, R. Leete	v. 3, 1993	PD	MD	1, 2	A	Population projections
POHEM, SES Modeling Group (Statistics Canada)	Prototype	NA	B	No	R	Micro-simulation: life cycle and characteristics of health conditions
POPGEN, B. Edmonston	v. 3.3, 1991	PD	B	No	R	Projection of migrants (4 generations)
POPILO, J. Sehqal (OIT/ILO)	v. 2, 1986–87	PD	MD	No	A	Population and labor force projection (Lotus)
POPSHOW, G. Immerwahr	v. 1, 1990	NA	MD		T	Teaching: basic multitask course
PopSyn, D. Bogue	1989	C	MD	1	R, A	Integrated package: indirect estimates, projection, stable populations (Lotus)
POPULATE, R. McCaa	v. 2.9, 1991	S	MD	1	R	Reverse projections
PRODEM, Celade	v. 1, 1989	PD	MD	2	R	Population projections
PRODEMOG, L. Petrioli	v. 1, 1989	PD	MD	No	T, R, A	Model life tables
PRODISA, Beatrix and Leo Brecht	v. 1.3, 1991	PD	MD	1	A	Biographical analysis by discreet period
PROFILE, E. van Imhoff	v. 1, 1991	PD	MD	No	R	Age adjustment of curves

(continues)

TABLE 141–1 *(continued)* **Demographic Software for Microcomputers**

Software name or acronym and author	Version and year	Distribution	User interface	Help	Major use	Main objective
PROGEM, A. Oskam	v. 2, 1991	NA	MD	No	A	Forecast of local populations according to housing conditions
PROJNPCII, K. Hill	NA	NA	MD	No	R, A	Population projections
PYRAMIDS, Applied Population Laboratory	v. 1, 1988	PD, C	MD	1	T, A	Teaching: population pyramid
QFIVE, United Nations	v. 1, 1989	C	MD	1	T, R, A	Indirect estimates: child mortality (Brass)
QCINCO, N. B. Dworak	v. 1.29, 1991	PD	MD	2	R, A	Indirect estimates: child mortality
RAMAS/AGE, S. Ferson, R. Ackakaya	v. 2, 1992	C	MD	1	T, R	Modelling of fluctuations in population age structure
RAMAS/SPACE, S. Ferson, R. Ackakaya	v. 1.2, 1992	C	MD	1	T, R	Models of population space structure for biological preservation
RAMAS/STAGE, S. Ferson	v. 1.2, 1992	C	MD	1	R, A	General modelling of population dynamics (matrix)
RAPID, Glastonbury The Futures Group	v. 1, 1991	PD	MD	No	T, R, A	Integrated projection and planning
REGPRO, R. Hakkert (PNUD/UNDP)	v. 3, 1989	PD	MD	No	T, A	Regional projections
ROOT, E. Lelièvre (INED)	v. 1, 1988	NA	MD	No	R, A	Estimates of hazard rates
SFM, E. L. Soroko	v. 2, 1991	NA	B	No	R, A	Micro-simulation: families
SIP, S. Bertino, E. Sonnino	v. 1, 1995	S	MD	No	R, A	Macro-simulation: reverse projection
SISCODEM, G. Calot	prototype, 1995	NA	MD	No	R, A	Integrated package: period and cohort analyses; structure and dynamics, projections
SPDP, Liao Hong-Zhi, Luo Ping	v. 2, 1989	PD	MD	No	R	Population projections
STEP, S. T. Moreland	v. 1, 1988	PD	MD	No	A	Projection: general education and cost (host)
SURVIVAL, D. P. Smith	v. 3, 1991	PD	B	No	R, A	Model life tables (by cause)
TARGET, J. Bongaarts	v. 2, 1989	PD	MD	No	A	Birth control: estimates of future trends (various contraceptive methods)
TCM, Glastonbury The Futures Group	v. 1, 1991	PD	MD	No	A	Model of targeted cost (family planning)
TDA, G. Rohwer	v. 3, 1991	NA	CL, (C)	No	R	Estimate of transition models
TED, S. T. Moreland	1991	PD	NA	No	A	Integrated projections (host)
TM1, G. Greene, Y. Ofosu (OIT/ILO)	v. 4, 1990	PD	MD	No	E	Simulation: population growth, labor force, and demand for social services
UN GAME, United Nations	v. 1, 1987	PD	MD	1	E	Introductory game for training on population and development models (host)
WORKERS (PEOPLE), R. Leete	v. 1, 1992	PD	MD	1, 2	A	Schooling, household, and labor force projections

A, application; B, batch; C, commercial; CL, command language; NA, information not available; MD, menu driven; PD, public domain; R, research; S, shareware; T, teaching; 1, context sensitive; 2, access to online help.

ized in a hierarchical and structured manner according to level (e.g., global, context, methodology).

4. Targeted Users

Programs are distinguished according to whether they are intended for teaching, research, or the application of demographic methods. About 15% of the programs have been designed to respond simultaneously to all these needs, and 44% have been designed to respond to various combinations of these three targets, with the combination of application and research (26%) being the predominant one.

TABLE 141–2 Demographic Software According to Year of Publication, Distribution Type, User Interface, Online Help System, Major Use, and Main Objective

Year of publication	%	Distribution type	%
1986	3.5	Shareware	4.6
1987	5.7	Public domain	44.8
1988	10.3	Commercial	20.7
1989	12.7	Public domain and commercial	2.3
1990	11.5	Information unavailable	27.6
1991	33.3		
1992	8.1		
1993	1.1		
1994	4.6		
1995	3.5		
Information unavailable	5.7		

User interface	%	Online help system	%
Batch	8.1	Context sensitive	26.4
Command language	5.8	Access to online help	15.0
Menu-driven	83.9	Both	2.3
Batch and command language	1.1	All help online	56.3
Command language and menu	1.1		

Major use	%	Main objective	%
Teaching	16.1	Population projection	19.5
Application	9.2	Retro projection	2.3
Research	12.6	Integrated projection	9.2
Teaching and application	10.4	Integrated projection and planning	8.1
Teaching and research	6.9	Macro-simulation and development	5.8
Application and research	26.4	Indirect estimates and other analysis techniques	10.3
Teaching, application and research	18.4	Biographical analysis, transitions, and multistate tables	11.5
		Model life tables	10.3
		Micro-simulation	8.1
		Fertility and family planning	5.7
		Education	6.9
		Integrated packages	2.3
Total number of programs	87		

5. Software for What Purpose?

Projections are dominant (34%) in the issues and problems dealt with in demography. The available programs are mainly aimed at long-term projections. Most of them are designed to make analytic population projections (i.e., cohort component method) for a given country or territory or for a group of countries, territories, or regions, whose results are aggregated *ex post* to provide more general results. Generally, the input (i.e., mortality, fertility, and migration) and output options are very detailed, and the methodology is traditional.[1] With other software (11%), it is possible to make projections on specific categories of the population, such as the labor force, students, or health workers.[2]

With regard to the wider subject of relationships between population, policy, and development,[3] it is possible to distinguish the programs that associate traditional projections to a development model (3%) and macro-simulation programs (3%), which by allowing multiple scenarios, show how different hypotheses can affect socioeconomic results.[4]

[1] Richard Leete's People program includes numerous traditional procedures, among which is the possibility of using the Brass models for mortality and fertility. For an evaluation of numerous projection programs, see McGirr and Rutstein (1987), Hermalin (1988), and McMurray (1996).

[2] Projections for these subpopulations can be carried out within general projection programs (e.g., Meds) or by specific programs that are based on the input and output of general programs (e.g., Hrp and Lf-Emp, linked to Nproj, and Workers, linked to People).

[3] Bogue (1996) has evaluated these programs. Arminger and Galler (1991) have compared Lipro and three other programs.

[4] These programs (e.g., Pdpm and Game, developed by the United Nations; Tm1, developed by BIT) frequently have the objective of training specialized staff in developing countries.

Other software (37%) is distributed more or less equally between the following four subjects:

1. Indirect estimates of mortality[5] and fertility, associated (or not) with other methods of analysis such as procedures for age adjustment of demographic curves
2. Analysis of transitions and multistate tables
3. Procedures for creating and analyzing model life tables (all causes or by cause)
4. Micro-simulation (particularly with regard to family)

Other software packages cover subjects linked to the specific analysis of fertility and family planning (2%).

Some software packages (20%) have been specifically designed for teaching. Some of them cover specific subjects,[6] but others have multiple functions and cover the different areas included in a basic course,[7] and there also are macro-simulations.[8]

From the prospective point of view, significant progress could be made if a general effort was made to integrate the functionalities of the software produced regarding different research objectives to enable users to develop their own analyses in sequences or in parallel. The Futures Group provides an interesting example of the combining of programs placed separately on the market between the 1980s and the beginning of the 1990s due to the creation of new procedures. The Spectrum program is a succession of models of population policies used for projecting requirements for reproductive health and the consequences when these needs are not fulfilled.[9]

[5] This category includes packages that can produce indirect estimates and different types of projections at the same time (e.g., Mortpak, developed by the United Nations; Cairo, designed by M. Sivamurthi; MCPDA, developed by the Westinghouse Institute of Resource Development).

[6] For example, Fert by the Applied Population Laboratory (University of Wisconsin), Piramides Animées by Henri Leridon (INED), and Pyramids by HPN Technologies.

[7] For example, Popshow by George Immewahr, Pas/Rup by Eduardo Arriaga, Prodemog by Luciano Petrioli, and Ifs by Barry Hughes.

[8] JAVA Demography, which simulates exponential demographic growth, enables the user to modify the characteristics and the dynamics of the population and to see the long-term effects. The objective of the World Health Chart program is to visualize the development of world health; it makes it possible to use international data for training, lobbying, and generating hypotheses. The PopTrain software (version adapted to Windows by NIDI in 2001) is a program for teaching demography in secondary schools.

[9] The models included in the Spectrum system are DemProj, FamPlan, AIM (AIDS impact model), Rapid (resources for the awareness of population impacts on development), PMTCT (prevention of maternal-to-child transmission, NewGen (reproductive health for adolescents), and BenCost (financial benefits and costs of family planning programs).

IV. CRITERIA FOR EVALUATING DEMOGRAPHIC SOFTWARE

Evaluation is an essential and very delicate phase in the production of any software. It aims at ensuring that the objectives have been fulfilled. The criteria proposed at the beginning of the 1990s by the IUSSP Working Group, which was greatly inspired by the evaluation studies of the late 1980s, are still relevant today:

Range: What can the software do? Does it incorporate significant variants of the method provided? Can it deal with different data structure?

Portability: Can it function on different machines? Does it include interfaces with other software?

Documentation

Exactitude and precision: Is the software likely to lead to incorrect results with bad data? Does it include warning systems?

User-friendliness

Effectiveness: Is the calculation time reasonable?

Help: Does the distributor provide appropriate assistance if there is a problem?

Beyond the general evaluation criteria of demographic software, specific requirements for teaching software must be emphasized. We focus more on this topic in the following section.

1. Evaluation of Teaching Software

For software to be considered as useful for teaching, it should fulfill a certain number of basic conditions. It should be very easy to use, it should not pose more problems than the subject covered would normally pose, navigation within the software should be easy, and the user should always be informed of the possible options available. Effective reactivity must be guaranteed, and the software should be capable of neutralizing any potential user errors (e.g., typing errors during data entry). Good communication should be established with the user, the screen should not be overburdened, and the text should be in large characters, legible, and accompanied when possible with good-quality graphics. The possibilities provided by micro-computing should be employed to their full potential by using animation to clarify many of the concepts presented as graphics, for example. The exactitude and relevance of the content should be verified. Moreover, the educational objectives of the subject proposed by the software should ensure, through the quality of the interaction between the student and the software, the possibility for the student to play an active role and to have control over the learning process.

To make a comprehensive evaluation, it is also necessary to test the results of the use of the software on the targeted population. When we wish to evaluate software that has been designed for teaching, it is essential to consider what can be achieved before and after the computer by other methods of communication. Why use a computer? Does it provide additional help to students in acquiring skills or a certain qualification? For example, there is no benefit from using a computer to present long sequences of linear texts; it is much better to read a book. The crucial point for investing in the design of teaching software is whether it will provide a real advantage over traditional courses and manuals. Such an advantage can be educational or economic and organizational. For example, it can enable students, aided by a good teaching program, to learn a series of basic concepts necessary for a university course, reducing the requirement of teachers to a minimum of intervention on some aspects. Another, more specific evaluation criterion is based on the exact subject of the software in question.

2. Computerized Tutoring

Software aimed solely at teaching objectives can be separated into two categories: tutoring and simulation. In the first case, they are self-sufficient teaching tools, although sometimes this condition may not be entirely fulfilled. There is a lack of effective teaching software that gives a sufficiently active role to students and that personalizes the learning process through knowledge testing or, better still, "problem solving" or "world in reduction" experiments. This means that the progress of scientific research in the field of teaching software has not been used to its full potential by the demographers and developers who design demographic teaching software.

It is not known whether these programs have been greatly used in teaching or whether it is possible to evaluate the impact on students' preparation and the manner in which teaching is organized. Such an evaluation is desirable because it would make it possible to establish whether new resources should be invested in developing programs of this kind. Multimedia teaching is expected and desired. However, the necessary resources must be invested to take full advantage of the possibilities available, because computers are not themselves a guarantee of a better or a cheaper education.

3. Simulation Software for Teaching

Among the macro- and micro-simulation software, some programs have explicit teaching objectives, and others are halfway between this objective and others. Their importance for teaching demography is undeniable because it is possible to experiment with the relations between demographic changes and, for example, the size, structure, and dynamics of families; the size, structure, and dynamics of the population; educational needs; and demand for employment and health structures. The educational aim of these packages is, however, not clearly visible in their formulation. Greater interaction with experts in education and computing involved in the development of simulations in teaching would help to make these programs more relevant to the targeted objectives.

Demographic projection software packages are the easiest ones to use with complete self-sufficiency. This is an extremely useful characteristic in the professional world and in the context of teaching.

Application programs offer considerable advantages, as much for researchers as for professionals. To promote greater use of them, more attention should be paid to encouraging autonomy and to making them easier to use.

The authors responsible for developing teaching software should include characteristics that facilitate learning and use, including the following:

Ease of navigation (e.g., possibility of going back, accessing the main menu, using a glossary, going to a given page)

High-quality graphics and animation possibilities

Pages that are not overburdened and enough of them for a presentation (i.e., to understand a concept, it is not necessary to go to a new page; a new screen should correspond to new content, even if it is linked to the previous page)

Help at different levels, as much with regard to procedures (what do I do next?) as to methodology (explanation of formulas) and to content (what is this?)

Providing the possibility of evaluating the results (by giving a few examples of results that have already been interpreted)

Allowing experimentation with futuristic or improbable hypotheses

CONCLUSIONS

The production of demographic software appears to have entered into a professional phase, in which it has become possible to consider it as a new specialization in demography, *computing demography*, which is deserving of more human and financial resources. Whoever is interested in producing software now has access to the necessary information for identifying the international network of potential users, as well as what already exists on the market, and is aware of the essential conditions required and the experience already acquired on the subject in other fields.

Statistical and mathematical software are a natural point of reference for demographic software. Educational software should also be included when it constitutes demographic software for teaching. Each of these disciplines (i.e., statistics, mathematics, and education) have great experience in software design and have accumulated ability and expertise that needs to be added to that which we already possess to be able to produce tools that take full advantage of the latest developments in this area. A new field of research is now open in software production, testing, evaluating performance, and monitoring their evolution. All this requires time, will, and patience, essential virtues if we wish to produce software that is easy to use, that functions perfectly, and that potential users are aware of and can easily acquire (Pinnelli, 1994).

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

ARMINGER Gerhard and GALLER Heinz P., 1991. *Demographisch relevante Modellrechnungen, Simulations- und Analyseverfahren auf der Basis empirischer Erhebungen.* Wiesbaden, Bundeninstitut für Bevölkerungsforschung (BiB), 102 p. (Materialien zur Bevölkerungswissenschaft Heft 72.)

ARRIAGA Eduardo E., 1992. *Demographic software as a didactic and analytical tool.* The Hague, IUSSP and NIDI, 14 p. (Paper presented at the IUSSP/NIDI Expert Meeting on Demographic Software and Micro-computing, The Hague, 29 June–3 July.)

BOGUE Donald, 1996. Software for population and development research, *in*: Ralph Hakkert and Frans Willekens (eds.), *Evaluation of demographic software*, p. 6–53. Groningen, University of Groningen, Faculty of spatial sciences, 189 p.

BURCH Thomas K., 1992. *Do we need special demographic software?* The Hague, IUSSP and NIDI, 19 p. (Paper presented at the IUSSP/NIDI Expert Meeting on Demographic Software and Micro-computing, The Hague, 29 June–3 July.)

GERLAND Patrick, 1992. *Software development: past, present and future trends and tools.* The Hague, IUSSP and NIDI, 28 p. (Paper presented at the IUSSP/NIDI Expert Meeting on Demographic Software and Micro-computing, The Hague, 29 June–3 July.)

GREENFIELD Patricia M., 1984. *Minds and media. The effects of television, computers and videogames.* Cambridge (Massachusetts), Harvard University Press, 232 p.

HAKKERT Ralph, 1985. Quasi-component models for small area population projections: examples from the State of Sao Paulo, Brazil, *in*: IUSSP (ed.), *International Population Conference, Florence, 1985, vol.4*, p. 163–174. Liège, Ordina Éditions, 556 p.

HERMALIN Albert I., 1988. *Microcomputer software for population and development planning.* Washington, National Academy of Science, 34 p. (Paper presented at a National Academy seminar.)

HOLMES Peter, 1988. *Using a microcomputer to teach statistics with a particular reference to demography*. Rome, Università "La Sapienza," Dipartimento di Scienze Demografiche. (Paper presented at the workshop on "Il calcolatore nell'insegnamento della Demografia," Roma, 4 November, 1988.)

KEYFITZ Nathan and FLIEGER Wilhelm, 1971. *Population facts and methods of demography.* San Francisco, W.H. Freemen., 613 p.

LEETE Richard, 1996. Towards tomorrow's demographic software, *in*: Ralph Hakkert and Frans Willekens (eds.), *Evaluation of demographic software*, p. 1–5. Groningen, University of Groningen, Faculty of spatial sciences, 189 p.

MCGIRR Nancy J. and RUTSTEIN Shea O., 1987. Six microcomputer programs for population projection: an evaluation, *Asian and Pacific Population Forum*, vol.1(5), p. 9–17 + appendices p. 22–28.

MCMURRAY Christine, 1996. Software for population projections: a review of six packages, *in*: Ralph Hakkert and Frans Willekens (eds.), *Evaluation of demographic software*, p. 104–125. Groningen, University of Groningen, Faculty of spatial sciences, 189 p.

OLSON David R., 1979. *Linguaggi, media e processi educativi.* Turin, Loescher, 224 p.

PINNELLI Antonella, 1994. Computer e formazione dei demografi, *Induzioni*, no. 8, p. 37–48.

PINNELLI Antonella, CESARENI Donatella, and BIRINDELLI Anna-Maria, 1996. Software for teaching and training in demography, *in*: Ralph Hakkert and Frans Willekens (eds.), *Evaluation of demographic software*, p. 54–102. Groningen, University of Groningen, Faculty of spatial sciences, 189 p.

PINNELLI Antonella and PORRI Mario, 1989. *Esercitazioni di demografia utilizzando i fogli elettronici su personal computer*. Rome, Università "La Sapienza," Dipartimento di Scienze Demografiche, 77 p.

PONTECORVO Clotilde, TAFFAREL Lorenzo, and ZUCCHERMAGLIO Cristina, 1989. Il computer come strumento di educazione alla riflessione metalinguistica, *in*: Antonio Calvani (eds.), *Scuola, Computer, Linguaggio*, p. 161–231. Turin, Loescher, 231 p.

SHUELL Tom J., 1992. Designing instructional computing systems for meaningful learning, *in*: Marlen Jones and Philip Winne (ed.), *Adaptive learning environments*, p. 19–54. Berlin, Springer Verlag, 408 p. (NATO A.S.I. Series.)

SRINIVASAN Krishnamurthy, 1992. *Note on use of software packages in the teaching of demography*. The Hague, IUSSP and NIDI, 7 p. (Paper presented at the IUSSP/NIDI Expert Meeting on Demographic Software and Micro-computing, The Hague, 29 June–3 July.)

VAN DE WALLE Étienne and KNODEL John, 1970. Teaching population dynamics with a simulation exercise, *Demography*, vol. 7, no. 4, p. 433–448.

WILLEKENS Frans J. and HAKKERT Ralph, 1996. *Directory of demographic software.* Groningen, University of Groningen, Faculty of spatial sciences, 330 p.

C H A P T E R

142

Centers of Demographic Research and Teaching: Origin, Development, and Characteristics

DIONISIA MAFFIOLI
Università degli Studi di Bari, Bari, Italy

A study of the centers of demographic research and teaching can illustrate the place this discipline occupies at the heart of the social sciences and enlighten us on the value, both pedagogic and political, accorded to studies on population as a key to understanding social relations and dynamics or as a means of influencing their evolution. The characteristics of these centers; their scientific, pedagogic, and operational activities; their place within or outside the university; and their integration in all kinds of public or private structures reflect the contribution of demography to scientific and social development and the expectations of the scientific community and of society. Naturally, contributions and requests are strongly interactive and variable in time and in space. They are closely linked to demographic events and to currents of opinion that they provoke, and, in the matter of teaching, and they also depend on opinions about the aims and content of school and university education. An examination of the origins and the current situation of the centers of teaching and research in demography is an exercise that is more stimulating and instructive than could be imagined *a priori*. The inventory of existing institutional and human resources in the matter of demographic research and teaching constitutes a strong base for thinking about the future of the discipline in a moment of its history that can appear as a critical turning point.

To study demographic centers of research and teaching and their activity at an international level, it is necessary to have access to a set of basic information whose collection demands patient and costly labor, exceeding the capacities of an individual initiative and demanding the mobilization of institutions and specialized agencies. Among the latter, special mention must be made of the Commission for International Cooperation in National Research in Demography (CICRED), an institution that, to promote scientific cooperation between countries, has adopted the goal of identifying the existing potential at the international level in the matter of research and teaching of demography and therefore of establishing and updating a repertory of centers. Between 1994 and 1997, CICRED realized an almost complete census of demographic centers, along with a special inquiry on the research programs of many of them (Gendreau and Huix-Adamets, 2003; Collomb, 2001). This allowed CICRED to establish a directory of centers of demographic studies and researches (Annuaire des Centers d'Etudes et de Recherches en Démographie [ACERD]), of which CICRED has already published two editions (1995, 2000). Since then, the directory has become available on the net and is continuously updated. It recent years, more and more centers have created their own web sites, greatly facilitating access to detailed information about their features, functions, activities, and scientific resources. The tools for the construction of an almost comprehensive picture of the situation are therefore available.[1]

Less than 20 years ago, the situation was quite different. The information available was scanty and incomplete, and specific literature was rare. This was partly because questions about scientific and pedagogic organization are not very attractive for researchers, but also because of the difficulties to be overcome to gather a sufficient documentation. Only recently have different international institutions and some researchers within them fully interested themselves in the subject and consequently arranged for the collection of information. A series of initiatives were organized (e.g., seminars, meetings, publications, directories) aiming to analyze and confront the existing situation in different parts of the world (Baldwin and Roy, 1997; Baldwin, 1996, CICRED-UPAS-ECA, 1995; CICRED, 1990, 1989; ECE, 1992; Tabutin, 1988; Jain, 1986; Cagiano de Azevedo, 1985; Martine, 1985; Farooq, 1985; De Graft Johnson, 1985; Federici, 1974, 1973, 1969, 1968, 1966).

The analyses presented here are based on these historical materials and on information provided by CICRED's ACERD. The whole of the documentation examined permits us to assess the existing structures in the field of population studies and suggests an initial general conclusion. The positioning of demography in the university structure appears quite varied and multiform as far as thematic and disciplinary links, teaching programs, and the ties with other disciplines are concerned. In a similar way, outside the university, studies of population or some of their aspects are developed in extremely varying contexts. Thanks to their vast thematic range, as much of a social nature (sociologic, anthropologic, cultural, economic, geographic) as biologic (epidemiologic, genetic, sanitary), population studies are combined with a number of other disciplines in centers devoted to research, teaching, scientific promotion and dissemination, collection of data, or the direction of programs of intervention in different sectors (e.g., birth control, health, family). From the juridical administrative point of view, these centers can be equally private companies or public enterprises or administrations; have a local, national, or international vocation; and be autonomous or depend on larger organizations. The research that they practice can extend from pure research to more or less applied studies to so-called operational research. The classic example of this latter category, more common in developing countries, is that of centers of family planning or of reproductive health, where scientific interest (when it exists) is quite marginal in relation to operational objectives. The extreme heterogeneity of demographic centers is accompanied by a marked geographic differentiation in numbers and characteristics.

This variety of forms and positioning, as much of the discipline as of the centers dealing with it, is an outcome of the history of the academic and scientific development of demography and of demographic events, which at different periods have aroused the interest of researchers and the preoccupations of politicians at the international or national level. To better understand the current situation, it is useful to retrace the origins and history of the principal institutions all over the world that have devoted themselves to the study of population. Only the main features of the narrative are addressed, with attention drawn to events that marked fundamental changes, because it is very difficult to summarize a multiform and geographically diversified history in a few pages.

[1] Without wanting to *a priori* define the notion of "demographic center," which would engage discussions as complex as lacking of general interest, criteria must be established for the collection of objective information and their quantitative analysis. The definitions adopted for the establishment of any repertory will therefore be specified when they are referred to.

I. THE ORIGIN OF INSTITUTIONS DEVOTED TO DEMOGRAPHY

1. From the Origins to the Foundation of IUSIPP in 1928: A Long History

Born with statistics, in which population was the first and the principal object, demography was for a long time considered as one of its branches. From the point of view of theoretical reflection, the birth certificate of demography can be seen in the work of Johann Peter Süssmilch (1707–1767), which constituted the first global and organized approach to the subject (Livi, 1940, p. 5). However, recognition of the new discipline's autonomy at an academic and pedagogic level appeared much later, in the second half of the 19th century. The first general assertion of the unity and independence of population studies dates from the first International Congress of Demography, organized in Paris in 1878 by Jacques Bertillon, Arthur Chervin, and Pierre-Emile Levasseur.

This Congress, which assembled researchers free of all official representation of their respective governments and which marked the adoption of the term *demography*, forged in 1855 by Achille Guillard for the designation of the new discipline, aimed at providing an exclusive space for the discussion of all subjects related to the "mathematics of population." Until then, these subjects could not be treated systematically except in the context of International Congresses of Statistics.[2] In this framework, population questions had been able to play a primary role, but only through the intermediary of official government representatives of participating countries (IUSSP, 1985; Dupâquier and Dupâquier, 1985). However, contrary to expectations, the Congress of Paris did not lead to the creation of a permanent organization of researchers on the subject. For a long time, the meetings of the new international association of statisticians, the International Statistical Institute (ISI), founded in 1885 with exclusively scientific aims, remained the only important forum where themes of population could be discussed, an increasingly unsatisfactory solution, because the field of statistical interests and methods was rapidly growing and the space left to demographic debate proportionately decreasing. During these same years (straddling the 19th and 20th centuries), demography had also found its space within international congresses of hygiene (which took the name of "Congress of Hygiene and of Demography"). This new alliance opened up broad horizons of research and gave a new impetus to the study of the biologic foundations of the demographic evolution, according to the openly naturalist vision of social phenomena that prevailed at the time. However, this sort of "strategic withdrawal" (Dupâquier and Dupâquier, 1985) translated the difficulty of the evolution of population studies into a disciplinary individuality.

The initial strong link with statistics and the proximity of interests with certain biostatistical sciences were not exclusive. The intrinsic multidisciplinary character of the object of study (i.e., population and its issues) favored the plurality of convergences and gains. If contiguity with the biostatistical sciences was linked beyond certain affinities of content and interests to a naturalist conception of the origins and future of human populations, seriously underevaluating the historicity of their evolutions, it can be affirmed that there also existed, since the beginnings, an even more intimate relation with the social sciences (Sgritta, 1994; Scardovi, 1994; Sylos-Labini, 1987). Demography had taken form within a melting pot of ideas common to all the social sciences, drawing inspiration from the same positivist conceptions, by observing populations and societies in the same theoretical perspectives, exploiting the same basic statistical materials with the same methods, so well that "the origins of demography and sociology, besides those of economy and statistics, appear, from several points of view, as essentially coincidental"[3] (Sgritta, 1994, p. 215), and for a long time, the disciplinary perspectives were hardly distinguishable.

Placed at a crossroads where the scientific contributions from different spheres of knowledge converged and drawn from the fields of human sciences and natural sciences, population studies attracted personalities from very heterogeneous cultural and disciplinary backgrounds and interests. The range of scientific interests of researchers who, until the first decades of the 20th century, can be placed in the category of demographers or who made important contributions to the development of demography is eloquent. Among them can be counted mathematicians, statisticians, actuaries, economists, and a number of sociologists, anthropologists, geographers, biologists, hygienists, geneticists, physicians, and epidemiologists. Since the beginning, the field of demographic research extended to current canonical themes (e.g., reproduction, mortality, mobility)—to all aspects that attach demographic facts to the biologic substrata on which they are based and to the social tissue of which they are an expression.

Until the end of the 1920s, there did not exist scientific organizations devoted to demography, nor was this discipline the object of independent university teaching. Most of the promoters of demography, whatever their cultural training, held administrative responsibilities consisting of coordination and management in the field of official statistics. It was to them, more than to university academics, and to the efforts they made to resolve concrete problems of analysis and interpretation of data, that the first theoretical and methodologic advances were due. Teaching of demography during the entire 19th century and the first 2 decades of the 20th century, when it did find a place

[2] The first of these Congresses was organized in 1853 in Brussels by Adolphe Quetelet under the patronage of the Belgian government. It had the specific object of harmonizing national statistics between themselves. This requirement of international cooperation, which followed the birth of the first national offices of statistics (since 1827) and the first national scientific organizations in the field (i.e., the Manchester Statistical Society was the first one founded in 1833, while the Statistical Society of London was founded the following year in 1834) gave way to numerous international conferences of a semi-official nature. The rejection in 1878 of this initial form of international organization of statisticians arose from difficulties born from the ambiguity of governmental positions with respect to the resolutions adopted in the frame of the conferences. In 1885, a society of an exclusively scientific character was founded in London. This was the International Institute of Statistics (ISI), which renewed different forms of international collaboration on new bases. The ISI published international statistics of population until 1939, from which date this function was filled by specialized agencies of the League of Nations, which itself was soon replaced by the United Nations Organization.

[3] "Le origini della demografia e della sociologia, come peraltro quelle dell'economia e della statistica, appaiono, sotto molteplici riguardi, sostanzialmente coincidenti" (Sgritta, 1994, p. 215).

in university studies, was done in the context of other disciplines to which demography was closely linked, because of the history of its disciplinary evolution. In particular, the teaching of demography existed within the faculty programs or schools teaching statistics. According to the conceptions of the period, statistics was also retained as a fundamental element in the economic and sociopolitical or geographic training. Instruction of demography could thus be found in certain programs organized within these disciplines. This strong integration between statistics and demography at the level of teaching has left traces in the university texts and manuals of the period. The chapters dealing with statistical methods and their applications are generally accompanied by a substantial discussion of demographic themes.[4]

It was in 1927 that the emancipation of demography and its separation from statistics really took shape at the international level during the organization of the first World Population Conference in Geneva. This initiative was taken in the context of malthusian preoccupations. The idea of organizing the Congress came from Margaret Sanger, an avant garde activist of the birth control movement, wishing to prove that the growth of the human population would be "creating social, economic, and political situations that threaten to profoundly alter our present civilization" (IUSSP/UIESP, 1985, p. 4). Despite this link, the Congress did not have a militant nature but was based on criteria of rigorous scientific neutrality. On the basis of agreements made on the occasion of the Geneva Congress, a permanent organization independent of governmental institutions was founded in 1928: the International Union for the Scientific Investigation of Population Problems (IUSIPP, now IUSSP), which set itself the mission of promoting and coordinating population studies in all their forms. This organization became the intellectual center of demographers and population specialists throughout the world.

This genesis of an institution with strictly scientific aims evokes the risks always underlying the intersection of scientific studies and preoccupations of a social and political order—legitimate and even appreciable preoccupations in principle, but carrying a serious danger of non-objectivity. The study of population can with difficulty be ideologically neutral, as is nature. The orientations of research are often suggested by the existence of contingent problems, for which solutions can be sought in harmony with considerations of an ethical or political nature, and it is for the rest quite licit and useful that the results of research offer elements for a governmental action founded on scientific knowledge. The entire history of the disciplinary assertion of demography is marked by the search for equilibrium between scientific independence with its risks of isolation, and an investment of its competence in the service of the *polis*, or society, with the dangers of subordination to practical necessities or to a political ideology. The contrast between these two contradictory aspirations characterized in many respects all the principal phases of the development of demography: from the initial search for scientific autonomy from state organizations that troubled its beginnings as a part of statistics to the creation of big institutions affiliated with the United Nations for education and research in developing countries, initially in the setting of a neo-malthusian ideology of population growth and its relations to economic and social development, until the current debate on the contribution that scientific approach must make in the search for political solutions to problems engendered by the levels of fertility.

2. From the Foundation of IUSIPP to the Second World War: The Assertion of Autonomy

Creation of the International Union for the Scientific Investigation of Population Problems in 1928 was a fundamental step in the world organization of population studies, with numerous fallbacks and consequences at the national level. It was on its trail that the first institutions were constituted that were devoted exclusively to the study of population. In the universities, the first courses specifically devoted to demography appeared and began to spread within the framework of training in statistics or the social sciences.

Sixteen countries (of which 12 are European: Belgium, Czechoslovakia, Denmark, France, Germany, United Kingdom, Italy, Netherlands, Spain, Sweden, Switzerland; 4 are American: Argentina, Brazil, Canada, United States) founded a National Committee of the Union in the years immediately after the creation of this institution. Some remained active very long, even after detaching themselves from the Union (e.g., the Italian Committee for the Study of Population Problems [CISP]), or still are active today (e.g., Dutch Society of Demography). Others were slowly

[4] For Italy, this affirmation is documented by a vast analysis led by Annunziata Nobile (1989) on Italian texts destined at the time to university teaching of demography. No similar study is known for other countries, but the bibliography established by the Dupâquiers in their *Histoire de la Démographie* (1985), although referring only very marginally to the production of university teaching in proper, supports the opinion according to which the situation described must be general.

transformed giving birth, in different ways, to the institutions forming the present configurations.

This period saw the appearance of certain journals that today are or have been among the principal means of expression of contemporary demography. *Population Index* was founded in 1933 by Frank Notestein (its final form[5] shaped by Ansley Coale) with the help of the American National Committee of the Union (created in 1928) and of the Population Association of America (PAA), which was born independently in 1931[6] (Hodgson, 1991). In 1964, *Demography*, the current journal of the PAA, was born. In 1933, the British journal *Population* appeared, supported by the British Population Society, but it, like the association itself, had a very short life and was replaced in 1947 by *Population Studies*, in the framework of the Population Investigation Committee, founded in 1936 by David Glass and Eugene Grebenik. The Italian journal *Genus* appeared in 1934, published by the CISP. The latter was created in 1928 by Corrado Gini, first as a section of the IUSIPP and then as an autonomous committee. The French journal *Population* appeared in 1945 as an official publication of INED, founded the same year by Alfred Sauvy on the ruins of the French foundation for the study of human problems created in 1941 by Alexis Carrel (Girard, 1995; Bogue, 1993; Crimmins, 1993; Keyfitz, 1993; Hodgson, 1991). Only recently, with the ramifications and the fragmentation of demography and with the arrival of new issues and the enrichment of the methodologic tools, more specialized reviews began to proliferate, devoted to particular aspects of the discipline. The most well known is *Population and Development Review*, founded in 1974 at the Population Council by Paul Demeny; it covers a vast horizon of themes and perspectives linked to population developments. Without citing all of them, mention must be made of *Theoretical Population Biology*, *Population Research and Policy Review*, *Population and Environment*, and *Social Network*. Today, the *Revue des Revues Démographiques*, edited by the CICRED, lists about 90 journals in the world (15 of which are in the United States) that are totally or partially devoted to population issues.[7]

Between the 1920s and 1930s, demography began to acquire the status of an autonomous university discipline, by positioning itself in different ways in the university systems, according to a set of local conditions. The current variety of positions of demography in the framework of university education resulted from the initial heterogeneity and cultural diversity of the scholars devoted to demographic themes. In each country, the independence of demography followed different paths that were marked by the scientific personality of those who strove to introduce the subject into university programs. The general cultural heritage that underlies the construction of the different pedagogic systems, and evolutionary rhythms of other disciplines, especially sociology, was also very important. The links between population studies and sociology were generally very close, even at the level of university organization, where sociology, born of currents of thought such as positivism in France, empiricism in England, and historicism in Germany, had had a precocious development or had enjoyed growth at the right moment. This happened in the United States in the interwar period, when in Russia, Italy, and Germany, dictatorial regimes attempted to suppress the teaching of sociology. In the first case, the accord between the two disciplines was favored by the natural intersection between the objects of study and the common recourse to empirical methods. In the latter, contacts remained feeble, even nonexistent. In Italy for example, the long cultural predominance of an idealist philosophy proved to be an obstacle to the development of sociology, and it made its discourse abstract and foreign to the empirical and statistical methodology characteristic of demographic analysis (Sgritta, 1994, p. 227). In such conditions, reinforced by the political context, no privileged relationship developed between demography and sociology, whereas the links between demography and statistics were permanently reinforced, as much at the level of research as of university education. These different relations between demography and allied disciplines have not failed to influence the frontiers and orientations of population research, as well as the type of education given to demographers of different countries, in a varied but enduring manner.

3. From the Second World War to the Present: A Short but Tumultuous History

On the brink of the Second World War and after the recreation of the Union in 1947 under the name of the International Union for the Scientific Study of the Population (IUSSP), the broad structures of research and teaching of demography were instituted and permanently installed in Europe and in North America. Even in this geographic area (the only one that had until then contributed to the disciplinary maturity of

[5] Alas, *Population Index* does not exist anymore.

[6] Sponsored by Margaret Sanger, it found itself at the beginning in competition with the National Committee of the Union.

[7] In reality, among the journals considered by the CICRED, at least 10 deal with different themes (e.g., statistics, economy, epidemiology, sociology) but frequently devote particular attention to related population issues.

demography), the situation was still quite uneven, and in certain countries where the studies on population had nonetheless borne fruit among scholars, there did not yet exist specialized centers of research on population nor independent university teaching of demography. Overall, however, the spaces for academic insertion, the necessary disciplinary alliances, and the relations with governmental institutions became clearly defined.

With the exceptions of some cases where, after the war, demography knew a more or less long period of stagnation or of crisis,[8] the 1950s and 1960s were most often years of definitive affirmation and of university expansion. The different demographic evolutions that played a prominent role in the socioeconomic transformations of many countries oriented the interest of researchers to themes likely to awaken the concerns of politicians and public opinion: the low fertility in Great Britain and in France, the international migrations into the United States, the internal migrations in Italy, and especially, the acceleration of world population growth, destined to have an immense influence on the development of demography on the scientific plane and in institutional settings. The latter experienced the full consolidation of its structures of research and teaching in the developed countries, and it began to implant itself in the developing countries and very rapidly spread there, although it was still almost completely absent in the immediate aftermath of the war. However, this did not signify that all the developed countries, especially European states, would henceforth be equipped with pedagogic structures in sufficient number and quality. It can be said that the opinion advanced at the beginning of the 1970s by Nora Federici (1974, p. 27), according to which, in many a case, an acceptable situation had not yet been reached at the level of the training of professional demographers remains valid, even if the demand for this kind of competence in the European labor market is quite limited.

In the second half of the 20th century, the emergence of Asian, Latin American, and African countries on the international scene and the epic of their development was accompanied by an acceleration of their population growth that was without historical precedent, long felt to be a danger as much for the global equilibrium of the planet's resources as for the perspectives of each countries development. The population dynamics of the Third World became the focus of an international scientific debate, and important means were mobilized for the diffusion of knowledge of the subject and research on paths that would permit the modification of the course of events. The most widespread opinion in the 1950s and 1960s was favorable to interventions for the promotion of birth control, particularly in countries such as the United States or Sweden, where there already existed a long tradition of activism on this issue. It was in this climate that a movement of institutional reinforcement and creation of new centers of population studies in developing countries took place. The principal actors were the big international institutions, but significant contributions also came from bilateral and multilateral cooperation. In the United States, enormous financial resources, as much private as public, were mobilized in the context of bilateral cooperation. This was also the case in countries such as the United Kingdom, France, Belgium, which had interests and special ties inherited from their colonial past with the Third World. A number of other countries had offered their support to the projects of multilateral cooperation, whether coordinated or not by international organizations.

In the United States, these international orientations had extremely important effects at the national level, as John and Pat Caldwell demonstrated (1986) in their well-known book on the contribution of the Ford foundation to the growth of centers of American population studies. The abundance of funds directed to population research in the Third World countries and to collaboration with researchers of these countries[9] gave the centers of research and demographic training an unprecedented force and independence, especially within the sociology departments of the universities, and it enabled their numbers and scientific prestige to increase. Even in the absence of other equally detailed case studies, it seems that an analogous, if more limited, process also took place in some universities of other developed countries (e.g., Great Britain, France, Belgium, Canada, Australia). In any case, this is what is suggested by the abundant literature available in these countries on the demography of the Third World, the importance of specific methodologic works in the collection and analysis of data of developing countries

[8] Essentially, it was Italy and Germany that were concerned where there was defiance toward a discipline that had somehow compromised with regimes backing populationist and racist policies. It had consequently attracted to itself a serious scientific discredit. This was also the case of Soviet Russia, where the ostracism against the discipline decreed by the regime in the 1930s would continue (Vichnevski, 1995).

[9] This availability was born essentially from the orientations of private foundations. such as the Ford Foundation and similar organizations, until the 1950s. It was extended in the following decades thanks to governmental institutions. The Kennedy Administration officially declared itself concerned about the control of population growth in developing countries, and it was open to finding financial means to achieve this end (Caldwell and Caldwell, 1986, p. 98).

(e.g., demographic surveys, analysis of defective data), and the institution of numerous courses aimed specifically at satisfying the particular needs of students coming from the developing world.[10]

The United Nations manifested impressive activity in the field of demography, directly and through its regional agencies (i.e., Economic Commission for Africa [ECA],[11] Economic and Social Commission for Asia and Pacific [ESCAP],[12] Economic and Social Commission for West Asia [ESCWA],[13] and Economic Commission for Latin America and the Caribbean [ECLAC][14]), all strongly implicated in the conception and application of population policies, or through their specialized agencies, many of whom (e.g., Food and Agriculture Organization [FAO], United Nations Educational Scientific and Cultural Organization [UNESCO], World Health Organization [WHO],[15], International Labor Organization [ILO][16]) handled issues closely linked to demographic questions. With regard to these various international organizations, two fundamental aspects should be emphasized. The United Nations, especially through its Population Division and its Statistics Division, fixed as its essential goal a task earlier entrusted to the ISI of producing internationally comparable statistical data and disseminating them. From 1945, the Statistics Division yearly published the *UN Demographic Yearbook*, and the Population Division has published an impressive quantity of works of demographic analysis, whether in its series of *Population Studies* or in other collections, along with the *World Population Prospects*, which are regularly updated. They also carried out another crucial mission, that of helping scientific authorities of developing countries in the institution of an educational system in demography through the creation of five important organizations with a regional vocation. They are: the International Institute for Population Studies (IIPS) in Bombay in 1956; the Latin America Demographic Center (LADC) in Santiago in Chile in 1958; the Cairo Demographic Center (CDC) in Cairo in 1963, with inter regional competences for the countries of the Middle East and of North Africa; the Regional Institute for Population Studies (RIPS) in Accra in 1971 for the Anglophone countries of sub Saharan Africa; and the Institut de Formation et Recherche Démographique (Institute of Training and Demographic Research [IFORD]) in Yaoundé in 1972 for the Francophone countries of sub-Saharan Africa. To these must be added the creation in 1988 of the Population and Human Resources Development in Africa of the Institute for Economic Development and Planning[17] (PHRDP/IDEP) in Dakar.

To financially support all these activities initially entrusted to the UN Population Division, a relatively independent fund of the United Nations was created on the basis of voluntary contributions of state members (see Chapter 120). In the 1970s, the United Nations Fund for Population Activities (UNFPA) was born, and it quickly took the leadership of most of the population activities of the United Nations, even if its original mission (i.e., the institution of services of family planning and more generally of reproductive health in developing countries) remained its first priority (UNFPA/FNUAP, 1977, 1987, 1995). The UNFPA invested extensively in the development of research and training in demography through a Global Program of Training in Population and Sustainable Development, intended for persons engaged in administrative or governmental responsibilities in developing countries.

The fear of the consequences of a population explosion, especially in the field of economic development, and the search for a solution by birth control policies first appeared among researchers of developed countries (e.g., especially the United States, Sweden), whose debates and research issues they nourished. These concerns, the same from which the principal theories about the relations between population dynamics and society were born and developed, from the malthusian and neo-malthusian propositions to the framework of demographic transition and the microeconomic theories of the family, were adopted very slowly by the scientific and political personnel of developing countries, usually under the pressure of the dominant ideology and as a result of long-term international debate. Three world governmental conferences organized by the United Nations on the theme of population and development (Bucharest in 1974, Mexico in 1984, Cairo in 1994) marked this debate and reflected the evolution of demographic-economic

[10] A look through the issues of the IUSSP Newsletter can provide a good idea of the topic.

[11] Headquarters of the Economic Commission for Africa are in Addis Ababa, Ethiopia.

[12] Headquarters for the Economic and Social Commission for Asia and Pacific are in Bangkok, Thailand.

[13] Headquarters for the Economic and Social Commission for West Asia are in Beirut, Lebanon.

[14] It then became Economic Commission for Latin America and the Caribbean (ESCALPC), with headquarters in Santiago, Chile.

[15] Headquarters for the World Health Organization are in Geneva, Switzerland.

[16] Headquarters for the International Labor Organization are in Geneva, Switzerland.

[17] "Population, ressources humaines, et développement en Afrique de l'Institut de développement économique et de planification."

thought toward a world consensus that oriented the strategies and activity of international cooperation, one of whose ambitions was to reinforce the bodies of research and teaching of demography in the Third World countries.

In this process of increasing awareness of population problems at the world level, the United Nations played a fundamental role supporting the dissemination and institutionalization of demography in the less developed countries, and it remains today at the center of a network of specialized agencies and international institutions (including the World Bank), which, in a more or less coordinated way ensured a technical-scientific assistance and financing for the research activities and teaching of demography. In particular, the five demographic regional centers previously mentioned developed competences, organized trainings at different levels for nationals in their geographic areas, published didactic works and other pedagogic instruments conceived in relation to specific local needs, locally organized occasions for meeting and discussion among researchers, cooperated with national institutions in the field of collecting data and improving their quality, and in certain cases, even undertook research and published the results in books or articles. In short, these centers constituted focal points for the diffusion of demographic knowledge and had, thanks to their activities of support, an important effect of multiplying efforts made by the different countries to reinforce or create their own research institutions and courses of demography. Once they had accomplished their mission, or because the source of their international financing had been reduced,[18] these centers had to merge with local institutions that they had had the mission of creating or helping. In this way, their skills, especially in the matter of teaching, were progressively absorbed by the national institutions.

The birth of more or less independent institutions, more precocious in Asia and in Latin America and much later in Africa, was effectively intensified in the second half of the 1970s, after the Bucharest Conference. A complete and well-documented panorama of the structures and pedagogic initiatives of the period, drawn up on this occasion by Nora Federici for the United Nations, underlined the great deficiency that still reigned at the beginning of the 1970s, although geographically uneven. If the universities that offered classes of demography in the departments of economy, sociology, geography, and statistics were already numerous and if several attempts had been made to insert elementary notions of demography in secondary teaching or in adult literacy classes, the countries where there existed a complete post-university training program, expressly destined to form demographers, were rare. This was true even of Asia, where the interest in family planning had manifested earlier, drawing attention to the usefulness of research and teaching of demography. The lack was more understandable in Latin America, where opposing points of view clashed on the problem of population growth, and in Africa, whose population growth was not yet as lively and which had benefited from only a very thin development of its university structures (Frederici, 1974). But growth was soon accelerated everywhere.

In the decade of 1975 to 1985, many countries of Latin America, Asia, and Africa equipped themselves with structures of research and training in demography and were able to offer courses leading to Masters' degrees and doctorates.[19] In general, it is estimated that the possibilities offered at the national or regional level were henceforth adequate for the training of traditional demographers, especially for the needs of official statistics and collection of data (Cagiano de Azevedo, 1985; Miro and Potter, 1980). Nonetheless, an important effort remained to be made to reinforce education on the theme of "population and development" with more interdisciplinary contours and more diverse and complex applications, essentially answering the needs of socioeconomic planning and population policies (Farooq, 1985, p. 300). These questions of frontiers are well in the line of disciplinary maturity of demography. After a period of concentration on the creation and improvement of data collection systems of the corpus of methods necessary for the "construction of a formal architecture of functioning of the demographic system"[20] (Livi Bacci, 1994), priority was given to the study of relations between the demographic processes and transformations of society.

The history of the next 20 years until today is that of a progressive reinforcement of university structures in developing countries, accompanied by improvement of the quality of studies and, in some cases, of

[18] According to a close observer of the evolution of the situation at CELADE, the crisis that hit these big centers from the middle of the 1980s was "possibly as a result of the fertility decline, which led to decreasing interest on the part of donors" (Martine, 1985, p. 102).

[19] In 1985, in his examination of the situation in Latin America, George Martine affirms that "Overall, however, the doctoral programmers that are beginning in Brazil and Mexico are as strong as many offered in developed countries, with the advantage of being tailored to problems specific to the developing countries' problems (1985, p. 110).

[20] "Costruzione di un'architettura formale del funzionamento del sistema demografico" (Livi Bacci, 1994).

catching up on levels of excellence. Without wishing to generalize the exceptional cases of China or of India, where demography had extraordinary success, it can be observed that in nearly all the countries of the world (and in all the big countries), there are now several centers of population studies within or outside universities.

The case of developed countries was different, especially in Europe, where an unprecedented proliferation of centers of population studies took place, perhaps partly created by the phase of expansion experienced by universities. Nonetheless, because of the particular fluidity of the context attributable to restructuring in university systems, the future of research and teaching of demography is not without problems. In this case, the place occupied by demography in the officially recognized university programs demands attention.[21]

II. THE CURRENT SITUATION

CICRED's ACERD, available on the Internet, listed 708 centers of research or demographic training as of September 2003, 147 of which are in Africa, 84 in Latin America, 18 in Australia and Oceania, 76 in North America, and 217 in Europe.[22]

The importance of these figures (among which CICRED judged it right to include, in some cases, the headquarters of international institutions present in different countries[23]) depends on the definitions adopted and on the care that persons contacted took to provide information. Cases of omission cannot be excluded nor can doubtful insertion of centers in the repertory. Even without wishing to give the figures presented later more meaning than they have,[24] we can base our ideas on the general picture provided by this enumeration, which CICRED considers as "quasi-exhaustive" (Gendreau and Huix-Adamets, 2003, p. 2). In particular, Table 142–1 leaves no doubt about the reality of the reinforcement and the dissemination of the demographic discipline during the past 25 years. Comparison with two of the earlier repertories of CICRED (1973 and 1995) shows a quadrupling of the number of centers in 30 years, with much more rapid growth in places where they were earlier more uncommon (i.e., a multiplication by 11 in Africa). Even if a part of this growth can be attributed to an improvement of the comprehensiveness of the repertories, the founding dates of the centers are there to prove the actual expansion; in Africa, about 75% of the centers whose foundation date is given were created after 1970, as is the case for 57% in Latin America, 73 % in Asia and Oceania, 69% in Europe, and 54% in North America. If the history of demography is already old by some centuries, its scientific and institutional autonomy date back only to the 1930s, and its general dissemination at the world level began only in the 1960s.

ACERD reports of the names and addresses of different centers allow us to distinguish university institutions from others. University institutions can be distinguished from others. For the former, the university status (i.e., department, institute, or other) and the type of institutional disciplinary alliance can be identified.[25] For the latter, a typology distinguishing public or private institutions, governmental offices, departments of national institutes of statistics, international organizations, and others can be established.[26]

The university is the principal institutional place of research and teaching of demography (Table 142–2). Without wanting to give an exaggerated weight to numeric aspects, most of the centers devoted to this discipline reside within academic structures.

Outside the university, institutions and organizations of quite varied kinds covering a large range of scientific and operational needs are found. In some

[21] The International Symposium 1945–1995 that celebrated INED's 50th anniversary in Paris in October 1995 (Chasteland and Roussel, 1997) reviewed the disciplinary evolution of demography in a number of industrialized countries. In most of the cases, the stages of institutional organization of the discipline were recalled (see the papers of Alain Desrosières on France; Gérard Frinking on the Netherlands; Andras Klinger on Hungary; Christopher Langford on Great Britain; Roland Lee on the United States; and Anatoli Vichnevski on Russia), as well as future perspectives (see the papers of Maire Ni Bhrolchain on Great Britain; Samuel Preston on the United States; Frans Willekens on the Netherlands; Graziella Caselli on Italy; Alexandre Avdeev on Russia; Jean-Claude Chasteland and Louis Roussel on industrialised countries in general). For the history of demography in Italy, we also refer to Massimo Livi Bacci's paper (1994).

[22] CICRED considers a demographic *center* a "unit of research or national studies on population, whose existence is proved by the title of center and by a specially nominated head and consisting of at least two full-time people (or its equivalent of part time)" (Gendreau and Huix-Adamets, 2003, p. 5).

[23] In most cases, the international institutions interested in questions of population are not considered to be "centers of demography," but they are ranked apart in the category of "other useful addresses."

[24] A report of the United Nations Population Division on the programs of population studies in various developing countries identified 181 centers of teaching in Asia, 111 in Latin America, and 75 in Africa, very high estimates that the authors of the report themselves considered as over-estimated figures (Baldwin, 1996).

[25] According to CICRED's definition, a department of sociology could constitute a "center of demography," just like a department of demography.

[26] The classification of centers presented in Tables 142–2 and 142–3, which is uniquely based on the denomination of centers, can contain arbitrary elements that cannot be controlled. It can, however, be considered as significant in its broad lines.

TABLE 142–1 Lenters of Population Research and Teaching of Demography, by Region (Several Years)[a]

Years	Africa	Latin America	Asia and the Pacific	North America	Europe	Total
Absolute values						
1973[b]	13	24	61	22	40	160
1995[c]	110	45	160	75	198	508
2003[d]	147	86	182	76	217	708
Relative values (%)						
1973[b]	8	15	38	14	40	100
1995[c]	19	8	27	13	33	100
2003[d]	21	12	25	11	31	100
Average year of creation[e]						
Year	1978	1967	1975	1971	1972	1973

[a] The CICRED considers *centers of demography* any institution where at least two people work full time on population questions (see footnote 22).
[b] Data from CICRED (Gendreau and Huix-Adamets, 2003, p. 7).
[c] Data from CICRED Repertory 1995 (Maffioli, 1997, p. 303).
[d] Data from Gendreau and Huix-Adamets (2003, p. 7).
[e] Data from Gendreau and Huix-Adamets (2003, p. 12).

TABLE 142–2 Centers of Population Research and Teaching of Demography Listed by CICRED'S ACERD, Classified According to the Denomination of Institutions, by Region

Institutions of affiliation	Africa	Latin America	Asia	Oceania	North America	Europe
Absolute values						
Universities	49	37	78	11	50	104
Public and private organizations	30	27	52	2	17	70
Ministries and governments	33	10	17	3	6	13
Among whom, bureau of statistics inside ministries or equivalent	*12*	*4*	*4*		*2*	
National institutes of statistics	18	8	8	2	1	20
International organizations	12	2	7		2	4
Scientific associations	4		2			3
▪ Total	146	84	164	18	76	214
Relative values (%)						
Universities	33.6	44.0	47.6	61.1	65.8	48.6
Public and private organizations	20.5	32.1	31.7	11.1	22.4	32.7
Ministries and governments	22.6	11.9	10.4	16.7	7.9	6.1
Among whom, bureau of statistics inside ministries or equivalent	*8.2*	*4.8*	*2.4*		*2.6*	
National institutes of statistics	12.3	9.5	4.9	11.1	1.3	9.3
International organizations	8.2	2.4	4.3		2.6	1.9
Scientific associations	2.7		1.2			1.4
▪ Total	100.0	100.0	100.0	100.0	100.0	100.0
Number of countries	50	24	30	5	3	36
Population (in millions)[b]	796	520	3680	31	316	728
Centers/countries	2.9	3.5	5.5	3.6	25.3	5.9
Centers/population (for 10 millions)	1.8	1.6	0.5	5.8	2.4	2.9

[a] Nationals and foreigners.
[b] The data on the population of countries are taken from United Nations (2003).
Data from the repertory CICRED's ACERD, consulted in April 2004.

cases, we found entities entirely devoted to research of a broad range in all the sectors of demography. They enjoy public financing and are strongly linked to the university with which they cooperate on scientific and pedagogic levels. The most well known examples are the National Institute for Demographic Studies (INED) in France, the Netherlands Interdisciplinary Demographic Institute (NIDI) in the Netherlands, the Institute for Research on Population (IRP, recently rebaptized the Institute for Research on Population and Social Policies [IRPPS]) in Italy, and the Max Planck Institute for Demographic Research (MPIDR) in Germany. Similar centers constituted in the framework of academies (e.g., Academy of Sciences, Academy of Economy, Academy of Medicine) are common in Eastern Europe. Most often, it is a matter of independent centers with public or private financing (e.g., foundations) that are inclined toward some particular aspects of the study of population, such as the family, migrations, health, and family planning.

In many countries, units of population research can be found at the heart of national institutes of statistics and are responsible for the census and vital statistics. This respects an old tradition and at the same time answers the basic need of ensuring the link between the collection of data and the orientations of research. Another type of population centers, very widespread in Africa and to a lesser extent in Asia, are those that function within ministries with economic competences (especially economics, planning, development, cooperation) or with social competences (health, work, family, social policy) or other governmental structures. Their principal interest is applied research, aimed at making available to the government knowledge useful for the integration of demographic variables in development plans or even for the formulation, application, and evaluation of policies in the matters of health, reproduction, family, and distribution of population on territory. Their presence in many African countries (where they constitute 12% of the centers identified by ACERD) are proof of the growing interest in population dynamics as a development factor shown by political decision-makers.

Another category of centers is constituted by international organizations, for the most part established within the United Nations. They represent about 4% of centers of ACERD (8% in Africa), but their importance goes well beyond this numeric aspect. They have played an essential role for a long time in the dissemination of demographic knowledge, and they are equally important in the field of harmonization and production of data. It is enough to mention here the backing given, also in recent years, to important data collection operations such as the World Fertility Survey (WFS), the Demography and Health Surveys (DHS), and the censuses.

Mention must finally be made of a small number of scientific associations that do not fir into any of the preceding categories. They do not directly exercise teaching or research functions but can develop activities of great importance for the promotion, coordination, and orientation of scientific discourse, especially through the organization of occasions of exchanges and debates among researchers. The history of the development of the demographic described has shown the role of catalyst played by the IUSSP. Along with the IUSSP, there exist a number of national scientific associations (the most prestigious is the Population Association of America, which supports the journal *Demography*) or regional groups, such as the European Association for Population Studies (EAPS, supporting the *European Journal of Population*) and the Union for African Population Studies (UAPS) that take important initiatives relating to their geographic zones.

III. UNIVERSITY CENTERS OF TRAINING AND DEMOGRAPHIC RESEARCH

1. Status and Organizational Structures

The configuration of centers of research and teaching of demography within universities is closely linked to the acknowledged ideas about the aims of demographic training. The specific professional competences of the demographer correspond to a restrained demand in the labor market, which limits university programs for obtaining a Bachelor's degree in demography, as is the case for economics or for sociology. This is why almost no university program is entirely devoted to demography. Instead, the teaching of demography aims at providing highly specialized professional preparation (i.e., postgraduate study) or, on the contrary, complementary education in the framework of more general learning itineraries, especially in the universities of developed countries. In other words, it is most often intended for producing future teachers or researchers, but it also aims at creating a general awareness of population issues among those operating in various fields in which demographic processes have an influence. The training of professional demographers with competences ranging from the assessment of demographic parameters useful for the planning of public and private services to the formulation, application, and evaluation of population policies is less widespread, except in the

United States, but these programs are expanding in developing countries (Palloni, 2002; Maffioli, 1994).

These conceptions of the profession of demographer have repercussions in the academic positioning and scientific organization of centers interested in demography. Departments or other autonomous university institutions exclusively devoted to demography are rare. Centers or individual researchers devoted to this discipline often are integrated in departments of sociology, statistics, economics, geography, history, epidemiology, or genetics. The situation varies from one country to another or even from one university to another within the same country, according to the disciplinary alliances that were forged during the course of history. All the same, the recent developments in world populations and the effects they have had on the public opinion and on policy makers, including through the activity of international agencies and the flourishing of specific studies, have reinforced the means and prestige of many research and demographic training institutions, so that in universities where organization is more flexible, they have succeeded in acquiring some independence.

To analyze these aspects of university positioning of demography, the centers listed by ACERD have been classified according to the disciplines figuring in their denomination. In the analysis that follows, the centers whose name explicitly mentions demography or one of its branches, alone or in association with other areas of studies, are opposed to those whose name makes an exclusive reference to other disciplines (see notes to Table 142–3). The former are considered as having the unique objective or priority of demographic research, contrary to the latter, where the studies of population represent only a more or less important part of scientific activity. The degree of independence of the first kind of centers has also been taken into account; centers that are directly affiliated to their parent University and do not constitute a section of an intermediary body were defined as autonomous.

Tables 142–3, 142–4, and 142–5 provide a view of the spread of demography in the most varied contexts. Among ACERD's centers of demography, a considerable proportion (41% at the world level, 52% in Europe) are structures whose principal object of studies is not demography; these are departments dealing with sociology (10%), geography (9%), statistics (8%) economics (4%), or other disciplines, only a part of whose scientific activity concerns population. Moreover, it can be observed that the centers devoted principally or uniquely to demography or its branches (59%) are often also closely linked to other disciplines, in the sense that they do not constitute autonomous university entities but form part of structures with larger scientific interests, among whom sociology predominates, Only about 27% of the centers (36% in North America) constitute autonomous units specifically devoted to demography or to demography associated with another subject, and they are often authorized to bestow degrees.

Particular mention must be made of centers (15.8%) in whose denomination demography figures in association with other disciplines (e.g., demography and social sciences, population and development, population and resources). The number of these centers is expanding. According to an earlier analysis based on the CICRED data of 1995 (Maffioli, 1997), less than 10 years ago, only 27 centers of this kind could be counted, compared with 52 listed in ACERD in April 2004. This tendency toward the creation of units of research and teaching where demography is in the foreground but in a larger framework of scientific interests, results from a new university strategy that refuses marginalization within large organizations and autonomy that carries the risk of isolation; it also seems to correspond to an evolution of research toward issues that are less strategy on the core relations internal to demographic processes and more focused on the study of relations between these processes and other aspects of social dynamics. In principle, the closest alliances forged in these mixed centers should have multiple positive effects. The interdisciplinary collaboration should be beneficial to research and to teaching, by encouraging scientific exchanges and maybe permitting to escape from the intellectual isolation from which may affect demography because of its role of a minor discipline in relation to the principal field with which it is allied. The teaching of demography, placed in the foreground but not closed in on itself and open to contents chosen with respect to programs of studies in which it is enrolled, should permit validation of the discipline's educational potentialities, as much on the professional level as on the cultural level. In the same way, research oriented toward exogenous and larger issues favoring interdisciplinary aspects should contribute to the advancement of knowledge and also consolidate the relations between demography and allied disciplines. However, it is necessary to ensure that this circular process is effectively realized in the daily practice of teaching and research of mixed centers and does not remain at the level of aspirations.

2. Teaching in University Centers

If research on questions of population has developed in a great variety of institutions, the teaching of demography is handed over in an almost exclusive way to university centers with diverse aims. It aims at

TABLE 142–3 University Centers of Research and Teaching of Demography Listed by CICRED's ACERD, According to Their Denomination[a]

Discipline displayed in the denomination	Africa	Latin America	Asia	Oceania	North America	Europe	Total
Absolute numbers							
Demography (or its fields) only	16	14	46	4	26	36	142
Among whom, autonomous structures	*12*	*10*	*16*	*1*	*11*	*20*	*70*
Demography and other	12	4	7	2	13	14	52
Among whom: autonomous structures		*1*	*4*	*1*	*7*	*7*	*20*
Social sciences, sociology, anthropology	7	4	4	1	4	14	34
Economics, management, econometrics		2	1	1	2	6	12
Statistics		1	4		1	8	14
Development and resources	1	2					3
Geography, town planning, urbanism, environment	9	3	5		1	13	31
Epidemiology, public health medicine, etc.	1	1	4	3		8	17
Others[b]	3	6	7		3	5	24
▪ Total universities	49	37	78	11	50	104	329
Relative values (%)							
Demography (or its fields) only	32.7	37.8	59.0	36.4	52.0	34.6	43.3
Among whom, autonomous structures	*24.5*	*27.0*	*20.5*	*9.1*	*22.0*	*19.2*	*21.3*
Demography and other	24.5	10.8	9.0	18.2	26.0	13.5	15.8
Among whom, autonomous structures		*2.7*	*5.1*	*9.1*	*14.0*	*6.7*	*6.1*
Social sciences, sociology, anthropology	14.3	10.8	5.1	9.1	8.0	13.5	10.3
Economics, management, econometrics		5.4	1.3	9.1	4.0	5.8	3.6
Statistics		2.7	5.1		2.0	7.7	4.3
Development and resources	2.0	5.4					0.9
Geography, town planning, urbanism, environment.	18.4	8.1	6.4		2.0	12.5	9.4
Epidemiology, public health, medicine, etc.	2.0	2.7	5.1	27.3		7.7	5.2
Others[b]	6.1	16.2	9.0		6.0	4.8	7.3
▪ Total universities	100.0	100.0	100.0	100.0	100.0	100.0	100.0

[a] The centers of *demography* (in the CICRED meaning) constituted within universities or equivalent higher institutes are classified according to the disciplines displayed in their denomination. Those that are directly affiliated to the university and do not form part of an intermediary body are considered autonomous. For example, under the rubric of "sociology," departments or faculties of sociology are included, which constitute centers of demography because they shelter demographers. The rubric "only demography" features departments of demography and other institutes of population studies, autonomous or organized in the frame of larger university entities (e.g., faculty, departments) devoted to other disciplines. The rubric "demography and others" includes institutes of demography and social sciences, statistics and demography, population and development, and population and resources in general (often in Africa) inserted within departments or faculties whose scientific field is larger.

[b] Including mixed situations (essentially economics and sociology or sociology and statistics).

From CICRED's ACERD data, consulted in April 2004.

TABLE 142–4 University Centers of Population Research and Teaching of Demography According to Their Denomination and Their Degree of Autonomy, by Region (%)

Discipline displayed in the denomination	Africa (%)	Latin America (%)	Asia (%)	Oceania (%)	North America (%)	Europe (%)	Total (%)
Demography (or its fields) alone or mixed	57.1	48.6	67.9	54.6	78.0	48.1	59.0
Among whom, with mixed denomination	*24.5*	*10.8*	*9.0*	*18.2*	*26.0*	*13.5*	*15.8*
Among whom, autonomous structures	*24.4*	*29.7*	*25.6*	*18.2*	*36.0*	*26.0*	*27.4*
Others	42.9	51.3	32.1	45.4	22.0	51.9	41.0
▪ Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

From CICRED's ACERD data, consulted in April 2004.

TABLE 142–5 University Centers of Population Research and Teaching of Demography, with Nondemographic Denomination, According to the Reference Discipline, by Region (%)

Discipline figuring in the denomination	Africa (%)	Latin America (%)	Asia (%)	Oceania (%)	North America (%)	Europe (%)	Total (%)
Social sciences, sociology, anthropology	33.3	21.1	16.0	20.0	36.4	25.9	25.2
Economics, management, econometrics		10.5	4.0	20.0	18.2	11.1	8.9
Statistics		5.3	16.0		9.1	14.8	10.4
Development and resources	4.8	10.5					2.2
Geography, town planning, urbanism, environment	42.9	15.8	20.0		9.1	24.1	23.0
Epidemiology, public health, medicine	4.8	5.3	16.0	60.0		14.8	12.6
Others	14.3	31.6	28.0		27.3	9.3	17.8
▪ Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

From CICRED's ACERD data, consulted in April 2004.

offering access to the levels of the highest specialization (a research doctorate in particular) and providing specialized knowledge in the framework of programs of exogenous studies.[27] Complete training in demography is given in general only by departments exclusively devoted to this subject. Demography as a complementary program is common and assumes varied forms and contents, sometimes weakly connected to the branch of affiliation.

A university education entirely focused on demography is not generally foreseen, but the idea is beginning to take root, especially in developing countries, although in a nonsystematic manner and with variable contents, often linked to available didactic competences rather than a precise program of teaching.

A more flexible kind of education, with the promise very differentiated programs leading to a broad range of degrees, is in the process of being instituted in an attempt to satisfy very different needs. A number of classes, not necessarily leading to formal degrees but offering the minimum of necessary demographic knowledge for the accomplishment of their missions to persons implicated in administrative activities or governmental programs related to questions of population, are now available in universities and other institutions and international organizations.

Classes on demography given at university centers generally offer a large range of subjects, particularly in developed countries. The contents have tended to diversify over the years in the course of the discipline's evolution, with enrichment of its methodologic corpus and the acquisition of an increasingly broad body of knowledge in different fields related to demography. Besides basic teaching on the foundations of population dynamics (e.g., demographic analysis, population theories, mathematical demography, other technical aspects), subjects such as historical demography, population economics, social demography, demography of the family, regional demography, and many more testify to the consolidation of the discipline's more recent advances and also to the exploration of new approaches that have not yet found their definitive systematization. Some fields of teaching are still more fluid, mixing education and research in progress, because they correspond to the growing interest of researchers for the study of relations between behaviors and demographic processes and the social and cultural contexts in which they manifest themselves, such as gender and population, population and development, or population and environment. These newer kinds of courses, which are in the trail of recommendations adopted at the World Conference on Population and Development in Cairo in 1994, have spread rapidly, especially in places where the organization of classes is the most flexible and the necessary teaching skills are available.

The teaching of *applied demography* (i.e., business demography) has developed over the past 20 years, particularly in the United States (Keyfitz, 1987a). This new branch of the discipline aims at giving professional competency by teaching students to apply

[27] The information on the principal programs of European demographic training is taken up in the *Réseau européen de formation en démographie* of the Catholic University of Louvain's Institute of demography (Louvain-la-Neuve, Belgium). The European Economic Commission (EEC) of the United Nations in Geneva publishes study programs of European and North American institutions. Besides, for different years, the *IUSSP Newsletter* published demographic teaching programs of institutions of different countries that submitted the information. For developing countries, a work of the Division of population of the United Nations on the program of demographic studies (Baldwin, 1996) as well as an extensive research conducted earlier by Stephen Baldwin (1986) provide information on the type and activity of centers identified as offering demographic training and on the content of the teaching.

demographic techniques (especially the techniques of forecasting) to very diverse fields of economic activity: marketing, management of personnel and human resources of an enterprise, follow-up of the production of goods and services, cohort analysis for all kinds of lucrative or charitable ends, and for private or public organizations.

The international debate on the modalities of transmission of knowledge on the matter of population has been very lively in the past decade. It highlighted the widespread dissatisfaction, linked to a sometimes too rigid architecture of teachings that did not leave enough room for new ideas or methods, did not adapt itself sufficiently to the educational needs of different categories of students, or did not always satisfactorily answer the questions raised by increasingly complex phenomena. The multiplication and diversification of teachings observed are beginning to answer to these needs.[28]

The geographic imbalances of distribution of pedagogic resources are in the process of diminishing. Although the most prestigious universities of developed countries remain the most attractive for numerous students coming from developing countries, the situations of research and of teaching in these countries have improved. In many cases, the reduction of international financing has temporarily produced difficult situations. However, very often, local institutions, especially universities, have succeeded in appropriating competences previously ensured by the centers created by the international organizations (e.g., CELADE, CDC, IIPS). In Latin America and in Asia, alongside still quite archaic situations that prevail in certain countries, a number of excellent centers have emerged in the field of demography, capable of ensuring education that accords importance to formal demography and to demographic statistics. China, going from zero in 1979, equipped itself with a dense network of centers of demography through the country's universities (in April 2004, ACERD listed 32 centers of demography). India is not very different, with at least 35 centers. India and China are at the head of the world honor's list of countries having the most population centers, respectively occupying second and third place, just after the United States (60 centers) and before France (31 centers) and Great Britain (29 centers). Africa is less well furnished with population centers, but initiatives abound, as much at the national as at the regional level, partially supported by the centers of the United Nations. The African Interuniversity Network for the Development and Study of Population (RIADEP) is a network of institutions linked to European and American universities, and the Center of Study and Research on Population and Development[29] (CERPOD) was established in Bamako, Mali.

CONCLUSIONS

From this study of the centers of teaching and research in demography, based on the very rich data provided by CICRED and on a systematic exploration of information available on the Internet, a panorama of half-tones finally emerges.

The reasons for satisfaction are numerous because of the growth spurt that demography has had since the 1950s that led to the recent multiplication of centers of research and teaching for the benefit of a discipline whose field has not ceased to grow in response to increasingly complex and diversified social demands.

However, there has been insufficient acknowledgment of the kind of training capable of ensuring the transmission of useful knowledge. In particular, it appears that demographic expertise does not fully succeed in incorporating itself in professional sectors that could benefit from it for practical purposes. It would be appropriate to undertake a more precise analysis of problems posed by the recruitment into the labor market of persons trained in demography. On the pedagogic level, it would also be good strategy to further develop the professional aspects of demographic training in view of adapting the teaching content to the needs of society.

University centers of teaching and research in demography and the modalities of their incorporation in academic structures reflect the eclectic positioning of the discipline, in the sense of its aptitude for developing different functions and contributing to a variety of subjects according to the programs of education with which it is allied. However, these effects also reflect the marginal nature of the discipline, which rarely constitutes the heart of autonomous university education and often has only a complementary position. The promises of the development of the discipline

[28] The International Symposium, which celebrated INED's 50th anniversary, devoted an important place to questions about training and research (Chasteland and Roussel, 1997). Special sessions devoted to teaching of demography were also organized during the annual meeting of the PAA (New Orleans, 1996) and the General Population Conference of the IUSSP (Beijing, 1997). Numerous other initiatives were taken at the same period in different parts of the world (Bogue, 1997). In recent years, the IUSSP's study group has coordinated its acitivity with that of by Graziella Caselli. It culminated in a seminar, Demographic Training in the Third Millennium, organized at Rabat in Morocco, on May 15–18, 2001, and the results were published in *Genus* in 2002 (Burch 2002; Caselli, 2002; Palloni, 2002).

[29] Centre d'étude et de recherche sur la population et le développment.

in the university framework are great, but so are the limitations. In this contrasted picture, the future lies in demographic research that is open to social issues and capable of explaining social phenomena. Scientific advances in this direction would permit it to consolidate the place of demography among social sciences, stressing at the same time its eclectic nature and its centrality for specific professional trainings.

Acknowledgment

This chapter was translated from the French version by Arundhati Virmani.

References

Cited in the Text

Avdeev Alexandre, 1997. Avenir de la démographie en Russie, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 369–396. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Baldwin Stephen C. (ed.), 1996. *Population and Development Curricula: A Selective Developing Country Review*. New York, United Nations, Department for Economic and Social Information and Policy Analysis, 87 p. (POP-TSS-96-4.)

Baldwin Stephen C. and Roy Krishna, 1997. Perspectives from developing countries, *in*: Donald J. Bogue (ed.), *Defining a new demography: curriculum needs for the 1990's and beyond*, p. 11–45. Chicago, Social Development Center, 64 p.

Bogue Donald J., 1993. How demography was born, *Demography*, vol. 30(4), p. 519–521.

Bogue Donald J. (ed.), 1997. *Defining a new demography: curriculum needs for the 1990's and beyond*. Chicago, Social Development Center, 64 p.

Burch Thomas K., 2002. Teaching demography: ten principles and two rationales, *Genus*, vol. 58(3–4), p. 21–34. (Special issue "Teaching demography in the early 21st century," edited by Graziella Caselli.)

Cagiano de Azevedo Raimondo (ed.), 1985. *Population, aid and development*. Liège, IUSSP, Derouaux, 404 p. (Proceeding of an international meeting on aid and cooperation in the field of population and development: issues and strategies, Florence, 1985.)

Caldwell John C. and Caldwell Pat, 1986. *Limiting population growth and the Ford Foundation contribution*. London, Frances Pinter, 194 p.

Caselli Graziella, 1997. Réflexions sur l'avenir de la démographie en Italie, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI^{e} siècle*, p. 295–397. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Caselli Graziella, 2002. Teaching demography: a number of priorities from the IUSSP Working Group, *Genus*, vol. 58(3–4), p. 11–17. (Special issue "Teaching demography in the early 21st century," edited by Graziella Caselli.)

Chasteland Jean-Claude and Roussel Louis (eds.), 1997. Un demi-siècle de démographie dans les pays industriels. Quelques réflexions sur un bilan, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 57–93. Paris, INED and PUF, p. 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

CICRED, 1989. *Conference of population research centers in Asia and the Pacific, Bali, Indonesia. Assessment of population research potentiality*. Paris, CICRED, 51 p.

CICRED, 1990. *Regional directory of population research potentialities. Asia and the Pacific*. Paris, CICRED, 146 p.

CICRED, 1995. *Annuaire des Centers d'Etudes et de Recherches en Démographie. ACERD 1995*. Paris, CICRED, 121 p.

CICRED, 2000. *Directory of population studies and research Centers/Annuaire des centers d'études et de recherches en démographie. ACERD 2000*. Paris, CICRED, 164 p.

CICRED, UAPS, and ECA, 1995. *Meeting of African population research centers, Addis-Ababa, 6–8 December 1994/Réunion des centers africains de recherche démographique, Addis-Abéba, 6–8 décembre 1994*. Paris, CICRED, 175 p.

Collomb Philippe, 2001. *Activity Report 1997–2001, CICRED General Assembly, Salvador de Bahia, 21 August 2001*. Paris, CICRED, 143 p.

Crimmins Eileen M., 1993. Demography: the past 30 years, the present, and the future, *Demography*, vol. 30(4), p. 579–591.

de Graft-Johnson Kweku T., 1985. Regional and operational needs, including training, data and research needs: Africa, with particular emphasis on data collection, human resources and training, *in*: Raimondo Cagiano de Azevedo (ed.), *Population, aid and development*, p. 129–168. Liège, IUSSP, Derouaux, 404 p. (Proceedings of an international meeting on aid and cooperation in the field of population and development: issues and strategies, Florence, 1985.)

Desrosières Alain, 1997. Démographie, science et societé: le cas français, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI siècle*, p. 57–93. INED-PUF, Paris, p. 434. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Dupâquier Jacques and Dupâquier Michel, 1985. *Histoire de la démographie. La statistique de la population des origines à 1914*. Paris, Librairie Académique Perrin, 462 p.

ECE, 1992. *Directory of Demographic centers in Europe and North America.* Geneva, European Economic Commission of the United Nations, 265 p. (GE.92-23919.)

Farooq Ghazi M., 1985. Teaching and training programs in population and development: Needs, approaches and assessment of effectiveness, *in*: Raimondo Cagiano de Azevedo (ed.), *Population, aid and development*, p. 299–324. Liège, IUSSP, Derouaux, 404 p. (Proceedings of an international meeting on aid and cooperation in the field of population and development: issues and strategies, Florence, 1985.)

Federici Nora, 1966. Rapport sur la recherche démographique en Europe, *in*: Conseil de l'Europe (ed.), *Première conférence démographique européenne*, vol. 2, p. 1–22. Strasbourg, Conseil de l'Europe.

Federici Nora, 1968. Le tendenze della demografia e le prospettive del suo sviluppo, *Genus*, vol. 24(1–4), p. 1–25. (Publishe in English: Trends in Demography and Prospects for its Development, *Genus*, vol. LVII(3–4), July–December 2001, p. 13–21.)

Federici Nora, 1969. L'insegnamento della Demografia in Italia, *in*: Centro nazionale di prevenzione e difesa sociale (CNPDS) (ed.), *Le scienze dell'uomo e la riforma universitaria*, p. 215–230. Bari, Laterza, 477 p. (Proceedings of the conference "Scienze sociali, riforma universitaria e società italiana" 17–19 November 1967.)

Federici Nora, 1973. L'enseignement de la démographie, l'information démographique et les roles des démographes, *in*:

UIESP/IUSSP (eds.), *International Population Conference, Liège 1973. Proceedings/Congrès international de la population, Liège 1973. Actes, vol. 1*, p. 97–104. Liège, IUSSP, 494 p.

Federici Nora, 1974. Training in the field of population. New York, United Nations/Nations unies, 43 p. (Background paper prepared for the World Population Conference, Bucarest, 1974.)

Frinking Gérard M., 1995. *50 ans de Démographie aux Pays-Bas*. *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI siècle*, p. 119–133. Paris, INED-PUF, p. 434. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Gendreau Francis and Huix-Adamets Silvia, 2003. Panorama institutionnel et scientifique de la recherche démographique dans le monde, CICRED, 58.

Girard Alain, 1995. Le premier age de *Population* (1946–1962), *Population*, vol. 50(6), p. 1335–1347.

Guillard Achille, 1855. *Éléments de statistique humaine ou démographie comparée*. Paris, Guillaumin, 376 p.

Hodgson Dennis, 1991. The ideological origins of the Population Association of America, *Population and Development Review*, vol. 17(1), p. 1–34.

IUSSP/UIESP, 1985. *The IUSSP in history. From Margaret Sanger to Mercedes Concepción/L'UIESP dans l'histoire. De Margaret Sanger à Mercedes Concepción*. Liège, IUSSP/UIESP, 45 p.

IUSSP, 1985—Demographic Science and its Teaching in Italy. IUSSP Paper n.28, p. 70.

Jain Anruda K., 1986. *Review of institutional capacities, human resources and development assistance in sub-Saharan Africa*. New York, The Population Council, 76 p. (Report n° 5.)

Keyfitz Nathan, 1987. La demografia accademica negli Stati Uniti: situazione attuale e prospettive, *in*: Eugenio Sonnino, Antonella Pinnelli, Dionisia Maffioli, and Annunziata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 383–401. Rome and Milan, Dipartimento di Scienze Demografiche and Franco Angeli, 566 p.

Keyfitz Nathan, 1993. Thirty Years of Demography and *Demography*, *Demography*, vol. 30(4), p. 533–549.

Klinger Andràs, 1997. 1945–1995: half a century of population research, past achievements. The experience of Hungary, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 281–293. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Langford Christopher M., 1997. Demographic research in Great Britain 1945–1995, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 135–156. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Lee Ronald D., 1997. History of Demography in the U.S. since 1945, *in:* Jean-Claude Chasteland and Louis Roussel (ed.), *Les contours de la démographie au seuil du XXIe siècle*, p. 31–55. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Livi Bacci Massimo, 1994. La demografia, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 3–23. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Livi Bacci Massimo, 1995. *1945–1995: un demi-siècle de démographie, l'expérience de l'Italie*. Paris, INED, 6 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Livi Livio, 1940. *Trattato di demografia. Le leggi naturali della popolazione*. Padova, CEDAM, Casa Editrice dott. Antonio Milani, 395 p.

Lombardo Enzo and Maffioli Dionisia, 1987. L'insegnamento preuniversitario della demografia. Alcune osservazioni, *in*: Eugenio Sonnino, Antonella Pinnnelli, Dionisia Maffioli, and Annunziata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 307–374. Rome and Milan, Dipartimento di Scienze Demografiche and Franco Angeli, 582 p.

Maffioli Dionisia, 1994. Organizzazione accademica, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 63–104. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Maffioli Dionisia, 1997. Centri di ricerca e di insegnamento della demografia, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Démographie: analyze et synthèse. Causes et conséquences des évolutions démographiques. Actes du Séminaire de San Miniato (Pisa), 7–19 December 1997. Volume 1*, p. 291–310. Rome, Paris, Louvain-la-Neuve, Dipartimento di Scienze Demografiche, INED, SPED, 318 p.

Maffioli Dionisia, 2004. Insegnamento e ricerca in demografia, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.), *Demografia: analisi e sintesi. Cambiamenti demografici, politiche di popolazione, dati e metodi. Atti del Seminario internazionale di Cortona, 19–21 June 2003: Volume 2*, p. 239–265. Sienna and Rome, Dipartimento di Studi Aziendali e Sociali, Università degli Studi di Siena et Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," 331 p. (Materiali di studi e di recerche, no. speciale.)

Martine George, 1985. Population studies in Latin America: Evolution, perspectives and future needs, *in*: Raimondo Cagiano de Azevedo (ed.), *Population, aid and development*, p. 95–127. Liège, IUSSP, Derouaux, 404 p. (Proceedings of an international meeting on aid and cooperation in the field of population and development: issues and strategies, Florence, 1985.)

Miro Carmen and Potter Joseph E., 1980. *Population policy: Research priorities in the developing world*. London, Frances Pinter, 197 p.

Ni Bhrolchain Maire, 1997. Future Prospects for the population research in the United Kingdom, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI^e^ siècle*, p. 337–368. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Nobile Annunziata, 1987. Testi italiani di demografia: evoluzione dei contenuti, *in*: Eugenio Sonnino, Antonella Pinnnelli, Dionisia Maffioli, and Annunziata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 236–251. Rome and Milan, Dipartimento di Scienze Demografiche and Franco Angeli, 582 p.

Palloni Alberto, 2002. Rethinking the teaching of demography: new challenges and opportunities, *Genus*, vol. 58(3–4), p. 35–70. (Special issue "Teaching demography in the early 21st century," edited by Graziella Caselli.)

Preston Samuel, 1997. Where is U.S. demography headed? *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI^e^ siècle*, p. 235–250. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Scardovi Italo, 1994. Demografia e biologia, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.) *Demografia*, p. 63–104. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Sgritta Giovanni-Battista, 1994. Demografia e sociologia, *in*: Massimo Livi Bacci, Gian Carlo Blangiardo, and Antonio Golini (eds.), *Demografia*, p. 63–104. Turin, Fondazione Agnelli, 582 p. (Collana "Guide agli studi di Scienze Sociali in Italia.")

Sonnino Eugenio, Pinnelli Antonella, Maffioli Dionisia and Nobile Anunciata (ed.), 1987. *Demografia: scienza, insegnamento, professione*. Rome and Milan, Dipartimento di Scienze Demografiche et Franco Angeli, 582 p.

Sylos-Labini Paolo, 1987. Le forze che regolano i movimenti della popolazione secondo il fondatore della teoria economica moderna, *in*: Eugenio Sonnino, Antonella Pinnnelli, Dionisia Maffioli, and Annunziata Nobile (eds.), *Demografia: scienza, insegnamento, professione*, p. 94–105. Rome and Milan, Dipartimento di Scienze Demografiche et Franco Angeli, 582 p.

Tabutin Dominique, 1988. Enseignement et recherche en Afrique. Bilan, problemes et perspectives, *in*: Étienne van de Walle, Mpembele D. Sala Diakanda, and Patrick O. Ohadike (eds.), *L'état de la démographie africaine*, p. 139–156. Liège, UIESP, 156 p.

UNFPA/FNUAP, 1977. *Inventory of population projects in developing countries around the world, 1975–76*. New York, United Nations Fund for Population Activities/Fonds des Nations Unies pour les Activité en Matière de Population, 528 p.

UNFPA/FNUAP, 1987. *Inventory of population projects in developing countries around the world, 1985–86*. New York, United Nations Population Fund, 787 p.

UNFPA/FNUAP, 1995. *Inventory of population projects in developing countries around the world, 1994*. New York, United Nations Population Fund, 684 p.

United Nations, 2003. *World population prospect. The 2002 revision*. New York, United Nations, Department for Economic and Social Affairs, Population Division, 2 vol., 781 + 939 p. + CD-ROM. (ST/ESA/Ser.A/225.)

Vichnevski M. Anatoli, 1997. La démographie soviétique après la seconde guerre mondiale, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 157–187. Paris, INED et PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Willekens Frans J., 1997. The future prospects for population research seen from Netherlands, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXI[e] siècle*, p. 329–336. Paris, INED et PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Other Publications Consulted

Federici Nora, 1970. Gli orientamenti di ricerca nelle Scienze umane e qualche riflessione sul possibile sviluppo della demografia, *Genus*, vol. 26(3–4), p. 305–312.

Federici Nora, 1971. La Demografia, *in*: Albino Babolin (ed.), *Le scienze umane in Italia, oggi*, p. 215–230. Bologna, Il Mulino, 291 p.

Federici Nora, 1977. Rimeditando sulle scienze statistiche e sulle scienze demografiche, *Rivista della Società Italiana di Economia, Demografia e Statistica*, no. "Studi in memoria di Livio Livi e Lanfranco Maroi," p. 387–400.

Federici Nora, 1989. Gli studi demografici fra discipline biologiche e discipline sociali nei 50 anni di attività della SIS, *in*: SIS (ed.), *Atti del Convegno "Statistica e Società," Pisa, 9–10 October 1989*, p. 253–293. Rome, Società Italiana di Statistica, 458 p.

Golini Antonio, 1985. New trends of demographic research in the developed countries, *in*: Département de démographie, *La démographie en perspective. Visages futurs des sciences de la population*, p. 77–101. Louvain-la-Neuve, Université Catholique de Louvain, 160 p. (Chaire Quetelet 1984.)

Loriaux Michel and Vichneskaia Tania, 2004. La démographie vue par ceux qui la font, *in*: Graziella Caselli, Jacques Vallin, and Guillaume Wunsch (eds.) *Demografia: analisi e sintesi. Cambiamenti demografici, politiche di popolazione, dati e metodi. Atti del Seminario internazionale di Cortona, 19–21 June 2003: Volume 2*, p. 267–295. Sienna and Rome, Dipartimento di Studi Aziendali e Sociali, Università degli Studi di Siena et Dipartimento di Scienze Demografiche, Università degli Studi di Roma "La Sapienza," 331 p. (Materiali di studi e di recerche, no. speciale.)

Nobile Annunziata, 1989. Per una storia dell'insegnamento della demografia nelle università italiane dalle origini al 1961, *in*: Carlo A. Corsini (ed.), *Da osservazione sperimentale a spiegazione razionale. Per una storia della Statistica in Italia*, p. 67–86. Pisa, Pacini Editore, 168 p.

Ricolfi Luca, 1992. Dove va la Sociologia? *Bollettino della Società Italiana di Statistica*, vol. 26(July), p. 12–46. (Proceedings of the Giornate di Studio della SIS "Crisi di Teorie, crisi di indicatori," Giardini Naxos, 24–25 October 1991.)

Urzua Raul, 1985. The uses of demographic knowledge for policies and planning in developing countries. Problems and issues, *in*: Raimondo Cagiano de Azevedo (ed.), *Population, aid and development*, p. 181–217. Liège, IUSSP, Derouaux, 404 p. (Proceedings of an international meeting on aid and cooperation in the field of population and development: issues and strategies, Florence, 1985.)

C H A P T E R

143

Demography Viewed by Demographers: Considerations Based on the Results of an International Electronic Survey Carried Out in 2000

MICHEL LORIAUX AND TANIA VICHNEVSKAIA

Institut de démographie, Université catholique de Louvain, Louvain-la-Neuve, France

Institut national d'études démographiques (INED), Paris, France

I. DEMOGRAPHY IN SEARCH OF ITSELF

Demography is frequently catalogued as a social science, even though its links with the life sciences are just as strong. Moreover, it can be considered relatively old and surprisingly young. It is old because its date of birth is known exactly, if we accept the publication in 1662 of John Graunt's *Natural and political observations* as a date of reference. It is young, insofar as its introduction as a university discipline occurred much later, and its first entry into academic circles dates back barely a century (and much less in many countries), confirming Alfred Sauvy's (1946, p. 7) reference to demography as "une science sauvage" (i.e., a wild science"), because it was not developed initially within university circles but rather began within statistical institutes to fulfill the need for administrative data on population trends. Demography was initially an analytic method aimed at understanding the mechanisms of population changes, and would probably have remained confined to a panoply of statistical and mathematical tools (e.g., life tables, projection methods) if it had not taken the risk of venturing out of its protective cocoon to deal with the delicate issues of the causes and consequences of the observed population movements and changes in structure or, in other words, issues of explaining and understanding phenomena.

The risk of being entirely colonized by its invasive neighbors (e.g., psychology, sociology, anthropology, economics) was even greater because this discipline was originally relatively poor in content (i.e., limited to the three traditional phenomena of fertility, mortality, and migration) and in conceptual frameworks and theories, such as that of the quite unique demographic transition.

Fortunately, the need to understand the why and the how of population trends, as well as the linkages that they have with other societal changes (e.g., technologic, social, cultural, economic, political), has resulted in extending the field of demography and increasingly making it a discipline that acts as a bridge between formal demography and the other sciences and that participates as a relay between the various disciplines that study major events connected to human life, such as birth and death, and that can communicate in an interdisciplinary manner through demography (e.g., social science in its relationship with biology and medicine). It is in this sense that David Coleman (2000) can define demography as "the statistical study of the processes of reproduction, migration and death in the human species, their interrelations with the distribution and dynamics of population, and their biologic, environmental, and socioeconomic causes and consequences” (Coleman, 2000, p. 27).

However, demography definitively remains in the hands of demographers or those who so identify themselves. Beyond intellectual debate about whether it is a science or a method, whether it is closer to natural sciences or cultural sciences, and whether it is one or several, it is useful to ask how demographers see their own discipline and its development and which population issues are of real interest to them in their professional lives.

The community of demographers is not only a relatively vague unit, but it is also a group on which very little information is available, with the exception of some celebrities whose biographies and scientific paths are quite well known.

The idea of carrying out a survey within this community was nearly self-imposed, when two well-known Institut National d’Études Démographiques (INED) demographers, Louis Roussel and Jean-Claude Chasteland, both approaching retirement age, decided to focus on their discipline by studying its development from the period when they were its honorable representatives (one as a famous family sociologist, and the other for having directed the United Nations Population Division) and when demography had a valued status and considerable recognition, both academically and politically, up to the beginning of the 21st century, when rumors of loss of specificity and even a crisis within the discipline were spreading from one end of the planet to the other, leading to real concern among some people about its survival.

II. A RATHER PECULIAR SURVEY

The survey, some of the results of which are presented here, was originally designed as a qualitative survey with the aim of examining the so-called demographic crisis (Chasteland *et al*., 2004).

Very quickly, the small group of key people formed by Jean-Claude Chasteland and Louis Roussel[1] redirected the project toward a quantitative opinion survey whose originality lies notably in the method of observation, insofar as it was one of the first attempts (in this sector) to carry out a survey using the internet. A web questionnaire was developed and placed on the network by several servers (the gateways used were those of INED, the United Nations Population Division, NIDI, and Moscow State University).

For the respondents who could not use the HTML form, two alternative methods were provided, an MS Word version sent by electronic mail and a questionnaire completed by hand and sent through the post. The questionnaire was available in four languages (i.e., French, English, Spanish, and Russian), which was quite uncommon (Fig. 143–1). A total of 636 responses were collected in the period between October 1999 and July 2000, and a server breakdown probably caused

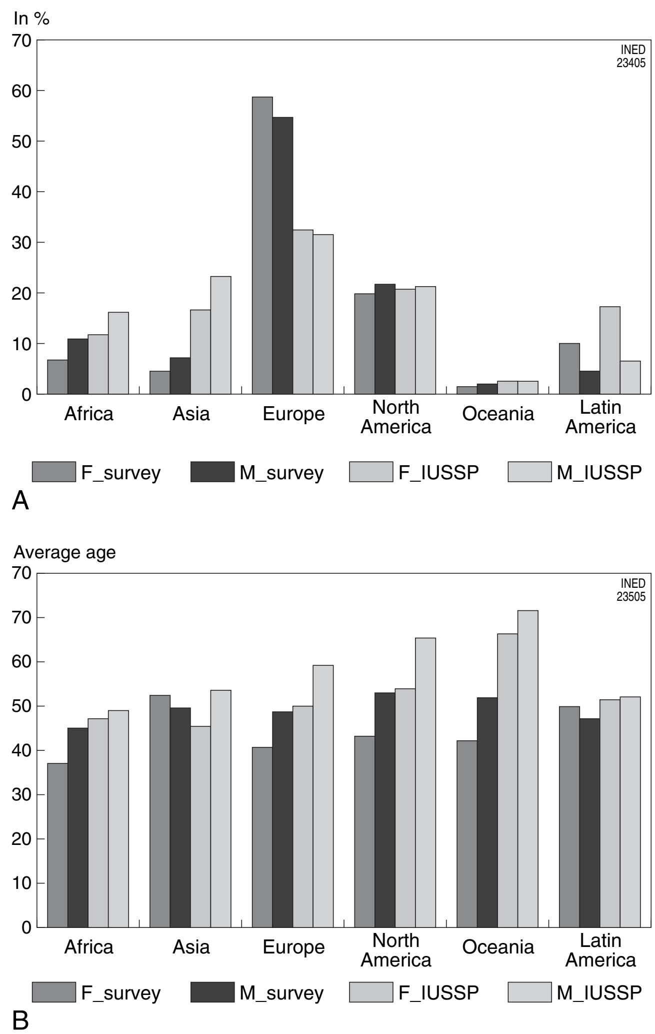


FIGURE 143–1 Languages and delivery method of questionnaires, as well as regional localization and respondents’ membership in large international associations.

[1] The group was made up of Alexander Avdeev (Moscow State University), Máire Ní Bhrolcháin (University of Southampton), Jean-Claude Chasteland (INED), Thérèse Hibert (Paris), Michel Loriaux (Université catholique de Louvain), Giuseppe Micheli (Catholic University of Milan), Anatole Romaniuc (University of Alberta, Canada), Louis Roussel (INED), and Tatiana Vichnevskaia (INED). Moreover, the group gained from the contribution of several other individuals at different phases of the research, notably Hubert Gérard (Université catholique de Louvain), François Héran (INED), and Alain Desrosières (INSEE).

the decrease in the anticipated size of the sample because some potential respondents were unable to send their replies through the web.

This technical hitch undoubtedly caused some errors, but beyond this unfortunate (and unforeseeable) occurrence, the issue of representativity would have been raised regardless due to the method of observation used. The reference population is vague and not precisely defined, and using the web to distribute and collect the questionnaires is similar to an old type of postal survey with a higher risk of representativity errors than is usually encountered in surveys using direct interview of a nominative list of respondents. Admittedly, it was partly the randomness of encounters on the web that led researchers all over the world to read the questionnaire before deciding to respond to it, but in this coincidence can hardly be compared with a random selection from a table of random numbers, as postulated by probabilistic theory.

Various selection errors could have occurred, without it really being possible to correct them *a posteriori* as there was no exhaustive sampling frame available and the only ones that could have been used (e.g., membership lists of the large demographic associations, such as the International Union for the Scientific Study of Population [IUSSP], the European Association for Population Studies [EAPS], the Population Association of America [PAA], and the Union for African Population Studies [UAPS]) were themselves only partial and were far from completely covering the whole of the field in question. These potential distortions led the originators of the study to abandon any global statistical inference and even, by precaution, to avoid any mention of a representative sample or even any kind of sample, preferring to speak of a reference group. As they wrote "The sample is, to some extent, representative of itself" (Chasteland *et al*., 2004), which does not necessarily limit the scope of the study, insofar as it is more concerned with revealing the diversity of situations and the relations between the variables than with precisely describing the levels of the phenomena.

Nevertheless, to better inform our readers, for whom the traditional probabilistic survey based on a random selection of statistical units from an exhaustive list of the observed population remains the preferred method of collecting representative information, we have carried out a comparison between our database and the membership list of the IUSSP, which is the only worldwide association of demographers. This comparison relates to three traditional identification variables that were used to a considerable extent in the Démographie 2000 survey as segmentation criteria for our reference group (Fig. 143–2).

Numerically speaking, the Démographie 2000 survey did not suffer from an excessively low number of respondents (636) compared with the total number of IUSSP members (1672), because it represents 38% of the total, while being aware that not all of the respondents were members of the association (only 331, or about 20% of members).

In terms of geographic origin, although North America is proportionally correctly represented (a little more than 20%), Europe is, as we would expect, over-represented (56% compared with 32% in the IUSSP membership database), and the other regions of the world are under-represented, particularly Asia (6% versus 21%). For Africa, whose limited accessibility to the Internet could have caused a massive under-representation, the deficit is moderate (9% versus 14%).

The distribution by sex in both databases is very similar; in both cases, about 40% are women, although the numbers are slightly greater in the survey than in the IUSSP list, and the male-female differences are *grosso modo* correctly respected, regardless of the continent of birth. However, the average age of the respondents is systematically lower than that of European Union members, with gaps of sometimes as much as 10 years (notably among European men and American women).

Despite the observed differences, no attempts were made to rectify the sample insofar as the reference populations themselves were not exempt from selection errors and could present significant distortions from one source to the other (the various existing professional associations of demographers). For example, we know that, mostly because of the IUSSP's procedure for becoming a member, the average age of this association's members is far higher than that of demographers. This is one reason that explains why almost 50% of the respondents were not IUSSP members (and 30% declared that they were not members of any international association).

Démographie 2000 is undeniably original because of the communication method used, although it poses the same problems (no more, no less) as any more traditional survey in which the target population is not precisely delimited and constitutes to some extent a vague group with imprecise borders.[2] There is therefore no real reason for calling into question the validity of the information collected here any more than for

[2] This population was defined as "the group of researchers and specialists who feel involved one way or another with demography or population studies or, although not demographers, who are members of an association of demographers, or who work within a demographic organization or project" (Chasteland *et al*., 2004, p. 13).

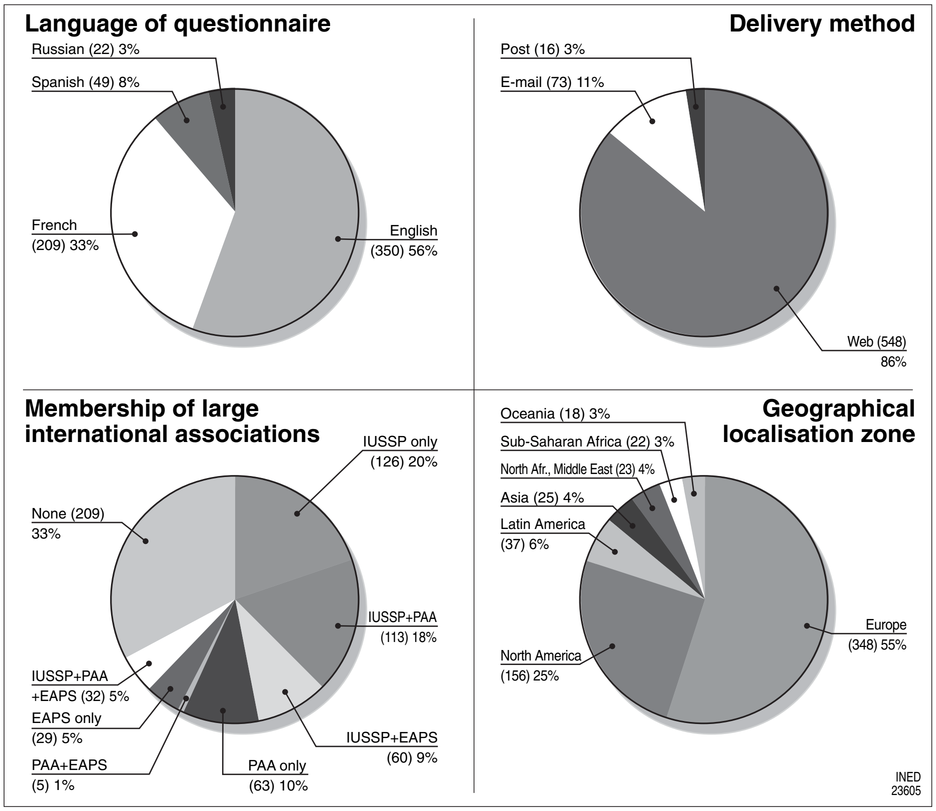


FIGURE 143–2 Compared distribution of the Démographie 2000 sample and of the Union for the Scientific Study of Population (IUSSP) database of members (enrolled before 2002), according to sex, average age, and continent of birth. **A:** Comparison of structures by sex and continent of birth (in percentages). **B:** Comparison of average age according to sex and continent of birth (in years).

other surveys that often have very high rates of non-responses and that as a result also have high risks of selection errors.

This survey on demography and demographers at the crossroads of the second and third millennia was all the more imperative because no similar exercise had ever been undertaken, with the exception of the much smaller survey carried out in 1995 by Francesco Billari and Giuseppe Micheli on the participants of the European conference organized by EAPS in Milan (Billari and Micheli, 1995).

As Jean-Claude Chasteland (2004) points out, unlike their colleagues from other scientific disciplines, demographers seem to be less interested in periodically carrying out general and critical assessments of their discipline. We can mention only a few special issues of journals such as *Demography*, *Population Studies*, and *Population* published at anniversaries celebrating their 30th, 40th, and 50th year of publication, respectively, as well as the *Chaire Quetelet*, which is devoted to themes such as causality and explanation in demography (Tabutin *et al*., 1999; Duchêne *et al*., 1989), the role of qualitative approaches (Gérard and Loriaux, 1988), or the role of projections and forecasts (Wattelar and Duchêne, 1999), in addition to *Les Contours de la démographie au Seuil du XXI ème siècle*, published by INED at the time of its 50th anniversary celebrations (Chasteland and Roussel, 1997).

> That journals devoted to population issues wait 30 or even 50 years before carrying out these assessments is either an indication of the difficulty of the exercise or of demographers' more or less conscious resistance to launch into introspective exercises. This reticence could be explained by the discipline's unresolved identity problems within itself and in relation to other social sciences (Chasteland, 2004, p. 12).

Such were the considerations that led the promoters of Démographie 2000 to fill a gap that seemed worrying to them by asking demographers the world over to describe their professional practices and gathering their perceptions of current and future problems for the discipline, rather than undertaking an analysis of demographic literature as had previously been done by the few authors concerned with establishing an inventory of the discipline of demography.

With two such radically different processes, we did not wish to attempt to systematically compare their results, but a table has been produced by one of the contributors of Démographie 2000, Giuseppe Micheli, who through a 4-year examination of *Population Index* at 10-year intervals (1963, 1973, 1983 and 1993) describes the demographic publications listed by this bibliographic journal according to several categories of scientific orientation (Table 143–1). We discover that

> the central themes of the discipline (fertility, mortality, nuptiality, family, migration) have constantly gained importance . . . and that as a result they have overtaken more demographic areas, more conceptual approaches (logical and methodological), and especially interdisciplinary and cross-sectional themes that can nevertheless contribute to understanding the processes studied, and naturally themes that we consider, *stricto sensu*, as being demographic (Micheli, 2004, p. 206).

III. A DISCIPLINE TORN BETWEEN ITS HISTORICAL DIVISIONS

Today, most people working on population issues define themselves as demographers (67%), and they have also had specific demographic training. However, double (or triple) training disciplines are common, with sociology largely in the lead, followed far behind by mathematics or statistics and economics (Fig. 143–3). It can therefore be said that self-taught people, who were previously in the majority, have been largely replaced by professionals who have benefited from specific training in population science.

Because of this eclecticism of original disciplines, most respondents do not demand a strict definition of demography based on the concept of demographic analysis (formal demography: only 8% of responses) and prefer to rally around the broad concept of population studies (84%), whether it is carried out in collaboration with other disciplines or not (Fig. 143–4)?

Only a small proportion of respondents continue to believe that the discipline's role should consist of "explaining demography using demography" (self-explanation: 5%), with one-fifth believing the opposite ("explaining demography using other factors"), whereas the others (75%) are more undecided but also

TABLE 143–1 Demographic Studies in 1963, 1973, 1983, and 1993 Listed by Population Index According to Research Theme

Themes Demography	1963	1973	1983	1993
Mortality	9.4	9.0	16.5	11.5
Fertility	7.6	16.8	19.9	23.6
Nuptiality and family	3.8	1.8	8.1	7.6
Migration	7.7	12.0	13.2	15.9
Subtotal	28.5	39.6	57.7	58.6
Population studies				
Regional population studies	6.1	6.2	2.8	2.2
Spatial distribution	7.2	6.3	4.1	3.0
Evolution in population growth and size	5.1	5.1	7.1	4.9
Characteristics of populations	8.2	5.2	4.9	8.3
Population statistics	3.2	3.1	4.0	3.9
Sub-total	29.8	25.9	22.9	22.3
Theories and methods				
General population studies and theories	6.9	5.6	2.8	2.3
Research and analysis methods, models	4.0	6.3	2.2	1.5
Subtotal	10.9	11.9	5.0	3.8
Interdisciplinary and cross-sectional studies				
Interaction between demographic or economic and natural resources	12.7	8.9	4.9	5.8
Demographic and non-economic interactions	5.3	3.1	2.1	4.7
Historical demography and history of demography	5.1	6.1	3.7	1.4
Policies	7.7	4.5	3.7	3.4
Subtotal	30.8	22.6	14.4	15.3
Total	100	100	100	100

From Golini (1985) and Micheli (2004, p.207) for the 1993 update.

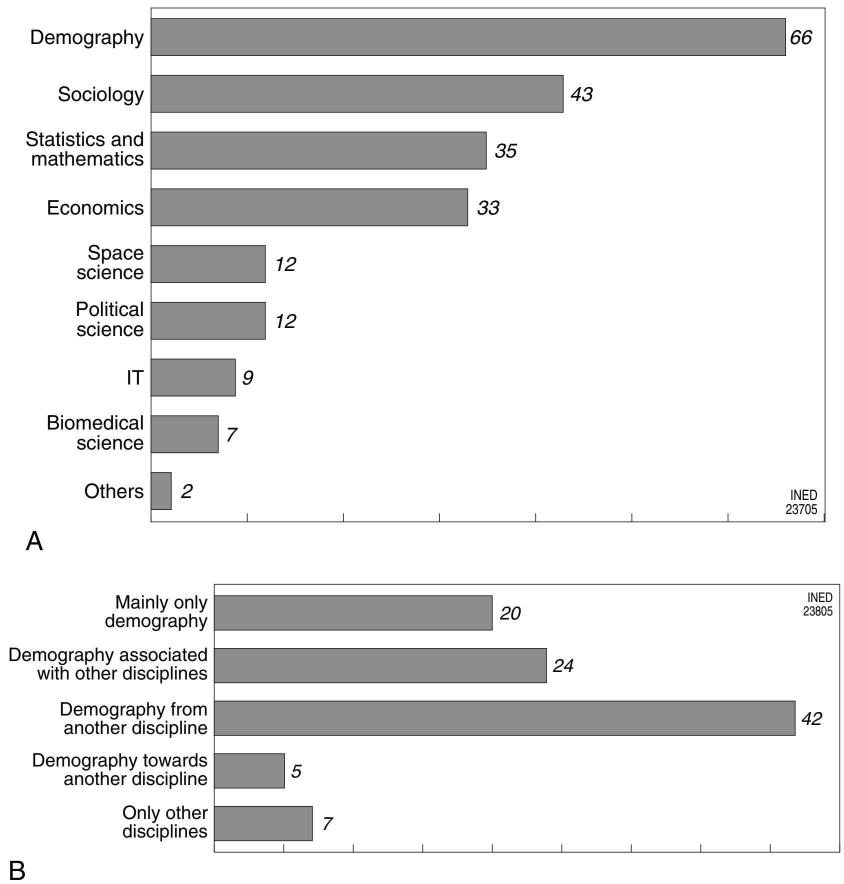


FIGURE 143–3 Respondents' training disciplines and professional history. **A:** Main discipline of training (in percentages). Most respondents (74%) have trained in two disciplines, and some (34%) have trained in three or more. **B:** Professional histories since the beginning of careers.

doubtless more realistic and opt for a dual reciprocal determination ("both"), thereby confirming that demography is no longer confined to a narrow conceptual framework, without a window to the outside world, but plays its role fully as an interdisciplinary science, accepting external influences on the phenomena that it observes, convinced that population issues can significantly affect society's other phenomena.

The distribution of the respondents is appreciably similar regarding whether the discipline is of a descriptive or explanatory nature; only 6% consider it as being essentially descriptive and 16% as explanatory, compared with 77% who consider it as being both (1% with no opinion).

This clear unanimity is strengthened by the question about whether demography belongs to social sciences or to natural sciences; however, this time, almost no respondent opts for natural sciences alone and only 22% for a combination of both, but most (76%) opt for an exclusive social sciences orientation.

Figures 143–4 and 143–5 demonstrate these choices. They are two-dimensional graphs that illustrate the ties between the elements of the survey introduced into the analysis by software called *Réseau-Lu*,[3] which is not unlike the more traditional multivariate analysis programs but which nevertheless introduces additional factors for flexibility and facility of interpretation.

For example, graph D in Figure 143–4 establishes the dominant attachment of demography to social sciences, with a minority opting for a combination of social sciences and natural sciences (and the near-elimination of natural sciences alone due to the lack of numbers). We see that the respondents who chose the social sciences option have themselves been trained in social sciences (e.g., sociology, geography, history) (see Fig. 143–5), which is relatively logical, whereas those who chose the dual option are physicians, epidemiologists, mathematicians, statisticians, actuaries, biologists, anthropologists, specialists in environmental science, or political economists.

[3] Designed and developed by Andreï Mogoutov, AGUIDEL (www.aguidel.com). For a description of the statistical method, see Mogoutov and Vichnevskaia (2004).

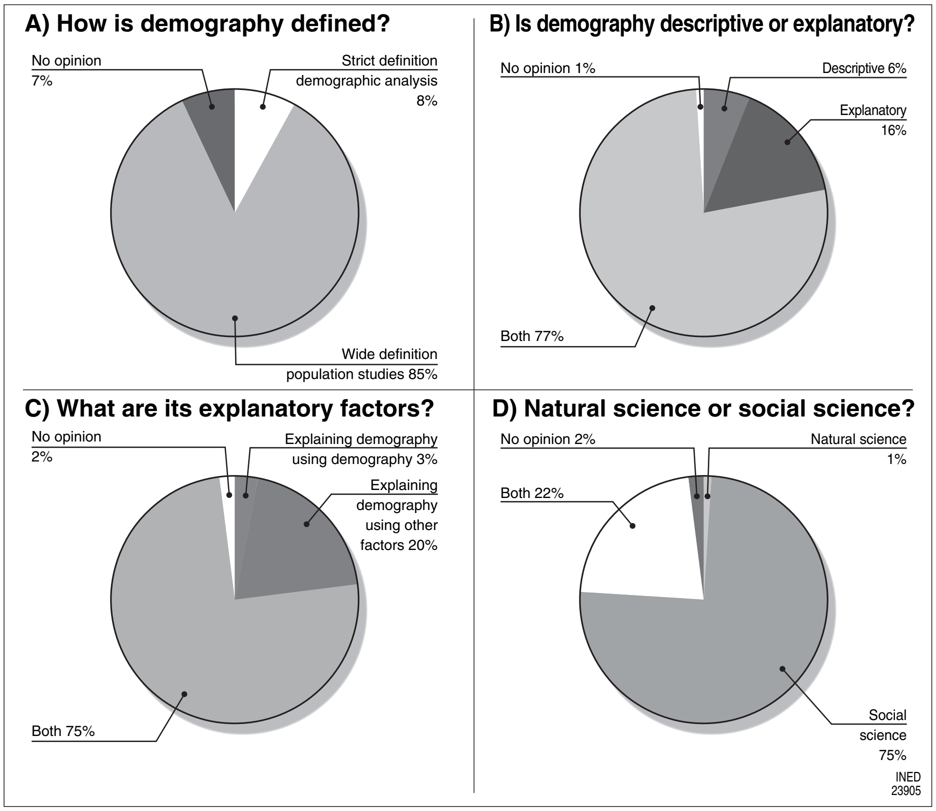


FIGURE 143–4 Overview of the discipline.

Curiously, no discipline acts as a bridge between these two poles, which are clearly disassociated, and demographers, who are nevertheless in the majority in terms of training, do not themselves have any relations with these to the benefit of a simple connection with the least-mentioned pole ("natural sciences"), which is a similar situation to that of economists and lawyers who generally decline to give an opinion (see Fig. 143–5).

The situation is significantly different if we put the training disciplines in correspondence with the conception of demography as a descriptive or explanatory science (see Fig. 143–5). Several subnetworks disassociated from each of the three possible options (i.e., descriptive, explanatory, or both) surround the nearest disciplines, but this time, there are quite numerous transition points that appear and that link the subnetworks. While actuarial scientists unilaterally opt for the descriptive aspect of the discipline and biologists for the explanatory aspect, other disciplines (e.g., medicine, geography, genetics) are simultaneously attached to both poles.

The choice of the dominant option ("both at the same time") is that of demographers, sociologists, environmentalists, town planners, and to a great extent, mathematicians and statisticians, whereas members of other disciplines reported a preference for a combination ("descriptive/both"); this is notably the case for historians, psychologists, and anthropologists.

If we look at the choice of the explanatory principle (see Fig. 143–5), we find a situation that is quite comparable to the previous one insofar as the division of opinions is not as clear as that regarding the scientific membership of disciplines. Although demographers almost unilaterally opt quite for the dual determination of phenomena, sociologists are split between this and explaining demography through demography, and economists express a dual choice, but this time in favor of explaining demography through other factors (in addition to dual determination).

Many specialists from other disciplines remain clearly monofactorial; this is the case of historians, geographers, geneticists, and political economists who prefer the option of explanation by external factors. Only psychologists and mathematicians or statisticians opt in favor of endogenous explanation (explaining demography through demography), as well as some choices of no opinion.

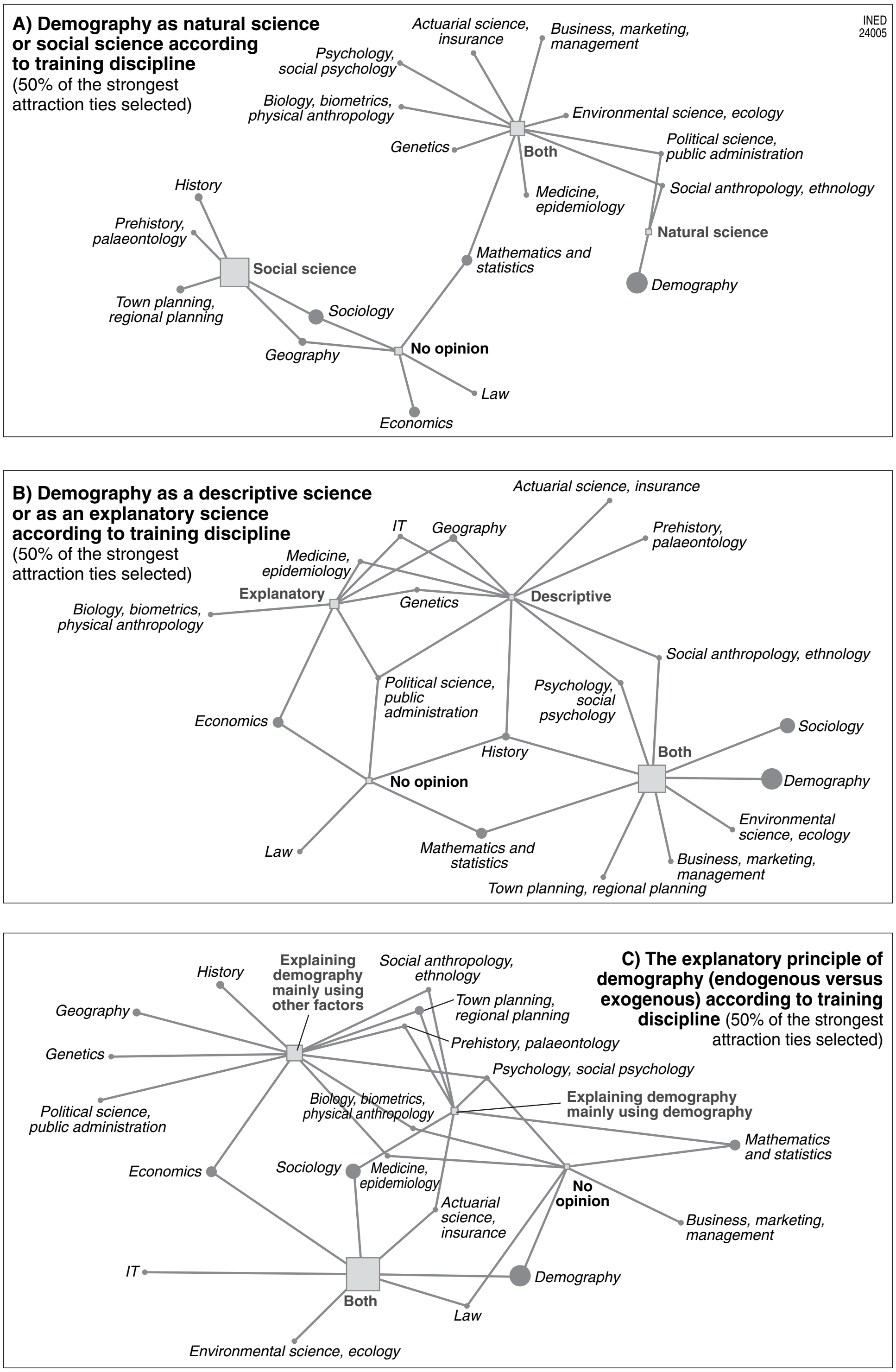


FIGURE 143–5 View of demography according to respondents' training discipline. **A:** Demography as a natural science or as a social science according to training discipline (50% of the strongest attraction ties were selected). **B:** Demography as a descriptive science or as an explanatory science according to training discipline (50% of the strongest attraction ties were selected). **C:** The explanatory principle of demography (endogenous versus exogenous) according to training discipline (50% of the strongest attraction ties were selected).

IV. A STROLL THROUGH DEMOGRAPHERS' GEOGRAPHIC, LINGUISTIC, AND INSTITUTIONAL SPACE

Like everyone, demographers (or experts in population science) are born somewhere, sometimes migrate from their countries of origin to other countries, work within different types of institutions, and write their reports or their work in certain languages (which are not always their mother language).

Among the respondents to the Demography on the Threshold of the Year 2000 survey, 24% write only in English and 8% in French, whereas taking into account bilinguals and trilinguals (or beyond), English is used by 60% of them, which makes it incontestably the dominant language among demographers.

The distribution between types of work organization is more balanced. Universities are in first place (25% of responses) but are closely followed by research centers (20%) and, trailing significantly behind are international organizations (14%), national statistical institutes (13%), national or federal administration (11%), nongovernmental organizations (NGOs) (9%), local administrations (8%), and private foundations (5%).

The graphs of Figure 143–6 illustrate this basic information. Figure 143–6 shows the type of organization matched against a question mentioning the regions of the world on which the respondents' work is based. A first glance at this graph shows considerable segregation between the regions, such as Europe, which undertakes self-investment through the population studies carried out in its universities and research centers, and regions, and such as Africa, which receives aid in this field only through international organizations and NGOs. Moreover, the management structure of population studies is very different in Europe and North America, because research funding in the latter region comes from universities and from private foundations and NGOs.

Graph B in Figure 143–6 is more complex because it results from a simultaneous projection of several variables: the country of birth, the usual country of work, the temporary country of work, the country of training, and the languages used for scientific production. The geographic and linguistic components are highlighted in this graph. We can distinguish a French political-linguistic pole that represents the Francophone world, including France and its former colonies or protectorates, with the slightly eccentric position of Belgium and Canada.

Another pole structured around Spanish, and to some extent Portuguese, unites most Latin American countries, even though Spain has only a marginal position. The Scandinavian countries, including Estonia, and the Austro-Hungarian world are placed closely to the countries of Eastern Europe, which are structured around Germany (but not necessarily around the German language). The same Eastern European countries are also attracted by the former Soviet space, partly due to the Russian language.

The central part is made up of European countries, among which are multilingual countries such as Belgium and Switzerland that are attracted by several poles in a structurally similar way. The most developed pole remains that of the English language dominated by the United States and the United Kingdom, even though it is at the same time the least specific one. Surrounding the United Kingdom is a group of countries that have traditional historical links with it (e.g., India, Egypt, Kenya).

Graph 6C provides a complimentary view of the respondents' migration space by summarizing migration for initial training, i.e. any movement from the country of birth and the country of first university training. All the countries of birth mentioned are represented, and the lack of a link of most of them to another country confirms that a large number of respondents received their training in their country of birth. However, France and the United States distinguish themselves as two considerably attractive poles, and at the same time they gather students from countries or regions with which they have traditional links.

Nationals of some countries such as Egypt, Lebanon, Argentina, and Germany are shared between the two dominant poles that, in addition, have little exchange with each other. Other centers, such as Belgium, Germany, and the United Kingdom also appear as centers for training and exchange at a European level, whereas Canada, the Netherlands, Austria, and Hungary form an exchange subnetwork for initial training. Russia has a relatively isolated position and attracts only some countries that were part of the former Soviet empire; it has not created any significant links with other European centers.

V. MULTIPLICITY OF STUDY THEMES

To think that demographers exercise all their talents in virtually identical fields would be completely wrong. Démographie 2000 asked respondents to choose between 23 study themes and to answer this question for three different dates of reference: "currently," "10 years ago," and "20 years ago." Because of age, not everyone could provide answers to all three dates, and the number of respondents inevitably decreases over

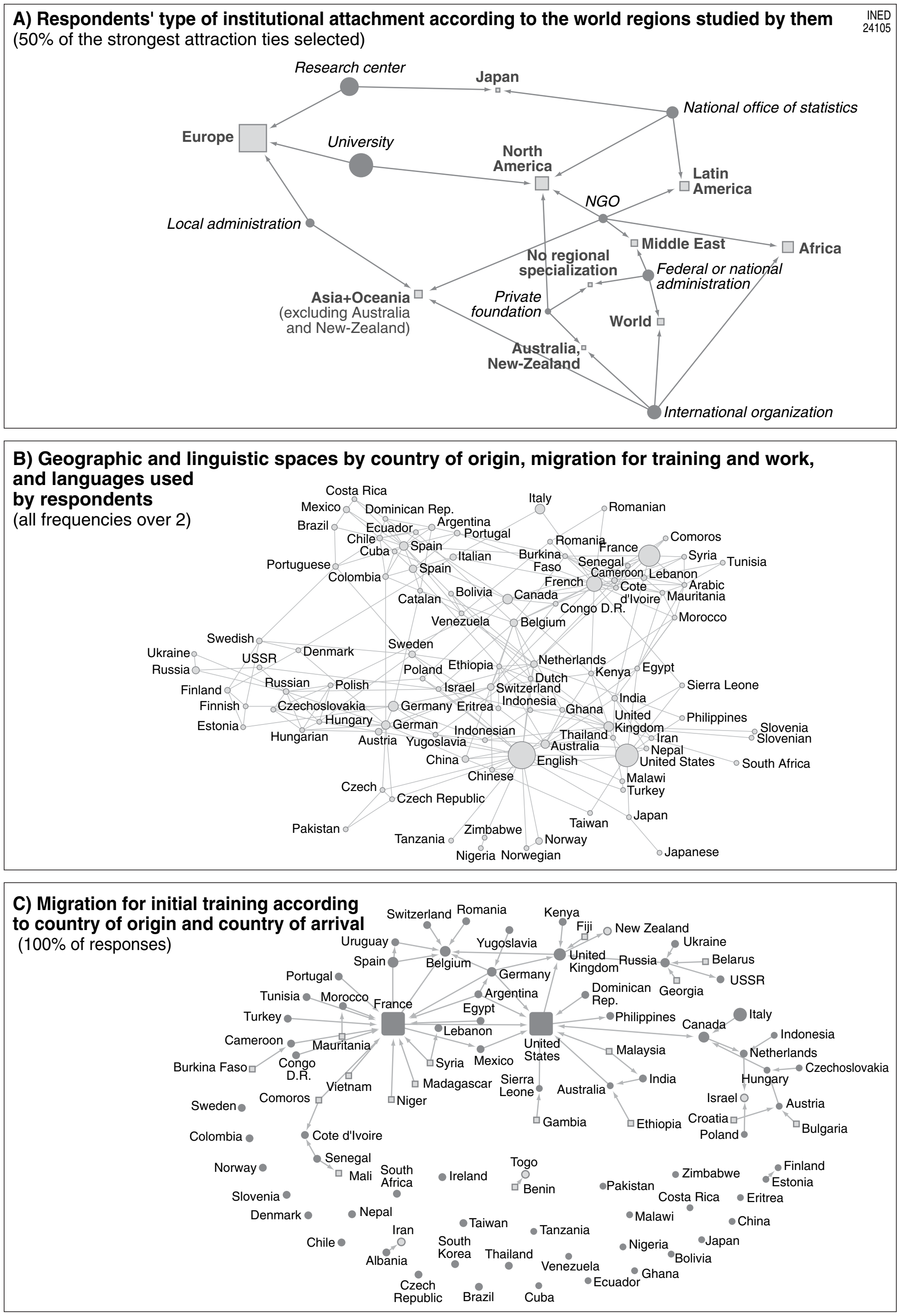


FIGURE 143–6 Country, place, and languages of work: a multivariate summary. **A:** Respondents' type of institutional attachment according to the world regions studied by them (50% of the strongest attraction ties were selected). **B:** Geographic and linguistic spaces by country of origin, migration for training and work, and languages used by respondents (all frequencies more than two). **C:** Migration for initial training according to country of origin and country of arrival (100% of responses).

time: 621 valid answers for the current period, 468 for "10 years ago," and only 320 for "20 years ago."

Nevertheless, the order of study themes has not been upset with the progress of time, and within the hit parade of themes, the same ones have remained more or less in the top ten for the three dates. As can be expected, studies on fertility and fecundity have remained the uncontested leaders over the last 2 decades (Fig. 143–7). However, the study of mortality, morbidity, or epidemiology have progressed over time, going from fifth position (20 years ago), to third (10 years ago), and finally reaching second place (currently). This is not the case for the third fetish phenomenon of demography, migration, which is currently only in 10th place with regard to internal migration and the spatial distribution of populations and is in 12th place with regard to international migration, which may be surprising. However, studies on nuptiality and family structures have progressed well and moved up from ninth position 20 years ago to fifth place currently, whereas for the study of population in developing countries, the reverse has happened because it seemed to be of greater interest 20 years ago (fourth place) than it is now (ninth place).

At the bottom of the scale we find, without any significant variations from one period to another, economic demography (currently in 18th place), the study of populations in relation with natural resources (19th), anthropologic studies and case studies (20th), cultural demography (21st), the history of demography (22nd), and epistemology (23rd).

The examination of the co-occurrences of study themes from graph A in Figure 143–8 confirms that the traditional fields of demography are strongly connected among themselves and form a clearly defined pole: fertility, mortality, nuptiality, population of developing countries, and population policies. At the opposite end, a more technical-methodologic pole emerges, with methods and modeling, prospects and projections, and migration and economic demography. Between the two, a third horizontal pole groups the less popular poles, but it is nevertheless connected to the other two, such as historical and cultural demography, anthropologic studies, or epistemology.

Two other graphs (see Fig. 143–8) make it possible to verify whether there is a clear differentiation in the choice of study themes according to the two major

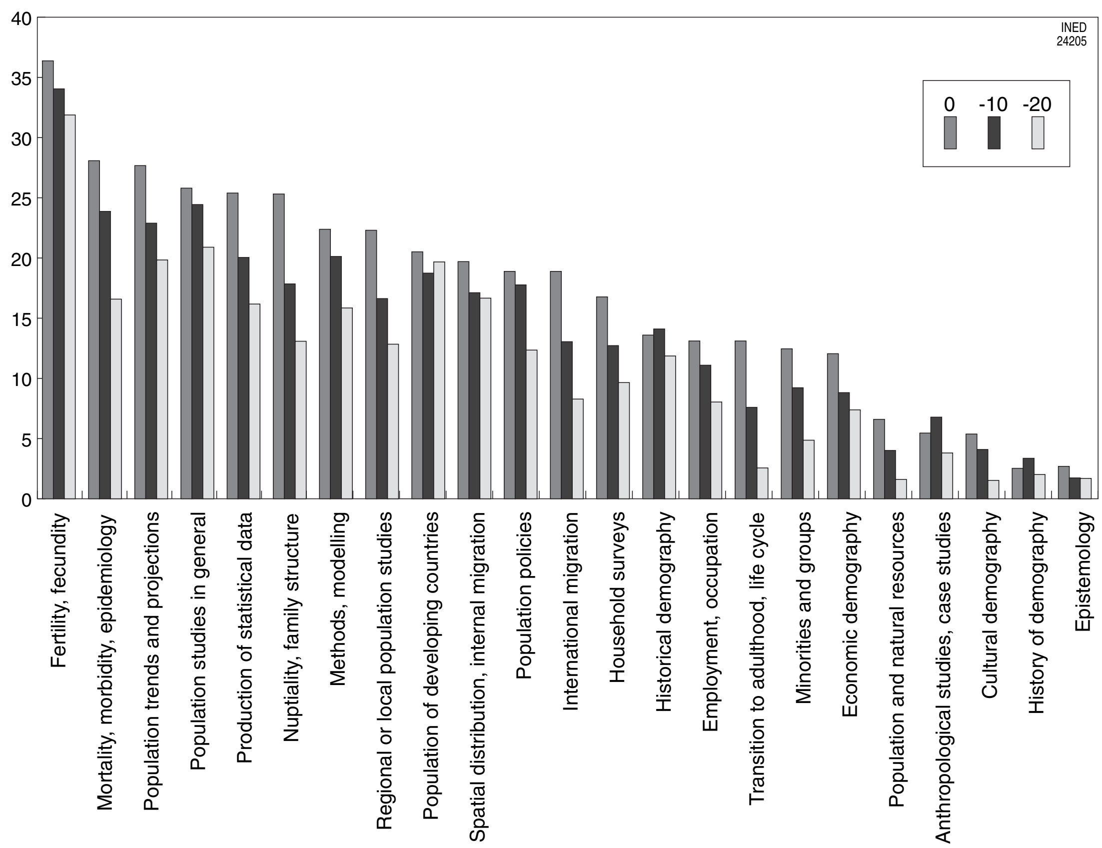


FIGURE 143–7 Classification of study themes identified in three reference periods (descending order from current responses).

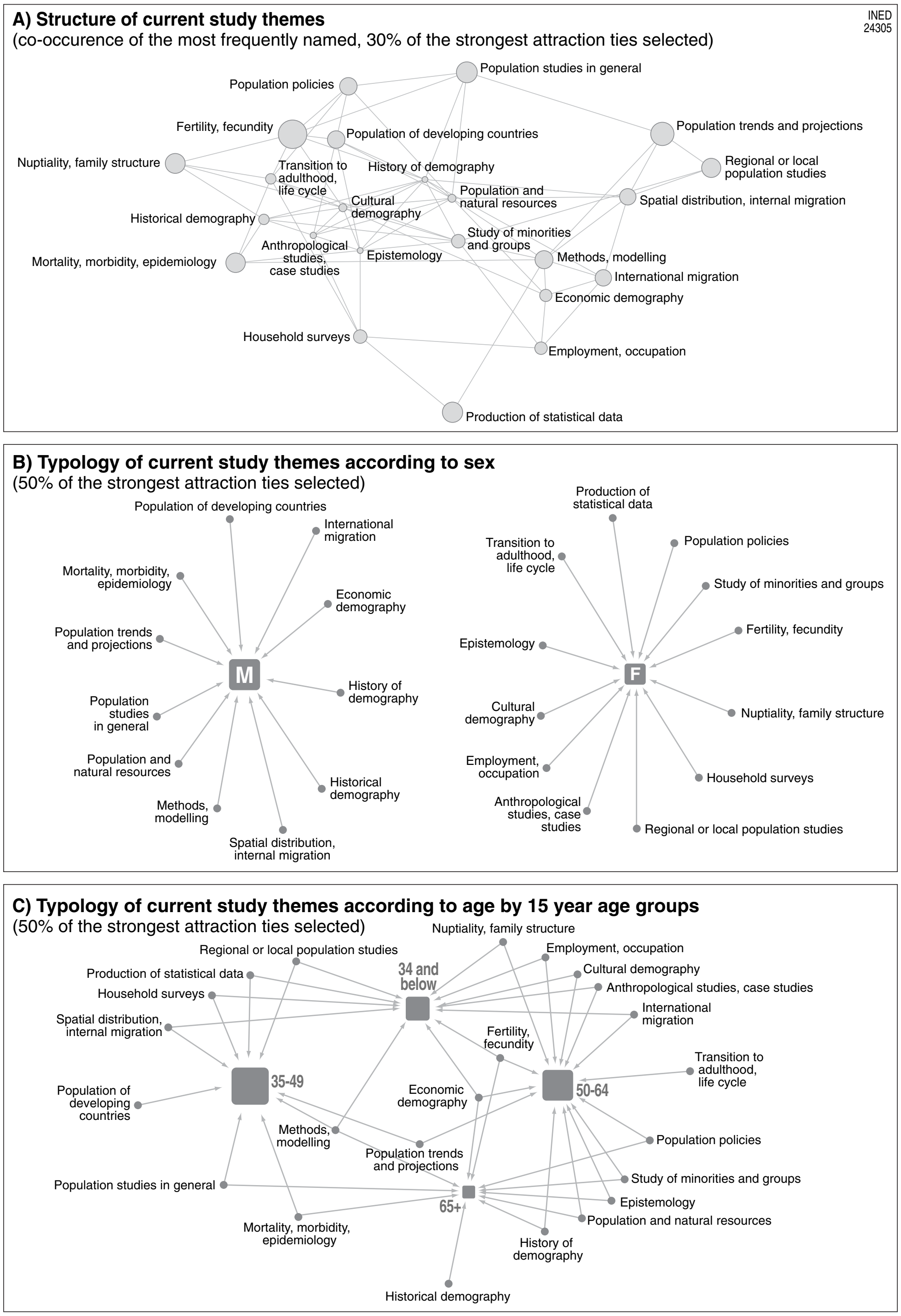


FIGURE 143–8 Interdependent structures of study themes according to sex and age. **A:** Structure of current study themes (co-occurrence of the most frequently named, 30% of the strongest attraction ties were selected). **B:** Typology of current study themes according to sex (50% of the strongest attraction ties were selected). **C:** Typology of current study themes according to age by 15 year age groups (50% of the strongest attraction ties were selected).

variables of social identification: sex and age. The structure by sex (see Fig. 143–8) is particularly interesting, insofar as it places men and women in complete dualization. Women are more concerned by fields connected to fertility, nuptiality, life cycles, household studies, the production of statistical data, the study of minorities and groups, and cultural demography. Men clearly state their preference for more technical themes such as demographic trends, methods and models, economic demography, natural resources, and the spatial distribution of population.

From the point of view of age distribution (see Fig. 143–8), the separations are not as clear as with the male-female divide, but the associations of preference are nevertheless unambiguous and reveal substructures that are easy to identify. For example, the oldest demographers (who are also the smallest group) state preferences for historical demography and the history of demography, the study of minorities and groups, but also population studies in general and population policies, even though they share these choices to a great extent with their nearest age group of 50 to 54 years.

At the bottom of the age scale, we observe that the youngest group (aged 34 or younger) and the demographers of average age (35 to 49 years) partly share the same choices; they are in favor of regional or local population studies, household surveys, the production of statistical data, and internal migration. The study of nuptiality and family structure, as well as international migration, are the prerogative of the more mature demographers (50 to 64 years) who nevertheless share these study themes with the youngest ones (aged 34 or younger). Each age group, apart from the first, has at least one study theme that it does not share with any other age group: population in developing countries for those between the ages of 35 and 49 years, the study of life cycles and transition to adulthood for those between 50 and 64 years, and historical demography for those 65 years old or older.

VI. DEMOGRAPHY'S MAJOR FACTORS OF CHANGE

Demography, like any other science, is subject to internal and external influences that contribute to its development and shape its current physiognomy. The choice of eight large factors of change was proposed to the respondents. The classification differs slightly according to whether first choices or multiple choices (four cumulated responses) are referred to, but it is clear that the introduction of individual or longitudinal data has been an important internal factor of development, as the development of information technology has been an equally decisive external factor (Fig. 143–9).

Other factors have played a more discreet role, such as the introduction of new demographic concepts, the increased availability and comparability of macro-demographic data, or the development of human and financial resources. Nevertheless, beyond a simple classification, these data indicate that the interaction of certain external developments such as computerization and the advent of micro-computing, with internal factors, such as some developments in demographic analysis and the availability of individual data, has resulted in a general dynamism of the discipline that otherwise could have sunk into a kind of inertia and regressive stability. In any case, it would be difficult to disassociate the two types of influence, external and internal, because of their considerable reciprocal overlaps, even though by grouping the different types of causes within the questionnaire we observe the slight advantage of external causes over internal causes.

Whatever the case, if demography is changing, it is undoubtedly because of the changes in influences at the level of its scientific approaches and work methods. The discipline is also changing simply because the world is changing and it is therefore constantly faced with new situations that did not exist previously or did not have the same significance or intensity.

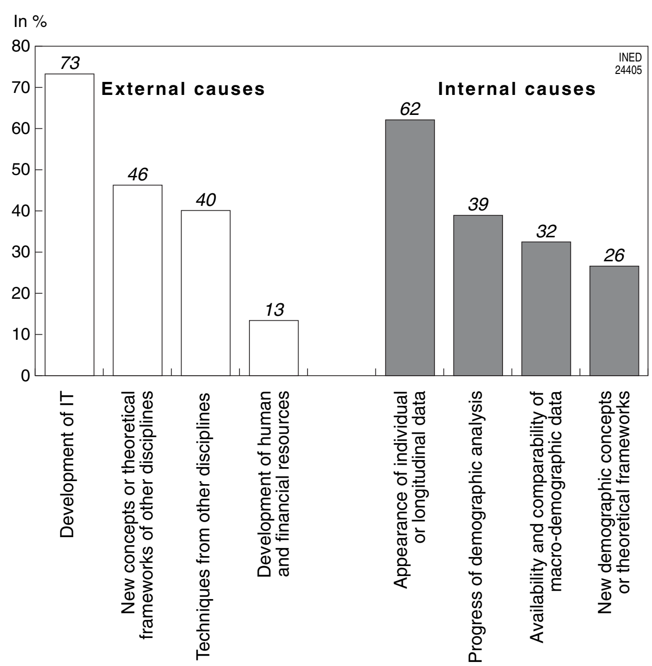


FIGURE 143–9 Factors responsible for the development of demography: internal or external causes?

A group of 26 large phenomena or societal changes that have affected the world over the past few decades were proposed to the respondents and they were asked to specify which one of them had had an impact on their personal practice of demography (five levels of impact were proposed: none, low, medium, high, very high), as well as generally on demography as a discipline.

Among these factors, which are inevitably all external, even if some of them have more demographic connotations than others (e.g., "the spread of contraceptive methods," versus "the decline of the state"), the three emerging phenomena are without contest "Rapid population growth in developing countries," "population aging," and "persistent low fertility in developed countries."

"The development of international migration" is in fifth position, preceded slightly by "the increased role of women." However, the change in women's roles, in fourth place, comes to the top of the list when it is in the context of demography in general and the impact on the respondents' practices. All the more economic or political factors are grouped together near the bottom of the list, even though "environmental threats" and "economic and financial globalization" are in the middle of the list, residing in 11th and 12th place.

An analytic graph of the co-occurrences of the various phenomena mentioned and described in Fig. 143–10, completes the vision that demographers could have of the impact of societal developments on their work (Fig. 143–11). Despite the relative complexity of the relations map, several poles are quite clearly detached. This is the case on the left of the subnetwork, which integrates nearly all the political-economic phenomena (e.g., unemployment, Third World debt, the welfare state crisis, globalization, new forms of poverty); only "the collapse of the Soviet bloc" and "European unity" appear decentralized on the right of the nodal point of unemployment.

In the bottom right corner is a second subnetwork organized around developed countries and their issues: low fertility, aging, and the increase in life expectancy. In the middle and toward the top, a third subnetwork expresses the demographic-epidemiologic problems that face developing countries, notably African countries, with the correspondence of themes such as rapid population growth, fertility decline, the spread of modern contraceptive methods, the AIDS epidemic, and the development of health systems.

On the right, a group of phenomena related to the female condition with changes in the roles of women, the increase of co-habitation outside marriage, and adolescent pregnancies bears witness to a significant problem in both the North and the South and to the existence of a bridge between the problems perceived as being dominant in developing countries and those that have a greater effect on countries that are in the second demographic transition.

VII. BETWEEN PARADIGMS AND METHODS: PUTTING DEMOGRAPHERS' TOOLS TO THE TEST

Demography is generally considered to be a quantitative discipline, dominated by statistical models and methods, so much so that some demographers would sometimes wish for it to be assimilated into a general

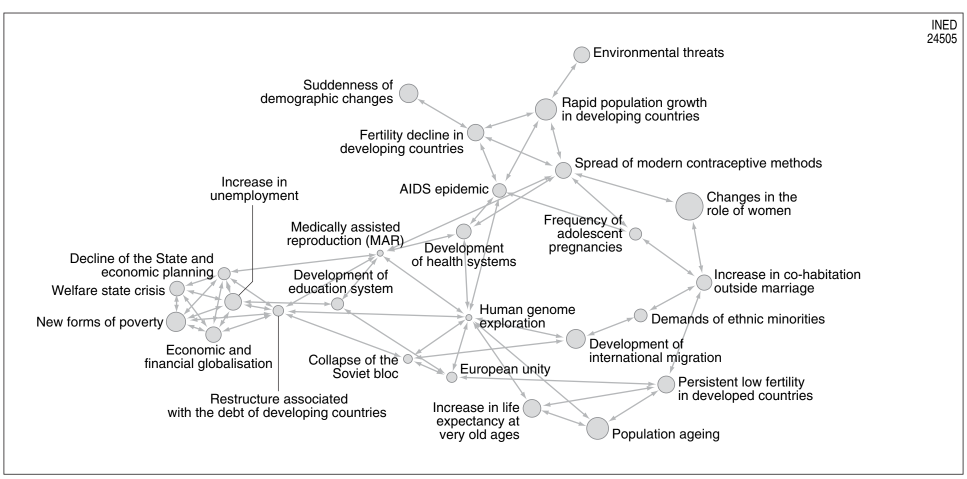


FIGURE 143–10 Cartography of societal phenomena with a high impact on respondents' professional work (20% of the strongest attraction ties were selected).

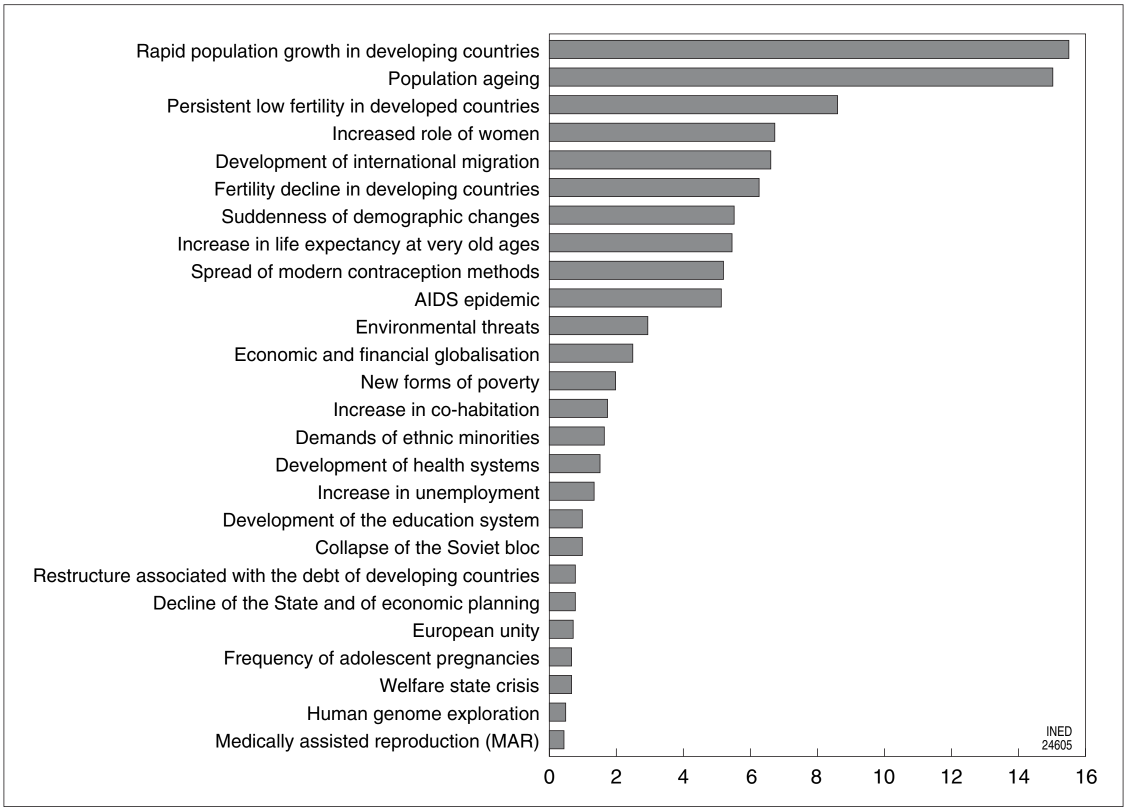


FIGURE 143–11 The impact of recent societal changes on demography as a discipline (classification in percentages by descending order of themes mentioned).

group of methods. However, it would undoubtedly be a dishonor to demography to reduce it to a science without a heart or a soul, providing a perspective of humans and the world through a heap of numbers and curves.

It is true that its isolation from other disciplines and the fact that it is locked away into its narrow field of analysis create a feeling of insularity that has occasionally protected demography from external influences harmful to its development, but that has also delayed the rise of awareness of the necessary interdependence of all social phenomena, of which population issues are only one integral part. Is this still true today?

Apparently not, at least if we judge from the wide diversity of paradigms, methods, and techniques of which modern demographers claim ownership and which are to a great extent borrowed from other disciplines or from the statistical-methodologic sources common to all sciences (see Fig. 143–11). This is the case with regard to regression methods, logistic models, life course analysis, or multilevel analysis, which were created and developed within other disciplines; however, this is much less true or not at all true with regard to life tables, stable populations, or indirect estimation methods that are of fundamentally demographic origin.

But what do the respondents think of them? With regard to paradigms, it is not surprising that the first one mentioned is the "demographic transition" theory (78% for the discipline in general and 53% for the respondents' personal research). However, we may find it strange that "changes in the roles of the sexes" is in second place (52% and 44%), and practically *ex æquo* with "family life cycle" (50% and 44%), and ahead of "health transition" (46% and 29%), or even "evolution in mortality at very old ages" (31% and 17%) placed last but one ahead of "sustainable growth" (29% and 14%).

It was not possible to decide between four or five of the proposed paradigms: spatial analysis, micro-macro relations, inter-generational transfers, health transition, or even "social networks," and the two categories "for demography in general" and "for personal work" are practically the same.

Perhaps the beginning of an explanation is found in the fact that some of the proposed items are concepts rather than analytic frameworks and that they do not

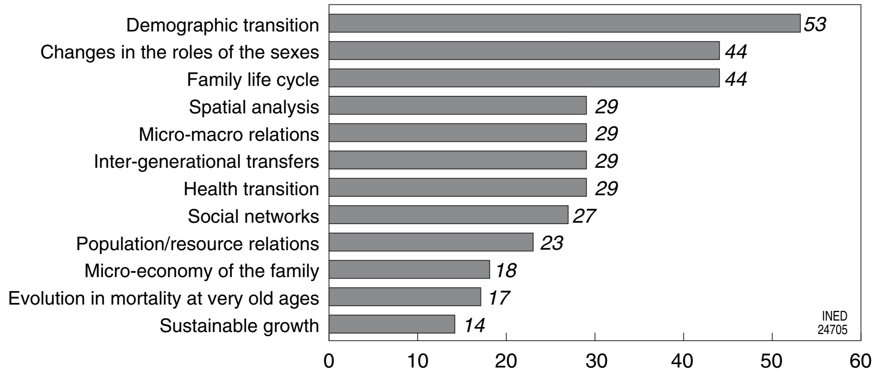


FIGURE 143–12 Classification of the main paradigms used by demographers in their personal work.

necessarily reach their outcomes in the formulation of a processed theory, as is the case for the demographic transition.

It must be emphasized that not all the respondents accord the same importance to the different paradigms proposed, according to their personal characteristics, especially sex; men refer more frequently to what could be called a heavy artillery made up of the demographic and health transitions, the evolution of mortality, sustainable growth, and population-resource relationships, and women are more attracted by "softer paradigms," such as family micro-economics, family life cycle, changes in the roles of the sexes, the analysis of social networks, or inter-generational transfers (trends not illustrated here).

If we leave paradigms and concentrate more on methods and techniques, we observe that the dominant values remain at the level of demography in general, "life tables" (74%), and "data collection methods" (60%). Other relatively traditional methods and techniques fall to the bottom of the list, such as "standardization through standard population" (39%), "modeling variables related to fertility" (39%), "logistic models" (26%), and "markovian models" (16%). Around the middle of the list are "stable populations" in seventh place at 43%. It is interesting to observe the breakthrough of the methods and techniques that have only recently been introduced into demography, such as "APC[4] models" (53%), "life course analyses" (40%), "survival analyses" (36%), and "multilevel analyses" (35%) (Fig. 143–12).

It is also true that the classifications can differ significantly according to point of view: that of researchers for their personal work, or that of the discipline in general (Fig. 143–13). This is why "life tables" receive 74% of the general vote but only 47% of the personal vote.

[4] For "age, period, cohort" (see Chapter 18).

The two graphs of Figure 143–14 provide an additional view of the relations between paradigms and various techniques and methods. The first graph (see Fig. 143–14) is structured in a diamond shape and reveals the close links between these two great dimensions. On the left of the map are the two classics of demographic analysis: "the demographic and health transitions" and "the evolution of mortality at very old ages"; on the side of the paradigms are life tables, stable populations, methods of standardization through standard population, and modeling of variables related to fertility. On the right are the methods and techniques, concentrated quite compactly around the paradigms "changes in the roles of the sexes," "micro-macro relations," "the analysis of social networks" and "spatial analysis," "multilevel and life cycle analyses," as well as "data collection methods."

At the top, another paradigmatic trio, "family life cycle," "intergenerational transfers," and "family micro-economics" attract the APC models, duration models, and survival analyses while keeping strong links with "life course analyses" that becomes a bridge between this subnetwork and that centers on "changes in the roles of the sexes."

The second graph (see Fig. 143–14) illustrates the relations between the researchers' choice of methods and techniques for their own work according to sex and age combined (binary variables: below the age of 50/age 50 and over). Among men, regardless of age, there is a strong attraction for life tables, APC models, stable populations, indirect estimates, and simulation methods. The markovian models are the prerogative of those aged 50 and over. On the female side, while age is relatively undiscriminating, women differ from their male counterparts by a sustained interest in data collection methods, life course analyses, and multilevel analyses, as well as in duration models and standardization by regression methods.

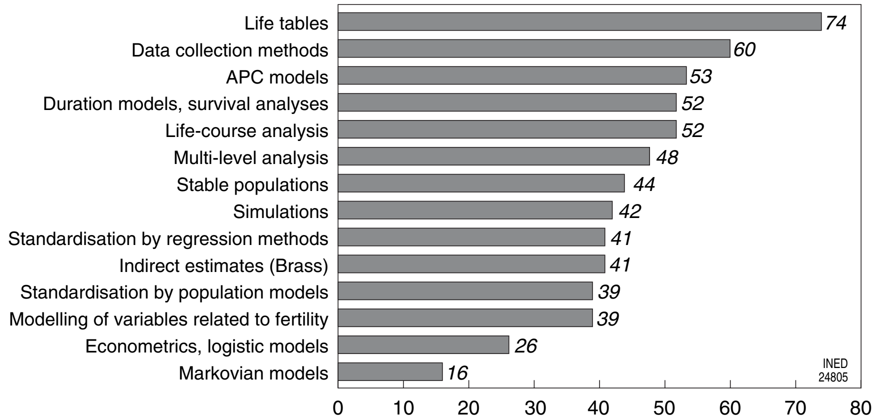
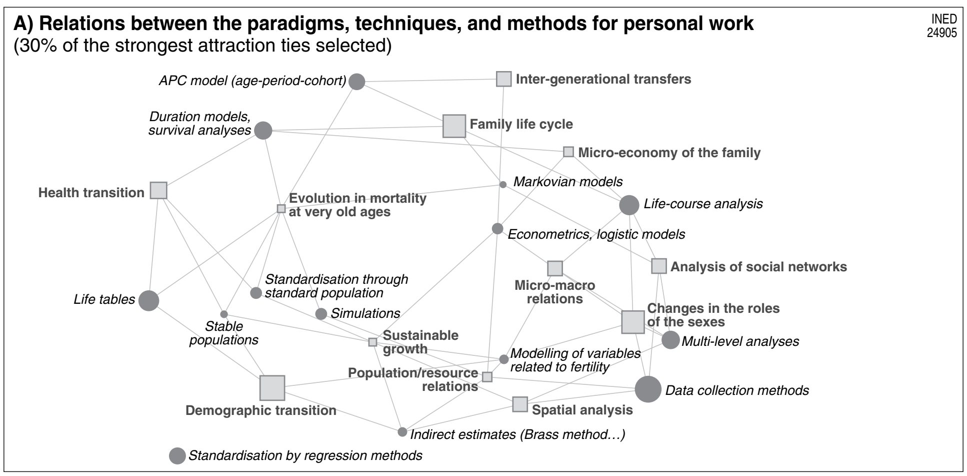


FIGURE 143–13 Classification of methods and techniques according to their importance to demography in general.



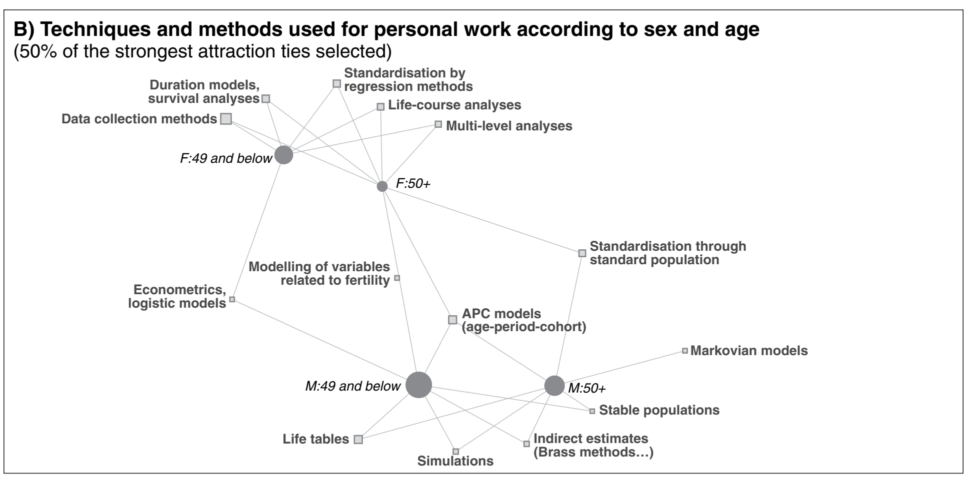


FIGURE 143–14 Paradigms, methods, techniques: what are the relationships and the preferences? **A:** Relations between the paradigms, techniques, and methods for personal work (30% of the strongest attraction ties were selected). **B:** Techniques and methods used for personal work according to sex and age (50% of the strongest attraction ties were selected).

However, the two groups distinguished by sex are not entirely isolated in their methodological preferences insofar as three nodal points ensure the connection between the men and the women, namely standardization by population models, modeling variables related to fertility, econometrics, and logistic models, although these methods and techniques appear nearer the bottom of the list by order of importance.

VIII. A DISCIPLINE IN CRISIS OR A DEVELOPING SCIENCE: DOES DEMOGRAPHY STILL HAVE A FUTURE?

We could not end this review of the professional characteristics and practices of demographers without also mentioning their perceptions about the future of their discipline, the question that initially motivated the Démographie 2000 survey: What are the fears and hopes of demographers concerning their discipline? Do they believe that there is a crisis specific to demography that could eventually call into question its existence as an autonomous discipline and reduce it to a mere store for methods for the use of other disciplines?

Let us immediately reassure ourselves: This hypothesis does not correspond at all to the dominant opinion of the demographers questioned. In the questionnaire, five levels of appreciation were proposed to the respondents: "is demography currently experiencing (in the fields in which you work personally) usual problems, serious problems, or a real crisis at the conceptual, technical, explanatory, ethical, political, or financial level?." The "crisis" response did not exceed 8% for any of these levels and was sometimes as low as 4% (at the conceptual and technical levels). However, 20% to 25% of the respondents believe that their discipline, although it is not experiencing a crisis, is having serious problems, with "usual problems" fluctuating between 50% and 70% (Fig. 143–15).

The question remains whether these usual or serious difficulties are specific to demography or whether they are generally shared with the other social sciences (Fig. 143–16). Although two-thirds of the respondents (68%) believe that the problems are common to all the social sciences, one-fifth (26%) consider that they are specific to demography, and the others (6%) prefer not to give an opinion (Fig. 143–16).

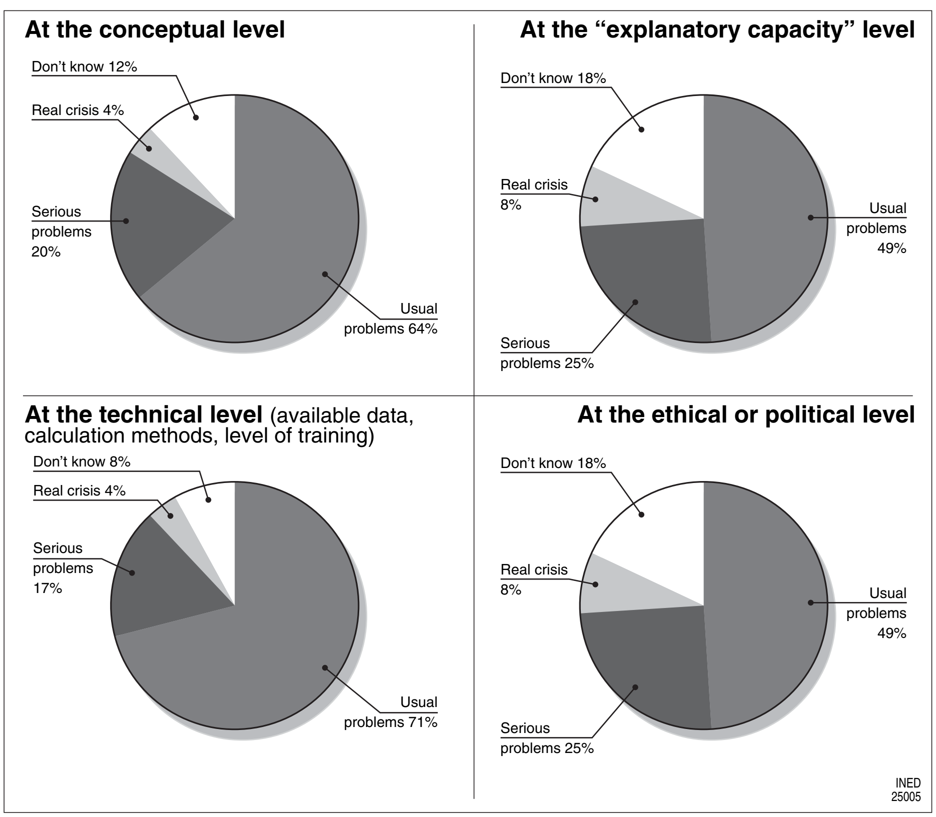


FIGURE 143–15 Crisis in demography or mere problems?

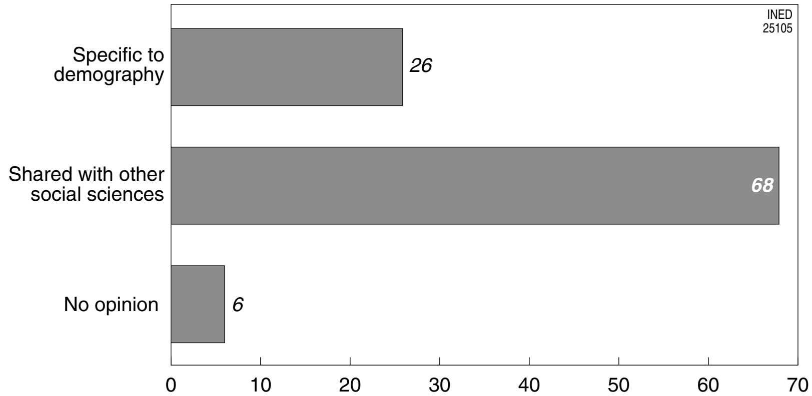


FIGURE 143–16 Problems specific to demography or problems shared with other disciplines?

Another danger facing demography is the possibility of exploding into a large number of relatively autonomous specializations (e.g., economic demography, family demography, demography of migration, mortality studies) to the point of losing its unity and finding itself split among other disciplines (e.g., statistics, economics, sociology, biology).

Do the respondents share this pessimistic vision? Only partly, insofar as they mainly think that in the future there will be an increase in autonomous specializations (52% agree compared with 30% who disagree; 18% did not respond), and they consider that the threat of demography looking inward is a serious risk that should be taken into consideration (50% versus 38% who consider that the risk is low or nil). The most pessimistic (26%) even fear that the discipline will collapse and that the fragments would be recovered by other powerful neighboring disciplines such as sociology, economics, and statistics.

The authors of the questionnaire thought that it would be interesting to find out what motivates the commitment of demographers. Fourteen reasons (and two free responses) were proposed to the respondents. The first choice is not surprising because it corresponds to a very common stereotype: "contribute to scientific progress through research and/or education" (19% of cumulated preferences). Three motivations then follow in nearly *ex æquo* order, with 9% to 10% of responses: "satisfy intellectual curiosity," (common) but more committed "to promote social integration and equality of conditions" and "to fight against North/South disparities." Immediately following are the objectives of "reducing inequalities in health and gender" (7% and 8% of responses).

"To control migratory flows," "to warn governments and opinions about excessive population growth or, on the contrary, about the threatened reproduction of generations," or "to safeguard the demographic bases of national identity" no longer appear to be very important objectives for most demographers.

Examining the graphs of Figure 143–17 reveals the existence of two clearly defined ideologic poles. The one on the left (see Fig. 143–17) groups the modalities corresponding to the most common motivations at the center of the pole, such as "to contribute to the progress of science" and "to satisfy intellectual curiosity" jointly with less common motivations that reflect concern for the preservation of national identity, the wish to regulate migratory flows, or the denunciation of threatened reproduction (in the North) and excessive population growth (in the South).

On the opposite side, a second pole reveals more innovative or more social values such as the reduction of social inequalities and those related to gender, the fight against racism and xenophobia, the search for equity between the generations, and the fight for integration and reducing North-South disparities.

The second chart (see Fig. 143–17), which compares the respondents' motivations with the type of organizations for which they work, highlights the different purposes of demographic research according to the place in which it is carried out. The universities, on the left of the graph, are mainly concerned with reducing inequalities between men and women and in health, as well as reducing racism and xenophobia and fighting against the pernicious use of demography. On the opposite side, the research centers want to contribute to scientific progress and satisfy intellectual curiosity, while warning of the risks of excessive population growth and the non-reproduction of generations. The local administrations and the national offices of statistics, which are structurally close on the graph, share the latter concern while wishing to ensure equity between the generations and regulating migratory flows. The international organizations and NGOs are

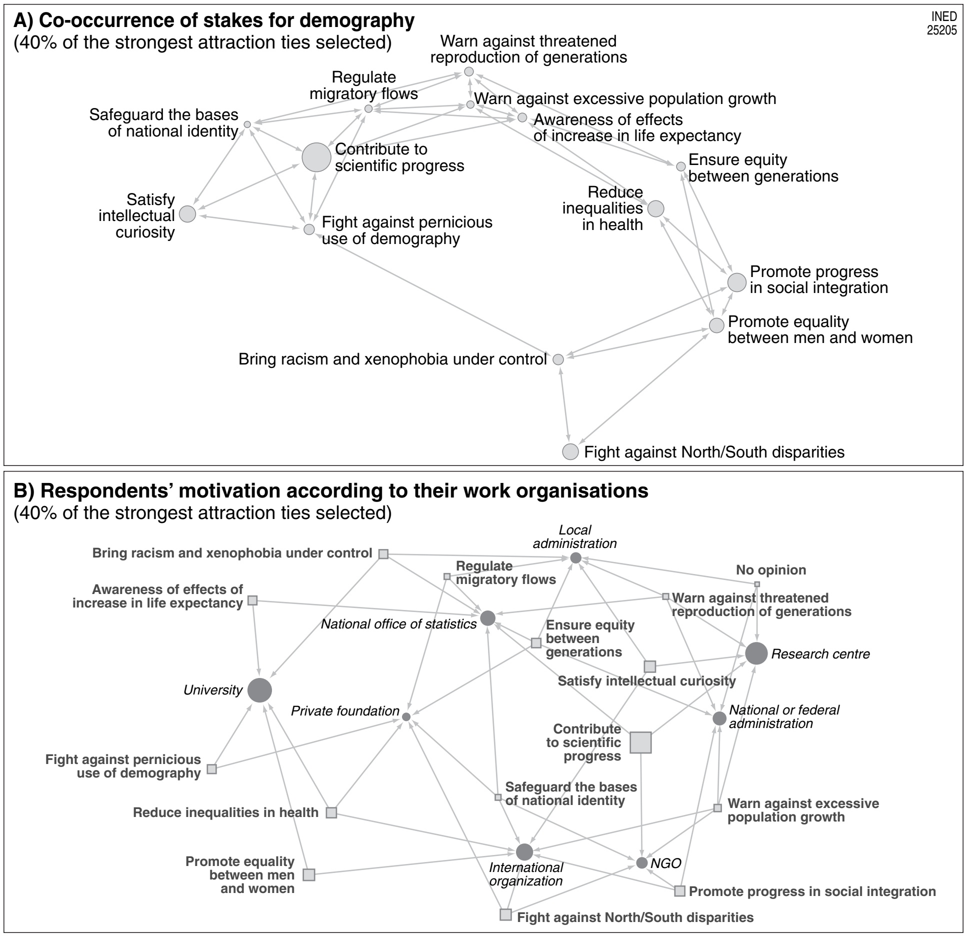


FIGURE 143–17 Demographers' motivation. **A:** Co-occurrence of stakes for demography (40% of the strongest attraction ties were selected). **B:** Respondents' motivation according to their work organizations (40% of the strongest attraction ties selected).

more concerned with excessive population growth and reducing North-South disparities, as well as wanting to promote social integration.

In any case, demographers have many challenges ahead of them that they will be able to rise to, alone or in collaboration with representatives from other disciplines, if they are not too timid, or as Alfred Sauvy said about aging, overcautious.

In the South, the previous priority themes of population growth and excess fertility are being supplanted by concerns regarding health. In the North, family and household life "remain at the center of social demography concerns, but the emphasis is moving away from traditional marital behavior toward new behavior" (Romaniuc, 2004). The massive intrusion of biogenetics into reproductive and health fields will no doubt contribute to the natural science aspect of demography, without weakening its attachment to the social sciences, as these problems of poverty and inequality between populations and within populations are increasingly intolerable, just as the increasingly pressing ecological constraint is imposing more activity in favor of sustainable development.

From the point of view of Western countries, all the attention of political leaders will for a long time remain focused on issues of immigration and low fertility insofar as they "call into question social cohesion, national identity, and the power struggle in the world" (Romaniuc, 2004).

All things considered, the vision sometimes evoked of demography as a discipline that is running out of steam and having difficulty finding its second wind could lessen, and even disappear, at the dawn of the 21st century, if demographers realize that beyond pure knowledge and scientific neutrality (which remain fundamental concerns), they also need to give more preference to knowledge applied to the service of humanity, without hesitating to intervene in public debates surrounding major regional, national, and world issues.

By abstaining for too long from making any normative judgments and taking any ideologic stands, demography has perhaps lost view of the fact that the essence of society is fundamentally political, and it will regain a full place within the social sciences (as demology?) only if it endorses the lucid sentence of Pareto when he stated in 1916 "that there is no scientific solution to the problem of human actions, nor for that of social organization."

Acknowledgment

This chapter was translated from the French version by Accenta Ltd.

References

Billari Francesco and Micheli Giuseppe A., 1995. Whither demography? A concise portrait of European participants to the 1995 Milan Conference, *in*: EAPS and IUSSP (eds.), *Évolution ou révolution dans la population européenne/Evolution or revolution in European population*, vol. 3, p. 323–341. Milan, Franco Angeli, 341 p. (European Demography Conference, Milan, 1995.)

Caldwell John C., 1997. Demography and social Sciences, *Population Studies*, vol. 50(3), p. 305–333.

Cambrosio Alberto, Kearing Peter A., and Mogoutov Andrei, 2003. Mapping collaborative work and innovation in biomedicine: a computer assisted analysis of antibody reagent workshops, *Social Studies of Science*, vol. 34(3), p. 325–264.

Chasteland Jean-Claude, 2004. Introduction générale, *in*: Jean-Claude Chasteland, Michel Loriaux, and Louis Roussel (eds.), *Démographie 2000. Une enquête internationale par Internet auprès des démographes*, p. 11–17. Louvain-la-Neuve, Academia-Bruylant, 368 p.

Chasteland Jean-Claude, Loriaux Michel, and Roussel Louis (eds.), 2004. *Démographie 2000. Une enquête internationale par Internet auprès des démographes*. Louvain-la-Neuve, Academia-Bruylant, 368 p.

Chasteland Jean-Claude and Roussel Louis (eds.), 1997. *Les contours de la démographie au seuil du XXI[e] siècle*. Paris, INED and PUF, 434 p.

Chasteland Jean-Claude and Roussel Louis, 1997. Un demi-siècle de démographie dans les pays industriels. Quelques réflexions sur un bilan, *in*: Jean-Claude Chasteland and Louis Roussel (eds.), *Les contours de la démographie au seuil du XXIe siècle*, p. 57–93. Paris, INED and PUF, 434 p. (Conference proceedings: 1945–1995. Un demi-siècle de démographie. Bilan et perspectives, Paris, 25–27 October 1995.)

Coleman David, 2000. Demography in an intellectual context: a subject in search of a home, *in*: Zdenek Pavlik (ed.), *Position of demography among other disciplines*, p. 27–35. Prague, Charles University in Prague, Department of Demography and Geodemography, 117 p.

Crimmins Eileen M., 1993. Demography: the past 30 years, the present, and the future, *Demography*, vol. 30(4), p. 579–591.

Duchêne Josianne, Wunsch Guillaume and Vilquin Éric (eds.), 1989. *L'explication en sciences sociales. La recherche des causes en démographie*. Louvain-la-Neuve, CIACO éditeur et ARTEL, 478 p. (Chaire Quetelet 1987.)

Gérard Hubert and Loriaux Michel (eds.), 1988. *Au-delà du quantitatif. Espoirs et limites de l'analyse qualitative en démographie*. Louvain-la-Neuve, CIACO Editeur, 669 p. (Chaire Quetelet 1985.)

Golini Antonio, 1985. New trends of demographic research in the developed countries, *in*: Département de démographie (eds.), *La démographie en perspective. Visages futurs des sciences de la population*, p. 77–101. Louvain-la-Neuve, Université Catholique de Louvain, Département de démographie, 160 p. (Chaire Quetelet 1984.)

Graunt John, 1662. *Natural and political observations made upon the bills of mortality*. London, printed by T. Roycroft for J. Martin, J. Allestry and T. Dicas, 150 p. (American edition by W. F. Wilcox, Baltimore, Johns Hopkins Press, 1939; French edition by Éric Vilquin, Paris, INED, 1977.)

Loriaux Michel, 1995. Les approches méthodologiques en sociologie de la population, *in*: Hubert Gérard and Victor Piché (eds.), *La Sociologie des populations*, p. 71–116. Montreal, Presses de l'Université de Montréal and AUPELF-UREF, 508 p.

Loriaux Michel, 1996. L'apport de l'école louvaniste de démographie: réflexions critiques sur le passé et nouveaux défis pour l'avenir des sciences de la population, *Population*, vol. 51(2), p. 405–416. (INED's 50th anniversary.)

Micheli Giuseppe A., 2004. Genius loci, le cloisonnement régional des démographies, *in*: Jean-Claude Chasteland, Michel Loriaux, and Louis Roussel (eds.), *Démographie 2000. Une enquête internationale par Internet auprès des démographes*, p. 205–248. Louvain-la-Neuve, Academia-Bruylant, 368 p.

Mogoutov Andreï, 1999. Réseau-Lu, Sistema di esplorazione di dati relazionali, *in*: Antonio M. Chiesi (ed.), *L'analisi dei reticoli*, p. 221–225. Milan, Franco Angeli, 255 p.

Mogoutov Andreï and Vichnevskaia Tatiana, 2004. Présentation de la technique d'analyse des données d'enquête Réseau-Lu, *in*: Jean-Claude Chasteland, Michel Loriaux, and Louis Roussel (eds.), *Démographie 2000. Une enquête internationale par Internet auprès des démographes*, p. 37–54. Louvain-la-Neuve, Academia-Bruylant, 368 p.

Pareto Vilfredo, 1916. *Trattato di sociologia generale. Firenze*. French trans. revised by the author: 1917–1919, Traité de sociologie générale. Paris, Payot, 2 vol., 1761 p. (Republished in 1968, *Œuvres complètes*, t. 12, Geneva, Droz, 1822 p.)

Pavlik Zdenek (ed.), 2000. *Position of demography among other Disciplines*. Pragues, Charles University in Prague, Department of Demography and Geodemography, 116 p.

Preston Samuel H., 1993. The contours of demography: estimates and projections, *Demography*, vol. 30(4), p. 593–606.

Romaniuc Anatole, 2004. Comment les démographes perçoivent l'avenir de leur discipline? *in*: Jean-Claude Chasteland, Michel Loriaux, and Louis Roussel (eds.), *Démographie 2000. Une enquête internationale par Internet auprès des démographes*, p. 248–295. Louvain-la-Neuve, Academia-Bruylant, 368 p.

Roussel Louis and Chasteland Jean-Claude, 1996. Un demi-siècle de démographie dans les pays industriels. Quelques réflexions sur un bilan, *Population*, vol. 51(2), p. 417–437. (INED's 50th anniversary.)

Sauvy Alfred, 1946. Introduction, *Population*, vol. 1(1), p. 5–8.

Tabutin Dominique, Gourbin Catherine, Masuy-Stroobant Godelieve, and Schoumaker Bruno (eds.), 1999. *Théories, paradigmes et courants explicatifs en démographie*. Louvain-la-Neuve, Academia-Bruylant, L'Harmattan, 672 p. (Chaire Quetelet, 1997.)

Wattelar Christine and Duchêne Josianne (eds.), 1999. *Le défi de l'incertitude. Nouvelles approches en perspectives et prospectives démographiques*. Louvain-la-Neuve, Paris, Institut de démographie de l'UCL, Academia-Bruylant and L'Harmattan, 436 p. (Chaire Quetelet 1995.)

Index

B

E

G

H

I

J

K

L

N

O

P

V